Comparison of sampling procedures for estimating botrytis bunch rot incidence in New Zealand vineyards

A.H. McKay, G.N. Hill and R.M. Beresford

*The New Zealand Institute for Plant & Food Research Ltd, Mt Albert Research Centre,
Private Bag 92169, Auckland 1142, New Zealand*

Corresponding author: Alistair.mckay@plantandfood.co.nz

Abstract Quantifying botrytis bunch rot (BBR) in vineyards is essential to the New Zealand wine industry. Simple random sampling (SRS), inverse sampling (IS) and adaptive cluster sampling (ACS) were compared using simulated disease incidence at two levels (1% and 3%), three clustering factors (random, low and high), and small (100 bays) or large (2500 bays) vineyard sizes. Sampling relative efficiency (*re*) was highest using ACS when disease was highly clustered at low incidence. Inverse sampling was investigated by repeatedly sampling from the simulated vineyards, which resulted in total sample number and variance of IS being greatest in highly clustered BBR at low incidence. IS resulted in a lower final sample number with less variance than ACS using simulated BBR that was randomly dispersed at 3%. IS or SRS using a sample size calculated from desired sample confidence can provide practical and accurate botrytis bunch rot sampling for disease management decision support.

Keywords disease survey, botrytis bunch rot, survey efficiency, adaptive sampling.

INTRODUCTION

Botrytis bunch rot (BBR) caused by *Botrytis cinerea* has a serious impact on New Zealand's wine grape industry by reducing vineyard yields and adversely affecting wine quality (Hill et al. 2010). BBR results in costly processes in wineries to reduce the effects of BBR on wine quality. Imposition of price penalties on growers occurs if more than 3–5% of the crop is affected and rejection of grape consignments if too much BBR is present (Hill et al. 2010). Accurate quantification of the amount of BBR affecting a vineyard before harvest is crucial for negotiations between wineries and contract grape growers. Accurate botrytis assessment is also required for the use of a weather-based botrytis risk prediction

model (Beresford & Hill 2008), the Botrytis Decision Support Model (BDSM). This model is designed to help winemakers and viticulturists to manage the impact of BBR and requires inputs of accurate disease measurements to predict late-season botrytis disease progress.

Theoretical aspects of sampling for estimation of plant disease incidence in various crops have been reviewed by Madden et al. (2007). A form of random sampling, such as simple random sampling (SRS) is recommended because it gives every sample unit, or grape bunch, an equal chance of selection. However, SRS can be very time consuming if disease incidence is very low, as occurs for BBR early in the season

when infections are in a latent state (McClellan & Hewitt 1973). For example, under SRS, if the desired sample reliability is half a 95% confidence interval at a fixed proportion of the mean incidence ($H = 0.5$) and an estimated disease incidence of 1%, a sample size of 1522 is required (Madden et al. 2007). A stratified approach that incorporates inverse sampling (IS) may offer an effective tool for BBR incidence estimation under these circumstances (Madden et al. 1996). With IS, sample size is not fixed, but rather sampling continues until a predefined number of diseased bunches (T) is recorded. The reliability of the sample is determined by the confidence interval required by the user and an initial estimate of disease. In the above example the calculated value for T is 15 (Equation 1). An additional stop condition can also be applied if T is not reached within a certain sample size, indicating disease incidence is lower than a predetermined threshold (Reardon et al. 2005).

