Research Article

Received: 30 August 2022 Revised: 18 October 2022 Accepted article published: 29 October 2022 Published online in Wiley Online Library: 11 November 2022

(wileyonlinelibrary.com) DOI 10.1002/ps.7257

Factors impacting the detection of weed seed contaminants in seed lots

Christopher E Buddenhagen,* Deborah Hackell, Harold V Henderson and Ben Wynne-Jones

Abstract

BACKGROUND: The setting and following of phytosanitary standards for weed seeds can lessen the impacts of weeds on agriculture. Standards adopted by seed companies, laboratories and regulators ensure the contamination rates do not exceed some thresholds. Globally sample size standards are set based on the amount needed to obtain a contaminant in a random sample of the seed lot, not detectability. New Zealand requires a 95% confidence that the maximum pest limit of 0.01% of quarantine weed seed contamination is not exceeded in an imported seed lot. We examined 24 samples each containing approximately 150 000 seeds of either perennial ryegrass (12 samples) or white clover seeds (12 samples) that were then spiked with seeds (contaminants) from 12 non-crop species (3–8 seeds of each). We considered factors that may impact detection rates: shape, color, size, and texture relative to the crop, and technician (including a commercial seed laboratory).

RESULTS: A linear mixed model fitted to the data indicated significant observer, crop, and seed color, shape, and size effects on detection. Detectability increased by 20% ± 7.7 (± standard error) when seeds had a distinct shape or color (28% ± 8.1), or were larger (23% ± 8.7) rather than smaller, relative to the crop. Commercial laboratory identifications were usually correct at the level of genus, and species for common weeds, but some misidentifications occurred.

CONCLUSION: Sample sizes for border inspections should be based on detectability of regulated weed seeds in the crop in combination with weed risk for the crop and location.

Supporting information may be found in the online version of this article.

Keywords: biosecurity; threshold; international seed testing association; phytosanitary; regulations; National Plant Protection Organisation

1 INTRODUCTION

Seeds are a key agricultural input and output and are a part of a complex global trade network.[1,10] However, as a global commodity, contaminated seed for sowing can provide a pathway for the introduction and establishment of pests, weeds and diseases with the potential to cause unwanted impacts across field, farm, regional or national scales.[1–3] These unintended introductions contribute to the homogenization of world floras and are a threat to global food security. Mitigating the impacts of weeds, pests, diseases or other invasive species in seed-for-sowing systems is a shared goal of farmers, seed producers and national/international plant protection organizations. Phytosanitary, quality and seed purity standards are used to mitigate these impacts and have a long history in the agricultural sector.[4–6] Farmers want to avoid any crop-weed competition and seed contamination such that the impacts are negligible, while government regulators want to avoid the introduction of biosecurity risk organisms while supporting free and fair enterprise.[7–9] At the field-level, management of pests, weeds and diseases can be mitigated by seed treatments or seed cleaning. The effectiveness of such measures is confirmed by seed inspection, including analytical purity tests, and seed certification, carried out by seed laboratories linked to regulatory authorities or commercial seed companies.[10] At a national/international level, the detection of regulated or problem seed contaminants can be mitigated through removal of seed lots from certification, destruction, cleaning or reshipping. These mitigation measures rely on the detection of weed seeds in seed lots. Factors contributing to successful detection of contaminant weed seeds in seed lots are therefore important and the focus of this work.

The belief is that these prevention measures, even if imperfect, are likely to reduce the establishment of unwanted weeds, pests or diseases within the balance of logistics and cost effectiveness. The evidence for pests other than weeds is that imperfect

* Correspondence to: CE Buddenhagen, AgResearch Ltd, Ruakura Research Centre, 10 Bisley Road, Hamilton, 3214, New Zealand. E-mail: chris.buddenhagen@agresearch.co.nz

† Better Border Biosecurity research collaboration (www.b3nz.org.nz).

AgResearch Ltd, Ruakura Research Centre, Hamilton, New Zealand

prevention efforts are cost effective.[11,12] Detections of new seeds of concern at the border prior to clearance in New Zealand, are rare, given the high inspection rates, with seed cleaning and inspections occurring offshore prior to shipment, and on arrival in New Zealand.[2,10] Post-border incursions (*i.e.*, detection of newly imported regulated weeds growing in fields New Zealand) are also rare.[2,10] It is possible that seed importers are taking extra steps to reduce seed contaminants to avoid problems (such as costly destruction or reshipping) due to the high inspection rates at the New Zealand border for seeds for sowing imported into New Zealand.

Seed sampling and inspection protocols are designed to obtain an unbiased submitted or 'working' sample of seed from the lot for analysis using methods set by the International Seed Testing Association (ISTA), where the seed lots may range in size from a few grams to 10–30 t.[13] The principle is that after the representative random sample from the seed lot is inspected a seed lot can be deemed to have contamination rates that are below an acceptable threshold and with a known certainty. Thresholds are estimated using binomial probabilities that a sample will by chance contain a contaminant.[13-15] For example, within New Zealand, inspections of a certain sample size of crop seeds allow inspectors to claim (if nothing is detected) with 95% confidence[17,18] that contamination rates are less than 0.5% or 1% for seed certification (all weeds can be considered), or less than 0.01% for regulated seeds in imported seed lots inspected at the border.[14,19] At the border there is an implied acceptance that some contaminants will be introduced with the crop, and as many common weeds are deemed non-regulated weeds, observed contamination rates can be significant considering sowing rates.[20] Within a 10-t seed lot, where the thousand-seed-weight is 2 g, the shipment could contain 5 billion crop seeds and with a contamination rate of just 0.005% a seed lot this big could contain 250 000 unwanted seeds (500 seeds per hectare if seeds are planted at 20 kg a hectare). An important caveat is that seed certification typically focuses on the overall varietal and analytical purity of the sampled seed lot but may be largely unconcerned about which species of weed seed are detected. Meanwhile, for inspections of imported seed for biosecurity purposes, the regulators focus on the mitigation of risk from regulated species but are unconcerned about many common non-regulated weed seeds (species are assessed as low risk) that may be present. A fair number of species are regulated for reasons other than their potential to be a weed, *e.g.*, wheat seeds as a contaminant of another crop, and these may need to meet treatment or cleaning standards, unrelated to their weed risk.

Standards adopted by regulators and used by laboratories focus on ensuring the contamination rates are not higher than some threshold level. Recommended sample rates focus on the sample size needed to obtain a contaminant in a random sample of the seed lot, not detectability. Detectability relates to the probability that an observer will find a non-crop seed in a sample when it is there. For hard-to-detect species, certainty about whether a seed lot is contaminated could be lower than indicated by a binomial threshold calculated from sampling rates. We hypothesize that detection rates will surely vary between species, depending on weed species of interest, and the seeds' size, shape, texture, and color relative to the crop seed it is found in. The International Seed Testing Association does occasionally release data about detection rates e.g.,[21,22] which are routinely assessed in proficiency tests which are part of seed laboratory accreditation processes.[23] The association expect laboratories to be able to identify seeds of specified contaminant species (83 weeds) and 47 crop species.[24] Across the laboratories globally, with inspection of samples of 25 000 crops seeds spiked with 2–4 seeds, rates of detection of weeds in perennial ryegrass and white clover ranged from 37–89% to 69–87%, respectively.[21,22] Here we explore factors influencing the efficacy of weed seed detection and the biosecurity implications of imperfect detection.

