Industry Viewpoint on Sampling

SAMANTHA THOMAS

MONSANTO, ASTA-VTS LEAD

Why does the seed industry sample?

- Samples are used for routine testing purposes
  - Quality tests: germination, vigor, hybridity, purity, seed health
  - Phytosanitary declarations
  - Retention
- Based on testing results, dispositions on use will be made
  - These could include:
    - Upgrade (to improve germination)
    - Seed Treatment (which will increase cost of goods)
    - Additional sorting
    - Do not use--DESTROY

What does the seed industry want from a sample?

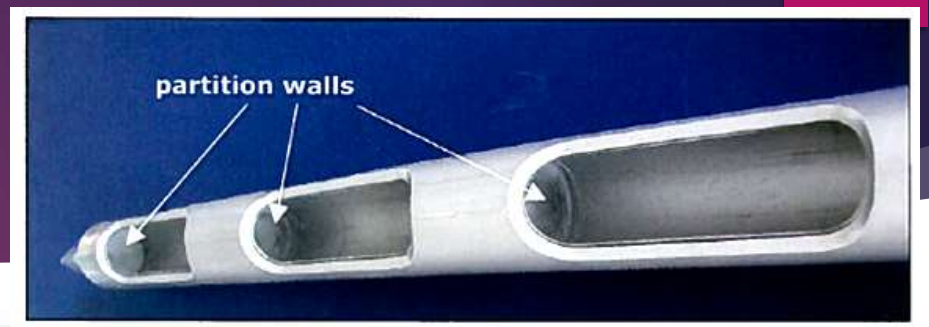
- The sample needs to accurately reflect the lot from which it comes from.

How do we sample?

- There are established protocols available
  - ISTA (International Seed Testing Association)
  - AASCO (Association of American Seed Control Officials)

- Either or both may be used routinely (for accreditation or declaration purposes)
  - Both provide guidance on how to sample from various materials
    - Seeds in an operational stream (such as on a conveyor towards packaging)
    - Bulk lot in a container
    - Packaged seeds