At a low disease incidence (<5%), the distribution of BBR may be spatially clustered because of the localised dispersal of initial inoculum (Ojiambo & Scherm 2010). Under these circumstances sampling based on random selection may not be the most efficient method for assessing disease incidence (Hughes et al. 1996). For clustered diseases, concentrating the sampling effort in locations where the disease is present, and not in locations where the disease is absent, may provide a more efficient sampling strategy with equal accuracy (Turk & Borkowski 2005). Adaptive Cluster Sampling (ACS) is a sampling approach that incorporates information about clustering through intensive sampling of areas close to diseased bunches observed through initial sequential sampling (Ojiambo & Scherm 2010). ACS is most efficient when populations are very rare and highly clustered and so may offer a useful tool for BBR sampling during early phases of an epidemic (Brown 2003). The efficiency of ACS depends on how the survey is designed, in particular defining the condition to adapt (CA), which is a disease incidence threshold that triggers adaptive sampling and the size of the neighbourhood

from which adjacent units are sampled (Turk & Borkowski 2005). Also, as with any sampling design, the size of the primary sampling unit influences the efficiency of the sampling design (Hughes et al. 1996). In some cases the choice of sample unit may be clear; for example, an individual plant in a field (Madden et al. 1996) or, in the case of BBR, a bunch of grapes. However, to account adequately for disease aggregation that develops in proximate bunches, a sampling unit of several bunches may be more appropriate. Sustainable Wine Growing New Zealand (SWNZ, www.nzwine.com) currently uses a 20 bunch sample size for estimation of botrytis and other diseases and pests. It is possible that sampling fewer bunches at more locations would offer a more accurate estimate of BBR incidence, particularly when disease is highly aggregated.

The aim of this study was to identify an optimal sampling strategy for BBR, in terms of sample size and sample efficiency, using hypothetical computer simulations. Simulations generally allow flexibility for evaluating several disease scenarios that may not be possible from empirical data. Likely scenarios of BBR incidence were simulated in large and small vineyard blocks under two rates of disease incidence and three levels of disease clustering. The relative efficiency of ACS compared with SRS was determined, and IS was evaluated for practical field assessment of BBR.

METHODS

Simulation

Vineyards that varied in size, disease incidence and clustering of infected bunches were generated stochastically. Two vineyard sizes were simulated: small (100 total bays, 10 rows $\times$ 10 bays) and large (2500 total bays, 50 rows $\times$ 50 bays). Disease frequency was simulated at two levels (1% and 3% incidence) and at three levels of clustering (random (non-clustered), low and high) (Figure 1). Based on information from commercial vineyards, a fixed estimate of 100 grape bunches per four-vine bay was applied to each simulated vineyard. Theoretically, each bay could contain between 0

and 100 diseased bunches. Clustered datasets were generated using the SpatStat analysis package written in the R statistical software platform (R 3.13.1) applying a realization of Matern's cluster process inside a defined coordinate window. Clusters were constructed by first generating "parent" points with intensity ($kappa$) using a Poisson point process. Then each parent point was replaced by a random cluster of points, the number of points in each cluster being randomized about a Poisson distribution (mu), and the points being placed independently and uniformly inside a disc of radius (r) centred on the parent point (Baddeley & Turner 2012). The assigned values for $kappa$, mu and r are given in the Appendix.

Figure 1 Examples of the simulated vineyard blocks showing the spatial distribution of botrytis bunch rot (BBR) (each dot represents a diseased bunch) at 3% incidence. (a) highly clustered ($kappa = 10$, $r = 0.05$, $mu = 30$) and (b) non-clustered ($kappa = 300$, $r = 1$, $mu = 1$).

Sampling approaches

Three sampling approaches were evaluated for estimating BBR incidence: (1) simple random sampling (SRS), (2) single stage sequential adaptive cluster sampling (ACS) and (3) inverse sampling (IS). For each sampling procedure, repeated samples were taken from the generated vineyard data and summary statistics for disease incidence estimates were calculated from 100 or 1000 sample runs. Two sampling designs, SRS and ACS, were tested using SAMPLE software (U.S. Geological Survey, Leetown Service Center, Kearneysville, WV) using

selected simulated vineyard datasets. The SAMPLE software package approaches ACS as follows: (1) an initial sample is conducted using conventional sequential design, (2) upon detection of disease above a predetermined numerical threshold (CA), a switch is made to neighbourhood sampling of the surrounding sample units, (3) if the CA is met in surrounding sample units, then step 2 is followed and (4) if neighbouring sample units do not meet the CA, then sequential sampling (step 1) is resumed until another sample unit meets the condition to adapt.