2 METHODS

To determine detection rates for non-crop seeds in seed for sowing samples, we spiked 24 samples with contaminant seeds of 12 non-crop (weed) species. We used 12 samples of seed for sowing 'crops' perennial ryegrass (*Lolium perenne* L.; variety 'Maverick', 300 g) and a further 12 samples of white clover (*Trifolium repens* L.; variety 'Quartz', 100 g) seed for sowing sourced from seed lots multiplied in New Zealand. We chose these two crop species because they provide a large proportion of temperate grazing forage in New Zealand.[25] Also the importation of basic seed of these two species for multiplication and reexport is a common profitable enterprise for New Zealand farmers.[10] New Zealand is the largest global clover seed exporter, providing 16% of the global supply and is the fourth largest ryegrass seed exporter, providing 11% of the global supply.[26,27] These samples were *five times* larger than the minimum working samples for 'Determination of other seeds by number' specified in Table 3 of Chapter 2 in The International Rules for Seed Testing 2021.[14,28] The purpose of the 'Determination of other seeds by number' is to identify and count all non-crop seed contaminants. The amounts we used (100 g for white clover and 300 g for ryegrass) represent approximately 125 000–150 000 or more crop seeds. This sample size is nominally five times larger than is used in proficiency assessments where ≥25 000 seeds are used. The one times ISTA amounts are 60 g and 20 g, respectively, for ryegrass and clover in the sampling chapter of the rules.[13] According to a local practice implemented by the Ministry for Primary Industries (MPI) this 'five-times ISTA' sample amount (described above) is normally inspected for all incoming perennial ryegrass and white clover shipments.[14]

To reduce the number of non-crop seeds (other than the spiked seeds) we searched for and removed seed contaminants we could find before we added known quantities of non-crop seed to each sample. Non-crop seeds for spiking were selected and classified for analysis based on their size (similar, larger, or smaller), shape and color (similar or distinct) compared with the crop (Table 1). Seeds were not selected based on whether they are likely to occur as contaminants of imported perennial ryegrass and white clover seed lots. Non-crop seeds were sourced from the field, or from a collection of seeds maintained for the writing of a weed seed guidebook in New Zealand.[29] Out of the 12 species, eight of the spiked contaminants in perennial ryegrass were known to occur in imported perennial ryegrass seed (*Alopecurus myosuroides* Huds*, Chenopodium album* L.*, Dactylis glomerata* L.*, Echinochloa crus-galli* (L.) P. Beauv.*, Elymus repens* (L.) Gould*, Festuca rubra* L.*, Poa annua* L. and *Vulpia bromoides* (L.) Gray); meanwhile only two (*Melilotus albus* Medik and *Chenopodium album*) were known to occur in imported white clover seed lots.[2] All spiked contaminants shown in Supplemental Figs. S1 and S2. All but two of the spiked species had known congeners previously detected in imported seed lots for each crop, *i.e.*, in perennial ryegrass seed the exceptions were, *Piptatherum miliaceum* (L.) Cross, and *Anthosachne kingiana* (Endl.) Govaerts and in white clover seed, *Eragrostis cilianensis* (All.) Vignolo ex Janch., and *Hypericum androsaemum*

Table 1. Classes of seed size, color, and shape relative to the crop used in the model in Table 3

Species	Crop	Seed size mm^2	Relative size	Relative Color	Relative Shape	Texture
Alopecurus myosuroides Huds,	Perennial ryegrass	7.02	larger	different	similar	rough
Amaranthus blitum (L.) Costea	White clover	1.77	larger	different	different	rough
Anthosachne kingiana (Endl.) Govaerts	Perennial ryegrass	12.20	larger	similar	different	smooth
Bromus inermis Leyss.	Perennial ryegrass	14.46	larger	different	similar	rough
Bromus tectorum L.	Perennial ryegrass	7.30	larger	different	different	smooth
Chenopodium album L.	White clover	0.81	smaller	different	different	rough
Chenopodium album L.	Perennial ryegrass	0.81	smaller	different	different	smooth
Cuscuta campestris Yunck.	White clover	1.09	similar	similar	similar	rough
Dactylis glomerata L.	Perennial ryegrass	3.74	smaller	similar	similar	smooth
Echinochloa crus-galli (L.) P. Beauv.	Perennial ryegrass	4.72	similar	similar	different	rough
Eleusine tristachya (Lam.) Lam.	White clover	0.85	similar	different	different	rough
Elymus repens (L.) Gould	Perennial ryegrass	16.60	larger	similar	different	rough
Eragrostis cilianensis (All.) Vignolo ex Janch.	White clover	0.34	smaller	similar	different	smooth
Festuca rubra L.	Perennial ryegrass	2.87	smaller	similar	similar	smooth
Hypericum androsaemum L.	White clover	0.31	smaller	similar	different	rough
Melilotus albus Medik.	White clover	2.34	larger	similar	similar	smooth
Piptatherum miliaceum (L.) Coss.	Perennial ryegrass	2.55	smaller	different	different	smooth
Plantago major L.	White clover	0.93	similar	different	different	rough
Poa annua L.	Perennial ryegrass	1.53	smaller	similar	different	rough
Ranunculus flammula L.	White clover	1.17	similar	similar	different	rough
Ranunculus sceleratus L.	White clover	0.68	smaller	similar	similar	rough
Sinapis alba L.	White clover	2.22	larger	similar	different	rough
Solanum mauritianum Scop.	White clover	1.93	larger	similar	different	rough
Vulpia bromoides (L.) Gray	Perennial ryegrass	6.86	similar	different	different	smooth

Note: White clover and ryegrass seeds had an area estimate of 1.06 and 5.33 mm^2, respectively, and were both classed as having a smooth texture. See Supplemental Figs. S1 and S2 for scaled color images of the seeds.

L. Only the spiked species *Alopecurus myosuroides* (added to perennial ryegrass seed), and *Hypericum androsaemum*, and *Solanum mauritianum* Scop. (added to white clover seed) were on the list of quarantine weeds for New Zealand.[30] Before sending the samples for inspection the seeds were all devitalized by heat treatment, (93°C for 23 h) higher than the required 85°C for 15 h specified for devitalization of imported seeds for human consumption.[31] The 24 seed samples were spiked with 3, 4, 5, 6, 7 or 8 seeds (*i.e.*, six levels were considered) of each non-crop species, such that each rate was duplicated in two seed samples. For the 12 species, the number of spiked seeds, *i.e.*, within the six levels, was randomly allocated to seed samples – the amount of perennial ryegrass and white clover seed for each sample is described above. Samples had identification codes so that spiked seed numbers could be tracked.

