CEBRA 170621: Proportional value of interventions across pathways and layers of the biosecurity system

Final Report

Andrew Robinson, CEBRA, The University of Melbourne

Michael Ormsby, The Ministry for Primary Industries, New Zealand

March 10, 2021

Contents

List of Tables

List of Figures

Executive Summary

Context

The biosecurity system faces increasing pressure from significant increases in goods and passengers, changing pathways and types of goods. All activities of the system need to work together cost-effectively to minimise the biosecurity risk to New Zealand under sharply constrained resources. In order to increase the efficiency of biosecurity investment and to identify opportunities for substantial improvement, the Ministry for Primary Industries (MPI) needs to determine the relative contribution of a continuum of biosecurity risk management activities toward overall system effectiveness.

Previously has been no framework to evaluate the comparative value of biosecurity activities implemented at intersecting sites across a matrix of entry pathways and across layers of the system. CEBRA Project 1606E (reported in Robinson et al., 2018a) initiated development of a framework through which MPI could summarize the actions of the biosecurity system against a pest, and Project 170621 (reported in Robinson et al., 2018b) completed the structure of the model.

The overarching objective of the suite of projects is to develop a simple, scaleable structural model of biosecurity pest risk that encompasses potential high-level regulatory responses in order to be able to assess the relative risk of different pests, and the likely utility of different regulatory responses against the risks presented by those pests.

Here, we present and assess the structural model as parameterised for gypsy moth (GM, European and Asian) using expert elicitation.

Model

The Pre-border and Border sub-model comprises the following components:

- Raw exposure — the rate per 1000 at which units that are arriving on six pathways are contaminated with GM;
- Interventions — three key activities that are performed or mandated by MPI to reduce the exposure and therefore the likelihood of incursion;
- Attenuation from incursion to establishment — propagules that arrive past the border may still fail to establish.

The Post-border sub-model focuses on the effect of post-border activities upon the consequences of incursion. It comprises the following components:

- Detection of the established pest — this is assumed to happen sooner or later, so the unknown is the area infested at time of delimitation. We use the Weibull distribution.
- Eradication of the population, the success probability and cost of which depend on the (unknown) area infested at time of delimitation.

Results

All of the following summary points must be read in the context of the substantial parameter and model uncertainty that arise from both the inherent unknown nature of the underlying process and the considerable number of unappetizing tweaks and corrections that were necessary to align the model predictions with the few available observations. *The conclusions of this exercise should be taken as indicative only*, being representative of the kinds of conclusions that would be possible when the model fitting exercise is tightened up.[1]

This report covers parameterisation and testing of the risk management model for GM. All economic data are presented as millions of NZ 2019 dollars. Key outcomes arising from the model simulation exercise are:

1. The biosecurity risk to New Zealand of establishment by GM, taking account of the current biosecurity measures, is *very low*.
    (a) The expected annual spacing of an establishment with the current biosecurity system is 6.5–2089.7 years[2], with best guess 18.7 years.
    (b) There has been exactly one known incursion in the past 30 years, a lone male in Hamilton in 2003.
    (c) The expected cost of a GM establishment is 9.1–93.0 ($NZ million), with best guess 40.1.
    (d) The expected annual cost of GM impacts (not including costs of measures[3]) is 0.0191–7.8060 $NZ million, with best guess 2.05.
2. Three pathways are implicated in transmission of GM egg sacs, namely: (i) sea containers, (ii) international marine vessels, and (iii) used vehicles and machinery.
    (a) Activities within the pathways are *high risk* or *low risk* depending on whether the activity passes through breeding areas for GM.
    (b) The source of greatest estimated risk, if unmitigated, is high-risk used vehicles and machinery (from Japan): the expected annual arrival rate of propagable GM egg sacs is 0.494–45.079, with best guess 12.1.
    (c) The biosecurity risk from this pathway is managed by inspection and heat treatment, which achieves a risk reduction of 70.7–99.2%, with best guess 89.9%.
    (d) The estimated annual value of heat treatment is therefore 0.0364–31.1255, with best guess 8.12 ($NZ million).
3. Two key post-border measures reduce the expected impact of an incursion and therefore the expected annual cost to NZ presented by GM.
    (a) Surveillance activities accelerate the detection of invading populations, which means that eradication will cost less and be more likely to succeed. The annual reduction to the cost of GM that is due to active surveillance is 0.00–6.06 million, with best guess 1.26 million.
    (b) Readiness for an incursion means having a plan that is based on the biology of the pest and knowledge of local climate and host distributions. The annual reduction to the cost of GM that is due to readiness *and* active surveillance is 0.00–6.62 million, with best guess 1.62 million.

Conclusions

The model allows us to estimate the value against GM of various activities, including surveillance, readiness, and border interventions. The biosecurity risk to New Zealand of establishment by GM, taking account of the current biosecurity measures, is *very low*.

Since the start of the project, similar exercises have been developed, notably IBRAM and the B3 project "Optimising biosecurity investment and effort across all invasion phases". These projects address the same challenge but from slightly different points of view.

[1]These processes are under improvement in current CEBRA work.
[2]All intervals are 90% coverage
[3]Measures to be considered in ongoing CEBRA project.

Recommendations

1. MPI should note the outcomes of the modeling exercise and verify whether the questions that can be answered using the outcomes have operational utility.

2. If border data and establishment data exist, then we need to find a proactive way to use them. How best to do that is an open question. MPI and CEBRA should consider this question in the continuation of the project. Progress in the project documented herein was hampered by the regulatory data being available after the expert elicitation exercise, necessitating contestable unilateral alterations to the parameters.

3. If quantities have a natural rank then we need to elicit them in a way that preserves the rank (e.g., given an invading population, does it ever make sense for eradication to be cheaper when we are *not* investing in active surveillance?) How best to do that is an open question that should be considered by future work.

Acknowgments

The authors are very grateful to the following expert scientists: Diane Anderson (MPI, NZ), David Gray (NRC, Canada), Andrew Liebhold (USDA, USA), Rory MacLellan (MPI, NZ), and Stephanie Sopow (Scion, NZ), for their generous and thoughtful engagement with the expert elicitation process. John Kean assisted with accessing the GERDA data. Nathaniel Bloomfield, Melissa Welsh and Rebecca Epanchin-Niell provided invaluable review comments and input. Craig Phillips and Shashi Sharma provided detailed and thoughtful reviews that greatly enhanced the quality of the report.

Chapter 1

Introduction

1.1 Background

The biosecurity system faces increasing pressure from significant increases in goods and passengers, changing pathways and types of goods. With this increasing pressure, all activities of the system need to work together to cost-effectively minimise the biosecurity risk to New Zealand under sharply constrained resources. In order to increase the efficiency of biosecurity investment and to identify opportunities for substantial improvement, the Ministry for Primary Industries (MPI) needs to determine the relative contribution of a continuum of biosecurity risk management activities toward overall system effectiveness.

Managing biosecurity risk can be considered from an investment point of view: one invests in particular portions of the biosecurity system to best avert the likelihood or the consequences of an incursion. However, the question about where to make the investment remains open and depends on the nature of the biosecurity threat. Some threats are relatively easily detectable at the border, for example, giant African snail, so border measures are efficient. Other threats cannot be easily detected by visual inspection, so border measures such as fumigation or heat treatment are preferred. Still other threats are more efficiently handled before they arrive at the border, for example by international conventions such as ISPM 15, which proscribes certain treatments for wood packaging, and others most efficiently handled post-border by means of sentinels such as sentinel hives or flocks.

Previously there has been no agreed framework or process available to evaluate the comparative value of biosecurity activities implemented at intersecting sites across a matrix of entry pathways and across layers of the system. CEBRA Project 1606E (reported in Robinson et al., 2018a), initiated development of a framework through which MPI could summarize the actions of the biosecurity system against a pest, and Project 170621 (reported in Robinson et al., 2018b) completed the structure of the model. At the same time, developments have been made toward comparable outcomes in IBRAM (Jamieson et al., 2016) and in the B3 project "Optimising biosecurity investment and effort across all invasion phases" (see, e.g. Welsh et al., 2020). We discuss these parallel projects in greater detail in Section 6.5.

The overarching objective of the suite of projects is to develop a simple, scaleable structural model of biosecurity pest risk that encompasses potential high-level regulatory responses in order to be able to assess the relative risk of different pests, and the likely utility of different regulatory responses against the risks presented by those pests.

This report covers the parameterisation and testing of the risk management model for the European gypsy moth (*Lymantria dispar dispar*) and Asian gypsy moth (*Lymantria dispar asiatica*) combined, treating them as the same species (GM). The report is structured as follows. The rest of this chapter provides an overview of the model structure as developed in the previous CEBRA projects. Chapter 2 reports the outcome of the expert-elicitation exercise for the pre-border and border components of the model, along with some comparison data, and Chapter 3 reports the outcome of the expert-elicitation exercise for the post-border component, also with comparison data. Chapter 5 reports the conclusions that we draw from the model fit.

Briefly, we assume that the biosecurity risk of a pest is the product of the likelihood and the consequences. The pre-border and border models focus on the impact of activities upon the likelihood, and the post-border model focuses on the impact of activities upon the consequences of establishment and spread.

1.2 Risk–Return Model

The structural risk-return model developed for this project comprises two main portions, namely: a pre-border/border portion that delivers a probability distribution function of the number of establishments per year (essentially the *likelihood* of damage occurring) and a post-border portion that delivers a probability distribution function of the expected damages conditional on establishment and spread (essentially the *consequences*). The product of these two components is the distribution of the annual damage from the pest. Note that for simplicity in computing damages we assume that establishment entails spread, and we do not separately consider spread from this point on.

1.2.1 Pre-border and border sub-model

The framework comprises four pathways (cargo, mail, passengers, transport) that are divided into 24 sub-pathways, namely (the pathways that are used for this GM model are noted in bold):

- Cargo
  - Animal Germplasm
  - Animal Products
  - Biological Products
  - **Containers**
  - Aircans
  - Fresh Produce and Cut Flowers
  - Live Animals
  - Nursery Stock
  - Plant Products
  - Seed and Grain
  - Vehicles and Machinery, **Used** and New
  - Wood and Wooden Products
- Mail
  - Articles
  - Bulk
  - Express
  - Letters
  - Parcels
- Passengers
  - Cruise, High Risk
  - Cruise, Low Risk
  - Air, High Risk
  - Air, Low Risk
- Vessels
  - Air, Fully Cleared
  - Air, Under Surveillance
  - **Marine, Fully Cleared**
  - **Marine, Under Surveillance**

The pre-border/border model also includes activities classified into four of the seven layers described in the MPI "rainbow diagram" of layers of the NZ Biosecurity system, namely:[1]

1. International Plant & Animal Health Standards ("International Agreements");

[1] The parenthesized labels were used in CEBRA report 1606E (Robinson et al., 2018a).

2. Trade Agreements & Bilateral Arrangements ("Bilateral Agreements");
3. Risk Assessment & Import Health Standards ("Import Health Standards"); and
4. Border Interventions ("Border").

The model is a simple accounting model that starts with raw exposure of a biosecurity risk along a pathway, which is then ameliorated by different measures classified as above.

1. Considering a pest, we identify all the pathways by which the pest might enter NZ, for example, air passengers, mail, or certain kinds of cargo. We may do this by examining interception records, if border inspection is performed, or by asking experts.
2. For each pathway, we identify the unit of intervention, which is defined as the level at which biosecurity decisions are typically made. For example, in used motor vehicles the unit is a car or motorbike. For cut flowers the unit is a consignment.
3. We determine the number of units of intervention (e.g., cars, containers) arriving at the border per unit of time, typically using Customs databases.
4. We then identify all the measures that are in place that may affect the exposure of the pathway to the pest — for example, the imposition of pest-free areas, border inspection, and so on. We may need to stratify the pathway by measures — for example, international passengers experience different measures depending on the country that issued their passport.
5. We estimate the multiplicative impact each measure has upon the exposure on the pathway – the impact is a number from 0 (no impact) to 1 (completely stops all exposure). We typically use expert elicitation for this estimation, and elicit uncertainties.
6. Then for the pest we define a minimal viable population, which is the minimum number of the pest at a given life stage that is required for a successful incursion.
7. We estimate the raw, baseline contamination of each pathway to the pest, that is, the rate per unit at which a propagable unit of the pest would arrive on that pathway if no measures were in place.
8. We compute the raw pathway exposure as the product of the baseline contamination and the number of units of intervention per unit time. This could also be considered a counterfactual or 'baseline' exposure.
9. Then to estimate the residual exposure of the pathway, we multiply the raw pathway exposure by each of the relevant measure impacts subtracted from 1. For example, if the raw pathway exposure were 0.9 and the impacts of two measures were 0.2 and 0.1, then the residual pathway exposure would be $0.9 \times (1 - 0.2) \times (1 - 0.1) = 0.648$.
10. We conclude by attenuating the expected number of incursions to reflect the fact that not all incursions will result in establishment. However, this attenuation rate is assumed to be independent of risk management activities.[2]

The model does not distinguish between measures taken at the border, such as inspection, and measures that are taken offshore, such as offshore certification.

Chapter 2 reports the outcome of the expert-elicitation exercise for the pre-border and border components of the model.

1.2.2 Post-border sub-model

This section describes the post-border sub-model in some detail. As mentioned earlier, we assume that the post-border activities affect the *consequences* of incursion but not the *likelihood*, so we condition on establishment of a minimum viable population, that is, we assume that a minimal viable population (MVP) of population has entered the country and established. We make the following assumptions and definitions.

1. Transitional Facilities are considered to be a part of the national border, so pests that arrive at but do not escape from transitional facilities have not entered the country.

[2] As reviewer CP pointed out, attenuation rate is not necessarily independent of management. For example, improperly irradiated organisms could still reproduce, but their fertility may be lower.

2. All damages from the pest before the pest had been detected are negligible.[3]
3. If eradication fails or is not attempted then the damages that arise from the incursion are independent of the area occupied by the incursion when it is delimited.
4. Eradication success or failure is instantaneous (but the costs of damages from slow eradication may be included in the cost of eradication).
5. If eradication is successful, then the only costs and damages are financial, that is, the impact on (i) human health, (ii) the environment, (iii) cultural values, and (iv) social amenity are negligible.[4]

We represent the area occupied by the incursion at the time of delimitation by the random variable x (which has probability density function $f_X(x)$ and cumulative distribution function $F_X(x)$), and the cutoff above which eradication will not be attempted as t. So, if $x > t$ then eradication is not attempted. If eradication is not attempted, or fails, then the damages are set at $\$_c$:

$$\$_c = \frac{\$_m}{d}, \tag{1.1}$$

where $\$_m$ is the expected annual loss due to pest management for the pest having reached its maximum spread and d is the discount rate.

In order to reduce complexity and avoid the need to model pest population spread explicitly, we assume that the full pest impacts take effect immediately upon failure to eradicate.[5] In the current study, we adopted an impact estimate for $\$_c$ from Harris Consulting (2003), updated to 2019 NZ dollars.

On the other hand, if $x < t$, then the costs and damages would comprise the cost of incursion response and the damages if the incursion response fails, so the expected loss would be:

$$\$_e(x) + (1 - P_e(x))\,\$_c, \tag{1.2}$$

where $\$_e(x)$ is the cost of the eradication attempt as a function of the area infested at x, and $P_e(x)$ is the probability of successful eradication as a function of x. Examples of these functions are

$$\$_e(x) = \exp(-0.46144 + 0.47376 \log x)$$

and

$$P_e(x) = 1/(1 + \exp(-0.55952 + 1.5919 \log x))$$

respectively, where x is in units of km^2 invaded (Epanchin-Niell et al., 2014).

In the current study, we obtained data from GERDA on the infestation size, and the cost and success/failure of response for GM incursions (Kean et al., 2019). These data were used to fit generalised linear models, the predictions from which were then scaled linearly in order to match the values obtained by expert elicitation (Sections 3.3.2 and 3.4.2). We used all available GM records from GERDA rather than constraining them by time.

We determine t using equations 1.1 and 1.2:

$$t = x \text{ } s.t. \text{ } \frac{\$_e(x)}{P_e(x)} = \$_c;$$

taking for example the models above, we seek t so that

$$\begin{aligned}\$_c &= \exp(-0.46144 + 0.47376 \log t) \times (1 + \exp(-0.55952 + 1.5919 \log t)) \\ &= t^{0.47376} \times (0.360249 \times t^{1.5919} + 0.630375),\end{aligned}$$

[3] As reviewer CP pointed out, this assumption is probably false for some pests, for example clover root weevil, but it is convenient because it simplifies the model and reduces effort for the experts.

[4] This assumption is also probably false for some pests, for example *Mycoplasma bovis*, but it is convenient because it means that the impacts of incursion in these four different sets of values only need to be considered in the light of failure or refusal to eradicate. If we do not make this assumption then we are obliged to estimate the impact of all incursions as a function of the area infested at time of delimitation in each of these four sets of values, which would be very taxing for the experts, instead of only those incursions for which eradication fails or is not attempted.

[5] As pointed out by reviewer CP, discount rate d would in reality have a much larger effect on the costs of slow-spreading pests compared to a fast-spreading pest.

with $\$_c$ expressed in \$US millions. We solve this equation for t by using a root-finding algorithm (see, e.g., Jones et al., 2014).

Then, let $\$_I(x)$ be the total costs and damages due to establishment as a function of the area infested at time of delimitation. It has two portions, depending on x and t.

$$\$_I(x) = \begin{cases} \$_e(x) + (1 - P_e(x))\,\$_c & \text{for } x < t \\ \$_c & \text{for } x \geq t \end{cases} \tag{1.3}$$

Following equation 1.3, the loss can be split into two portions that depend on the area infested at time of delimitation (x). The first portion is the average loss of incursion response (including the possibility of failure) as a function of incursion size, and the second is the fixed loss of pest management multiplied by the probability of detecting the incursion too late for response. Symbolically, the probability-weighted contribution to the loss due to an incursion can be written as

$$[\$_e(x) + (1 - P_e(x))\,\$_c]\, f_X(x) \text{ for } x < t$$
$$\$_c\,(1 - F_X(t)) \text{ for } x \geq t$$

We can calculate the total expected cost by integrating these functions across x, resulting in

$$E_x(\$_I(x)) = \int_0^t [\$_e(x) + (1 - P_e(x))\,\$_c]\, f_X(x)dx + (1 - F_X(t))\,\$_c, \tag{1.4}$$

which can be computed using mathematical programming algorithms or by Monte-Carlo simulation.

The parameters needed to compute this expectation can be estimated using information derived from official databases and/or the expert elicitation exercise. Estimates can be computed for different scenarios as elicited, for example, including or excluding active surveillance, which enables estimation of the value of an active surveillance program.

Alternatively, in a simulation exercise to capture the uncertainty (see Chapter 4), we can simulate a random instance of the incursion size at detection and compute the cost as a function of the random size. We compute these values using random draws from the parameter estimate distributions obtained from each expert, and then average them across experts to obtain an overall prediction instance. This process is repeated 10,000 times to obtain a predicted distribution that reports expert-level uncertainty. Chapter 3 reports the outcome of the expert-elicitation exercise for the post-border component.

1.3 Counterfactuals: Readiness and Surveillance

Two significant post-border investment options are supported by the model, namely: readiness and surveillance. We describe each in this section, but briefly, readiness is the suite of activities that may be taken before the arrival of a pest in order to make response more efficient, for example, choosing lures, designing trap networks, and so on. Surveillance is taken to mean the active surveillance for the pest, by which we mean for GM the establishment of trap networks in vulnerable host material. See Section 1.2.2 for further details on the post-border model

1.3.1 Readiness

The Readiness Group oversees and manages a comprehensive programme of readiness for the Ministry for Primary Industries. The readiness programme focuses on improving overall readiness in order to respond effectively, and includes maintaining generic readiness that supports all systems (including biosecurity, food safety, adverse events, animal welfare and trade).

MPI's readiness programme is based around seven areas of readiness, namely:

1. Plans

2. Processes

3. Information
4. Resources
5. People
6. Relationships
7. Performance and Assurance

Key programmes of work include:

- Response capability programme
- Response improvements and lessons identified framework
- Exercise programme
- Response information systems (Tiaki)
- Threat specific readiness (e.g. Foot and Mouth disease, Brown Marmorated Stink Bug)
- Government Industry Agreement readiness for biosecurity
- Partnering across other response systems within MPI (including food safety, trade, animal welfare, adverse events)
- Biosecurity New Zealand Response Services Project

In the model, readiness can affect the outcome by (i) increasing the probability of eradication at a given area occupied by the invasive population ($P_e(x)$), and (ii) decreasing the cost of eradication at a given area occupied by the invasive population ($\$_e(x)$).

In this exercise we elicited these impacts by first asking for the cost and probability of success for eradicating GM given an invading population that occupied 500 ha, and then asking the experts the same questions under the counterfactual scenario of an imaginary pest that was in every way identical to GM but about which we knew nothing, for example no lures, no traps, no spread data, etc. Comparison of the pest damages estimates under these two scenarios provides an estimate of the value of readiness.

1.3.2 Surveillance

Targeted surveillance is undertaken to try to detect the invading population before it becomes too expensive or unwieldy to eradicate. Consequently, surveillance can be seen as analogous to insurance against pest establishment.

NZ's GM surveillance program was initiated in 1992. As of 2012, it comprises about 1500 sticky delta traps accompanied by a commercially available pheromone to detect male moths, set up in grids and maintained from October to May each year (Ministry for Primary Industries, 2012).

In the model, surveillance can affect the outcome by changing the probability distribution that reflects the size of the invading population at the time at which it is detected and subsequently delimited. This probability distribution in turn affects the attempt decision, cost and success probability of eradication.

In this exercise, we elicited these impacts by first asking for the most likely area occupied by the invading population at the time of its detection (namely, the *mode* of the probability distribution) for the current surveillance setup, and then the same question under the counterfactual of no active surveillance. Comparison of the pest damages estimates under these two scenarios provides an estimate of the value of the active surveillance.

1.4 Expert elicitation

This section is copied with only modest alteration from CEBRA report 1505A Output # 1,[6] and was authored by Dr. Jan Carey. Any inconsistencies are the responsibility of the first author of the current document.