To compare ACS with SRS, the coefficient of variation (CV) of the mean estimate, final sample size (N) and relative efficiency (re), were calculated from 100 replicate re-sampling simulations. The re was calculated as the ratio of the variances of the sample means for SRS / ACS after 100 replicate sampling simulations (D.R. Smith, Aquatic Ecology Lab, Kearneysville, WV, USA, personal communication). For scenario testing, a sampling scheme of 10 bunches per sample unit and an initial sampling fraction of $f=0.01$ (1%) were used in order to conserve computing resources. Because of the low disease frequency and the possibility of missing sample units from too few diseased bunches per sampling unit at the lowest clustering, a conservative CA of two diseased bunches per sample unit was applied to the ACS simulation to instigate neighbourhood sampling.

IS was tested using a customised sampling program written using the R-statistical package (Ver. 2.13.1. The R Foundation for Statistical Computing), based on fixed criteria for the predefined number of diseased bunches (T), calculated using Equation 1 (Madden et al. 2007).

$$T = (1 - p') \left(\frac{Z_{\alpha/2}}{H} \right)^2$$

Equation 1

T = Predefined number of diseased bunches,
 p' = Initial estimate of disease incidence,
 $Z_{\alpha/2}$ = 100(1 - $\alpha/2$)th percentile of the standard normal distribution,
 H = Margin of error as a proportion of the initial sample estimate, e.g. ± 0.5 .

For disease incidence at 1% and 3%, a value for T of 15 was selected to give approximately $H = 0.5$ at $\alpha=0.05$. The relative efficiency of IS was calculated by the mode sample size of 1000 samples and the respective CV value. For better comparison with the ACS method using an initial sequential sampling approach, as well as a more realistic estimation of the empirical data collected from field surveys, the IS program was designed to have each sample performed without replacement.

Changing the size of the primary sample unit was tested at 10, 20, 50 and 100 bunches per bay for both comparison of ACS with SRS and for IS sampling simulations using highly clustered and non-clustered disease, and at 3% (high) and 1% (low) for the small vineyard size. The samples taken from the simulated vineyards using ACS were conducted with an initial sampling fraction of $f = 0.01$ (1%) and a CA of 2.

RESULTS

Relative efficiency of adaptive cluster for three levels of clustering

At a simulated disease incidence of 1%, using an initial sample fraction of $f = 0.02$ and a sample unit of 10 bunches, the relative efficiency (re) of ACS was greatest when applied to a highly clustered disease distribution in both the small (100 bays) and the large (2500 bays) vineyard simulations (Figure 2). An re value of greater than 1.0 indicates a sampling plan more efficient than an SRS design. In the smaller vineyard, simulation mean BBR incidence estimates were most consistent with actual incidence when disease was less clustered, but ACS accurately estimated BBR incidence when highly clustered in larger vineyard simulations (Table 1). Sample variance was greatest in the highly clustered BBR distributions, demonstrated by larger CV values. This was due to the number of 'misses' or sample units without diseased bunches encountered during the initial sequential sampling approach, as indicated by sample unit occupancy. The re of ACS was also a function of final sample size compared with the initial sample frequency. For the most highly clustered simulations of the

small and medium sized vineyards reported in Table 1, the final sample size was nearest to the initial sample size, indicating higher sampling efficiency. As expected, sample unit occupancy, or the proportion of sample units containing at least one diseased bunch, increased as the spatial distribution of disease became less clustered.

Figure 2 Relative efficiency of adaptive cluster sampling compared with simple random sampling for estimating botrytis bunch rot incidence in computer-generated vineyards of two sizes (small and large), for 1% incidence of infected bunches simulated at three levels of clustering (high, low and non-clustered).