We sent our samples to an ISTA accredited seed lab in New Zealand to analyze the samples in October 2021. This 'analysis' involves separating contaminants, identifying them to species or genus and returning the samples and the separated contaminants. The inspectors at this lab had several years of experience and regularly inspect samples from most large ryegrass and clover seed lots imported into New Zealand (of the same sample size). We asked that they inspect the whole sample (not a subsample), as described above, the samples were five times larger than the minimum working samples for the 'Determination of other seeds by number' specified in the rules (sample weights are mentioned above).[13] We were able to then reconcile their identification with the species we had added. To get an estimate of variability in detection between observers, we had our two technicians reinspect the same carefully reconstituted samples. They used a stereo zoom microscope (magnification range 10 × to 300 ×) to find non-crop seeds. For our technicians inspections took between an hour and 2 h for each sample. Our technicians regularly work with seeds and microscopes but do not routinely inspect seed lots. For the searches carried out by technicians in our lab (repeated on the same samples) we asked only that the non-crop seed be separated (no ID was required), and the detected seeds were reconciled with the species we had added, and detections were counted. The technicians (1 and 2) who did the inspections were naive to the number and type of species spiked into the samples. A third technician set up the experiment and spiked the samples, and reconciled the seeds found with the known seeds spiked. *Chenopodium album* L. and *Poa annua* L. were found in some samples at higher than expected based on our spiking. The *Poa annua* we spiked was distinguishable from the existing sample contaminants because the seeds we added were separated into individual seeds while the contaminants were clumped with two or more seeds per clump.

We compared the numbers of seeds detected by our technicians and the ISTA seed lab to those added for each sample and species in terms of color, shape, texture and size. In the case of the seed lab analyzed samples we were also able to compare their species identifications with the known species we added (Figs. 1 and 2). Size differences from the crop seed were estimated in imageJ (version 1.53), by estimating average seed area in square millimeters from photos of five seeds.[32] After setting the scale the 'make binary' and 'analyze particle' functions were used to estimate seed areas. Seed ratios of <0.8 relative to the crop were classed as smaller, and larger if the ratio was >1.3, seeds between these values were similar (Table 1). The R statistical platform was

used for data analysis. Data were manipulated and plotted using the 'tidyverse' package.[33,34] For our analyses of detection rates in relation to technician/lab and seed factors (Table 1) we used R packages for linear mixed-effects models lme4, lmerTest, and for summarizing model outputs emmeans.[35–38] The linear mixed model for detection rate specified fixed effects Relative size, Relative color, Shape, Texture, Crop, Technician and Crop by Technician interaction and random effects bag and weed species (Supplemental Fig. S3 shows the weed species random effects). The model specification process we followed involved stepwise removal of non-significant variables from the full model. We included the non-significant variable texture because we expected it to influence detection and wanted to show that it did not. We also tried the generalized linear mixed model (glmer function) with a binomial distribution which produced similar results in terms of the significant variables, but we chose the linear mixed model for the percent of detected seeds because the results were easily interpretable.

We used the binomial distribution (dbinom, function in base R) to calculate percent of shipments with contaminants that might be expected to be missed at different rates of contamination and inspected sample size. The probability that an inspection fails to detect a contaminant in a seed lot involves two aspects, (i) the probability that rare contaminants would not by chance occur in the random sample of the seed lot, (ii) the probability of missing it when it is in the sample because of imperfect detection rates. We considered the recommended rates of inspection for purity analysis, the minimum working samples for 'Determination of other seeds by number' (1 × ISTA) and five times that rate 5 × ISTA; discussed above.[13] The 1 × ISTA rate is designed to sample a minimum 25 000 crop or host seeds but the standard sample weights seem conservative, with published 1000-seed weights indicating the samples will often tend toward 30 000 seeds. This is the estimate of crop seed number that we used in our binomial calculations in Fig. 3. For example, for *Lolium perenne* 1000 seed weights can range from 1.38 to 3.86 g and for *Trifolium repens* the range is between 0.5 and 0.8 g.[39–41]

3 RESULTS

3.1 ISTA seed lab identifications

Weed seed identifications were carried out on samples by the ISTA accredited lab, whereas the technicians in our lab only separated non-crop seeds from the sample and did not attempt to identify contaminants to species. As noted in the methods section the species included were not necessarily the species expected to be found in perennial ryegrass and white clover seed lots. Rates of detection (Table 2) and identification accuracy varied between species and crops (Figs. 1 and 2). For white clover, of the 62% of seeds detected by the seed laboratory, 49% were correctly identified to species, and an additional 36% to genus. Species correctly identified most of the time in white clover were, *Chenopodium album*, *Sinapsis alba* L., *Plantago major* L. and *Melilotus albus*, as well as the distinctive regulated quarantine weed *Hypericum androsaemum*. In perennial ryegrass samples, of the 57% of seed found, 70% were correctly identified to species, and additional 15% to genus. Species correctly identified most of the time in perennial ryegrass were *Chenopodium album, Vulpia bromoides, Echinochloa crus-galli, Bromus inermis* Leyss., *Elymus repens, Poa annua, Pipteratherum miliaceum, Dactylis glomerata* and *Festuca rubra*, as well as the quarantine weed *Alopecurus myosuroides*. In perennial ryegrass, *Bromus tectorum* was identified as *Bromus sterilis*. The quarantine weeds *Alopecurus myosuroides, Hypericum androsaemum*, and *Solanum mauritianum* were found 59%, 65%, 87% of the time and identified correctly 56%, 61% and 0% of the time, respectively. We know that two