Ideally, decision-making should be informed by high quality data. However, empirical data are often sparse or lacking in fields such as conservation biology, environmental management (Martin et al., 2012) and biosecurity (Burgman et al., 2011). In such cases, managers and other decision-makers have little option but to rely on the opinion of experts (Sutherland, 2006). Increasing use of expert judgement (Martin et al., 2012) has stimulated the development of elicitation methods that aim to address recognised problems and thus improve the quality of judgement obtained from experts (Sutherland and Burgman, 2015). Structured expert elicitation is increasingly recognised as a valid approach in data-poor settings (Cooke, 2013).

1.4.1 On Expert judgement

The use of expert judgement is not without its difficulties. Individuals are known to be subject to a number of psychological frailties and cognitive biases, including the following:

- Framing — The way a question is presented or framed can influence the response received (Tversky and Kahneman, 1981; Plous, 1993).
- Availability — Judgements of probability may be influenced by the ease with which an event is recalled. Events may be judged to have a higher probability of occurrence than is really the case if they are recent, invoke strong emotions or have been widely reported (Tversky and Kahneman, 1973; Plous, 1993; Gigerenzer, 2004).
- Anchoring and adjustment — The tendency to base quantitative estimates on values that been previously suggested or estimated (Tversky and Kahneman, 1974) coupled with an inability to adjust sufficiently far from the anchor (Epley and Gilovich, 2006).
- Overconfidence — When experts have unwarranted confidence in their own judgments (Oskamp, 1965).

Using a group of experts instead of just one individual draws on a wider range of experience than could be achieved with just a single individual and avoids judgements that reflect only the cognitive biases of a single individual. An early example of what Surowiecki (2005) has labelled “the wisdom of crowds” dates back to an agricultural show in the early 1900s when 787 entries in a competition to judge the dressed weight of an ox were summarised by Galton (1907). The median of the entries was less than 1% from the true weight of 1198 lb.

Of course, group assessments are also subject to their own suite of biases, including:

- Group think — Judgements are unduly affected by a desire for agreement within the group (Janis, 1971).
- Dominance — Individual experts are unduly influenced by the views of a senior or dominant group member (Maier, 1967).
- Halo effects — Perceptions of the opinion of one expert are influenced by the perception of attributes of that expert unrelated to the subject under consideration (Thorndike, 1920; Nisbett and Wilson, 1977).

The use of structured expert elicitation can minimize the impacts of these biases.

1.4.2 Structured expert elicitation

If expert judgement is to be used to inform decision-making, then clearly it is desirable that the judgement be of the best possible quality. Quality in this context has been defined in terms of calibration between judgements and reality (O’Hagan et al., 2006) and the degree of precision and confidence associated with an estimate (Cooke, 1991). Some key elements of structured expert elicitation are described below:

[6] Ornamental Fish Import Reform — Health Monitoring Program

- The Delphi technique was designed to obtain a "most reliable consensus of opinion of a group of experts" (Dalkey and Helmer, 1963). It was developed at the RAND Corporation in 1950s as a predictive tool for military purposes. More recently, it has been widely used for expert elicitation in some disciplines such as medicine and social policy, but is used rather less in ecology and conservation science (Mukherjee et al., 2015).

  A standard Delphi assessment involves two or more rounds of estimates from a group of experts. Between rounds, an anonymous summary of judgements and the reasoning behind the judgements are provided to participants, who are then able to revise their estimates if they wish (Linstone and Turoff, 1975). The private nature of judgements (e.g., completion of a questionnaire) aims to reduce dominance and halo effects; individuals are not directly influenced by others in the group they might perceive as more expert.

- Questions asked of experts can be structured to minimise the experts' overconfidence. Soll and Klayman (2004) proposed a three-step format for estimates of quantities (lower, higher, mid-point for researcher-defined level of confidence) which reduced overconfidence compared to that associated with eliciting ranges (intervals) or point estimates.

  Speirs-Bridge et al. (2010) extended the format to four steps with the addition of a measure of confidence assigned by the expert. Providing lower and upper bounds separately (Steps 1 and 2) encourages experts to consider different lines of evidence before moving on to make their best estimate (Step 3). Because people are generally better at evaluating intervals than producing them (Teigen and Jørgensen, 2005), the assigning of their own level of confidence to their interval (Step 4) offers experts an opportunity to review that interval. Calculation of 'derived' intervals at a common level of confidence (generally 80%) is then necessary for the purposes of comparing and combining the intervals of individual experts (Speirs-Bridge et al., 2010).

- Not all group judgements will be accurate and well calibrated with the truth (Krinitzsky, 1993). Surowiecki (2005) identified four conditions that characterise 'wise crowds' able to generate good group judgements, namely: (i) diversity of opinion, (ii) independence, (iii) decentralisation (individuals draw on their own local knowledge) and (iv) aggregation (having a suitable means to generate a group judgement from multiple individual estimates). Where there are calibration questions to provide a measure of the performance of individual experts, the responses of experts might then be weighted to improve the quality of group judgements (Cooke, 1991).

1.4.3 Modified Delphi-style elicitation

Researchers associated with the Centre of Excellence for Biosecurity Risk Analysis (CEBRA; formerly the Australian Centre of Excellence for Risk Analysis, ACERA) developed a Delphi-style elicitation approach of private judgement / group discussion / private judgement that allowed for the sharing of evidence while also addressing some cognitive biases known to have a negative effect on group judgements (ACERA, 2010). The discussion phase is more direct than with the standard Delphi method. While summarised data are still presented free of personal identifiers, participants are able to communicate with one another, for example, via group emails or in a workshop setting. This provides an enhanced opportunity to test the credibility of particular estimates and the logic behind them. CEBRA's Delphi-style elicitation process also differs from the standard Delphi in not requiring consensus among the experts. Emphasis is placed on minimising language-based uncertainty (Carey and Burgman, 2008) and identifying any assumptions underlying the questions asked of the experts. Genuine differences of opinion that remain after discussion are acknowledged and incorporated in the final outcome, a key aim being to quantify the extent of uncertainty. Various aspects of CEBRA's approach to structured expert elicitation have been tested experimentally, with the following results:

- In experimental studies where the truth was known to researchers but not to study subjects, group judgements were generally found to perform as well as or outperform even the best-regarded individuals in terms of accuracy and calibration with the truth. Domains tested were the prediction of geopolitical events (Wintle et al., 2012) and questions of human health, biosecurity or ecology (Burgman et al., 2011).

- The accuracy of group averages improved substantially after the discussion stage in six separate workshops dealing with questions of human health, biosecurity or ecology (Burgman et al., 2011).
- No consistent relationships were found between elicitation performance and years of experience, publication record, self-assessment of expertise or any other demographic measure of expert status in questions of public health (Burgman et al., 2011), biosecurity (Burgman et al., 2011; McBride et al., 2012), ecology (Burgman et al., 2011; McBride et al., 2012), or geopolitical events (Wintle et al., 2012). This was in contrast to the initial expectations of the participating experts in the studies of Burgman et al. (2011).
- The four-point elicitation method of Speirs-Bridge et al. (2010) demonstrably reduced expert overconfidence when tested in the domains of ecology (Speirs-Bridge et al., 2010; McBride et al., 2012), and epidemiology (Speirs-Bridge et al., 2010).

1.4.4 Applications of expert elicitation

The use of quantitative data generated by expert elicitation is not without controversy. There are concerns that such data may be "biased, poorly calibrated, or self-serving" (Martin et al., 2012). Sutherland and Burgman (2015) emphasised that the quality of expert-generated data should be considered before using them in subsequent analyses or to inform decision-making. These authors go on to point out some of the strategies that can be employed to improve the quality of expert judgements, including the use of groups and a structured approach to the elicitation.

Some recent applications of quantitative data obtained by structured expert elicitation are outlined below:

- Expert opinion contributed to the development of a surveillance system for black rats on an island of high conservation value (Jarrad et al., 2011). The experts estimated parameters where data were lacking for the model used to optimise surveillance.
- Experts in fish health were used to establish likelihood ratios for risk factors that might influence the occurrence of infectious salmon anemia in Chilean salmon farms (Gustafson et al., 2014).
- About 20 experts of diverse backgrounds and affiliations participated in an elicitation exercise on the effects of collisions on North Atlantic right whales (Fleishman et al., 2016). The experts first contributed to the development of a conceptual model of collision effects, then provided estimates of the principal parameters needed to populate the model.
- A panel of 15 active koala researchers participated in a structured expert elicitation to estimate the total koala population of eastern and south-eastern Australia and changes in populations within different bioregions (Adams-Hosking et al., 2016). The entire elicitation process took six months, including a four-day workshop for the discussion phase.

1.4.5 Usage in current project

We used the IDEA structured expert judgement protocol to elicit estimates of the parameters needed to run the model. The question topics are presented in Appendix A. We needed answers to eight preliminary questions (Appendix B) and then 18 parameter questions (Appendix C and Appendix D). The parameter questions were structured as follows.

- Pre-Border / Border (12 questions)
  - Six approach rates (low/high risk on three pathways) (Table 2.1);
  - Three intervention impacts (Table 2.2);
  - Three attenuation rates (incursion → establishment) (Table 2.6);
- Post-Border (6 questions)
  - Two expected incursion size at detection (with/ without surveillance) (Table 3.2);

– Two cost of eradication at given size (ready / not ready) (Table 3.3); and

– Two success probabilities of eradication at given size (ready / not ready) (Table 3.4).

The question that we put to the experts was unfortunately ambiguous: we asked for their best guess. It is not clear whether the experts intended their response to be interpreted as a mean, a mode, or a median. Upon subsequent e-mail follow-up, we established that the experts interpreted the best guess as being the *mode* with one exception, who preferred the *mean*. The purpose of this report is demonstration rather than to guide policy, so we retained the mode in all cases.

Handling between-expert variation

The results from each expert vary considerably from one another, as can easily be verified in the graphical summaries (Appendix C). In order to come up with an overall model, it is necessary to find an average of the expert results. In this project we have followed the adage: *aggregate late*, meaning that we have computed the entire model for each expert and only taken averages of the key reporting statistics. This strategy is as opposed to taking averages of all the expert parameters and then running the model with those values.

The reason that we average the expert outputs (as opposed to the inputs) is that averaging the inputs creates bias, according to Jensen's Inequality (Jensen, 1906). Briefly, the mean of a curved function of a distribution is not the same as the curved function of the mean of the distribution. Therefore the expert opinions will be better reflected if we average the outputs than if we average and use their inputs.

An unwelcome side-effect is that the averages that we report do not necessarily align in expected ways. For example, we might expect that if the average cost of an incursion is $7.5 million and an intervention saves one incursion per year then the value of the intervention will be $7.5 million per year. However, if expert A's results are that the cost is $10 million and the saving is 1.5 incursions whereas expert B thinks that the cost is $5 million and the saving is 0.5 incursions then the expected cost is $7.5 million and the expected saving is 1 incursion per year — but the value is estimated at the average of $15 million and $2.5 million, which is $8.75 million — considerably higher than $7.5 million, which is the product of the averages.

1.4.6 Missing values

We used multiple imputation to allow for the modest number of missing values in the data. The algorithm used was multiple imputation via chained equations (MICE) as provided by the 'mice' package from R (van Buuren and Groothuis-Oudshoorn, 2011). We applied the algorithm by choosing a number of imputation instances (here, 5), generating 5 candidate fill datasets that differed in only the predictions of the missing values as obtained from the imputation algorithm, and split the simulations evenly among these 5 imputed instances. This provided 10,000 randomly simulated databases with the imputed values salted appropriately throughout, which in turn provided a realistic view of the uncertainty arising as result of the missing values.

Chapter 2

Pre-border and Border Sub-model

The Pre-border and Border sub-model comprises the following components:

- Raw exposure — the rate per 1000 at which units that are arriving on the pathway are contaminated with GM (Section 2.1.1);
- Interventions — the activities that are performed or mandated by MPI to reduce the exposure and therefore the likelihood of incursion (Section 2.1.2);
- Attenuation from incursion to establishment — propagules that arrive past the border may still fail to establish (Section 2.3);

In this chapter, the average expert-elicited values are provided in tables 2.1 and 2.2. The tables are constructed as follows. The best guesses, lowest and highest values, are all averaged across the all experts who responded to the question. A link is provided to each summary graph in the Appendix. For example, the summary graph for Question 9a in Table 2.1, which concerns the raw exposure to GM via high-risk ships, is in Figure C.1 and can be accessed by the hyperlink in the table.

The question numbering starts at 9 because there were eight preliminary questions that were used to prepare the framework.

The outcome of the Pre-border and Border sub-model is the expected number of establishments per year. In a simulation setting, the outcome of the sub-model is the simulated number of establishments per year. The values reported in this chapter are indicative; as mentioned earlier we construct end-to-end simulations using random draws from the elicited parameter distributions from the individual experts, and take the mean of the predicted costs.

2.1 Pre-Border Model Components

2.1.1 Raw exposure on pathways

The six pathways were identified based on interception records and verified with the experts. The three base pathways were ships, shipping containers, and used vehicles and machinery. Each pathway includes activity from high-risk and low-risk countries.

2.1.2 Interventions

The impact of interventions is reported in Table 2.2. It is not known how many high-risk vehicles undergo offshore certification and inspection as opposed to heat treatment. We used heat treatment as the default treatment across the pathway.

2.2 Pathway risk model

In this section we use the elicited values from the previous sections (namely, Tables 2.1 and 2.2) along with border data provided by MPI to compute the exposure along the different pathways that we have identified.

Table 2.1: Question 9: Expert-elicited Raw Exposure; number of units contaminated per 1000 in the absence of any biosecurity effort. Results are averages of expert-elicited values.

Question	Summary	Best	Low	High	Graph
9a	High-risk Ships	185.6	17.6	420.0	C.1
9b	Low-risk Ships	1.6	0.0	11.0	C.2
9c	High-risk Containers	310.0	21.6	552.0	C.3
9d	Low-risk Containers	2.6	0.0	20.0	C.4
9e	High-risk V'cles & Mach.	238.0	19.0	484.0	C.5
9f	Low-risk V'cles & Mach.	4.6	0.2	28.0	C.6

Table 2.2: Question 10: Expert-elicited Impacts of interventions; number of units free from contamination after intervention on 1000 contaminated units. Results are averages of expert-elicited values.

Question	Summary	Best	Low	High	Graph
10a	Offshore Cert & Insp. of HR Ships	346.0	40.0	570.0	C.7
10b	Offshore Cert & Insp. of HR V'cles & Mach.	710.0	218.0	916.6	C.8
10c	Heat Treatment of HR V'cles & Mach.	938.0	774.0	996.0	C.9
10d	Inspection of HR Containers	732.4	441.2	924.6	C.10

We can then compare the implications of the expert elicited values against observations from MPI and other sources.

In order to match the pathways to the interventions we have to identify seven pathways that are the pathways in Table 2.1 with the high-risk vehicles and machinery split into two sub-pathways: inspection and certification and heat treatment (note Table 2.2).

Table 2.3: Pathway-level exposure before and after intervention. *Approach* and *Impact* are percentages of pathway units that are contaminated and effect of interventions upon contaminated units, respectively. *Volume*, *Reduction* and *Residual* are numbers of arriving units, intercepted contaminated units, and incursions (unintercepted contaminated units) per year.

Path	Volume	Approach	Intervention	Impact	Reduction	Residual
Low-risk Containers	770900	0.26		0	0.0	2004.3
High-risk Containers	100	31.00	Inspection	73	22.7	8.3
Low-risk Ships	1918	0.16		0	0.0	3.1
High-risk Ships	658	18.56	Offshore Cert & Insp.	35	42.3	79.9
Low-risk V'cles & Mach.	104009	0.46		0	0.0	478.4
High-risk V'cles & Mach.	189323	23.80	Heat Treatment	94	42265.2	2793.7

2.2.1 Pathway risk model check using border interceptions

We can use MPI data to assess some aspects of the expert-elicited model as a procedure for external validation. Table 2.4 reports border interception data for high-risk vehicles and machinery. The historical interception rate from inspections of high-risk vehicles and machinery was about 100 per year (Table 2.4). Surprisingly, the contemporary number of high-risk vehicles and machinery consignments is approximately the same as the years from which these interception counts are drawn.

In order to compare these data with our expert-elicited values, we make the following ball-park assumptions.

Table 2.4: MPI data: AGM Interception data on vehicles by year.

Country	1999	2000	2001	2002	2003	2004
Japan	114	196	104	58	25	13

1. The average contamination rate of vehicles and machinery now is about the same as 15–20 years ago.
2. The expected effectiveness of inspection of vehicles and machinery is about the same as external inspection of shipping containers (close to 75%).

Then the counterfactual expert-elicited interception rate for HR vehicle and machinery inspection would be $189,323 \times 23.8\% \times 0.7324$ which is 33001 — **much higher than 100**, which is the ball-park figure based on the historical results.

This comparison suggests that (i) the raw exposure is much too high (Table 2.1), or (ii) the estimated quality (*effectiveness*) of inspection is much too high (Table 2.2). **Consequently we reduce the approach rate for all pathways by a factor of 400.**

This response is an unsatisfactory element of the modeling process that arose because the regulatory data was available only after the SEJ workshops were carried out. We could also or instead have altered (we resile from writing 'corrected'!) the efficacy of the inspection, and at the point in the analysis the difference between them seems arbitrary. However, we believed the efficacy to be better grounded than the approach rate, which we considered to have been blighted by anchoring.[1] That said, as reviewer CP pointed out, the selected inspection efficacy is substantially higher than reported in Work et al. (2005) and Eyre et al. (2018).

This correction results in a revised version of Table 2.3, presented in Table 2.5. If more border data become available then a more convincing comparison might be possible. Note that this correction is based on the averaging of inputs. Its effect will be altered at the point of assessing the model outputs (See Section 1.4.5).

Table 2.5: Pathway-level exposure before and after intervention, with correction. *Approach* and *Impact* are percentages of pathway units that are contaminated and effect of interventions upon contaminated units, respectively. *Volume* and *Border* are numbers of arrival and incursions per year.

Path	Volume	Approach	Intervention	Impact	Reduction	Residual
Low-risk Containers	770900	0.00065		0	0.0000	5.0109
High-risk Containers	100	0.07750	Inspection	73	0.0568	0.0207
Low-risk Ships	1918	0.00040		0	0.0000	0.0077
High-risk Ships	658	0.04640	Offshore Cert & Insp.	35	0.1056	0.1997
Low-risk V'cles & Mach.	104009	0.00115		0	0.0000	1.1961
High-risk V'cles & Mach.	189323	0.05950	Heat Treatment	94	105.6631	6.9841

2.3 Attenuation component

We expect that not all incursions will result in establishment. A number of incursions will fade away, some will fail to find host material, some will arrive in locations with unsuitable climate, and so on. The expert-elicited values for this attenuation are presented in Table 2.6.

[1]To justify this claim: the motivating questions were such as: "Consider high-risk used vehicles and machinery. Imagine 1000 random, representative units of used vehicles and machinery arriving in New Zealand without any biosecurity measures in place, that is, the raw, unrefined exposure. Think about the number of units of used vehicles and machinery that could be carrying at least one egg mass." Questions are asked in this manner because it is commonly held to be easier to think about integer numbers of real events than it is to think about proportions. We believe that the choice of the number of arriving units (here, 1000) anchors the expert in the sense that it is cognitively tricky to push into sub-units. Hence, other than 0, 1 is the smallest reasonable number (no expert answered in fractions or decimals for any of the questions). Therefore we contend that the way we asked the question influenced the expert response.

Table 2.6: Question 11: Establishment Rate; number of establishments arising from 1000 arrivals. Results are averages of expert-elicited values.

Question	Summary	Best	Low	High	Graph
11a	Establishment from Ships	37.0	2.0	186.2	C.11
11b	Establishment from Containers	111.6	4.6	352.0	C.12
11c	Establishment from V'cles & Mach.	132.0	5.2	393.0	C.13

2.3.1 Pathway risk model check using detected establishments

We can then compute the expected number of establishments overall and by pathway, using the developed model and the elicited parameter estimates. This is done in Table 2.7. The expected number of established incursions per year across all three conveyances is the sum of the final column: 1.649. **This is much higher than the number of recorded establishments according to MPI records, which is 1 in 20 years**.

Table 2.7: Pathway-level post-border annual establishment rates. *Establishment* is the percentage of establishments per incursion. *Established* is the expected number of establishments per year.

Path	Intervention	Residual	Establishment	Established
Low-risk Containers		5.0109	11.2	0.55921
High-risk Containers	Inspection	0.0207	11.2	0.00231
Low-risk Ships		0.0077	3.7	0.00028
High-risk Ships	Offshore Cert & Insp.	0.1997	3.7	0.00739
Low-risk V'cles & Mach.		1.1961	13.2	0.15789
High-risk V'cles & Mach.	Heat Treatment	6.9841	13.2	0.92190

This result suggests that (i) the raw exposure is too high (Table 2.1), (ii) the intervention impacts are too low (Table 2.2), (iii) the attenuation from incursion to establishment is insufficient (Table 2.6), (iv) that many establishments go undetected, or (v) some combination of these factors holds. **Consequently we reduce the establishment rate for all pathways by a factor of 100.**

This correction results in a revised version of Table 2.7, presented in Table 2.8. The expected number of established incursions per year across all three conveyances is the sum of the final column: 0.01649. This translates to an approximate rate of an establishment every 61 years. Note that, as above, this correction is based on the averaging of expert inputs. Its effect will be altered at the point of assessing the model outputs (See Section 1.4.5).

Table 2.8: Pathway-level post-border annual establishment rates, corrected. *Establishment* is the percentage of establishments per incursion. *Established* is the expected number of establishments per year.