Inverse sampling results

The sample size required to encounter $T = 15$ diseased bunches was greatest for the highest levels of clustering at 1% and 3%. However, as disease distribution became less aggregated, N became lower and the sample variance was reduced (Table 2), and for the low frequency of disease, the N exceeded that obtained by ACS (Table 1).

Effect of sample unit size on sampling efficiency

Sampling variance determined by the calculated coefficient of variation was greatest in samples of the highly clustered datasets using both ACS and IS sampling procedures (Table 3). For ACS, sample unit values of 50 and 10 bunches provided the least variability for sample estimates from highly clustered and non-clustered simulations, respectively. For IS, a 10-bunch sample unit demonstrated the least sample variability from

Table 1 Sampling efficiency and accuracy of adaptive cluster sampling at a low incidence (1%) of botrytis bunch rot (BBR) for three levels of disease spatial distribution (high, low or nil clustering) in small (100 bays) and large (2500 bays) simulated vineyards.

Simulated vineyard size	BBR clustering ¹	Initial sample size ² (<i>n</i>)	Relative efficiency ³ (<i>re</i>)	Final sample size (<i>N</i>)	Occupancy ⁴	BBR incidence (%)	CV ⁵
Small	High	29	1.11	43.3	0.04	0.66	0.48
Small	Low	29	0.60	48.1	0.08	0.96	0.33
Small	Nil	29	0.55	53.4	0.10	1.01	0.30
Large	High	149	2.23	191	0.02	0.99	0.55
Large	Low	149	0.19	1681	0.3	0.87	0.09
Large	Nil	149	0.18	1075	0.35	0.92	0.11

¹The spatial distribution of botrytis bunch rot disease in the simulated vineyard at three levels of clustering determined by the Matern's clustering process using a parent point generation of intensity kappa and dispersal of offspring points around the parent with radius *mu*.

²Initial sample size *n* with an initial sample fraction of *f* = 0.02 of the total vineyard size.

³Relative efficiency of ACS compared with SRS based on the ratio of the variance of the sample means for SRS / ACS after 100-replicate re-sampling simulations.

⁴Occupancy is the calculated estimate of sampling units meeting the *CA* = 2.

⁵Coefficient of variation calculated from mean estimates and their standard deviation across 100 replicated samples.

Table 2 Final sample size required to acquire the *T* value calculated for inverse sampling estimation of botrytis bunch rot estimation at two frequencies of botrytis bunch rot (BBR) disease incidence and three levels of spatial disease distribution (clustering) in small (100 bays) simulated vineyards at 95% confidence.

BBR frequency (%)	BBR clustering	<i>T</i> value ¹	Final sample size (<i>N</i>) ²	Calculated BBR incidence (%) ³	CV ⁴
1.0	High	15	200	0.70	0.30
1.0	Low	15	139	1.01	0.25
1.0	Nil	15	139	1.01	0.23
3.0	High	15	59	2.37	0.40
3.0	Low	15	48	2.92	0.27
3.0	Nil	15	46	3.05	0.23

¹*T*-value calculated for a 95% confidence interval with initial disease estimates of 1% and 3% using equation 1; results in *T* = 15 for both disease incidence frequencies.

²Mode value of sample size (10-bunch sample units) required to encounter *T* = 15 diseased bunches using 1000 samples using the IS procedure.

³Calculated using minimum-variance unbiased estimator of disease incidence from the mode sample number (*N*) of inverse sampling $p' = (T - 1) / ((N \times 10) - 1)$ derived from Madden et al. (2007).

⁴Coefficient of variation calculated from the mean estimates and their standard deviation across 1000 replicated samples.

Table 3 Sampling variance (reported as calculated coefficient of variation, CV) of adaptive cluster sampling and inverse sampling for 10, 20, 50 or 100 bunches in the primary sampling unit of a small simulated vineyard at botrytis bunch rot (BBR) disease incidence of 1% and two levels of clustering (highly clustered and non-clustered).