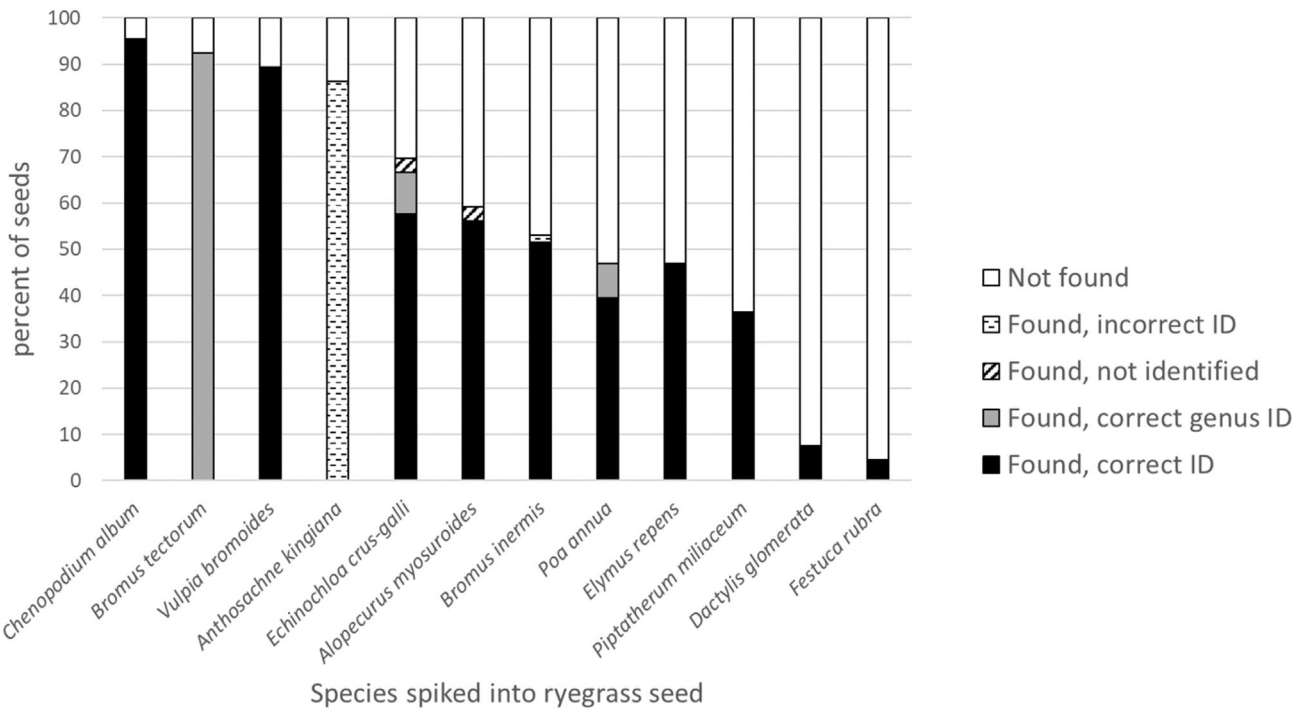

Figure 1. Seeds found and correctly or incorrectly identified by ISTA seed analysts during inspections of in 12 samples of 300 g (approximately 150 000 seeds) of perennial ryegrass (*Lolium perenne*) seed spiked with 3–8 seeds of the 12 species indicated (a total of 66 weed seeds of each species were spiked).

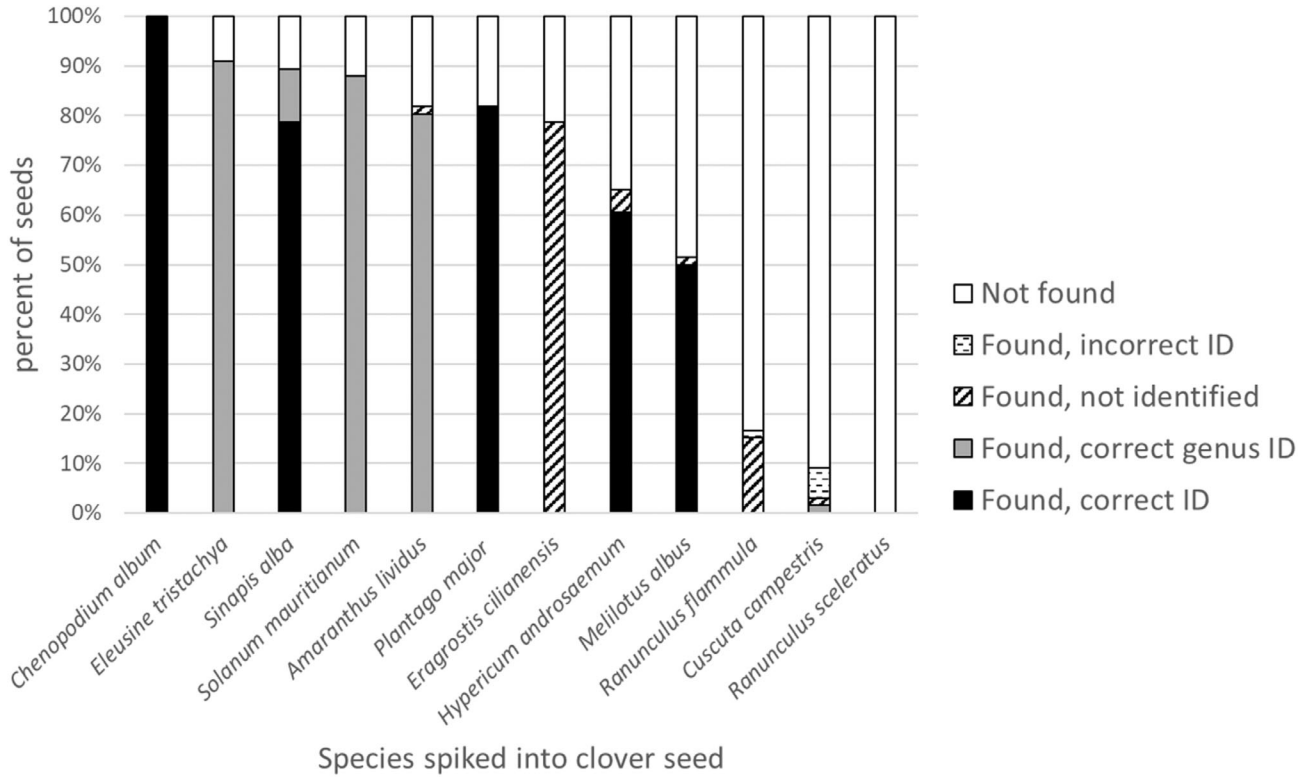

Figure 2. Seeds found and correctly or incorrectly identified by ISTA seed analysts during inspections of 12 samples each of 100 g (approximately 150 000 seeds) of white clover (*Trifolium repens*) seed spiked with 3–8 seeds of the 12 species indicated (a total of 66 weed seeds of each species were spiked).