Path	Intervention	Residual	Establishment	Established
Low-risk Containers		5.0109	0.112	5.6e-03
High-risk Containers	Inspection	0.0207	0.112	2.3e-05
Low-risk Ships		0.0077	0.037	2.8e-06
High-risk Ships	Offshore Cert & Insp.	0.1997	0.037	7.4e-05
Low-risk V'cles & Mach.		1.1961	0.132	1.6e-03
High-risk V'cles & Mach.	Heat Treatment	6.9841	0.132	9.2e-03

Chapter 3

Post-Border Sub-model

The Post-border sub-model focuses on the effect of post-border activities upon the consequences of incursion. It comprises the following components:

- Detection and delimitation of the established pest — this is assumed to happen sooner or later, so the unknown variable is the area infested at time of delimitation (Section 3.2); and
- Eradication of the population, the success probability and cost of which depend on the (unknown) area infested at time of delimitation.

3.1 GERDA data

We can assess the quality of the model by comparing some key outcomes with relevant data extracted from the research database GERDA (Kean et al., 2019). Here, the relevant data are all records of responses to Asian or European GM. Table 3.1 shows the number of reported incidents by GM sub-species and reporting country.

Table 3.1: GERDA data: number of reported incidents by subspecies and country.

Sub-species	CA	GB	NZ	US
Asian gypsy moth	1	0	0	17
European gypsy moth	2	1	0	71
Hokkaido gypsy moth, Dosanko gypsy moth	0	0	1	0

3.2 Incursion size at time of detection

Table 3.2 provides the average across the experts of the best estimate and the lower and upper limits for the answer to the question of the size at detection of the infesting population.

Table 3.2: Question 12: Incursion size at time of Detection (ha). Results are averages of expert-elicited values.

Question	Summary	Best	Low	High	Graph
12a	Incursion Size at Detection; Current Surveillance	1344.6	301.2	28080.0	C.14
12b	Incursion Size at Detection; No Surveillance	17380.0	2550.2	2055000.0	C.15

We assume that the PDF of the incursion size at the time of its detection is Weibull with shape parameter equal to 2. Following our earlier work and Table 3.2, and interpreting the best guess as the mode (which is plausible but not confirmed), we obtain an *average* incursion size distribution for the current surveillance system as Weibull($\alpha = 2, \beta = 1517.22$), and for the counterfactual of no surveillance system, Weibull($\alpha = 2, \beta = 19611.2$)[1]. The aggregated distributions, as used in the simulation experiment, are presented in Figure 3.1, along with the distribution from GERDA.

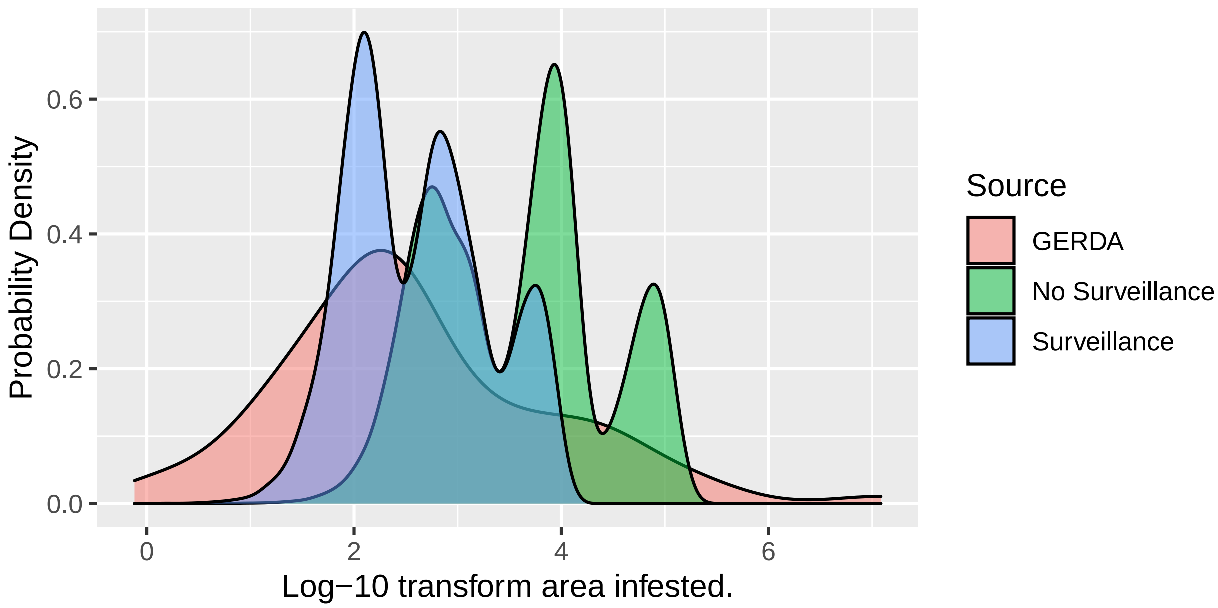

Figure 3.1: Density estimate of log (base 10) transformed areas infested at time of detection *or* maximum area treated from GERDA, overlaid with aggregated Weibull densities elicited from each expert.

3.2.1 Sub-model check using GERDA data

We can compare the elicited expected area infested at time of delimitation with data derived from the GERDA database. Figure 3.1 provides a smooth density plot of the base-10 logarithm of the estimated area of the incursion when delimited. The expert-elicited values are shown as a blue density under current surveillance and green density for no surveillance.

The distributions of the GERDA values and the expert-elicited values have only modest alignment. The larger number of smaller detected incursions in GERDA can potentially be explained by two factors. First, most of the values from GERDA are from the USA and are for European GM populations, whereas the experts were asked about both Asian and European GM populations. EGM females cannot fly so their spread is slower, as a rule. Second, the experts were asked about the extent of the invaded area, meaning the number of hectares occupied by the incursion. The GERDA data typically cover the maximum treated area, which under some circumstances could be smaller than the population, for example because the invading populations may have outlying individuals that are too sparse to produce a satellite population. Alternatively, the maximum treated area may be larger than the area occupied by the invading population in order to allow for a buffer.

3.3 Estimated cost of eradication

The average for the four experts who estimated the cost of response for a 500 ha incursion (that is, the incursion is delimited when the population size is 500 ha) was 3.25 million $NZ with the current level of

[1] *NB:* for $X \sim$ Weibull(α, β), $E(X) = \beta\Gamma(1 + 1/\alpha)$), and $\Gamma(1 + 1/2) = 1/2\sqrt{\pi} = 0.886227$.

readiness and 19.62 million $NZ in the counterfactual state of no readiness.

Table 3.3: Question 13: Eradication cost if the incursion is detected at size 500 ha. ($million). Results are averages of expert-elicited values.

Question	Summary	Best	Low	High	Graph
13a	Cost of Eradication with Readiness	3.2	1.0	12.4	C.16
13b	Cost of Eradication without Readiness	19.6	4.0	40.5	C.17

3.3.1 Eradication cost check using GERDA data

To compare the costs of different eradication programmes worldwide these have been converted to a common currency: the 2015 United States dollar (USD). We converted these to contemporary NZ currency by multiplying by 1.522.

Figure 3.2 provides GERDA data that shows the estimated eradication cost as a function of infestation size, with maximum area treated used when infestation size is missing. The green crosses show the expert-elicited cost for eradicating a 500-ha infestation under current readiness systems, and the red crosses show the cost under the counter-factual of no readiness. The means are represented by dots. Both points are within the span of the data but at the higher end. To use the expert data, we scale the fitted function by increasing the intercept (in the log scale) so that the fitted line passes through the green or red dot.

We noted above that we model the infestation size but the GERDA data includes infestation size and maximum area treated, which we would expect to be smaller. We decided to include these data because an incursion response in NZ is likely to be conservative in any case, and will likely treat a larger area than would be treated in the USA, which is the source of most of the GERDA data.

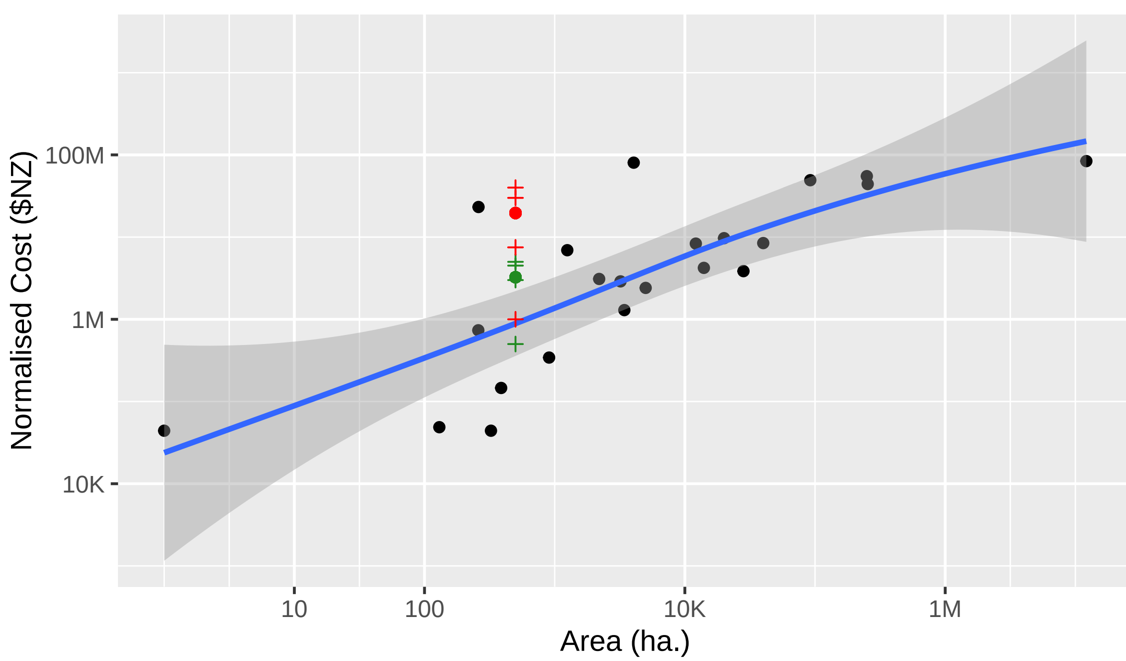

Figure 3.2: GERDA normalised cost data for all GM records with expert-elicited values for the cost of eradication of an invading population of size 500 ha. under current readiness (green cross) and the counterfactual of no readiness (red cross).

3.3.2 Eradication cost model

We constructed a model that predicts the committed cost of eradication as a function of the area infested at time of delimitation ($\$_E(x)$) as follows. We fit a simple allometric function to the GERDA data,

$$\log_e y_i = \beta_0 + \beta_1 \log_e x_i + \epsilon_i \tag{3.1}$$

where y_i was the normalised cost of eradication, x_i was the area initially infested *or* maximum area treated, and $\epsilon_i \sim N(0, \sigma^2)$ were independent errors.

We combined this model with the expert-elicited cost of eradication as follows. The experts provided a single value for $x = 500$ (ha). We used a ratio to scale the GERDA-predicted value at 500 ha to match the expert-predicted value.

$$\$_E(x) = \exp(\hat{\beta}_0 + \hat{\beta}_1 \log_e x) \frac{\hat{\$}_{500}}{\exp(\hat{\beta}_0 + \hat{\beta}_1 \log_e 500)} \tag{3.2}$$

where $\hat{\$}_{500}$ was the average expert elicited value (Table 3.3).

In this way the GERDA data provides the 'shape' of the function and the expert values are used to 'pin' the shape.

3.4 Estimated probability of eradication success

The experts estimated the success probability of response for a 500 ha incursion (that is, the incursion is detected at size 500 ha) as 86.9 with the current level of readiness and 77.4 in the counterfactual state of no readiness.

Table 3.4: Question 14: Eradication success probability if the incursion is detected at size 500 ha. Results are averages of expert-elicited values.

Question	Summary	Best	Low	High	Graph
14a	Success Prob. of Eradication with Readiness	86.9	56.0	100.0	C.18
14b	Success Prob. of Eradication without Readiness	77.4	40.0	95.0	C.19

3.4.1 Eradication success check using GERDA data

Figure 3.3 shows the eradication success probability estimated using GERDA data. As above, the green cross shows the expert-elicited probability under current readiness systems, and the red cross shows the probability under the counter-factual of no readiness for an invading population of size 500 ha. These expert-elicited results align very well with the GERDA data. Unfortunately, the fitted model from which we drew random combinations of slope and intercept (for the purposes of capturing the variability) had wide confidence intervals on the slope. This resulted in some random variants giving an increasing probability of eradication success with increasing area of invasive population at delimitation, which is counter-intuitive. We reduced the scale of the uncertainty to ensure that all instances reflected a negative slope.

3.4.2 Eradication success model

We constructed a model that predicts the success probability of eradication as a function of the area infested at time of delimitation ($P_E(x)$) as follows. We fit a simple generalised linear model to the GERDA data,

$$o_i \sim \text{Bernoulli}(p_i) \tag{3.3}$$

$$p_i = \text{expit}(\delta_0 + \delta_1 \log_e x_i) \tag{3.4}$$

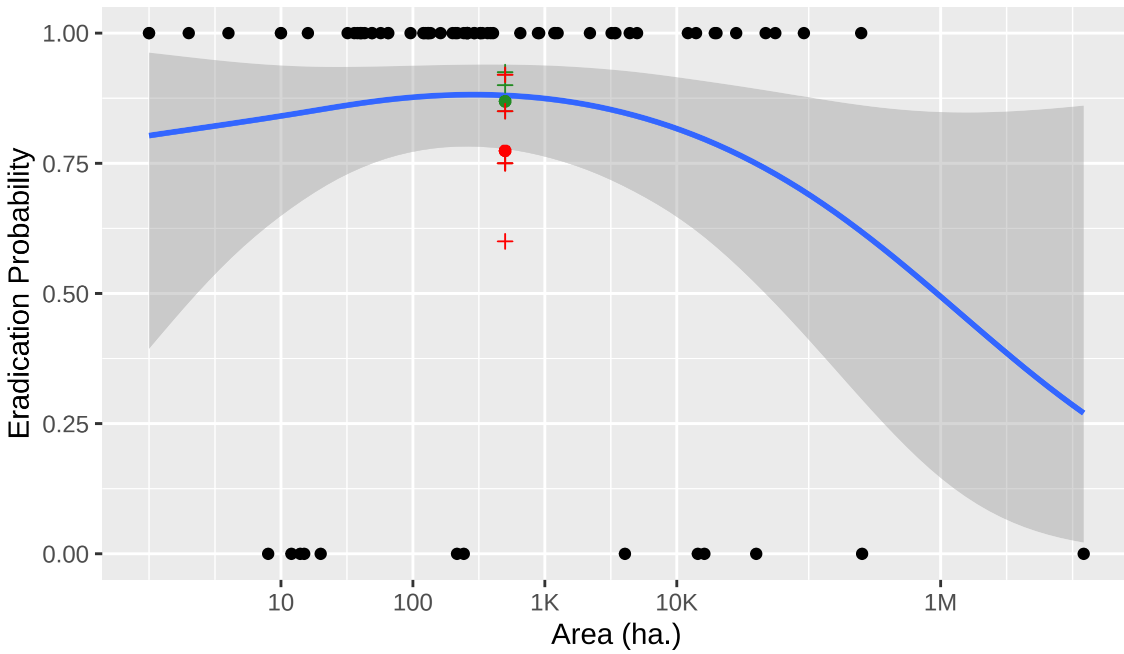

Figure 3.3: GERDA success probability data for all GM records with expert-elicited values of the probability of success of eradication for an invasive population occupying 500 ha. under current readiness (green cross) and the counterfactual of no readiness (red cross).

where o_i was the outcome of eradication (success = 1, failure = 0), x_i was the area initially infested *or* maximum area treated, and expit was the reverse logit function ($\text{expit}(x) = 1/(1 + \exp x)$).

We combined this model with the expert-elicited cost of eradication as follows. The experts provided a single value for $x = 500$ (ha). We used a ratio to scale the GERDA-predicted value at 500 ha to match the expert-predicted value.

$$P_E(x) = \text{expit}(\hat{\delta}_0 + \hat{\delta}_1 \log_e x) \frac{\hat{P}_{500}}{\text{expit}(\hat{\delta}_0 + \hat{\delta}_1 \log_e 500)} \tag{3.5}$$

where $\hat{P}_{500}$ was the average expert elicited value (Table 3.3).

In this way the GERDA data provides the 'shape' of the function and the expert values are used to 'pin' the shape.

Subsequent to the expert elicitation exercise, further discussion with the expert cohort led us to speculate that the problem of eradication of GM might be sufficiently well known in the modern era that eradication might be considered inevitable — that is, eradication would succeed with probability 1 regardless of the size at which the incursion is detected, with the current level of readiness. However, as reviewer CP noted, various future events could impede or frustrate an eradication response, including loss of social license to: apply pesticides (biological or chemical) in urban or natural environments; or widespread use pheromones in high density trapping grids or for mating disruption (which was recently prevented in California by public outcry). Success probabilities also seem partly dependent on the GM subspecies being targeted because some of their behaviours, including dispersal, differ.

3.5 Cost of management if eradication fails

Gypsy-moth incursions present a serious threat to the New Zealand forestry industry (see, e.g., Castedo-Dorado et al., 2016). Ross (2005) reported an upper-end estimate of $88 million annually ("It defoliates trees and New Zealand perceives it as a very serious threat to our forest and horticultural industries, indigenous flora, ornamental trees and urban parks. A recently completed Economic Impact Assessment estimated that once established damage could amount to $88 million per annum at the high end”). Alternatively, Harris Consulting (2003) gave a range of estimates of net present values of impacts from $5 million to about $399 million, in 2003, with 46 as the midpoint. We updated these estimates to current New Zealand dollars by multiplying by 1.39. The model assumes that GM’s national impacts occur immediately upon failure or refusal of the eradication attempt.

Chapter 4

Measuring and Tracking Uncertainty

4.1 Measuring Uncertainty

Uncertainty is inevitable in predicting future impacts of different activities against counter-factual scenarios. In order to most faithfully represent reasonable uncertainty about the overall outcomes, we captured the following sources of uncertainty:

1. Expert elicited estimates using triangular distributions (see Appendix C);
2. GERDA-based model estimates of eradication cost (see Section 3.3.2);
3. GERDA-based model estimates of success probability (see Section 3.4.2);
4. The area infested at time of delimitation using the Weibull distribution (see Section 3.2); and
5. Economic impacts of failure to eradicate using a triangular distribution (see Section 3.5).

In each expert-elicited case, we interpreted the best guess as being the *mode* of the triangular distribution and then scaled the uncertainty linearly to most closely match the experts stated uncertainty. When such scaling would have taken a random instance above or below a fixed biological or process limit, we then swapped a uniform random number between the set limits. (The alternative was to recast the triangular distribution so that the logical limits were the same as the realized limits, which would be less conservative).

4.2 Tracking Uncertainty

Simulations were then carried out by selecting random instances for all the needed parameters, then computing the costs using the algorithms as reported in the preceding chapters. A single incursion instance was generated that was then used for all of the four counter-factual scenarios in order to remove incursion-to-incursion variation from the comparison. Scenarios that randomly resulted in unreasonable outcomes — for example, an incursion being less expensive to eradicate under a scenario of no readiness than under a scenario of readiness — were dropped. Future work will provide a more elegant preclusion mechanism.

Chapter 5

Results and Discussion

The results reported in this chapter will differ systematically from those summarized above because these results account for the uncertainty in the data and between and within experts.

5.1 Establishment Rate

The expected annual spacing of an establishment with the current biosecurity system is 6.5–2089.7 years[1], with best guess 18.7. This value is computed by inverting the distribution of the simulated establishment rates and determining the 5th and 95th quantiles.

5.2 Value of intervention

This section reports the effects of the three interventions. Under the current system, the expected cost of an incursion is 9.1–93.0 NZ million, with best guess 40.1 million. Each of the interventions provided in Question 10 (Table 2.2) averts this cost by reducing the establishment rate of the pest. We summarized the estimated benefits (in terms of damages avoided) for each of the interventions in Table 5.1. This table shows us, for example, that heat treatment of about 190,000 units of high-risk vehicles and machinery entering every year saves an establishment every four years. The standard deviations of the simulations are not shown here, but the (considerable) variability of the outcomes can be assessed by reference to Figure 5.1.

In reading Table 5.1 it is important to keep in mind that we have computed the expected savings according to each expert and then computed the average. This means that we cannot expect that the results in the Table will necessarily be internally consistent. See Section 1.4.5 for an explanation and an example.

Table 5.1: Pathway-level intervention outcomes, with correction. *Approach* and *Impact* are percentages of pathway units that are contaminated and effect of interventions upon contaminated units, respectively. *Volume* and *Established* are numbers of arrival and establishments per year. *Reduction* is the annual number of pests intercepted by the intervention. *Saved* is the reduction in the number of incursions, and *Saved.MNZD* is in millions of NZ dollars.