Sample unit size (<i>n</i>)	Clustering level	CV (ACS) ¹	CV (IS) ²
10 bunches	High	1.01	1.18
10 bunches	None	2.29	1.10
20 bunches	High	1.09	1.57
20 bunches	None	2.09	0.96
50 bunches	High	1.02	1.46
50 bunches	None	1.14	1.06
100 bunches	High	1.17	1.58
100 bunches	None	1.84	1.02

¹Coefficient of variation calculated from mean estimates and their standard deviation across 100 samples.

²Coefficient of variation calculated from mean estimates and their standard deviation across 1000 samples.

highly clustered datasets and for non-clustered disease incidence estimates, 20-bunch sample units were the least variable.

DISCUSSION

These simulations show that, when BBR incidence is low ($\approx 1\%$) and highly clustered, an adaptive cluster sampling strategy provides the most efficient estimation of disease. However, when disease incidence is moderate ($\approx 3\%$) and not clustered, ACS is less efficient than SRS, given the parameters of the sampling approach used in this study. Conversely, based on the final sample size as a component of sampling efficiency, IS became more efficient as disease incidence increased and clustering decreased. Similar results have been observed to a greater or lesser degree in other comparisons of ACS, such as in disease incidence of agronomic crops (Ojiambo & Scherm 2010) or in population density estimation of fresh water mussels (Smith et al. 2009). Both of these studies demonstrated that sampling efficiency and accuracy were largely accounted for by population density (equating to disease incidence in the present study) and clustering rather than sampling design.

In the present simulations, use of IS as a type of stratified SRS resulted in a higher sample size required at low disease frequency than ACS with

a higher differential if the disease was highly clustered. However, when disease frequency was moderate, that trend was reversed, outlining some of the limitations of designing an adaptive cluster sampling plan. Selection of the critical value *CA* at which neighbourhood sampling is implemented using ACS is an important decision that affects the accuracy of ACS protocols (Turk & Borkowski 2005). In the present simulations, a *CA* value of 2 may have been too small at the higher disease incidence, resulting in a greater final sample size and a less efficient sample. Also, defining the size of the neighbourhood and selection of the initial sampling fraction have important implications on ACS efficiency, and changes in these factors are all interrelated (Turk & Borkowski 2005). Unfortunately there is no accepted procedure for choosing the critical value, neighbourhood size or initial sample frequency for adaptive cluster sampling (Brown 2003). Pilot sampling has been suggested as a method for assigning an appropriate value for *CA* (Brown 2003). However, this has drawbacks for practical application to vineyard sampling. In particular pilot sampling is time consuming and upon developing an ACS design, the same bunches may be repeat sampled (Turk & Borkowski 2005).

For ACS and IS, the size of the sample unit affected sampling variance. Based on the present IS

simulations, a sample unit size of 20 bunches would be preferred when disease is randomly distributed and a sample unit value of 10 bunches could be best under clustered conditions. Since disease clustering is likely to fall somewhere between randomly distributed and highly clustered, the current sample unit value recommended by SWNZ of 20 bunches is likely to provide adequate precision for most BBR scenarios.

Adaptive cluster sampling provides many options to optimise a field sampling plan for BBR of wine grapes. However, this flexibility complicates the development of a practical sampling plan that works across New Zealand's many growing regions and wine varieties. Application of a stratified sampling approach based on SRS with a defined stopping point, such as demonstrated here with IS, would add simplicity to BBR sampling and would effectively estimate disease for the majority of field epidemics with perhaps the exception of highly clustered botrytis at very low frequency. Under these circumstances, a modified ACS plan that incorporates a stop rule once the sample size meets a certain value may prove useful (Ojiambo & Scherm 2010). Such a plan would require further validation for use in BBR incidence estimation and currently may be beyond the required sample precision of the industry. Application of a stratified form of IS-based SRS methodology may provide winegrowers with the precision needed to run the BDSM accurately while at the same time limiting BBR sampling when unnecessary. In order to validate this approach, additional research has been initiated using IS protocols to sample empirical data collected from across a range of New Zealand winegrowing contexts.