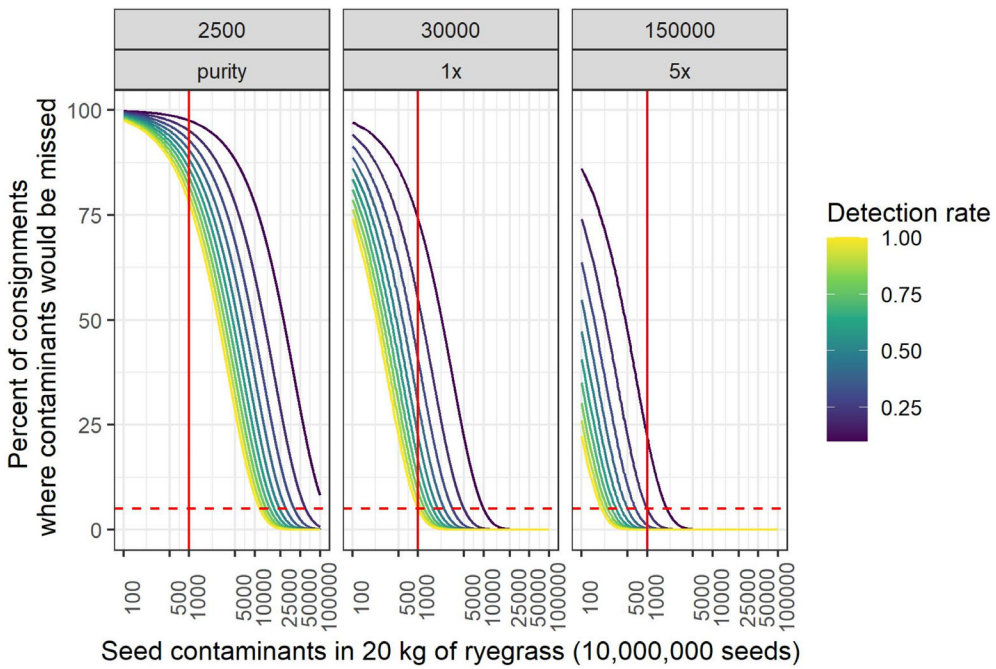

Figure 3. Binomial estimates of the proportion of consignments where contaminants would not be detected at three levels of inspection, assuming different contamination rates (x-axis) and detection rates (0.1–1 in increments of 0.1). Rates of contamination are expressed in terms of contaminants in 10 000,000 seeds, which is the approximate amount in a 20 kg bag of perennial ryegrass. In New Zealand, border inspections require 95% confidence that the maximum pest limit of 0.01% is not exceeded (vertical red line); the horizontal line indicates 5% are missed. The panels represent inspection rates relative to the ISTA sampling protocols (discussed in main text), purity analysis is sometimes carried out for the purpose of seed certification.

technicians at the ISTA accredited lab were involved in the inspections but could not get information about which samples they each inspected.

3.2 Variability in detection

Our samples had contamination rates between approximately 1:41000 and 1:16000 (*i.e.*, 3–8 seeds in approximately 150 000

Pest Management Science published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

Table 2. Number of seeds found out of the 66 that were spiked into 12 sample bags for each crop

Species	Crop	Seeds found		
		ISTA analysts	Tech 1	Tech 2
Alopecurus myosuroides Huds,	Perennial ryegrass	39	23	6
Amaranthus blitum (L.) Costea	White clover	54	46	53
Anthosachne kingiana (Endl.) Govaerts	Perennial ryegrass	57	18	17
Bromus inermis Leyss.	Perennial ryegrass	35	20	6
Bromus tectorum L.	Perennial ryegrass	61	37	47
Chenopodium album L.	White clover	66	60	62
Chenopodium album L.	Perennial ryegrass	63	17	54
Cuscuta campestris Yunck.	White clover	6	12	15
Dactylis glomerata L.	Perennial ryegrass	5	6	2
Echinochloa crus-galli (L.) P. Beauv.	Perennial ryegrass	46	34	18
Eleusine tristachya (Lam.) Lam.	White clover	60	55	59
Elymus repens (L.) Gould	Perennial ryegrass	31	9	6
Eragrostis cilianensis (All.) Vignolo ex Janch.	White clover	52	17	44
Festuca rubra L.	Perennial ryegrass	3	23	7
Hypericum androsaemum L.	White clover	43	10	27
Melilotus albus Medik.	White clover	34	37	34
Piptatherum miliaceum (L.) Coss.	Perennial ryegrass	24	2	6
Plantago major L.	White clover	54	30	39
Poa annua L.	Perennial ryegrass	31	0	2
Ranunculus flammula L.	White clover	11	14	38
Ranunculus sceleratus L.	White clover	0	0	11
Sinapis alba L.	White clover	59	48	59
Solanum mauritianum Scop.	White clover	58	59	55
Vulpia bromoides (L.) Gray	Perennial ryegrass	59	33	39

Note: Data distinguish two technicians in our lab *versus* results provided by the ISTA seed analysts.

Table 3. Inference on the parameters of the linear mixed-effects model on the percent of seed detected in samples using the lmer package in R. Satterthwaite's approximation is used to calculate degrees of freedom

Fixed effects	Estimate	Standard Error	Df	Pr(>|t|)
(Intercept)	18.0	10.76	17.5	0.11
Relative size (similar)	7.6	9.85	17.1	0.45
Relative size (larger)	23.2	8.69	16.5	0.016*
Relative color (different)	20.1	7.72	16.6	0.019*
Shape (different)	27.6	8.06	16.5	0.003**
Texture (rough)	−6.2	7.38	20.0	0.41
Crop (perennial ryegrass)	−30.4	7.75	24.6	0.0006***
Technician (2)	13.9	2.77	813.9	7 e-07***
Technician (Seed lab)	14.5	2.77	813.9	2 e-07***
Crop (perennial ryegrass): Technician (2)	−15.7	3.92	813.9	7 e-05***
Crop (perennial ryegrass): Technician (Seed lab)	15.5	3.92	813.9	9 e-05***

Random effects	Standard Deviation
Bag	3.7
Species	16.9
Residual	23.5

*** <0.001;
** <0.01;
*<0.05.

seeds or 5 × ISTA recommended working sample). The seed lab inspections at an accredited ISTA lab detected 5–95% of spiked contaminants for perennial ryegrass, and 0–100% for white clover (Table 2). The two technicians in our lab were able to detect 49% ± 2 (binomial SE for p = proportion from n = 792 seeds; sqrt(p (1-p)/n)) and 63% ± 2 of the white clover but performed less well in

Pest Management Science published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

Table 4. Predicted mean detection rates (%) for size, shape, color and texture from the model specified in Table 3, with the average standard error of the differences (SED), using the emmeans (expected marginal means) package in R.

Factor	Descriptor	Predicted Mean	Df	Letter Group	Average SED
Size	smaller	33.1	17.1	A_	9.3
	similar	40.7	17.5	AB	9.3
	larger	56.3	17.1	_B	9.3
Color	similar	33.3	17.8	A_	7.8
	different	53.4	17.0	_B	7.8
Shape	similar	29.5	17.0	A_	8.1
	different	57.1	17.7	_B	8.1
Texture	smooth	46.4	18.8	A_	7.4
	rough	40.2	18.6	A_	7.4

Note: Different letter groups within a factor indicate significant differences in the predicted means ($P < 0.05$).

the perennial ryegrass case (where most seeds were also grass species), detecting just 28% ± 2 and 27% ± 2 of spiked seeds. The overall detection rate for non-crop spiked seeds was 37% ± 10 (SE among the two technicians and the lab) and 58% ± 5 for perennial ryegrass and white clover, respectively. Importantly though, there appeared to be blind spots for some species, with differences between inspectors *e.g.*, in white clover the seed lab and Tech 1 failed to detect *Ranunculus flammula* L., but Tech 2 detected 11 of the 66 seeds (Table 2). The seed lab did not detect many of the *Cuscuta campestris* Yunck. seeds either. The ISTA lab was good at detecting *Poa annua, Piptatherum miliaceum, and Elymus repens* compared with the technicians (Table 2). Overall performance of the experienced seed analysts at an ISTA accredited seed lab was higher; the mean difference in detection was 18% higher (Table 2). *Chenopodium album, Eleusine tristachya* (Lam.) Lam. and *Solanum mauritianum* (regulated) were amongst the easiest to detect, while the *Bromus* spp., *Alopecurus myosuroides* (regulated), and *Hypericum androsaemum* (regulated) were intermediate, *Dactylis glomerata* and *Ranunculus flammula* were the most difficult (Table 2, Supplemental Fig. S3).