Pathway	Volume	Approach	Impact	Reduction	Established	Saved	Saved.MNZD
High-risk Containers	100	0.073	69.3	0.05	0.0000	0.00	0.00
High-risk Ships	658	0.053	31.7	0.11	0.0000	0.00	0.00
High-risk V'cles & Mach.	189323	0.062	89.9	104.69	0.0026	0.26	8.12
Low-risk Containers	770900	0.002	0.0	0.00	0.0022	0.00	0.00
Low-risk Ships	1918	0.001	0.0	0.00	0.0000	0.00	0.00
Low-risk V'cles & Mach.	104009	0.003	0.0	0.00	0.0005	0.00	0.00

[1] All intervals are 90% coverage

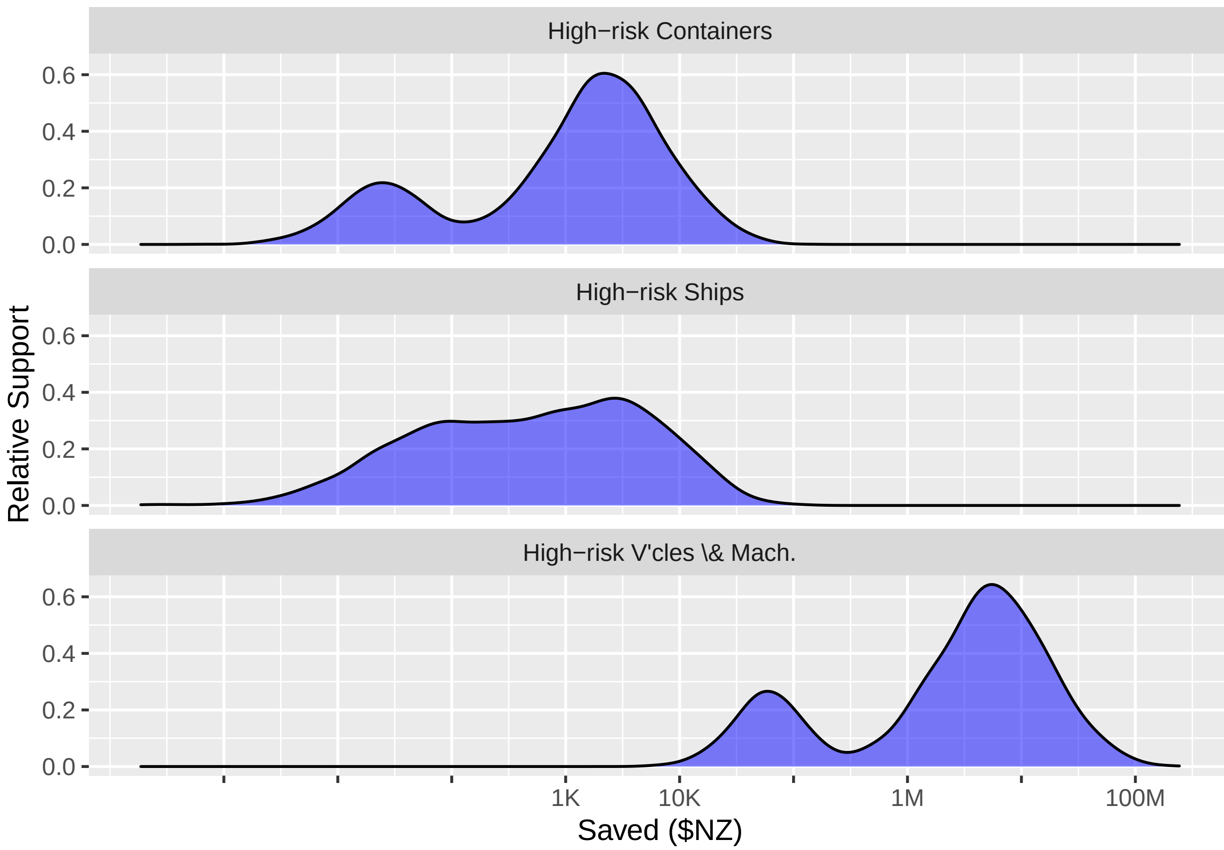

Figure 5.1: Predicted distributions of proportional value of different border biosecurity measures undertaken by MPI.

5.3 Counterfactual scenarios

This section provides the results of the comparison of the simulation model under the current situation with the counter-factuals, namely (i) no pest-specific surveillance, and (ii) no pest-specific surveillance and no pest-specific readiness. With the model as constructed, we can now assess the different outcomes of the counterfactual experiment.

The expected net present value of the cost of a GM incursion that is not eradicated is $64 million, taking account of pest management activities, updating the estimate from Harris Consulting (2003) to equivalent 2019 dollars. Table 5.2 provides the averages of the simulation results, which are presented in Figure 5.2. These results are presented on a per-incursion basis (as opposed to a per-year basis) but they show that an incursion under the current system is expected to cost about $40M (on average), and could cost up to $120M.

The additional expected cost of an incursion of a comparable pest against which there is no surveillance is about $18M, principally because without surveillance the incursion may not be detected until it has become large and expensive to eradicate.

The additional expected cost of an incursion of a comparable pest against which there is neither surveillance nor readiness is also about $18M[2] and could reach $75M (Figure 5.2), the reason here being that a late detection of the incursion creates a very expensive campaign that may in any case fail in the absence of appropriate readiness infrastructure, such as lures, traps, and the like.

[2]The apparent equivalence of the additional costs under the two counterfactual scenarios is accidental!

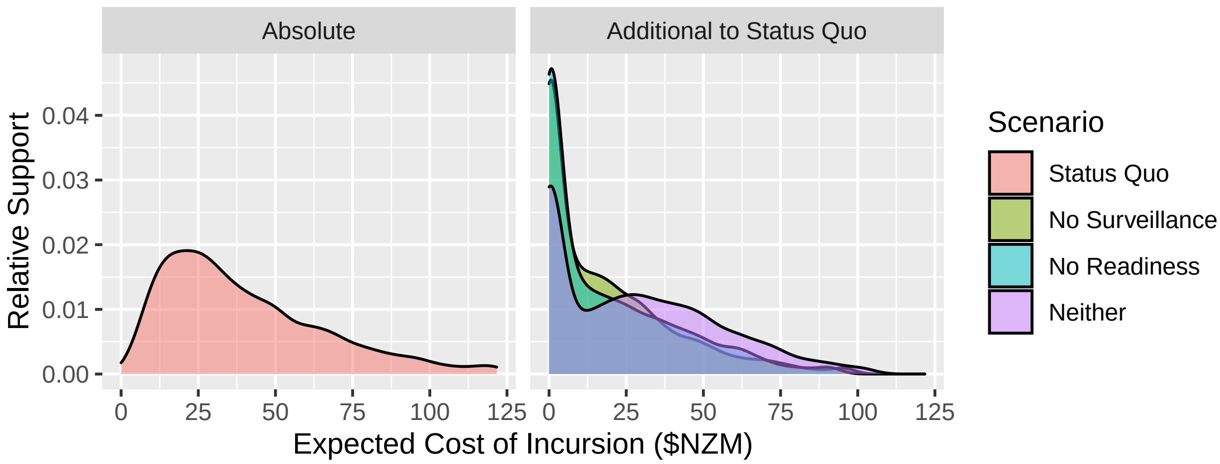

Figure 5.2: Predicted distribution of proportional cost of incursions under different counterfactual scenarios. The distributions are delimited at their 99% quantiles, creating the toe of the "boot".

Table 5.2: Total elicited and estimated cost of an incursion in $NZ (millions).

	With Surveillance	No Surveillance
With Readiness	40.11	58.07
No Readiness	58.47	68.05

5.3.1 Weighted benefits

The costs presented in Table 5.2 are per *established invasive population*, however, in order to translate these figures into more usable values, we need to scale them by the expected rate at which establishments occur. There is, after all, no value in reducing the impact of an event that will never happen. We estimate that — on average, with the current pre-border/border system in place — the annual probability that New Zealand will have a GM establishment is 0.00048–0.15449[3] (best guess is 0.053). Hence,

1. the elicited expected annual cost of GM with the current system is about 0.019–7.806 million (not including border measures) with best guess 2.05 million;
2. the annual reduction to this cost due to active surveillance is 0.00–6.06 million with best guess 1.26 million; and
3. the *joint* benefits of readiness and surveillance reduce the annual cost to that reported above by 0.00–6.62 million with best guess 1.62 million.

[3]This is the average rate after taking account of the uncertainty, so it necessarily differs from the value reported in Section 2.3.1; see Section 1.4.5 for an explanation.

Chapter 6

Conclusions and Recommendations

6.1 Caveats

We summarise the caveats that we place upon the outcomes of this modeling exercise in the following list. These caveats refer to the process of developing parameters for the model, rather than the structure of the model itself, which are provided in Section 1.2.

1. The pathway volumes that are used to compute the risk presented by each pathway are derived from different data resources and may not be directly comparable.

2. Considerable tweaking of the expert-elicited parameter estimates was required in order to align the model outcomes with observed or implied outcomes (for example, interception rate under historical inspections, and historical establishment records). In all cases this tweaking reduced the apparent threat of the pest, which would reduce the apparent value of the activities undertaken against the pest risk.

3. The estimates of eradication cost and success probability were made from models constructed using GERDA data and calibrated using expert judgement. The GERDA collection is incomplete, omitting a large number of successful eradication instances from the USA (pers. comm., Sandy Liebhold). We hope that the calibration provided by the experts will ameliorate this information gap.

4. The impacts of establishment were estimated based on Harris Consulting (2003), and then corrected to current $NZ. It is possible that the real value of the affected industries may have changed in the past 16 years. This could affect the outcomes in either direction.

6.2 Outcomes

The outcomes of the case study modeling exercise are reported in detail in Chapter 5. All of the following summary points must be read in the context of the substantial parameter and model uncertainty that arise from both the inherent unknown nature of the underlying process and the considerable number of unappetizing tweaks and corrections that were necessary to align the model predictions with the few available observations.

The conclusions of this exercise should be taken as indicative only, being representative of the kinds of conclusions that would be possible when the model fitting exercise is tightened up.[1]

Then, with these caveats, the outcomes of the model suggest that the cost of an establishment is high (e.g., $NZ 40M) but the predicted establishment rate under the current system is low (about 1 establishment every 20 years), and the historical establishment rate lower still (1 establishment in the last 40 years).

The consequence of the low inherent establishment rate is that some measures taken on low-volume risk pathways do not avert much GM impact. However, it is important to note that the impacts considered in

[1] These processes are under improvement in current CEBRA work.

this project are solely for GM; the activities considered may avert other impacts as well (for example, heat treatment of vehicles will also act against BMSB).[2] If, as proposed by the model, the high-risk vehicle and used machinery pathway represents the highest risk of entry for GM, namely 1 establishment every 4 years, and the cost of an establishment is about $40M, and if heat treatment of the 190,000 high-risk vehicle and used machinery consignments is 90% effective then it may save about $8M per year of potential GM impact. On the other hand, if, as proposed by the model, the high-risk container pathway represents a lower risk, at least in part because of its very much lower volume, then inspection of the 100 high-risk containers averts a much lower portion of the GM impacts, closer to $1K.

Another consequence of the low expected establishment rate is that the value of readiness and active surveillance in terms of averted GM impact is less; according to the model the value of each is about $18M on average per incursion, but incursions are unlikely so the expected value of both surveillance and readiness is between one and two million annually. That said, in 2019 the GM surveillance program detected an invasive poplar sawfly, demonstrating that it may have value against impacts from non-target pests (Ministry for Primary Industries, 2019).

6.3 Process

This section discusses the process of model parameterisation. If the current model is to be scaled to work for many pests then the elicitation load needs to be as light as reasonably possible. The current balance of complexity against simplicity of effort seems about right; expert elicitation takes a half-day face to face after appropriate preparation, and the questions are taxing but tractable. The exercise will prove more robust of more experts can be recruited. There is considerable literature on the value of a large and diverse pool of experts to the outcomes of expert elicitation. For this exercise we had only 5, which meant that modest differences in opinion created starkly different outcomes. A different set of experts would inevitable have led to different estimates.

An ongoing challenge is the reconciliation of the expert opinion with such regulatory data as are available. Due to timing constraints, the expert elicitation exercise took place before all the data were available, which meant that we needed to try to harmonise the results with the limited observations post-hoc, which was a clumsy and unsatisfying outcome. As the project scales up to multiple pests it will be important to ensure that all the available data are amassed before expert elicitation begins. This action is being addressed in the follow-up CEBRA project, 17062102.

6.4 Uncertainty

The uncertainty with which the model projections are made is high, with upper interval values being up to an order of magnitude higher than lower interval values. This considerable uncertainty reflects a range of contributing factors, including the uncertainty of the experts about future events, and indeed disagreements between the experts. The question as to whether policy-makers can use the model outcomes in the light of the uncertainty remains open. Further speculation is beyond the remit of this report.

6.5 Alternative Models

As noted earlier, during the time of this project there have been two other exercises with comparable goals, namely IBRAM (Jamieson et al., 2016) and the B3 project "Optimising biosecurity investment and effort across all invasion phases" (see, e.g. Welsh et al., 2020). These projects address the same challenge but from slightly different points of view.

IBRAM comprises Bayesian nets of pest arrival, entry, establishment and spread. The structure of the nets is comparable to the model developed here, with the following variations.

- arrival rates are monthly, to allow for seasonal effects

[2]A version of the model that assesses impacts of activities against multiple pests is under development in current CEBRA work.

- infestation rate (per consignment) and pests per infested unit are accounted separately
- post-arrival escape is modeled in several steps (e.g. transitional facility, fruit shop)
- post-invasion pest spread is modeled explicitly on a monthly timestep
- pre-eradication damages are accounted
- impacts are distributed spatially.

These variations provide more fine-grained predictions of pest behavior than the CEBRA model, for example, allowing for the possibility that pests will arrive at the wrong time of year, but the demands on data are concomitantly more extreme. The CEBRA model can be interpreted as a lighter-touch variant of IBRAM.

The B3 project developed two models, one of which focused purely on pest spread, and the other that was designed to optimize pest risk management across the continuum, using some of the data and sub-models from this report.

6.6 Recommendations

1. MPI should note the outcomes of the modeling exercise and verify whether the questions that can be answered using the outcomes have operational utility.

2. If border data and establishment data exist, then we need to find a proactive way to use them. How best to do that is an open question. MPI and CEBRA should consider this question in the continuation of the project. Progress in the project documented herein was hampered by the regulatory data being available after the expert elicitation exercise, necessitating contestable unilateral alterations to the parameters.

3. If quantities have a natural rank then we need to elicit them in a way that preserves the rank (e.g., given an invading population, does it ever make sense for eradication to be cheaper when we are *not* investing in active surveillance?) How best to do that is an open question that should be considered by future work.

Bibliography

ACERA (2010). Process manual — elicitation tool. Technical Report Project 0611 report, Australian Centre of Excellence for Risk Analysis, The University of Melbourne, Melbourne. `http://cebra.unimelb.edu.au/data/assets/pdf_file/0010/1290466/0611-process-manual.pdf` (note that the Elicitation Tool to which the report refers is no longer available from the web site).

Adams-Hosking, C., McBride, M., Baxter, G., Burgman, M., Villiers, D., Kavanagh, R., Lawler, I., Lunney, D., Melzer, A., Menkhorst, P., Molsher, R., Moore, B., Phalen, D., Rhodes, J., Todd, C., Whisson, D., and McAlpine, C. (2016). Use of expert knowledge to elicit population trends for the koala (*Phascolarctos cinereus*). *Diversity and Distributions*, 22(3):249–262. DOI: 10.1111/ddi.12400.

Burgman, M., McBride, M., Ashton, R., Speirs-Bridge, A., Flander, L., Wintle, B., Fidler, F., Rumpff, L., and Twardy, C. (2011). Expert status and performance. *PLoS ONE*, 6:e22998.

Carey, J. and Burgman, M. (2008). Linguistic uncertainty in qualitative risk analysis and how to minimize it. *Annals of the New York Academy of Sciences*, 1128:13–17.

Castedo-Dorado, F., Lago-Parra, G., Lombardero, M., Liebhold, A., and M.F.Álvarez-Taboada (2016). Designing cost-efficient surveillance for early detection and control of multiple biological invaders. *New Zealand Journal of Forestry Science*, 46(18):1–6.

Cooke, R. (1991). *Experts in uncertainty: Opinion and subjective probability in science*. Environmental Ethics and Science Policy Series. Oxford University Press.

Cooke, R. (2013). Expert judgement assessment: quantifying uncertainty on thin ice. *Nature Climate Change*, 3:311–312.

Dalkey, N. and Helmer, O. (1963). An experimental application of the Delphi Method to the use of experts. *Management Science*, 9:458–467.

Epanchin-Niell, R. S., Brockerhoff, E. G., Kean, J. M., and Turner, J. A. (2014). Designing cost-efficient surveillance for early detection and control of multiple biological invaders. *Ecological Applications*, 24(6):1258–1274.

Epley, N. and Gilovich, T. (2006). The anchoring-and-adjustment heuristic. *Psychological Science*, 17:311–318.

Eyre, D., Macarthur, R., Haack, R. A., Lu, Y., and Krehan, H. (2018). Variation in Inspection Efficacy by Member States of Wood Packaging Material Entering the European Union. *Journal of Economic Entomology*, 111(2):707–715.

Fleishman, E., Burgman, M., Runge, M., Schick, R., and Kraus, S. (2016). Expert elicitation of population-level effects of disturbance. In Popper, A. and Hawkins, A., editors, *The Effects of Noise on Aquatic Life II*, volume 875, pages 295–302. Advances in Experimental Medicine and Biology.

Galton, F. (1907). Vox populi. *Nature*, 75:450–451.

Gigerenzer, G. (2004). Dread risk, September 11, and fatal traffic accidents. *Psychological Science*, 15:286–287.

Gustafson, L., Antognoli, M., Lara Fica, M., Ibarra, R., Mancilla, J., Sandoval, O., Enriquez, R., Perez, A., Aguilar, D., Madrid, E., Bustos, P., Clement, A., Godoy, M., Johnson, C., and Remmenga, M. (2014). Risk factors perceived predictive of ISA spread in Chile: Applications to decision support. *Preventive Veterinary Medicine*, 117:276–285.

Harris Consulting (2003). Asian Gypsy Moth: Assessment of Potential Economic Impacts. Technical report.

Jamieson, L., Woodberry, O., McDonald, C., and Ormsby, M. (2016). Developing a generic biosecurity risk assessment model for imports. *New Zealand Plant Protection*, 69:186–199.

Janis, I. (1971). Groupthink. Psychology Today. 5: 43-76. In Shafritz, J., Ott, J., and Jang, Y., editors, *Reprinted in: Classics of Organization Theory*. 2015 Cengage Learning, Independence, KY., 8th edition.

Jarrad, F., Barrett, S., Murray, J., Parkes, J.and Stoklosa, R., Mengersen, K., and Whittle, P. (2011). Improved design method for biosecurity surveillance and early detection of non-indigenous rats. *New Zealand Journal of Ecology*, 35:132–144.

Jensen, J. L. (1906). Sur les fonctions convexes et les inégualités entre les valeurs moyennes. *Acta Mathematica*, 30:175–193. (in French).

Jones, O., Maillardet, R., and Robinson, A. P. (2014). *An Introduction to Scientific Programming and Simulation, Using R*. Chapman & Hall/CRC, Boca Raton, 2-nd edition.

Kean, J. M., Suckling, D. M., Sullivan, N. J., Tobin, P. C., Stringer, L. D., Smith, G. R., Kimber, B., Lee, D. C., Flores Vargas, R., Fletcher, J., Macbeth, F., McCullough, D. G., Herms, D. A., et al. (2019). Global eradication and response database. WWW. `http://b3.net.nz/gerda` (accessed 26 February 2019).

Krinitzsky, E. (1993). Earthquake probability in engineering - part 1: The use and misuse of expert opinion. *Engineering Geology*, 33:257–288.

Linstone, H. and Turoff, M. (1975). *The Delphi Method: Techniques and Applications*. Addison-Wesley Pub. Co., Reading, MA.

Maier, N. (1967). Assets and liabilities in group problem solving: the need for an integrative function. *Psychological Review*, 74:239–249.

Martin, T., Burgman, M., Fidler, F., Kuhnert, P., Low-Choy, S., McBride, M., and Mengersen, K. (2012). Eliciting expert knowledge in conservation science. *Conservation Biology*, 26:29–38.

McBride, M. F., Fidler, F., and Burgman, M. A. (2012). Evaluating the accuracy and calibration of expert predictions under uncertainty: predicting the outcomes of ecological research. *Diversity and Distributions*, 18(8):782–794.

Ministry for Primary Industries (2012). Gypsy Moth Surveillance Programme. WWW. `https://www.mpi.govt.nz/dmsdocument/3412/direct`, accessed 4 March 2021.

Ministry for Primary Industries (2019). Exotic insect discovered in new zealand for the first time. WWW. `https://www.mpi.govt.nz/news/media-releases/exotic-insect-discovered-in-new-zealand-for-the-first-time/`, accessed 4 March 2021.

Mukherjee, N., Hugé, J., Sutherland, W., McNeill, J., Van Opstal, M., Dahdouh-Guebas, F., and Koedam, N. (2015). The Delphi technique in ecology and biological conservation: applications and guidelines. *Methods in Ecology and Evolution*, 6:1097–1109.

Nisbett, R. and Wilson, T. (1977). Halo effect: evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, 35:250–256.

O'Hagan, A., Buck, C., Daneshkhah, A., Eiser, J., Garthwaite, P., Jenkinson, D., Oakley, J., and Rakow, T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. Wiley, London.

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, 29:261–265.

Plous, S. (1993). *The Psychology of Judgment and Decision Making*. McGraw-Hill, New York.

Robinson, A., Brockerhoff, E., and Ormsby, M. (2018a). Scoping the value and performance of interventions across the NZ biosecurity system. Peer-Reveiwed Technical Report 1606E, CEBRA.

Robinson, A., Welsh, M., Ormsby, M., Brockerhoff, E., and Reed, C. (2018b). Proportional value of interventions across pathways and layers of the biosecurity system. Peer-Reveiwed Technical Report 170621, CEBRA.

Ross, M. G. (2005). Response to a gypsy moth incursion in new zealand. A Paper presented at the IUFRO conference, Hanmer, 2004, and since updated July 2005.

Soll, J. and Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning Memory and Cognition*, 30:299–314.

Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G., and Burgman, M. (2010). Reducing overconfidence in the interval judgments of experts. *Risk Analysis*, 30:512–523.

Surowiecki, J. (2005). *The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. Doubleday, USA.

Sutherland, W. (2006). Predicting the ecological consequences of environmental change: a review of the methods. *Journal of Applied Ecology*, 43:599–616.

Sutherland, W. and Burgman, M. (2015). Use experts wisely. *Nature*, 526:317–318.

Teigen, K. and Jørgensen, M. (2005). When 90% confidence intervals are 50% certain: On the credibility of credible intervals. *Applied Cognitive Psychology*, 19:455–475.

Thorndike, E. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4:25–29.

Tversky, A. and Kahneman, D. (1973). Availability: a heuristic for judging frequency and probability. *Cognitive Psychology*, 5:207–232.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty — heuristics and biases. *Science*, 185:1124–1131.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211:453–458.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–67.