ACKNOWLEDGEMENTS

Funding for this study was provided by the New Zealand Ministry of Science and Innovation (Low Impact Disease Control). Support for the development of the inverse sampling methodology and critical editing of the final report was provided by Mark Wohlers of Plant & Food Research.

REFERENCES

- Baddeley A, Turner R 2012. Spatial Point Pattern analysis, model-fitting, simulation, tests. The R project for statistical computing. <http://www.r-project.org/>.
- Beresford RM, Hill GN 2008. Predicting in-season risk of botrytis bunch rot in Australian and New Zealand vineyards. In: DeGaris K, Krstic M, McCorkelle G, McLoughlin S ed. Breaking the mould – a pest and disease update. Proceedings of the ASVO seminar, 24 July 2008, Mildura, Victoria. Australian Society of Viticulture and Oenology Inc. Adelaide, South Australia. Pp. 24-28.
- Brown JA 2003. Designing an efficient adaptive cluster sample. *Environmental and Ecological Statistics* (10): 95-103.
- Hill GN, Beresford RM, Evans KJ 2010. Tools for accurate assessment of botrytis bunch rot (*Botrytis cinerea*) on wine grapes. *New Zealand Plant Protection* 63: 174-181.
- Hughes G, Madden LV, Munkvold GP 1996. Cluster sampling for disease incidence data. *Phytopathology* 86(2): 132-137.
- Madden LV, Hughes G, Munkvold GP 1996. Plant disease incidence: Inverse sampling, sequential sampling, and confidence intervals when observed mean incidence is zero. *Crop Protection* 15(7): 621-632.
- Madden LV, Hughes G, Van den Bosh F 2007. Estimating plant disease by sampling. In: Madden LV, Hughes G, Van den Bosh F ed. The study of plant disease epidemics. APS Press St. Paul, MN, USA. Pp. 279-318.
- McClellan WD, Hewitt WB 1973. Early Botrytis rot of grapes: time of infection and latency of *Botrytis cinerea* Pers. in *Vitis vinifera* L. *Phytopathology* 63(9): 1151-1157.
- Ojiambo PS, Scherm H 2010. Efficiency of adaptive cluster sampling for estimating plant disease incidence. *Phytopathology* 100(7): 663-670.
- Reardon JE, Berkett LP, Borkowski J 2005. Field evaluation of a new sequential sampling technique for determining apple scab "Risk". *Plant Disease* 89(3): 228-236.

- Smith DR, Rogala JT, Gray BR, Zigler SJ, Newton TJ 2009. Evaluation of single and two-stage adaptive sampling designs for estimation of density and abundance of freshwater mussels in a large river. *River Research and Applications* 27(1): 122-133.
- Turk P, Borkowski JJ 2005. A review of adaptive cluster sampling: 1990-2003. *Environmental and Ecological Statistics* 12(1): 55-94.

APPENDIX

Assignment variables for the development of clustered disease simulations using *R* statistical computing package “Spatstat Ver. 1.26.1”.

Vineyard size (number of bunches)	Disease incidence (%)	Level of aggregation	Parent point (κ)	Offspring points (μ)	Radius of distribution (r)
Small (10 000)	1.0	High	5	20	0.05
Small (10 000)	1.0	Med	20	5	0.1
Small (10 000)	1.0	None	100	1	1.0
Small (10 000)	3.0	High	10	30	0.05
Small (10 000)	3.0	Med	30	10	0.1
Small (10 000)	3.0	None	300	1	1.0
Large (250 000)	1.0	High	25	100	0.05
Large (250 000)	1.0	Med	50	50	0.1
Large (250 000)	1.0	None	2500	1	1.0
Large (250 000)	3.0	High	25	300	0.05
Large (250 000)	3.0	Med	150	50	0.1
Large (250 000)	3.0	None	7500	1	1.0