A linear mixed model of detection was fitted on the combined data (Table 3). In the model, with all inspections combined, the technician and crop:technician interaction was highly significant (Table 3). The number of seeds spiked per sample did not have a significant effect on detection, and it was removed from the model. The predicted mean rates at which seeds were detected for seed sizes that were larger, similar or smaller than the crop were, respectively, 56% ± 6, 41% ± 8, 33% ± 7, though the difference was only significantly different between the largest and smallest seeds (23% ± 8.7 SED, Table 3, Table 4). Seeds similar in color to the crop were significantly less likely (20% ± 7.7%) to be detected; the predicted mean was 33% ± 5 as opposed to those that were distinctly different in color 53% ± 6 (Table 4) and the linear model fit showed a *p*-value of 0.02 (Table 3). Seeds distinct in shape were (28% ± 8.1 SED) easier to find than those that were similar (Table 3), the expected marginal means estimate showed detection rates of 29% ± 7 for similar shaped seeds *vs* 57% ± 4 for distinctly shaped seeds (Table 4). There was no significant impact of texture on detectability (Tables 3 and 4).

3.3 Predicted detection rates at different thresholds

The rules for the 'Determination of other seeds by number' aim for at least 25 000 seeds but seed weights suggest that approximately 30 000 seeds are inspected, and if no species are detected, assuming a close to 100% efficacy of detection, then there is a 95% probability that the shipment has less than 0.01% for imported seed lots inspected at the border (Fig. 3). Obtaining a non-detection of quarantine species is required for offshore seed lots to be certified for export to New Zealand but because MPI inspects all imported seed lots at the NZ border to verify that they meet import requirements at the much higher 5 × ISTA sample size, it could detect weeds that meet that threshold, even with a 20% detection rate (Fig. 3). None of the non-crop seeds in this study had a 100% detection rate though *Chenopodium album* came close. Given the quarantine weed detection rates for the ISTA lab for *Alopecurus myosuroides, Hypericum androsaemum* and *Solanum mauritianum* of 59%, 65%, 88%, respectively, and assuming identifications were perfect (they were not in this study), >1 × ISTA would be needed to meet the standard (Table 2, Fig. 3).

4 DISCUSSION

The detectability of non-crop seeds in seed lots requires the observer to distinguish the contaminant seed from the crop seed. We showed that size, shape and color differences between the crop and the non-crop contaminant do significantly impact detectability – by as much as 20–30%. Our data did not show any detectability differences for seeds with a distinct texture despite that being a distinguishing feature for the identification of some seeds. A larger study with more samples that included more observers, crops and contaminants could allow us to make broader generalizations about seed detectability. The power of our model could also be improved by a balanced sample design for the number of species in each combination of crop, size, shape, color and texture class. Nevertheless, we think the data collection and analysis used here are robust relative to the claims we make about seed detectability. It makes sense that contaminants similar to the crop seed would be harder to distinguish and detect. In the ISTA proficiency rating tests they make allowances for species specific differences in the retrieval rate.[23,42,43] However, unlike this study they make no attempt to distinguish *which* seed or crop features may influence detection. Size, shape and color also impact algorithmic recognition of seed images[44] and automated seed cleaning[45] but have yet to be used to improve border inspection protocols. Furthermore, it is unsurprising that we detected individual differences in technician ability to detect species. Our results suggest that some species can be a kind of 'blind spot' for some observers, *e.g.*, *Ranunculus sceleratus* L. Compared to our technicians ISTA laboratory analysts detected 18% more seeds (across species and bags) implying they

benefited from methodological training and contaminant-crop familiarity.

Accurate identification of seeds requires practice and familiarity with many potential contaminant species. The ISTA accredited laboratory analysts identified most of the seed correctly, but not all. This may reflect the non-standard seed types that we added (discussed in methods). Familiarity may incline experts toward correct identifications for common weeds but may not help when faced with a new or less familiar species, especially if it looks a lot like a commonly encountered species. For example, the quarantine species *Solanum mauritianum* in this study was identified by ISTA laboratory analysts as the common crop weed *Solanum nigrum*. Easy-to-confuse species could be periodically subjected to extra identification checks (*e.g.*, growing plants from seed, peer review, DNA fingerprinting). Knowledge of the regulated weed species present in the source farms, importing region or country could help prioritize such efforts.[46]

In New Zealand the schedule of regulated quarantine weeds contains hundreds of species, the seeds of which may or may not occur in regularly imported seed lots. Our results show that the detectability of each seed type in a crop will vary from crop to crop, and then the correct identification of seeds also must occur, adding another level of difficulty and uncertainty. Meanwhile the accredited seed labs are only tested on 85 species in the proficiency testing, which may not prime them for detecting the specific regulated species for a country.[23,42,43] For example, only a few of the quarantine species in New Zealand are also part of the ISTA proficiency test, they are *Alopecurus myosuroides*, *Eragrostis curvula* (Schrad.) Nees, *Sorghum halepense* and *Tagetes minuta* L. Some genera on the ISTA list match to some species on the quarantine list: *Amaranthus* spp. (matches one species), *Cuscuta* spp. (three species), *Ipomoea* spp. (five species), and *Sporobolus* spp. (one species), meaning that the ISTA proficiency testing is supporting detection of 14 of the taxa on the quarantine list, which contains >400 taxa. Individual ISTA accredited laboratories will need to be well versed in a range of regionally important species that are not part of the ISTA proficiency testing program.

If a quarantine weed or other pest is likely to occur in a crop from a location, and its detectability is low because of differences in size, shape, or color, a crop specific inspection standard might be needed for high priority pests. For example, in this study the quarantine weed *Alopecurus myosuriodes* was detected 60% of the time in perennial ryegrass. Seed companies or regulators could require offshore inspection sample size to be raised to two times ISTA, or approximately 50 000 seeds to match the 0.01% threshold with 95% confidence, when no weeds are detected while accounting for the detection rate. It would lower the probability that the shipment is later rejected when it is re-inspected in New Zealand at the 150 000 seed standard (5 × ISTA). For seed lots right at the threshold contamination rate of 0.01% (1:10000), and inspected as required offshore at 1 × ISTA, but assuming imperfect detection rates of 60%, these would be incorrectly identified as free of pest 16.6% of the time, *i.e.*, using the R function (dbinom(0, 30 000, 1/10000*0.6)). Here we would expect five out of six consignments to be correctly identified as having the contaminant even though it is present at that threshold. In this case, if no quarantine species were detected, we would only be 95% certain the contamination rate is less than 1/6000 at the 0.6 detection rate. This is equivalent to the 100% detection after inspecting 18 000 seeds (30 000 × 0.6). This reflects the combined probabilities of failing to get contaminants in the sample, and of detecting them even if they were present. Still following the example (0.01% contamination) but with the higher inspection rate (five times ISTA) implemented in New Zealand (at the 60% detection) a zero detection of a quarantine weed implies that only one out of 8105 of shipments would be wrongly categorized as meeting the inspection threshold of 0.01% for the regulated pest. The inspection rate in New Zealand was set at five times the ISTA standard to address seed lot heterogeneity,[14] but it has the added benefit of improving detection. The final amount of ryegrass and clover seed inspected in New Zealand is usually six times the ISTA standard (1 × ISTA offshore plus 5 × ISTA in New Zealand). If no regulated species is detected, and we assume the detection rate is 60%, the implication is that we can be 95% confident any regulated seed is rarer than approximately 1:36000.