Welsh, M., Turner, J., Epanchin-Niell, R., Monge, J., Soliman, T., Robinson, A., Kean, J., Phillips, C., Stringer, L., Vereijssen, J., Liebhold, A., Kompas, T., Ormsby, M., and Brockerhoff, E. (2020). Approaches for estimating benefits and costs of interventions in plant biosecurity across invasion phases. *Ecological Applications*, Accepted.

Wintle, B., Mascaro, S., Fidler, F., McBride, M., Burgman, M., Flander, L., Saw, G., Twardy, C., Lyon, A., and Manning, B. (2012). The intelligence game: assessing delphi groups and structured question formats. In *Proceedings of the 5th Australian Security and Intelligence Conference, Novotel Langley Hotel, Perth, Western Australia, 3rd-5th December, 2012.* http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1025&context=asi.

Work, T., McCullough, D., Cavey, J., and Komsa, R. (2005). Arrival rate of nonindigenous insect species into the United States through foreign trade. *Biological Invasions*, 7:323–332.

Appendix A

SEJ Question Topics

Preliminary :

1. What is the minimal viable population, that is, the smallest size of the likeliest invasive life stage that may result in an establishment?
2. What are all the pathways / vectors by which the pest may enter New Zealand?
3. For each pathway, what are the units of biosecurity intervention?
4. On each pathway, what are the measures that may affect the biosecurity risk before or at the border?
5. What is the best metric of incursion size at the time at which the incursion is detected we can use to describe the incursion?
6. What is the largest incursion population size at which you think that eradication will still be attempted?
7. (Example question)

Preparation :

8. Determine number of units of intervention for each pathway

Face-to-Face :

9. Raw unrefined risk on each pathway
10. Effects of pre-border and border measures on risk
11. Establishment rate after incursion
12. Invading population size upon detection
13. Cost of eradication
14. Success probability of eradication

Appendix B

Preliminary Conference Call

Expert elicitation workshop: Preliminary Conference Call for Gypsy Moth.

8 am AEST 22 November 2018.

Attendees: Sandy Liebhold (USDA), Diane Anderson (Entomologist, MPI), Michael Ormsby (Pathologist, MPI), Rory MacLellan (Incursion Investigation, MPI), Stephanie Sopow (Entomologist, Scion), David Gray (Entomologist, NRC), and Andrew Robinson (CEBRA).

Agenda

1. Introductions (AR, 5 min)
2. Project background (AR, 5 min, see Attachment A.)
3. Expert elicitation protocol (AR, 10 min, see Attachment B.)
4. Preliminary Questions — for now (AR & MO, 20 min (actually 30))

   (a) What is the minimal propagable unit (MPU), that is, the smallest size of the likeliest invasive life stage that may result in an establishment? (AR)
   Sandy: Robinet et al. said 2–4 egg masses. 50/50. Establishment of A may differ from E.
   Diane: 1
   Mike: A v E — different?
   Rory: 1 egg mass? A E not different.
   Stephanie: single egg mass.
   David: Absent specific knowledge A E not different for now.

   (b) What are all the pathways / vectors by which the pest may enter New Zealand? (MO)
   Used vehicles, Sea containers, Ships, . . . , personal goods, nursery stock, wood.
   Sandy: hitchhiker; could be almost anything. Sheets of steel. New cars.
   Diane: nothing to add
   Rory: also used machinery (MO: yes, includes in used vehicles)
   Stephanie: anything is possible. New car parts wrapped in plastic.
   David: anything stored near a port has potential.

   (c) For each pathway, what are the units of biosecurity intervention? (MO)

   (d) On each pathway, what are the measures that may affect the biosecurity risk before or at the border? (MO)
   Mike: EGM might not arrive on those pathways.

Rory: most interceptions are not tested.
Mike: Use of lights; don't load at night.
Andrew: what's actually done?
David: nothing done in Canada that I know of to prevent export. Import: demand Asian ports conduct examination and cleaning of outbound ships to NA. Captain forwards ship log; if at designated Asian port in preceding 18 months then needs certification for clean and a small sample is examined on arrival in Canadian port.
Rory: IHS says inspection certificate mandatory if in port in high-risk season, else held offshore for inspection. Same for used vehicles and equipment from Japan.
Mike: Inspection and certification covered. Also treatment of vehicles from Japan are heat treated (for BMSB). Mildly certain that some pathways/ports consider lights.
Rory: inspection at origin and destination.
Sandy: potentially plant imports, but NZ's other plant quarantine policies and procedures. Domestically they move around on plants.
Rory: NZ IHS are stringent.
Andrew: anything else?
Stephanie: what proportion of personal goods are fumigated?
Rory: pretty sure everything inspected or fumigated.
Mike: there are many such pathways. Managed effectively. Managed for a range of pests.
Andrew: handle these separately.
Rory: plants, personal goods, inanimate objects.

(e) What is the best metric of incursion size at the time at which the incursion is detected we can use to describe the incursion? (AR)
Sandy: max trap capture (# males) and area of non-zero trap capture.
Rory: max trap capture.
David: consider two things: 500 adults in one trap different than 10 in 50 traps.
Mike: done modeling in FF. Looked at it as trapping grid. Question is: what population of gypsy moth will be in the area that will lead to eradication?
David: can't avoid discussion of risk tolerance. Does one adult trigger eradication? Don't know what capture means for population in truth.
Mike: these are connected. What is assumption around population size?
Sandy: North America knows statistically # males captured and probability that incursion will go away — e.g. 1 male 80–90% incursion fail. 95% at capture of 10.

(f) What is the largest incursion size at which you think that eradication will still be attempted? (AR)
Mike: scenario. Assume establishment in urban area. Assume can spray. Can do eradication. However, cannot spray native area. Cost limitation. May be no other means of eradication.
Rory: very high, spread over large area – half of North Island. Well studied. Can close pathways to avoid reinfestation.
Sandy: agree. Had good success in NA. Easy to eradicate. Know how to delimit and treat. Eradicated up to 50K Ha in size.
(Andrew: Ok to use infestation area for Question e?
Rory: Yes.
Sandy: Area the primary determinant. Max trap catch is determinant of incursion response method.)

5. Demonstration of expert elicitation protocol (AR, 15 min (actually 30)).
For a given pathway (ships), imagine 1000 random, representative units (or more, e.g. some suitable multiple of 10) of the pathway arriving in New Zealand without any biosecurity measures in place, that is, the raw, unrefined exposure. Think about the number of such units that could be carrying an MPU of GM (an egg mass).

(a) List the reasons that this number might be low. What factors mitigate against these units carrying an MPU of the pest in question? NB: please exclude ALL existing biosecurity measures for this question!
Andrew: might not visit ports, or might be a different time
Rory: temperature, time of year, lights
Sandy: year — most years not an outbreak.

(b) Record the lowest reasonable number of the 1000 random, representative units that you would expect could be carrying at least one MPU of the pest.
E.g. 0

(c) List the reasons that this number might be high. What factors mitigate in favour of these units carrying an MPU of the pest in question?
Andrew: did visit ports in critical time of year, or maybe it is a mast year.
Sandy: seasons, window, high-risk ports due to proximity to hosts.
Rory: amount of trade.

(d) Record the highest reasonable number of the 1000 random, representative units that you would expect could be carrying at least one MPU of the pest.

(e) Record your best guess at the number of the 1000 units carrying at least one MPU of the pest.

(f) What is your personal probability that the true value for any random set of 1000 units would be between your low and high values?

6. Wrap up and debrief (AR, 5 min).

Appendix C

Graphical Summaries

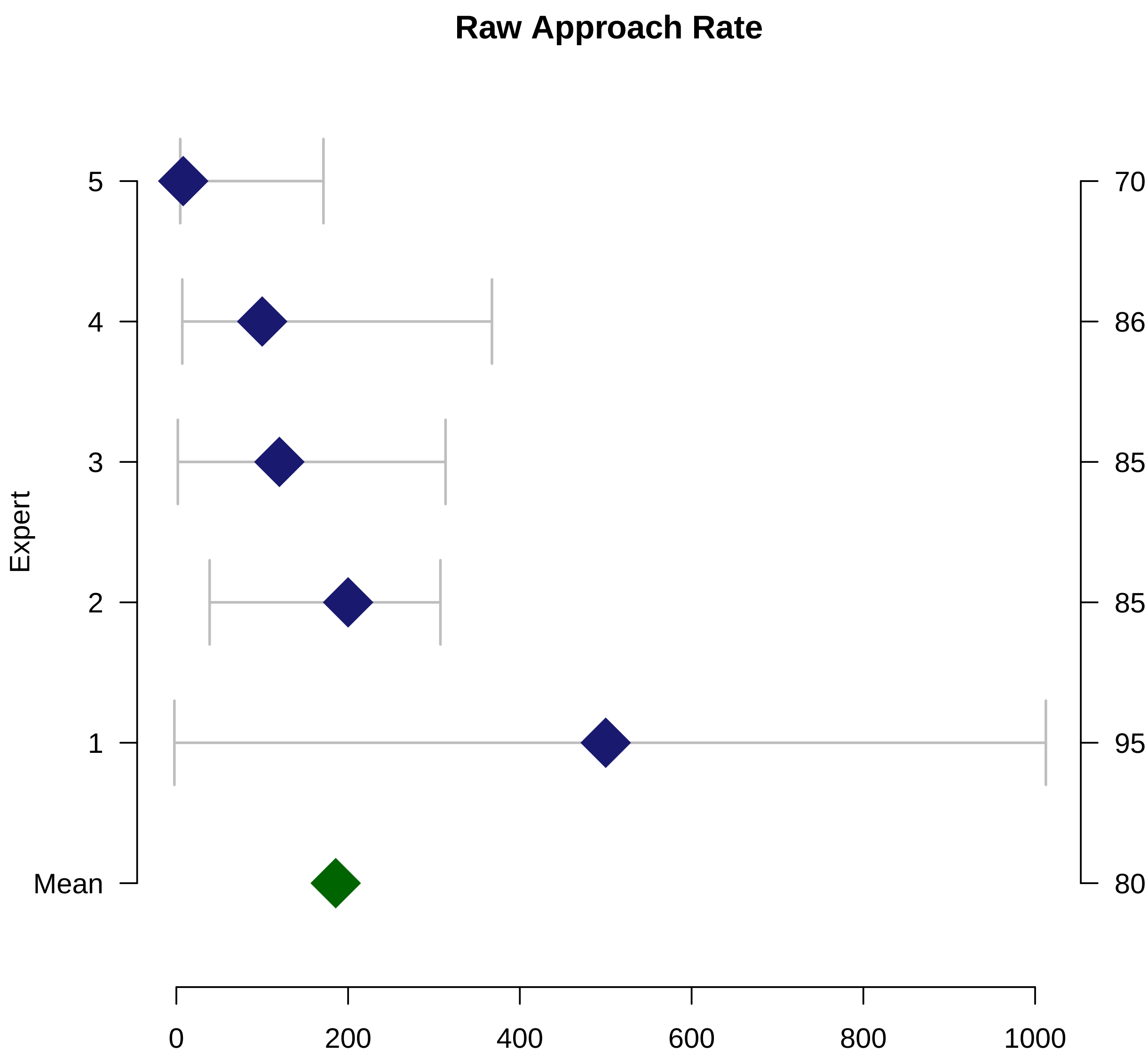

Figure C.1: Summary of expert elicited values for Question 9a: High-risk Ships.

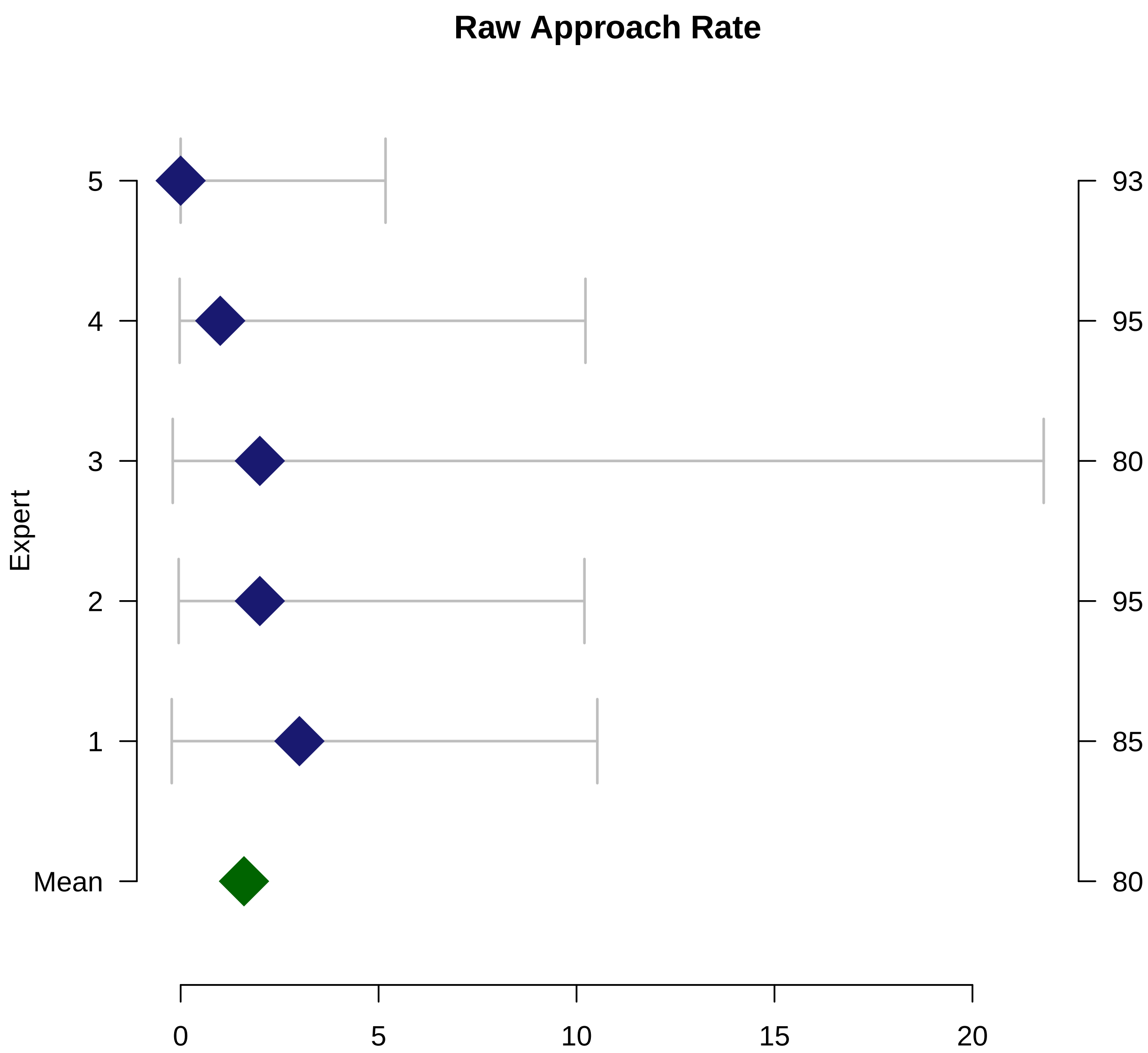

Figure C.2: Summary of expert elicited values for Question 9b: Low-risk Ships.

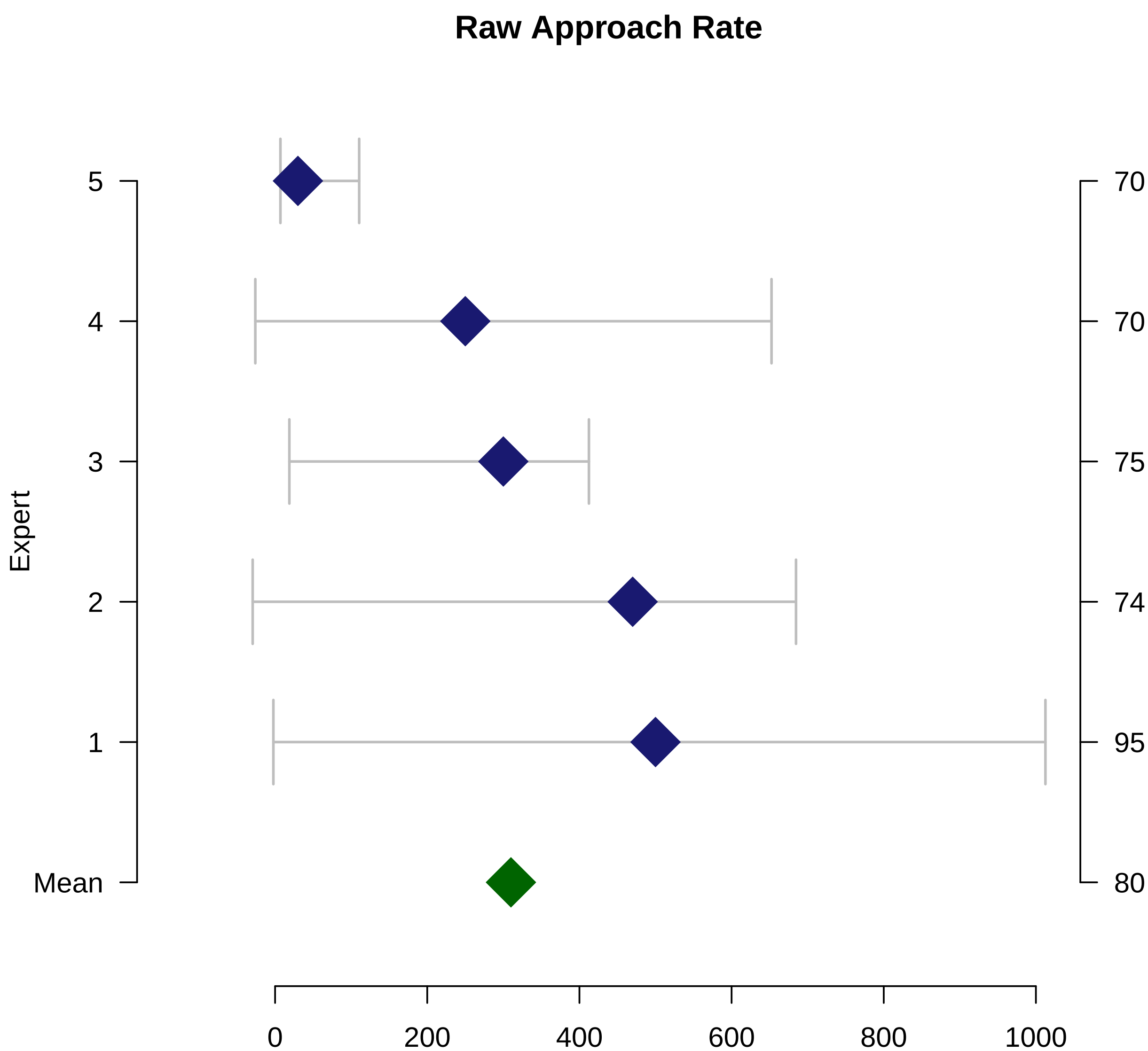

Figure C.3: Summary of expert elicited values for Question 9c: High-risk Containers.

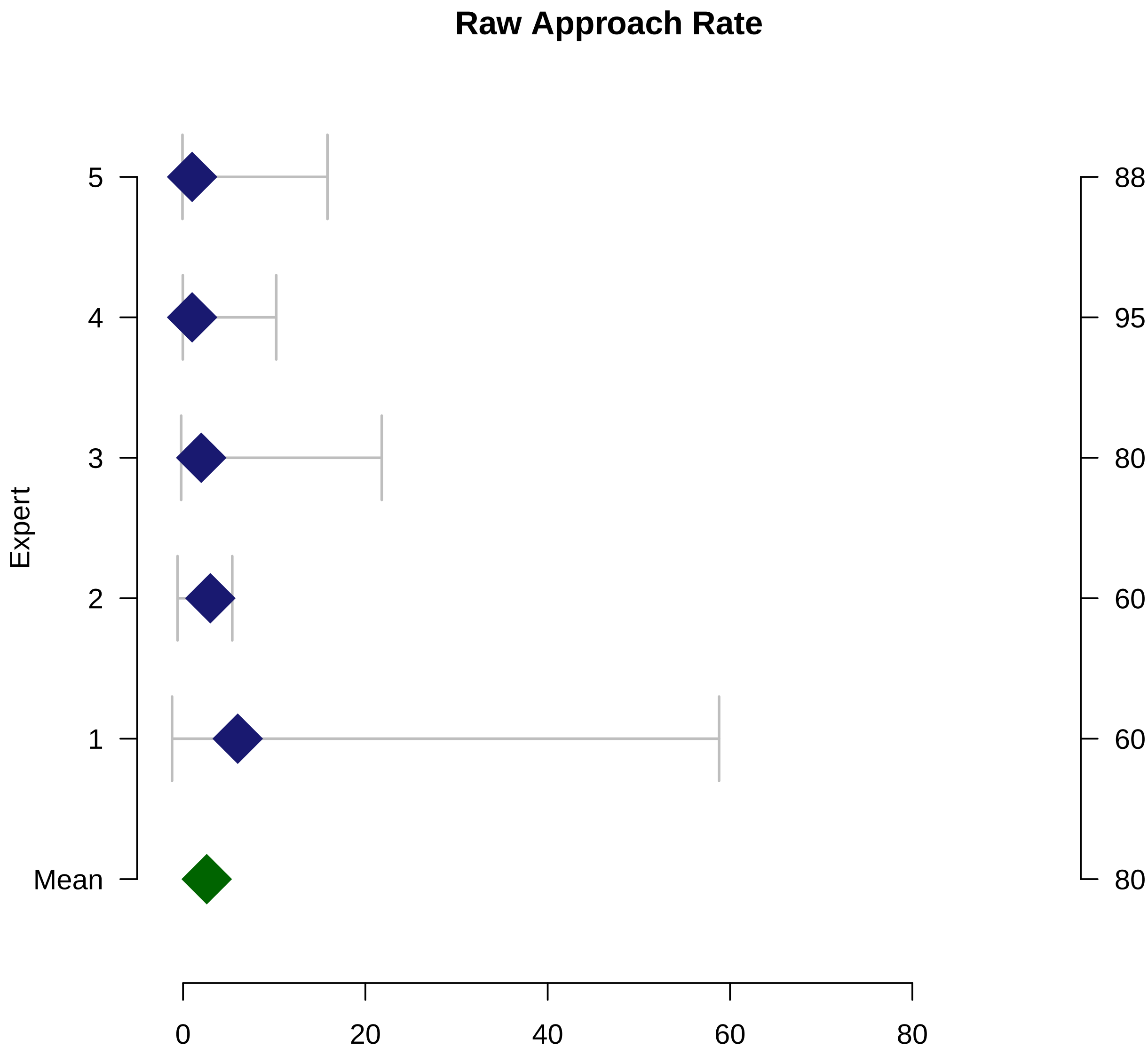

Figure C.4: Summary of expert elicited values for Question 9d: Low-risk Containers.