Seed Sampling



Submitted Sample Creation

Seed Lot/Batch Containers

Primary Samples

Composite Sample

Submitted Samples

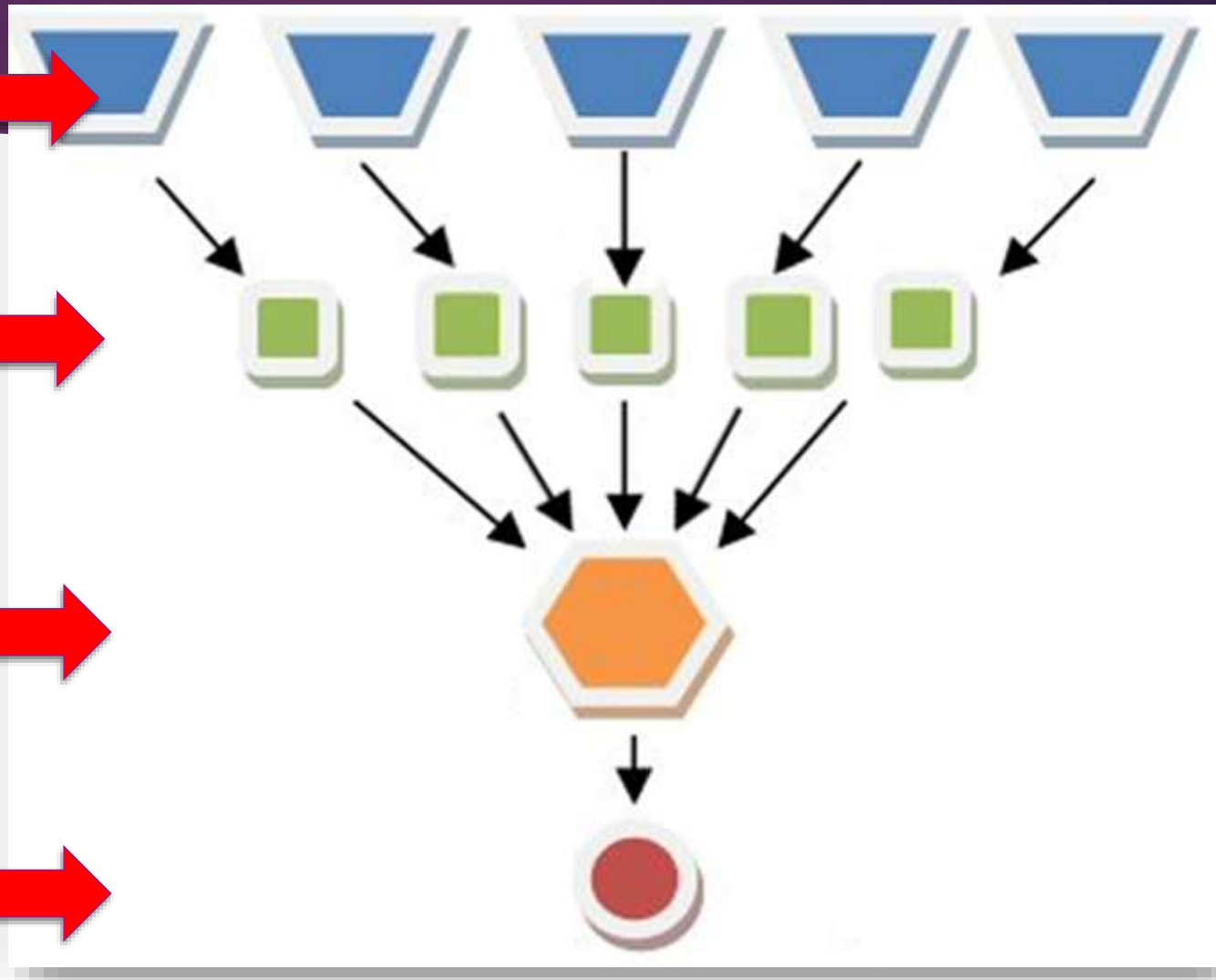


Diagram adapted from UC Davis image

Sample sizes

- For quality testing, the sample size that is needed for analysis can vary greatly
- Germination, vigor, purity, hybridity have standard samples sizes
  - Ex. 400 seeds for germination tests
- For vegetable seed health, samples sizes are a reflection of pathogen biology, epidemiology, thresholds for disease outbreaks
  - Sample sizes range from 2000 seeds to 30,000 seeds depending on host:pathogen combination
  - There may be several disease test needed per crop
  - Different test types result in multiple samples will be needed to complete
  - The composite sample needs to be adequate to generate all the samples required for testing

Example of Quality Testing Specifications

Quality Component	Test	Sample size
Germination	Germination	400
Vigor	Usable Transplant	400
Hybridity		200
Seed Health	Bacterial diseases	30,000
	Viral diseases	3000
Composite sample needs to exceed:		34,000
Additional seed needed for phyto declarations		

Again, given the use/no-use disposition, it is very important all of these samples represe

Background on Vegetable Pipeline

- Given the large samples sizes for seed health assays, there can be challenges with meeting the recommendations/standards
- For vegetables, there will be varietal/hybrid trialing with commercial growers as a part of product advancement
  - We breed for many reasons
    - Taste
    - Fruit size, shape, color, taste
    - Plant structure, plant/field stand
    - Fruit shelf-life, packing/shipping ability
    - Disease resistance

The goal is the ensure the new variety will be usable prior to generating a lot of product (seed)

Small seed lots

- Pipeline materials are from breeding/early stock seed trials
  - May be the product of a handful of plants
  - Therefore the seed quantity is very limited
    - A few thousand seeds or less
- Not possible for the standard samples sizes to be met
- Greatest challenge is the seed health sample sizes
- What options to address?

Small Seed Lots

- Options for sampling
  - Composite: pull a sample from many small lots to meet the standard sample size
    - Pro: meet the standard sample size
    - Con: will have to do additional testing (of different composites) if there is a positive result (as that indicates there was a positive seed but not sure which small lot it originated) [may lead to further reductions in available seed]
  - Percentage approach: take a standard percent (5 or 10%) of the total lot and analyze
    - Pro: will have an unique result for each lot
    - Con: accuracy of result in reflecting the lot quality can be questioned

Small seed lots

- Options (cont.)
  - ISPM 31: use hypergeometric analysis combined with desired confidence and impurity level to determine the sample size
    - Pro: statistical model (can clearly state confidence and impurity)
    - Con: not practically applicable

Table 1: Table of minimum sample sizes for 95% and 99% confidence levels at varying levels of detection according to lot size, hypergeometric distribution

Number of units in lot	P = 95% (confidence level) % level of detection × efficacy of detection					P = 99% (confidence level) % level of detection × efficacy of detection				
	5	2	1	0.5	0.1	5	2	1	0.5	0.1
25	24*	–	–	–	–	25*	–	–	–	–
50	39*	48	–	–	–	45*	50	–	–	–
100	45	78	95	–	–	59	90	99	–	–
200	51	105	155	190	–	73	136	180	198	–
300	54	117	189	285*	–	78	160	235	297*	–
400	55	124	211	311	–	81	174	273	360	–
500	56	129	225	388*	–	83	183	300	450*	–
600	56	132	235	379	–	84	190	321	470	–
700	57	134	243	442*	–	85	195	336	549*	–
800	57	136	249	421	–	85	199	349	546	–
900	57	137	254	474*	–	86	202	359	615*	–
1 000	57	138	258	450	950	86	204	368	601	990