In the cases where border inspections in New Zealand do detect regulated species the rate of contamination should be carefully estimated, to better inform regulators about the range of contamination rates that are missed by seed testing laboratories in the exporting country. If the distribution of regulated seed abundances were known it could be used to estimate how many seed lots are likely to have been missed, or at least if the miss rate is low enough to justify maintaining the current standards.

Apart from detection, the effectiveness of quarantine measures (*e.g.*, seed contaminant screening) as a biosecurity measure is defined as much by what species are regulated as not, and their potential for introduction and harm. Quarantine lists regulating seed for sowing contaminants in seed for sowing vary between countries, *e.g.*, Argentina (nine species and three genera); Canada (19 species one genus); USA (33 species and nine genera) and New Zealand (420 species and 13 genera).[30,47,48] The effectiveness of seed inspection as a risk mitigation also relates to the amount of seed traded (and planted), the frequency of inspection, and the detection rates. We have only addressed aspects of the latter. The benefits of seed inspection extend beyond preventing weed seed dispersal, seed certification measures address other aspects of seed quality including, varietal purity, germination potential, pests and diseases.[4] Despite these efforts, recognized pests and weeds of plants can appear in new areas where previously they were unknown, sometimes with serious consequences for crops or native vegetation and the economy of the country concerned. We show that the failure to consider the detectability of weeds means that the regulatory thresholds set by regulators are often not being met. In the case of New Zealand, the border verification protocol, inspecting all seed lots at a higher rate, mostly addresses the problem. Nevertheless, there have been at least two post-border detections (incursions) of the regulated species *Alopecurus myosuroides* in New Zealand after it established in fields sown with imported perennial ryegrass and linseed seed from the UK in 2007 and France in 2021, respectively.[10,49] These post-border incursions are being successfully managed, highlighting the value of post-border interventions. Unfortunately, there is no easy way to convert border detection rates to estimate the number of weed incursions that are avoided because of the seed quarantine inspection protocols that are in place. This is because each seed lot has its own individual set of circumstances related to provenance, weed control practices, harvest and seed cleaning procedures. Determining an acceptable amount of investment in border mitigation of biosecurity risk is difficult. Where it has been attempted the economic returns are estimated to be high (costs avoided are much higher than the costs incurred).[12] We suggest the mitigation efforts should be fit for purpose, and that detection rates should be considered.

4.1 Recommendations

Regulators should consider the crop specific detectability of high priority regulated weeds and pests when setting standards for the sample size that should be inspected.

Regulators could inspect larger samples to capture hard-to-detect species. New Zealand border inspection protocols require 95% confidence that the maximum pest limit of 0.01% of quarantine weed seed contamination is not exceeded in any imported seed lot, the local standard (5 × ISTA) examining approximately 150 000 seeds can be achieved even with a 20% detection rate.

Proficiency testing carried out by ISTA on 85 weed species is useful for comparing labs internationally, but locally regulated species could differ, and proficiency testing of the locally regulated species is important.

Regulators should try to accurately estimate and document the contamination rates of regulated species when they are detected to develop a weed contamination profile for that weed.

Occasional surveys of the weeds and weed seedbanks in the overseas fields of commonly imported crops could be used to set inspection threshold based on risk.

ACKNOWLEDGEMENTS

The research was supported by AgResearch Ltd *via* its contribution to the Better Border Biosecurity (B3) research collaboration, www.b3nz.org. We would like to thank Bridget Wise and Derrick Wilson for inspecting our samples. Comments on the manuscript were gratefully received from John Hampton, Craig McGill, Melanie Newfield, Stuart Wylie, Andrew Robinson, Mike Ormsby and Phil Rolston. Open access publishing facilitated by AgResearch Ltd, as part of the Wiley - AgResearch Ltd agreement via the Council of Australian University Librarians.

DATA AVAILABILITY STATEMENT

Data is provided in the manuscript for overall detection rates for different species.

SUPPORTING INFORMATION

Supporting information may be found in the online version of this article.