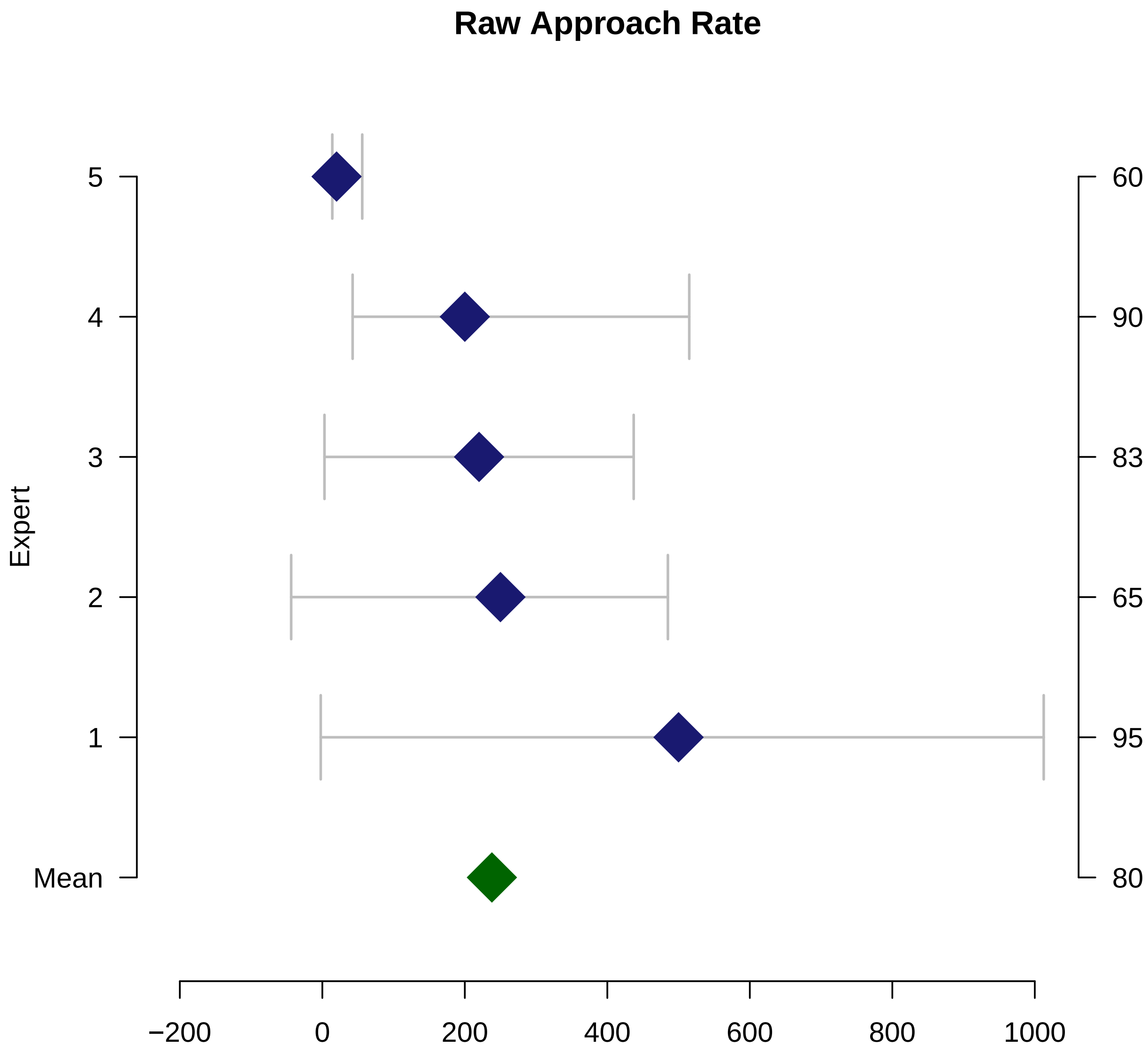

Figure C.5: Summary of expert elicited values for Question 9e: High-risk V'cles & Mach..

Raw Approach Rate

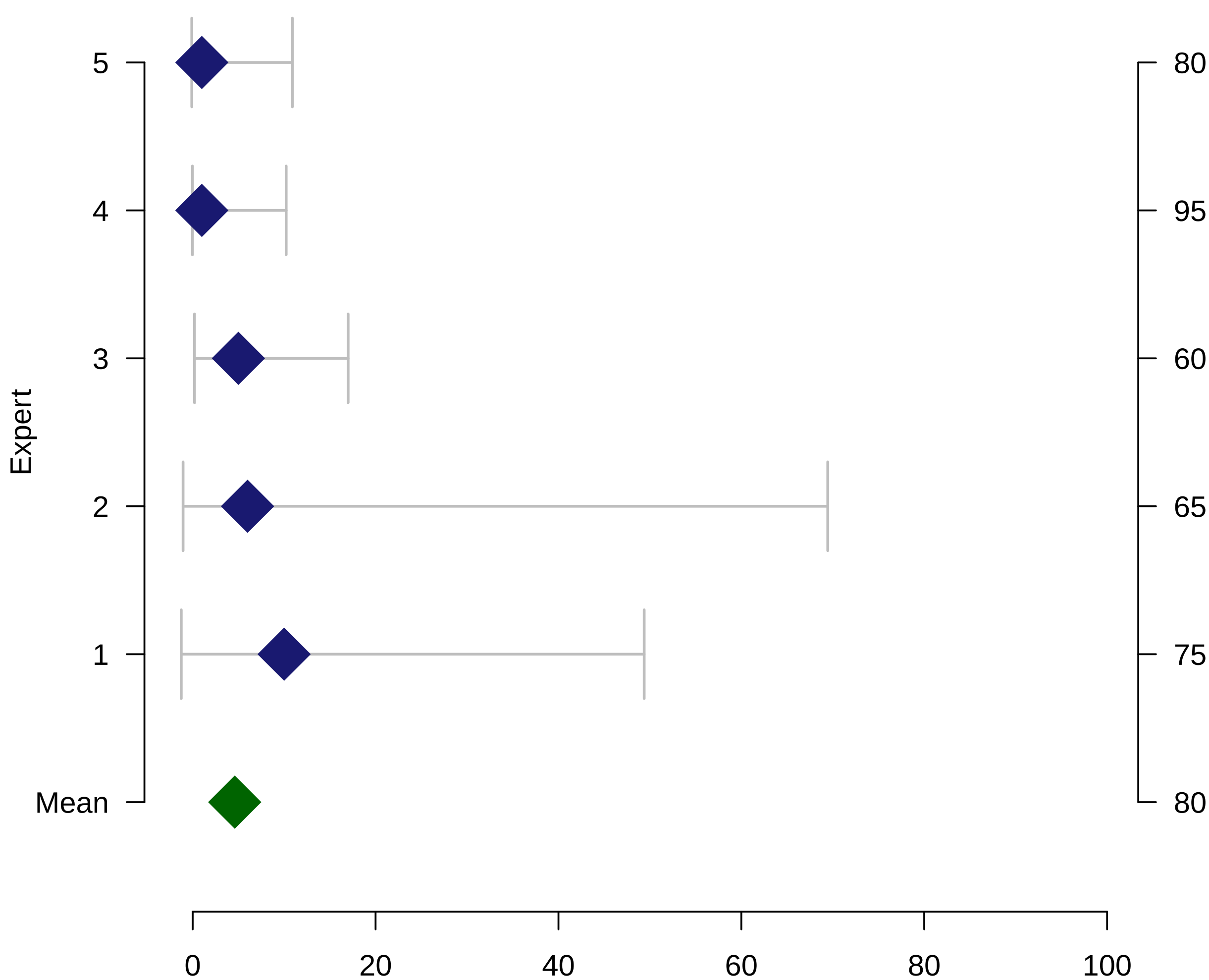

Figure C.6: Summary of expert elicited values for Question 9f: Low-risk V'cles & Mach..

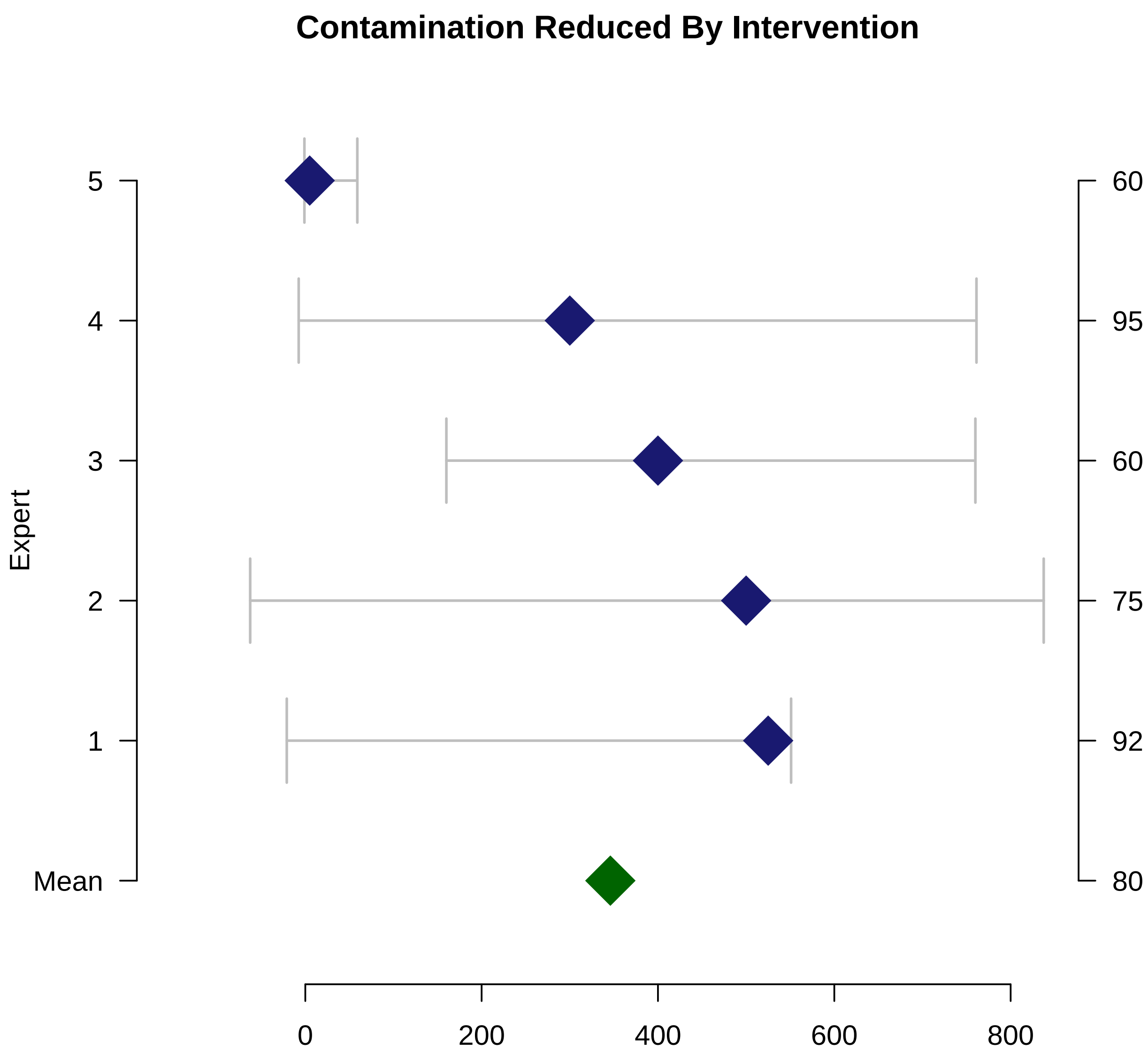

Figure C.7: Summary of expert elicited values for Question 10a: Offshore Cert & Insp. of HR Ships.

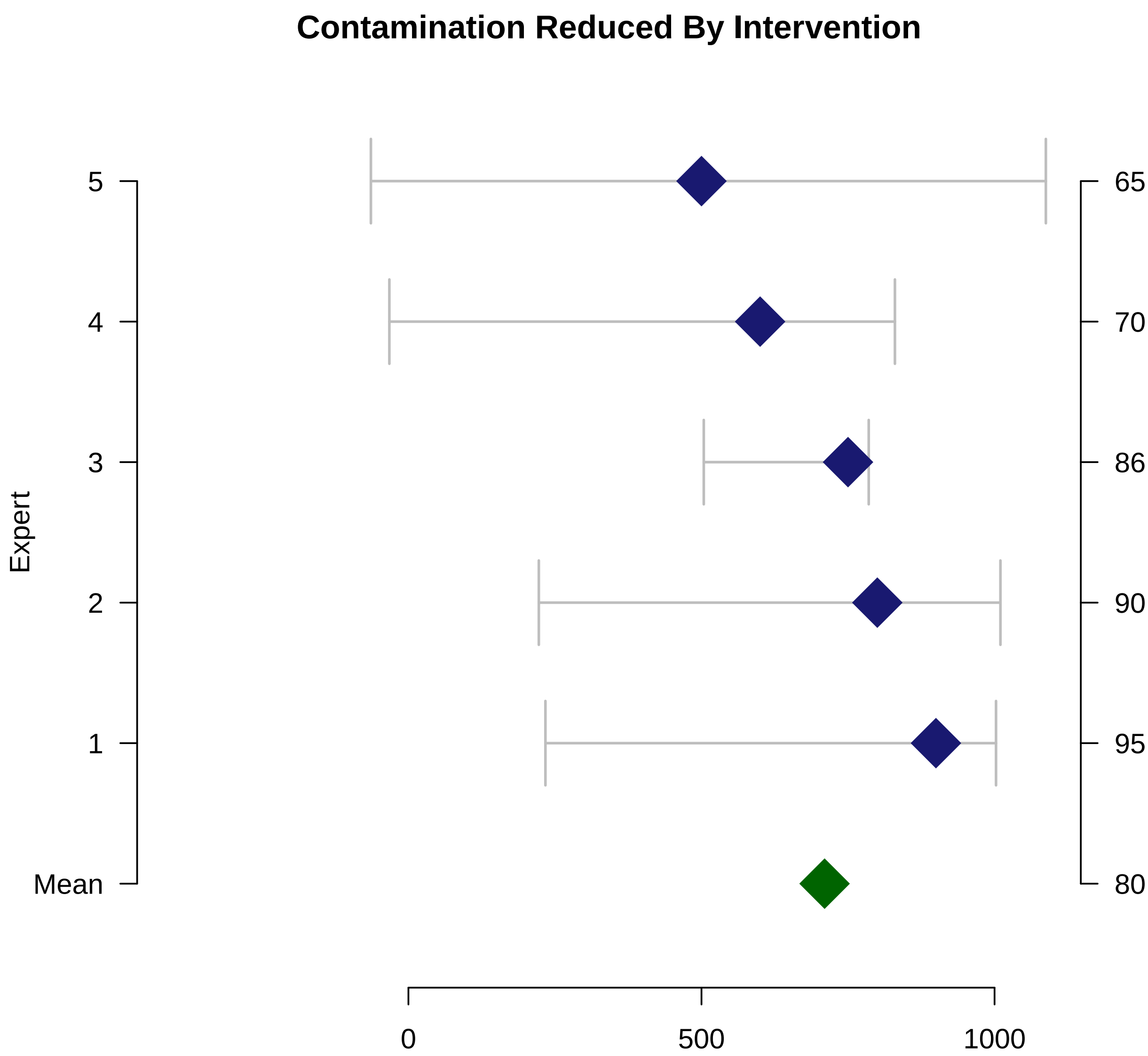

Figure C.8: Summary of expert elicited values for Question 10b: Offshore Cert & Insp. of HR V'cles & Mach..

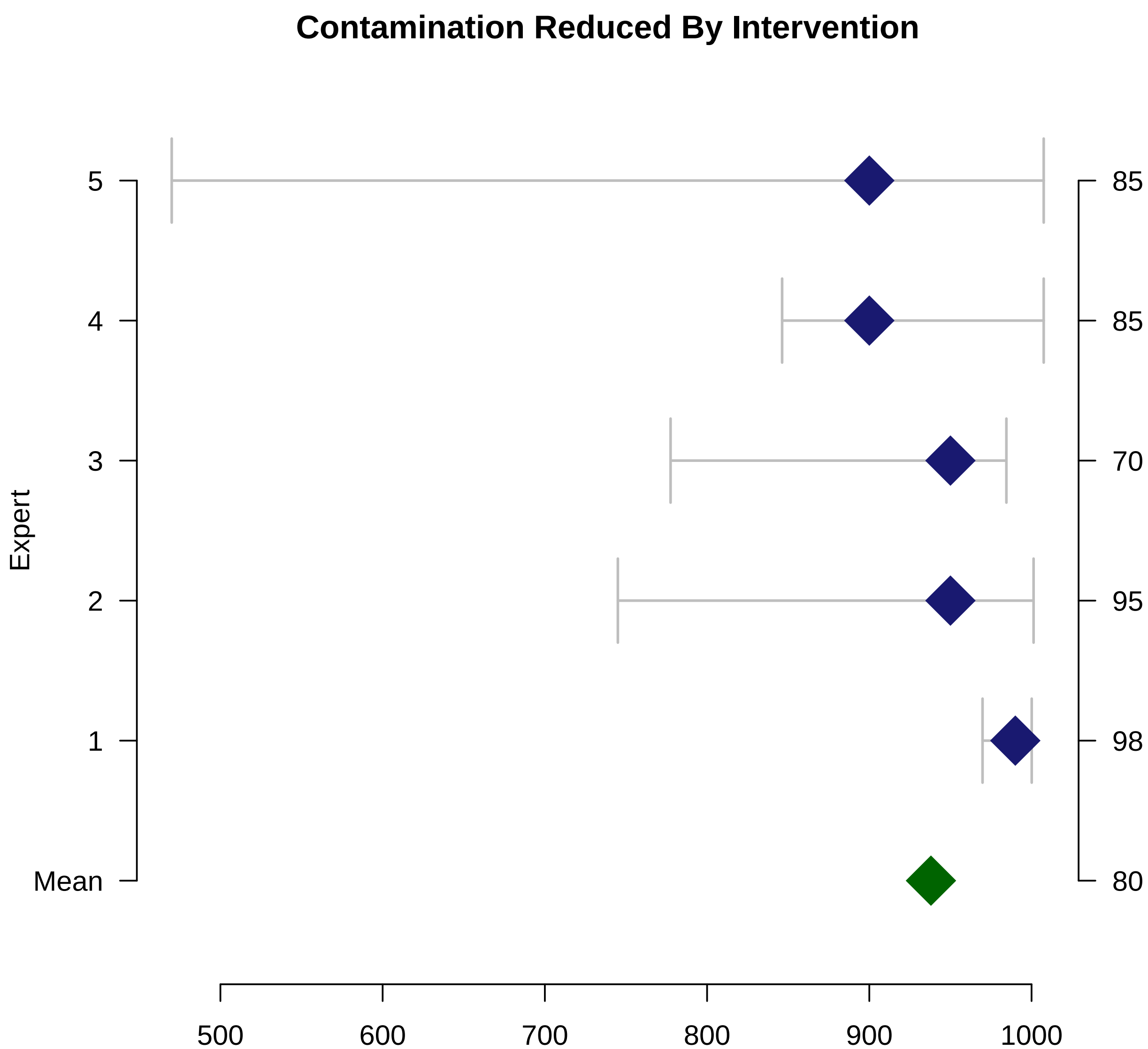

Figure C.9: Summary of expert elicited values for Question 10c: Heat Treatment of HR V'cles & Mach..