Basic Plant Pathology

- Basic plant pathology
  - Plants become infected in many ways
  - Some generalities based on pathogen type.
  - For seed-borne pathogens,
    - Viruses: move systemically = whole plant infected once it is introduced
      - Plants infected
    - Bacteria: may move systemically, may splash and move from location to location on the plant and to neighboring plants
      - Fruits, plants, siliques, pods can be infected
    - Fungi: may splash or move aerially from location to location on the plant and to neighboring plants
      - Fruits, plants, siliques, pods can be infected

Infection Unit Approach

- Sampling strategies do not account for the pathogen biology
- Current sampling strategies are based on probabilities and setting desired confidence rates
  - Zero tolerance by regulatory agencies
  - Basically the target is 1 infected seed in X lot size
  - *Given the biology of the pathogen target, there will NOT be just one infected seed in the lot*
- Infection unit: in practice the smallest part of the plant that will be infected
  - How to sample to detect that part? Fruit, pod, plant?

Infection unit

- Small ISF working group to define the appropriate way forward to present infection unit concept
  - Currently we are comparing models
  - Will be looking at existing supporting data to document infection units associated with each seed borne pathogen
  - Leveraging GSPP information (as Infection unit was used to set the sample size guidelines)
- Desired output: a publication (poster, presentation, paper) which establishes the infection unit concept and permits better conversations on appropriate sampling models

Model Developed

- WG met with Fritz Quandt (ISHI Statistician) who verified the model is sound
- Assumptions identified in the models for
  - Viruses (CGMMV)
  - Bacterial-Systemic (Cmm)
  - Bacterial-Blighter (Xcc)

Example: Assumptions specific to CGMMV

1	All seeds in the infection unit are contaminated	infection unit plant, 50% seeds infected	Can be addressed by research
2	1 infected seed can be detected in the test	Yes	Has been shown in validation
3	Number of seeds per unit are (more or less) equal	Yes, but different per crop	Describes the boundaries of the model: a. effects of nr of seeds on %, and b. effect of disease on seed size etc.
4	Number of seeds per infected and healthy unit is (more or less) equal	Yes, but different per crop	
5	First fruit is infected	Yes	Can be addressed by research
a	when do plants become infected / what plant-flower-fruit stage?		Research on moment of infection leading to infection of seed
b	spread within the plant	Homogeneous, and progressive	Look at current literature
6	The assumptions can be different for:		
a	varieties,	crops not varieties taken into account	
b	CGMMV isolates		There is no indication, but might need research to address to show that it has no influence
c	production systems		Lower importance to address—ReFreSH or Clean Corridor focus?

The statistics…

- For an infection unit, sample size is calculated using the Binomial distribution* $P = 1-(1-r)^n$, where
  - P = the probability of detection (POD)
  - r = the infection rate (i.e. infected seeds in the seed lot divided by the total seed number in the same seed lot)
  - n = sample size
- Sample size is calculated as follows:

  $n = \log(1-P) / \log(1-r)$ *Or* $n = \log(B) / \log(1-r)$ where (B=1-P)

SEED IS LIFE ISF International Seed Federation

* Note: To calculate sample size for small lots a Hypergeometric distribution is often recommended. The Binomial distribution is much simpler to use and sample sizes calculated using the two distributions do not differ very much

Use of infection unit in NSHAPP

- Assumptions
  - Infection unit for CGMMV in cucumber: a plant
    - A plant produces a minimum of 4 fruits, there are 25 seeds per fruit and all seed in an infected fruit are infected
    - A plant produces 2 infected and 2 healthy fruits (so as to account for delayed infections or uninfected seeds).
- Additional Notes
  - The model requires the number of female plants in a unit of production, seeds per plant
  - The epidemiological threshold underlying ISTA's ELISA test for the detection of CGMMV in cucurbit seed provides the upper limit of the sample size which is 2000 seeds

Calculation: a maximum 5% sample size for a small lot will result in 95% confidence in detecting CGMMV contamination

SEED IS LIFE
ISF
International Seed Federation

Next Steps

- Many assumptions need to be proven by research
  - Need researchers interested in the work
    - ISU has been suggested by ASTA
      - Infection unit concept in part of the NSHAPP/ReFreSH
    - Others?
  - Work to be completed (for each of the pathosystems)
    - Monitor of disease progression in/on plant
    - Monitor disease progression within a plant population
    - Monitor seed infection within each plant and among the plant population
    - Determine what sampling rate is adequate for each scenario to reach each desired confidence (90%, 95%, etc.)

Final Thoughts

- Seed industry applies standard practices on sampling when:
  - They are available
  - They are practical
- There are some challenges that remain for some of our processes but we are readily looking for answers and collaborators
- If you want more info or would like to collaborate, please reach out to me

- THANKS!