REFERENCES

1 Hulme PE, Trade, transport and trouble: managing invasive species pathways in an era of globalization. *J Appl Ecol* **46**:10–18 (2009).
2 Rubenstein JM, Hulme PE, Buddenhagen CE, Rolston MP and Hampton JG, Weed seed contamination in imported seed lots entering New Zealand. *PLoS One* **16**:e0256623 (2021).
3 Ikeda M, Nishi T, Asai M, Muranaka T, Konuma A, Tominaga T *et al*., The role of weed seed contamination in grain commodities as propagule pressure. *Biol Invasions* **24**:1707–1723 (2022).
4 Ebbels DL, *Principles of plant health and quarantine*. CABI Publ, Wallingford (2003).
5 Aveling TAS, Global standards in Seed Health Testing, in *Global Perspectives on the Health of Seeds and Plant Propagation Material*, ed. by Gullino ML and Munkvold G. Springer, Netherlands, Dordrecht, pp. 17–28 (2014).
6 Stewart AV, Goldson SL, Marris JW, Rubenstein JM and Rolston MP, Analytical purity of old New Zealand forage seed samples and detection of fungal and insect contaminants including ryegrass endophyte and argentine stem weevil. *New Zealand J Agric Res*:1–14 (2022). https://doi.org/10.1080/00288233.2022.2066140
7 Ministry of Agriculture and Forestry, *Balance in TRADE: Understanding how the Sanitary and Phytosanitary (SPS) Agreement Helps Protect New Zealand's People, Economy and Environment while Improving Trade Opportunities*. Ministry of Agriculture and Forestry, Wellington, New Zealand (2009).
8 MacLeod A, Pautasso M, Jeger MJ and Haines-Young R, Evolution of the international regulation of plant pests and challenges for future plant health. *Food Sec* **2**:49–70 (2010).
9 Black R and Bartlett DMF, Biosecurity frameworks for cross-border movement of invasive alien species. *Environ Sci Policy* **105**:113–119 (2020).
10 Buddenhagen CE, Rubenstein JM, Hampton JG and Rolston MP, The phytosanitary risks posed by seeds for sowing trade networks, ed. by Uludag A. *PLoS ONE* **16**:e0259912 (2021).
11 Leung B, Springborn MR, Turner JA and Brockerhoff EG, Pathway-level risk analysis: the net present value of an invasive species policy in the US. *Front Ecol Environ* **12**:273–279 (2014).
12 Dodd A, Stoeckl N, Baumgartner J and Kompas T, *Key Result Summary: Valuing Australia's Biosecurity System*. Centre of Excellence for Biosecurity Risk Analysis, Australia (2020).
13 Chapter 2: Sampling, *International Rules for Seed Testing*. International Seed Testing Association (ISTA), Bassersdorf, Switzerland. Vol. **2022**, pp. i-2–i-44 (2022).
14 Hampton JG, *Imported Seed Lot Heterogeneity: Implications for Biosecurity Sampling Protocols*. Lincoln University, Lincoln, New Zealand (2005).
15 Lane SE, Cannon RM, Arthur AD and Robinson AP, Sample size for inspection intended to manage risk within mixed consignments. *NB* **42**:59–69 (2019).
16 ISTA, Chapter 2: Sampling, *International Rules for Seed Testing*. International Seed Testing Association (ISTA), Bassersdorf, Switzerland. Vol. **2020**, pp. i-2–i-44 (2020).
17 Jovanovic BD and Levy PS, A look at the rule of three. *Am Stat* **51**:137–139 (1997).
18 Hanley JA and Lippman-Hand A, If nothing goes wrong, is everything all right?: interpreting zero numerators. *JAMA* **249**:1743–1745 (1983).
19 MPI, *Import health standard: seeds for sowing, Ministry for Primary Industries*. Ministry for Primary Industries, Wellington, New Zealand (2020).
20 Gervilla C, Rita J and Cursach J, Contaminant seeds in imported crop seed lots: a non-negligible human-mediated pathway for introduction of plant species to islands, ed. by Albrecht H. *Weed Res* **59**:245–253 (2019).
21 ISTA, *Trifolium Repens Proficiency Test*. International Seed Testing Association, Wallisellen, Switzerland (2017).
22 ISTA, *Lolium Perenne Proficiency Test*. International Seed Testing Association, Wallisellen, Switzerland (2017).
23 Rösch M, ISTA quality assurance programme? The proficiency test rating system. *Accreditation Qual Assur* **8**:372–374 (2003).
24 ISTA purity committee, Universal list of species, International Seed Testing Association (2021).
25 Caradus JR, Goldson SL, Moot DJ, Rowarth JS and Stewart AV, Pastoral agriculture, a significant driver of New Zealand's economy, based on an introduced grassland ecology and technological advances. *J R Soc N Z*:1–45 (2021). https://doi.org/10.1080/03036758.2021.2008985
26 Seed, clover, for sowing | OEC, *OEC - The Observatory of Economic Complexity*, 2022. https://oec.world/en/profile/hs/seed-clover-for-sowing [accessed 4 October 2022].
27 Seed, rye grass, for sowing | OEC, *OEC - The Observatory of Economic Complexity*, 2022. https://oec.world/en/profile/hs/seed-rye-grass-for-sowing [accessed 4 October 2022].
28 ISTA, *International Rules for Seed Testing*. International Seed Testing Association, Bassersdorf, Switzerland (2021).
29 James T, Popay I, Champion PD, Grbavac N and Rhode B, *An Illustrated Guide to Weed Seeds of New Zealand*. New Zealand Plant Protection Society, New Zealand (2012).
30 MPI, Schedule of regulated (quarantine) weed seeds, 2018. https://www.mpi.govt.nz/dmsdocument/7111.
31 MPI, *Import Health Standard: Stored Plant Products for Human Consumption*. Ministry for Primary Industries, Wellington, New Zealand (2018).
32 Schneider CA, Rasband WS and Eliceiri KW, NIH image to ImageJ: 25 years of image analysis. *Nat Methods* **9**:671–675 (2012).
33 Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R *et al*., Welcome to the tidyverse. *J Open Source Softw* **4**:1686 (1-6) (2019).
34 R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2020).

35 Bates D, Mächler M, Bolker B and Walker S, Fitting linear mixed-effects models using lme4. *J Stat Softw* **67**:1–48 (2015).
36 Lenth RV, Least-squares means: the R package lsmeans. *J Stat Softw* **69**: 1–33 (2016).
37 Kuznetsova A, Brockhoff PB and Christensen RHB, lmerTest package: tests in linear mixed effects models. *J Stat Softw* **82**:1–26 (2017).
38 Lenth RV, Emmeans: Estimated Marginal Means, aka Least-Squares Means (2022).
39 Burdon JJ, Trifolium repens L. *J Ecol* **71**:307–330 (1983).
40 Larsen SU, Andreasen C and Kristoffersen P, The variation in seed weight within and among cultivars of slender creeping red fescue (*Festuca rubra* ssp. *litoralis*), perennial ryegrass (*Lolium perenne*) and Kentucky bluegrass (*Poa pratensis*) and its importance for the composition of seed mixtures. *Seed Sci Technol* **32**:135–147, International Seed Testing Association (2004).
41 Heimbach U, Variability of thousand grain weights of seed batches of important arable and some horticultural crops, J Cultivated Plant. **70**:250–254 (2018).
42 Koji J, Milivojevi M, Stojadinovi J, Marinkovi M and Ivanovi D, The analysis of proficiency test results on the seed purity and the other seed determination for ten-year period. *J. process. energy agric* **15**:255–259 (2011).
43 ISTA, *International Seed Testing Association*. International Seed Testing Association (ISTA), Wallisellen, Switzerland (2021).
44 Luo T, Zhao J, Gu Y, Zhang S, Qiao X, Tian W *et al.*, Classification of weed seeds based on visual images and deep learning. *Inf Process Agric*: S2214317321000809 (2021). https://doi.org/10.1016/j.inpa.2021.10.002
45 Heo YJ, Kim SJ, Kim D, Lee K and Chung WK, Super-high-purity seed sorter using low-latency image-recognition based on deep learning. *IEEE Robot Autom Lett* **3**:3035–3042 (2018).
46 Robinson A, Burgman MA and Cannon R, Allocating surveillance resources to reduce ecological invasions: maximizing detections and information about the threat. *Ecol Appl* **21**:1410–1417 (2011).
47 SENASA, Listas de plagas cuarentenarias para Argentina, 2015. https://assets.ippc.int/static/media/files/reportingobligation/2015/02/03/listas_plagas_cuarentenarias_para_argentina.pdf.
48 USDA, Federal Noxious Weeds List, 2021. https://plants.usda.gov/java/noxious [accessed 17 February 2021].
49 Foundation for Arable Research, Black grass detection, 2022. https://www.far.org.nz/articles/1606/black-grass-detection [accessed 17 October 2022].