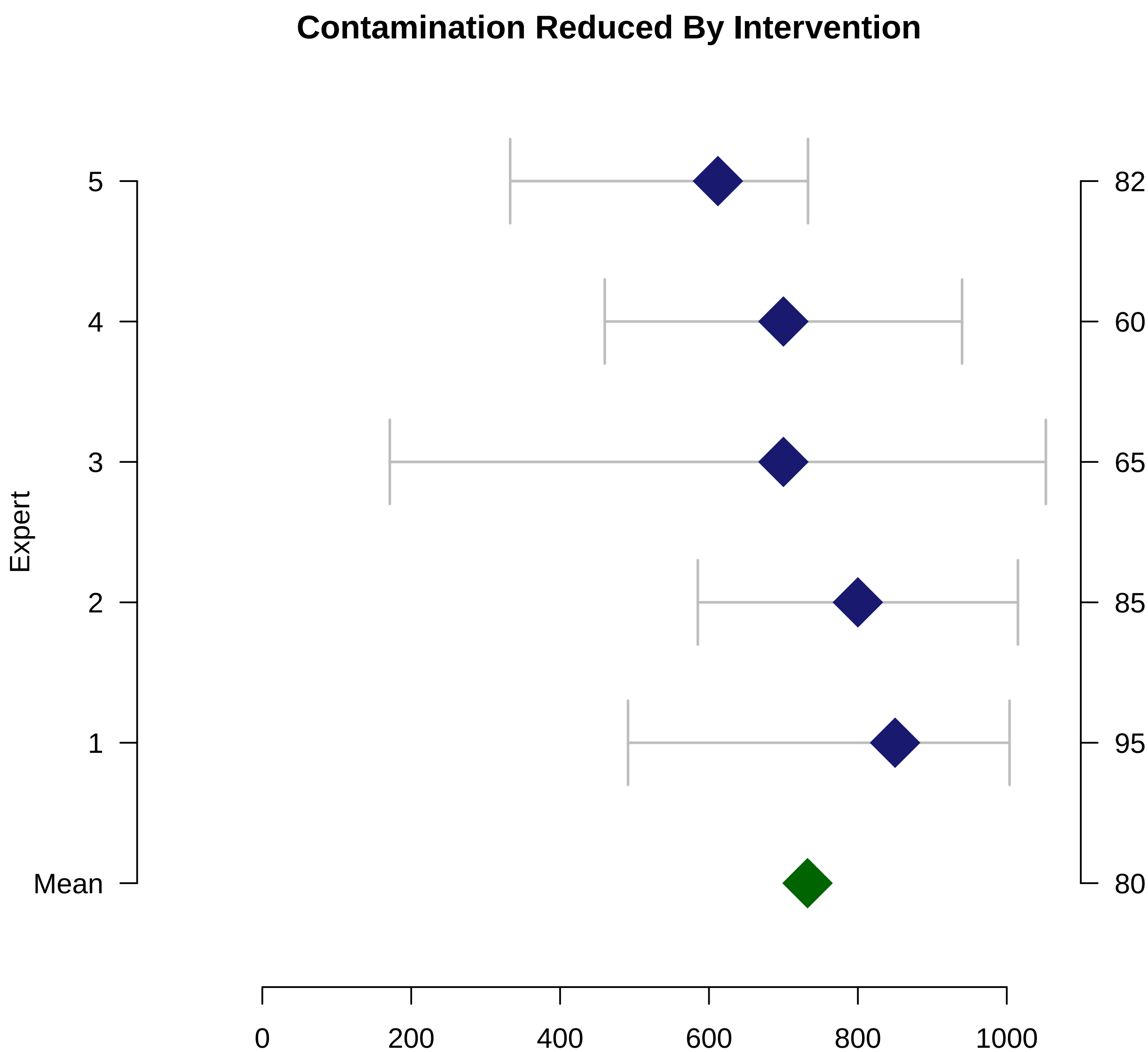

Figure C.10: Summary of expert elicited values for Question 10d: Inspection of HR Containers.

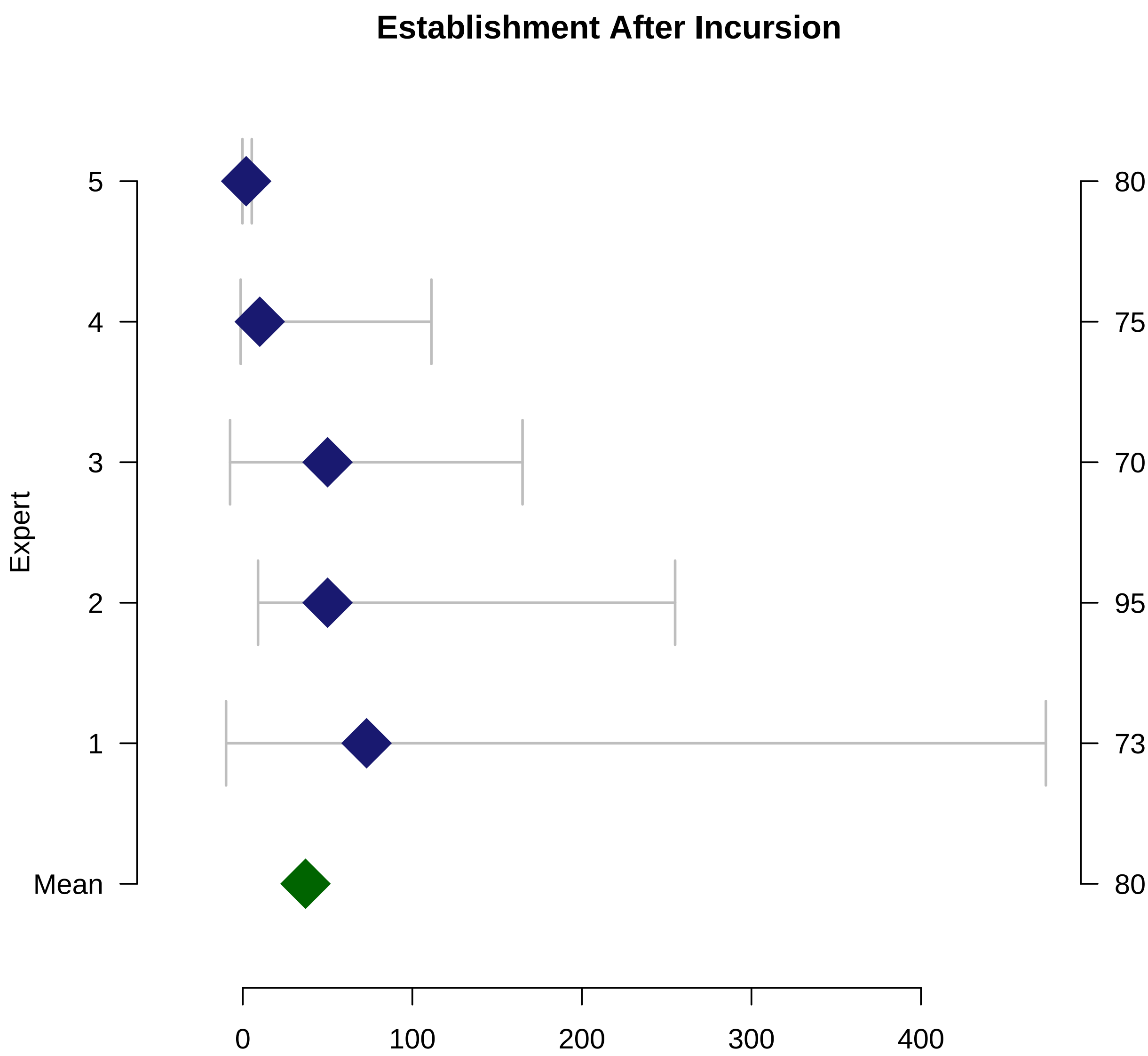

Figure C.11: Summary of expert elicited values for Question 11a: Establishment from Ships.

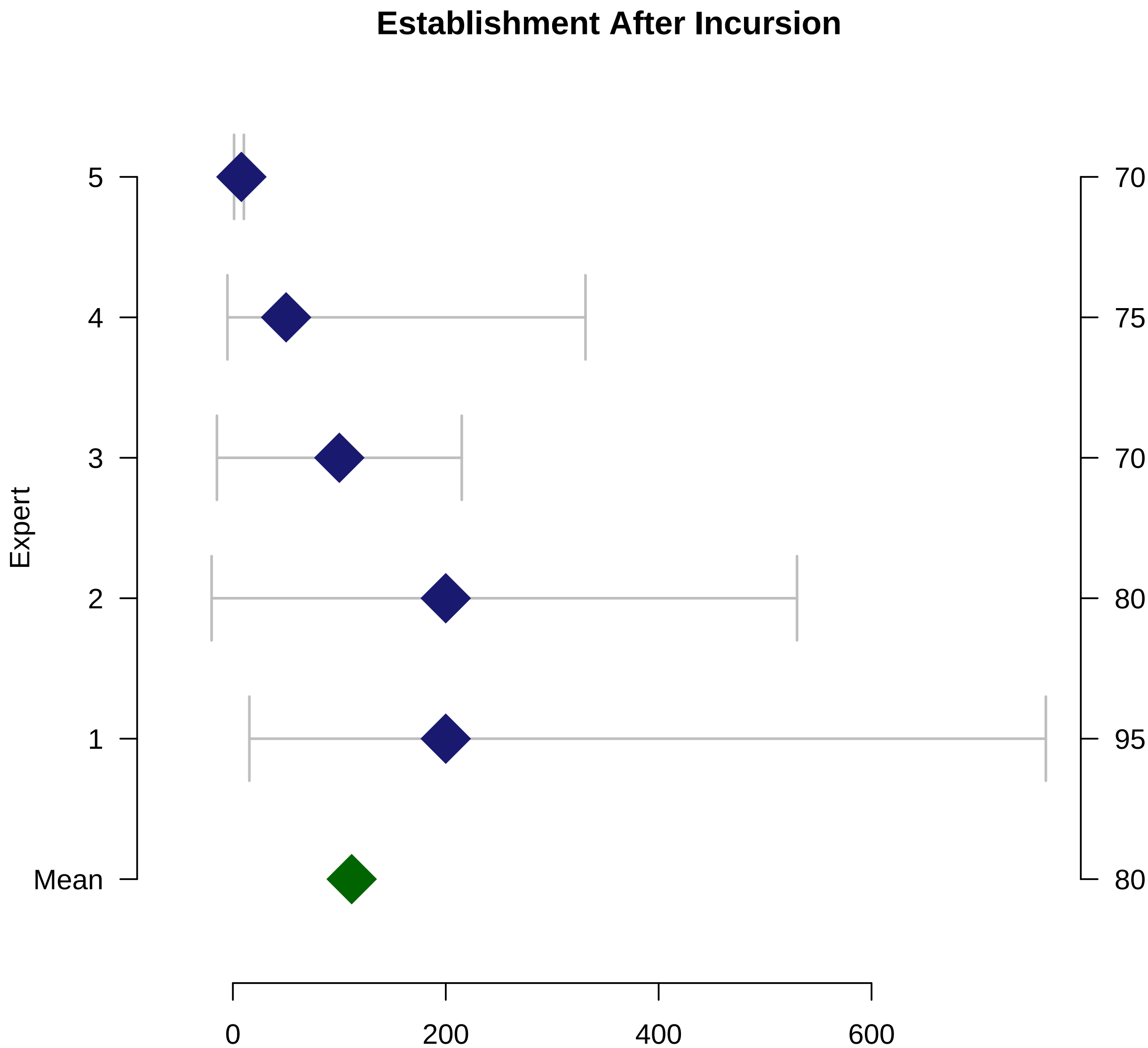

Figure C.12: Summary of expert elicited values for Question 11b: Establishment from Containers.

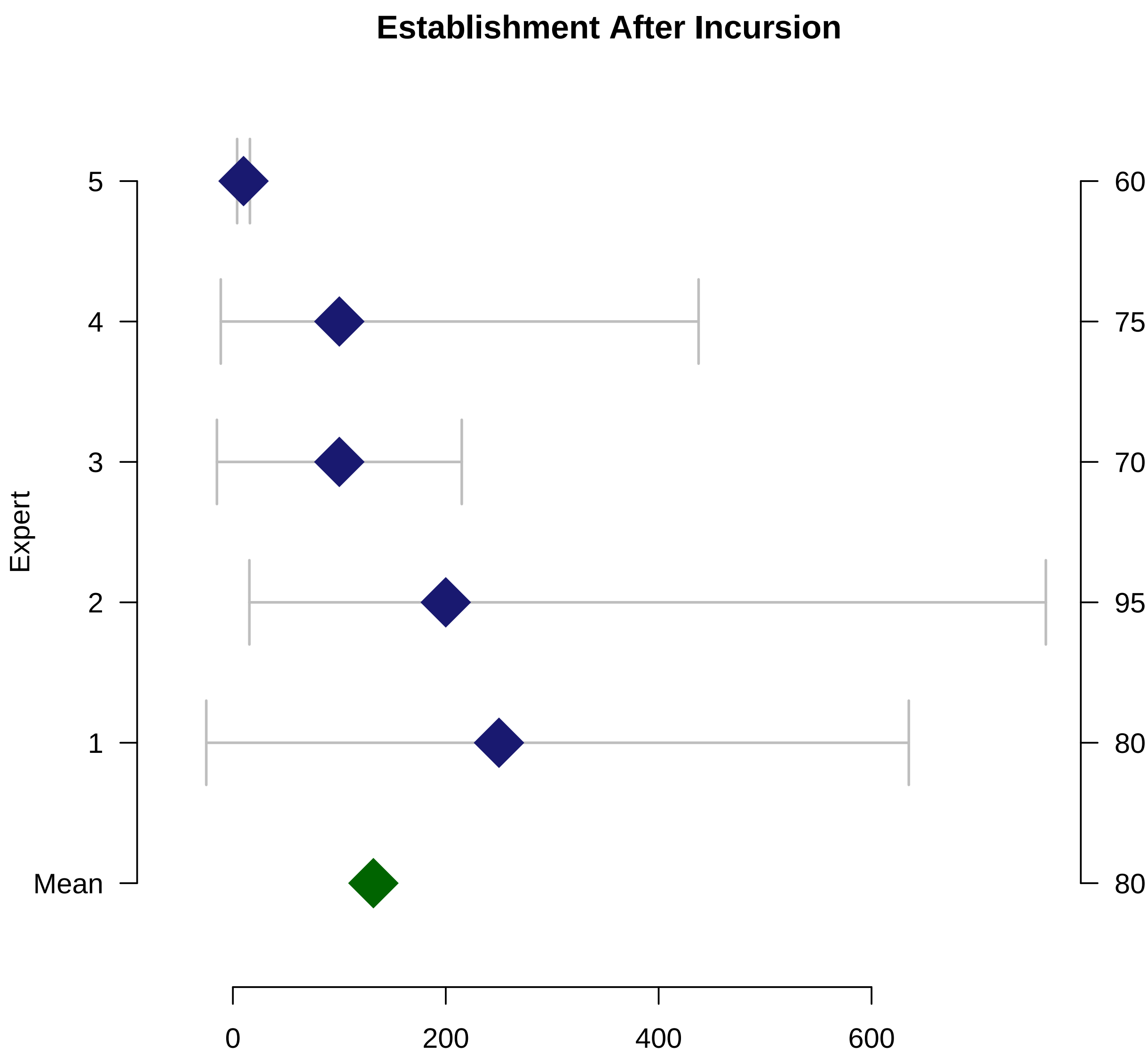

Figure C.13: Summary of expert elicited values for Question 11c: Establishment from V'cles & Mach..

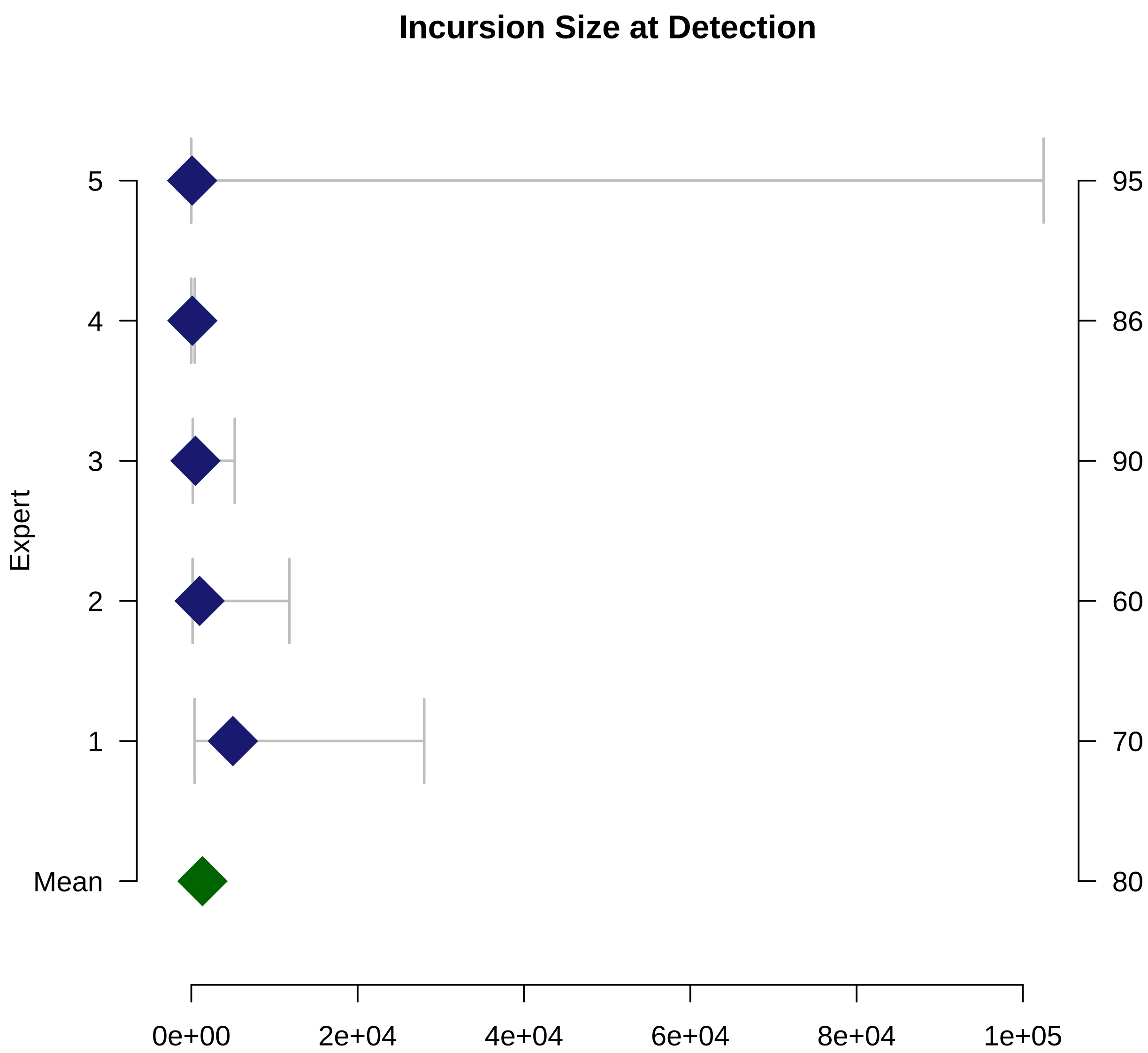

Figure C.14: Summary of expert elicited values for Question 12a: Incursion Size at Detection; Current Surveillance.

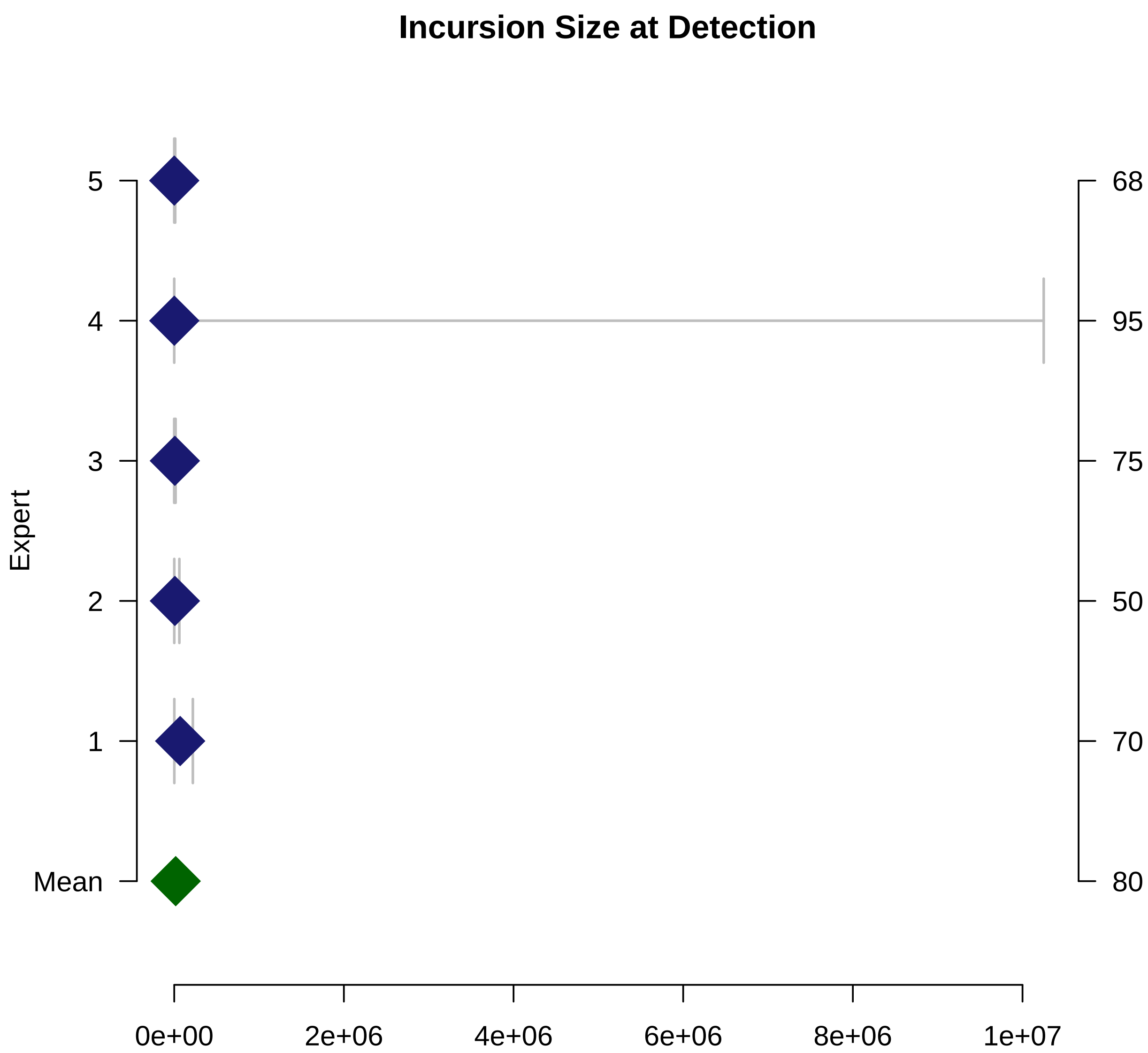

Figure C.15: Summary of expert elicited values for Question 12b: Incursion Size at Detection; No Surveillance.

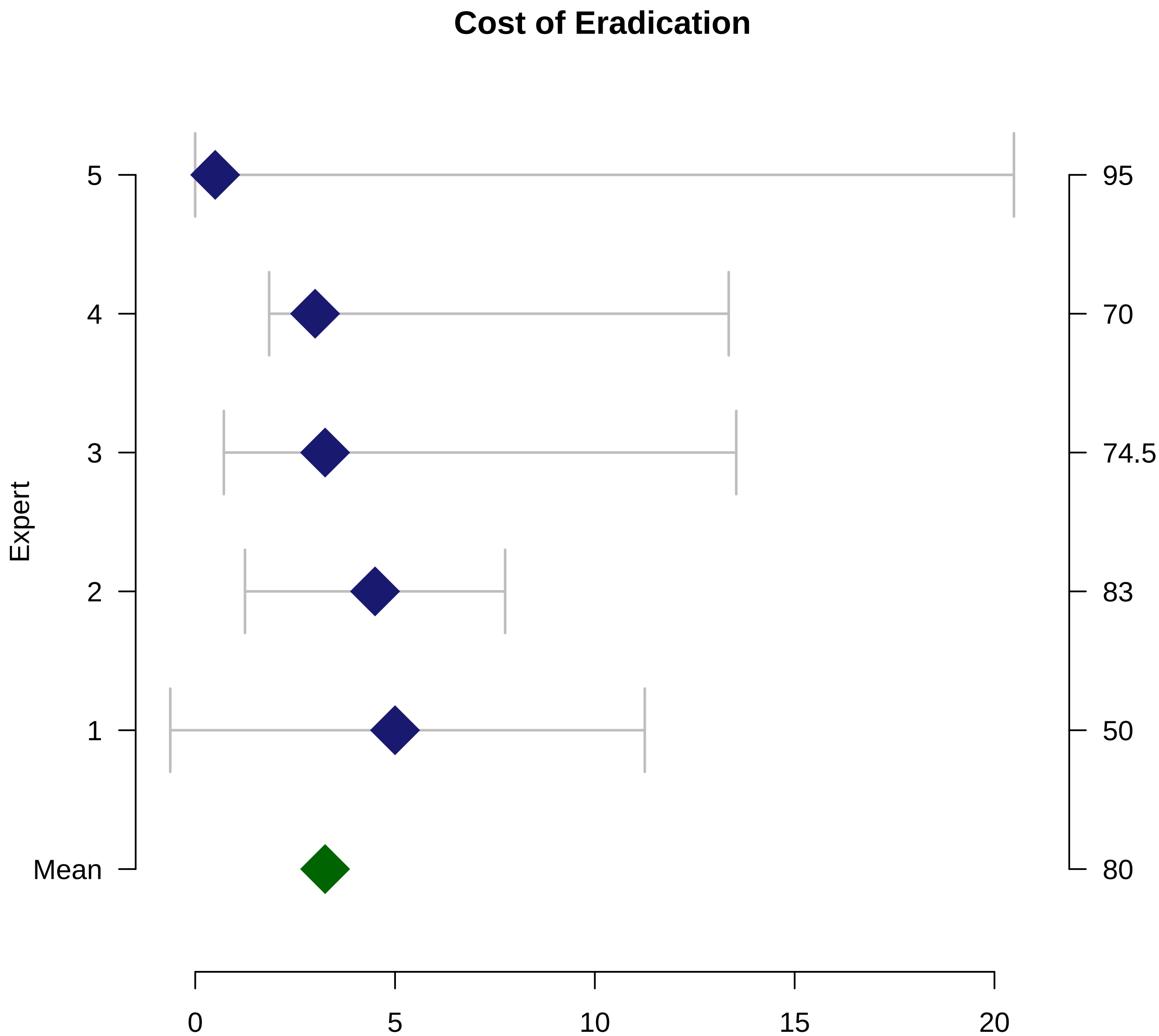

Figure C.16: Summary of expert elicited values for Question 13a: Cost of Eradication with Readiness.

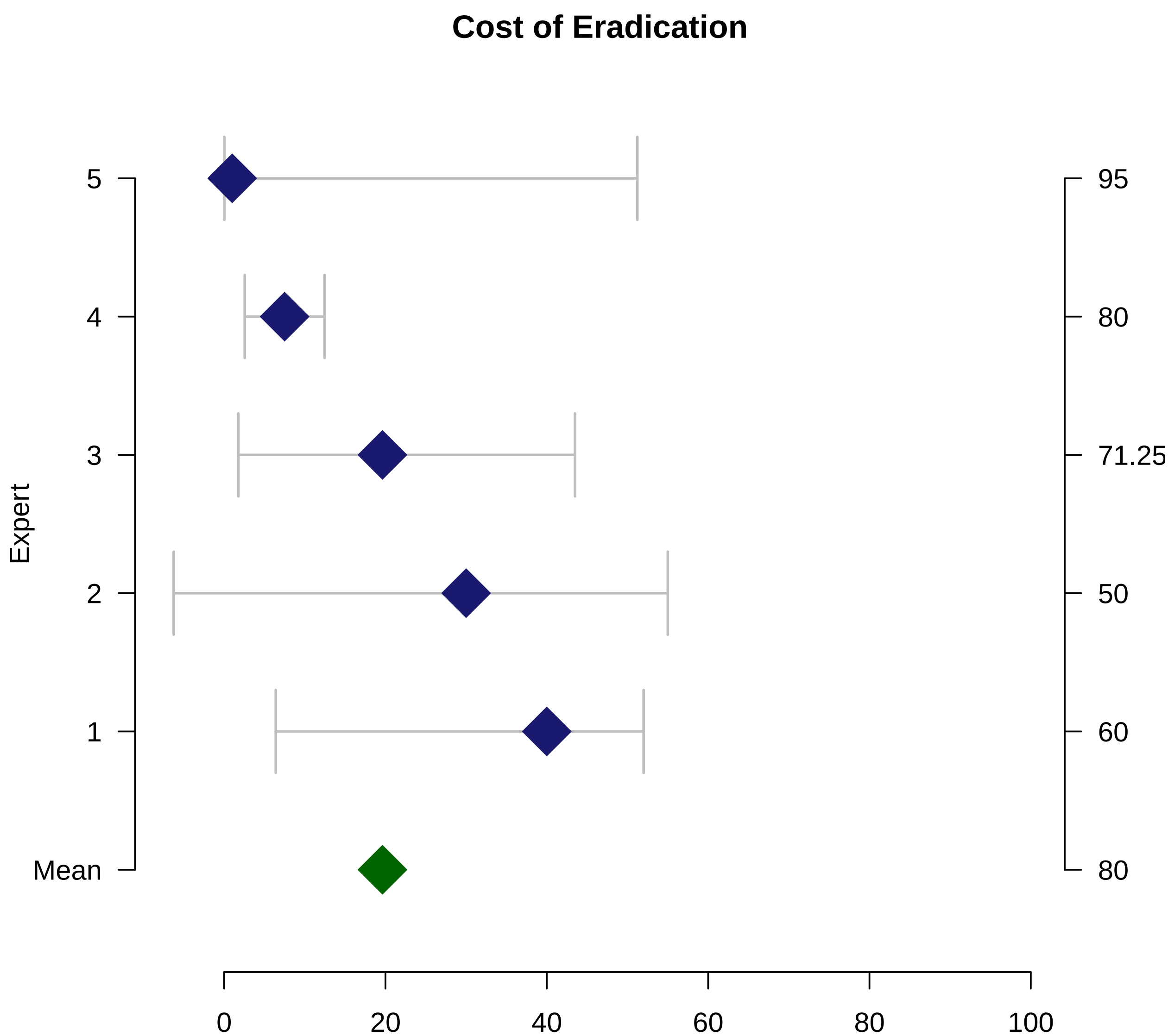

Figure C.17: Summary of expert elicited values for Question 13b: Cost of Eradication without Readiness.

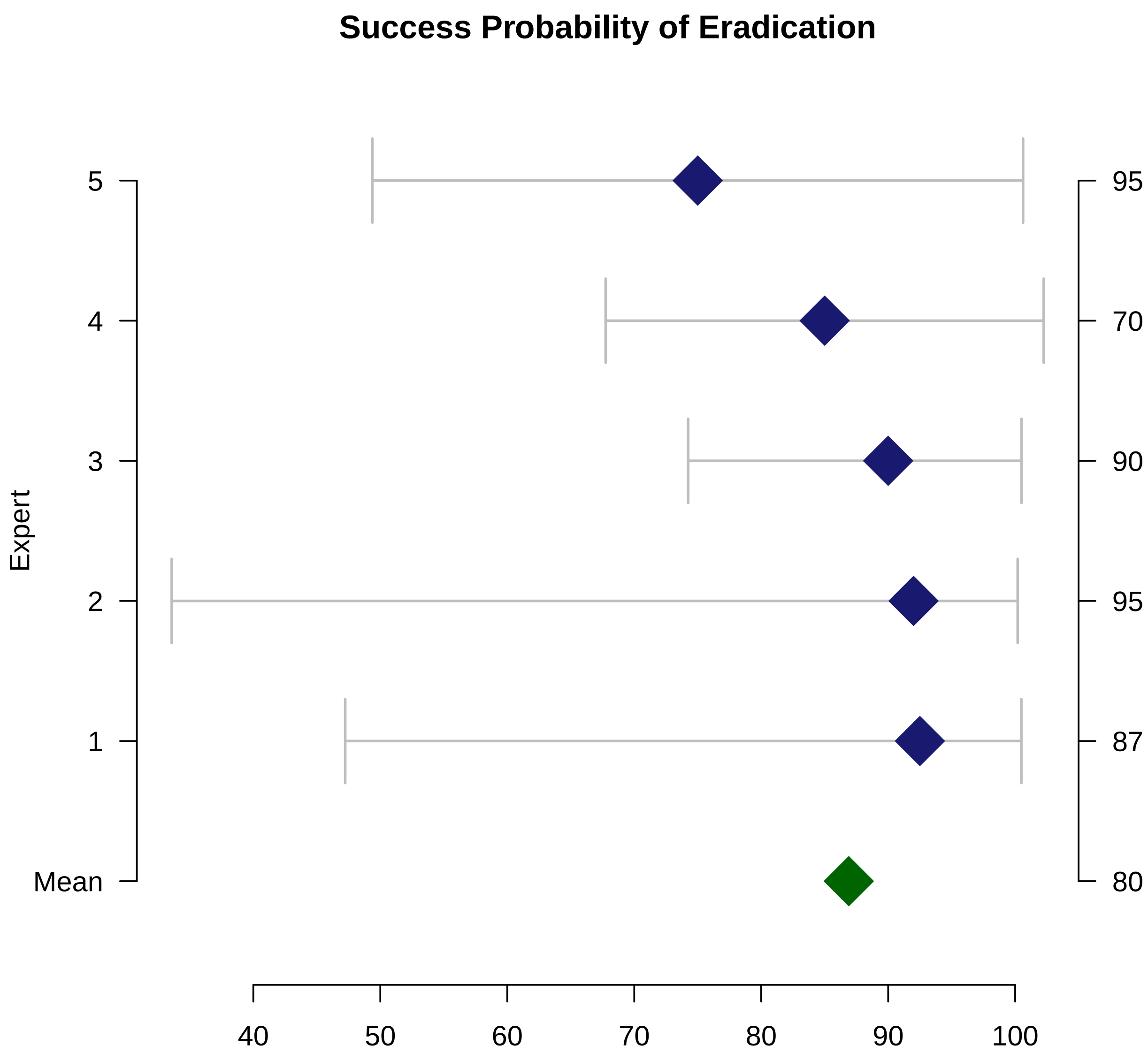

Figure C.18: Summary of expert elicited values for Question 14a: Success Prob. of Eradication with Readiness.

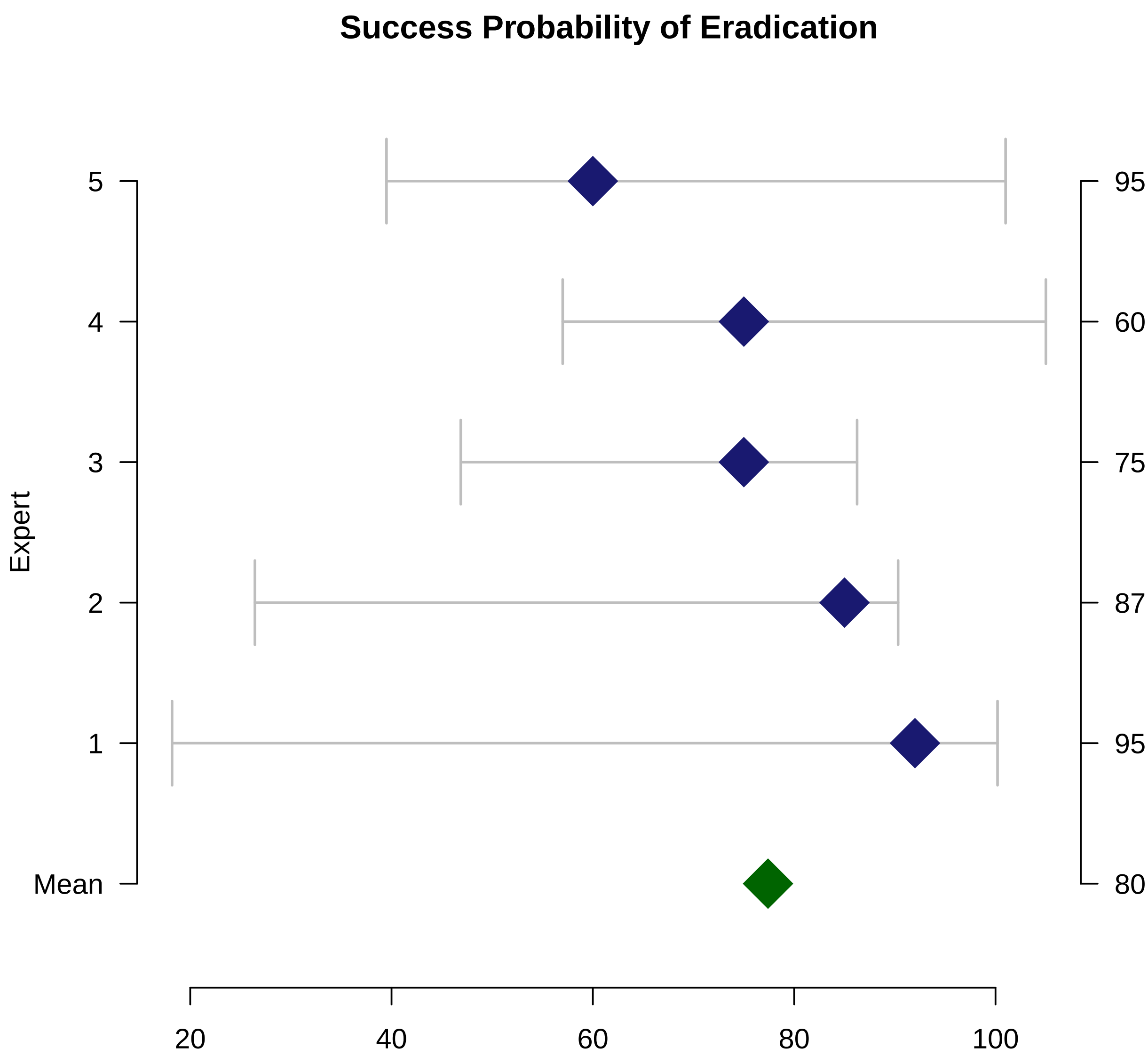

Figure C.19: Summary of expert elicited values for Question 14b: Success Prob. of Eradication without Readiness.

Appendix D

Conference Call Notes

David Diane Rory Stephanie Michael

Low – *High* should be like a confidence interval.

9a *Low*

might not be population outbreak that year visit might not coincide with flight area in port

High

High population in each ports Ship visit coincide with peak flight Ship visit multiple ports Ship longer stay in port Light at ports

Discussion: non-independent events. *High* number – ships coming over a season. Can't get 1000 ships into one port.

9b *Low*

No outbreak *Low* risk period of year Not high-risk areas

High

Unknown population in low-risk area Vessel could fail to report Atypical season that puts moth flight out of risk window. Host-tree availability within port

9c *Low*

Think of container being from inside the country. Don't worry about ship. Not necessarily on the same ship. Separate from ship infestation. More likely to be infested if port infested. Make them independent? Ignore ship.

See vessels. Population, timing.

High

Containers sit on port for weeks – at facility as well.

Wording – number that could be carrying. Interpreted as outside possibility. Answered in more a statistical probability sense.

9d *Low*

Same as low ships – low-risk areas, time of year, population fails.

High

Definition of LR times not germane to containers. May have come from distant area. Amount of time for container as oviposition site is huge. LR containers same as LR ships don't expect to come from risky times and places. Ships declare ports but containers could be from anywhere. Making assumptions about location. How can we know the risk of containers? Make an assumption about where it's been located from goods. LR ship is LR because it e.g. visited port outside window. Container can't have designation because don't know location in moth flight. MPI makes assumptions but higher uncertainty. Most containers don't move too far from port. We know about container movements and pattern of movements. Although LR many more containers therefore more risk? Correct for volume later.

9e *Low*

Ship vs container vs vehicle – sheer size. More surface on ship.

High

Same as containers – stored outside. More nooks and crannies for eggs, less likely to be removed by elements and predators.

9f *Low*

Same as vessels and containers. *Low*-risk ports, population fail, low-risk time of year.

Definition – how do we define low-risk vehicle? Is there a definition of low-risk vehicle or container distinct from low-risk vessel? Definition should be the same but confidence could be less.

Vehicles could have travelled very far. Lifespan of egg mass could be 9 months. What's a LR vehicle? Thinking about origin, e.g. Japan.

Compare pathways – similarly characterized vessel brief, container longer, used vehicle and machinery longer and nooks and crannies.

Question 10.

10a

Certifying that it doesn't need inspection or that it has been inspected.

Low

RM has done offshore. Looking for a needle in a haystack. Very hard to do. 4-5 hours 2-3 people. DG has done in Port of Vancouver. As close to impossible as one can get. RM and DG no doubt dedicated to purpose. Give even less confidence for inspection by export authorities.

High

Inspectors are trained, given tools (mirrors), know where to look. DG: R, were you given access to all areas of vessel? E.g. containers? RM: Yes but I inspected bulk carriers / oil tankers. DG: We weren't allowed into unsafe areas. Single egg mass may not be alone. What if you find one? Inspection becomes more vigorous. Finding the last egg mass is still difficult. In NZ, can hold offshore if many masses. In Canada, one egg mass means stay offshore and clean, and apply for reinspection.

10b

Low

Nooks and crannies – difficult to inspect entire item. Should be easier than a vessel. Inspection might end early. Commonly found on the insides of wheels.

High

Greater efficiency – smaller object, rate of detection should be higher.

10c

Low

Treatment failure – undetected. Pockets where heat didn't reach specified temperature.

High Facility should be certified and audited Question: where are egg masses most likely to be laid? Easily exposed or in a nook? Mostly exposed. But no data. Not very outside but should be accessible. Does circulation take place? Yes, car raised, have fans, open boot and bonnet.

10d

Low

Could be full of material, small objects? Assumption: external inspection. Internal separate; please ignore. Some are reefers; most likely to be 4-sided inspection only. Not bottom. 6-sided very difficult. GM do tend to lay egg masses on underside of objects.

High

Much easier to inspect than vehicle or vessel – few nooks and crannies. What if one is detected? – 6-sided clean. If multiple egg masses, only need to detect one.

Q 11

11a clarify egg mass has evaded intervention

Low

Only way to get off ship is by hatching and ballooning at the right time. Depends on time of year – winter. Not all egg masses viable. Assume viability. Q 9? Or Q 11? What factors reduce viability? Transit aspects? Temperature extremes? Cold down to -32. Do quite well in North America. Not likely to be killed by temperature common in transit. Can affect development. Some egg masses tucked away. Cooked by direct sunlight – up to 60-70 degrees.

High

Wind is blowing right towards lovely oak forest Vessels could be highly contaminated – vast majority could hatch while in port.

Why are the highs so low? Percent of hatching survival, finding food, pupal survival, finding a mate.

11b

Low

Containers have same percentages as ships. Containers can get hot – reducing viability.

High

Containers moved inland, and stay for a while. Could sit next to a forest. Survival would increase quite a bit.

11c

Low

Same reasons as vessels and container. How do the larvae get out if concealed in a nook? Crawls. How far? Crawl upward.

High

Used vehicles and equipment tend to stay.

12a Hectares

Low

Trapping program. Can detect in first year. NZ monitoring does a good job. 1530 traps in urban network. Could be detected really early. If picked up in trap then response is to flood area. Respond to single live male. Do have a very good passive surveillance system as well. Beyond a certain size.

High

In an area not trapped or surveyed.

12b Counterfactual – don't exclude passive surveillance.

Passive surveillance in play. Obvious defoliation

High

Could take 10 years for defoliation – traps better for detecting low levels. *Low* population density – no one sees it. Native moths that look similar. Get lots of false positive reports. Traps all in urban areas.

In Oregon only trapped HR areas; remote incursion grew to large area. NZ doesn't have oak forests. Therefore not deemed as likely for native forests. There are plantings.

13a

Low

Based on values from a paper using rate of 166.73 per ha for treatment. Assumed treat once and done. Infestation size in paper? Do have link. Fair proportion will be fixed cost. Probably just spraying. What was actually spent in Hamilton incursion. Single moth. Spent 5 million. Sprayed at least a 500 ha zone. 1253 ha sprayed. Gets back to how to define the area. May trap males over 500 ha but not all be treated.

Treatment area or incursion area? Treatment zone 500 ha. Trapping zone might be larger. Within 500 ha might be non-suitable habitat. Going for treatment zone.

High

Something goes wrong – have to continue to treat. Public objections that drag out, force legal action. Aerial treatment might be impossible.

13b Counterfactual – pest the same as GM but unknown. Don't have lure or effective treatment. Think about timing. More time to build, therefore longer to remove. Part of this is having personnel in place.

Low

Might be a treatment available and might be a lure – just by good luck.

High

Takes up to 1-2 year to get chemical application registration approvals. Noone trained. Ongoing surveillance to determine the effect of the treatment.

14 Don't use probability.

14a Use 13a as a budget limit.

Low

Inadequate delimitation. Could have been inaccurate delimitation. 2-3 populations. Sandy thinks always 100% success. Might spend a lot of money. But, ok. Prepared or not, it's pretty easy to eradicate. Take a particular scenario. Question 13 has made some assumptions – e.g. one treatment. Q 14 asks: based on investment in 13, what is probability?

High

Readiness

14b