EVALUATION OF THE ASSESSMENT OF DOTHISTROMA NEEDLE BLIGHT IN STANDS OF PINUS RADIATA

J. B. VAN DER PAS, M. O. KIMBERLEY, and D. J. KERSHAW

Forest Research Institute, New Zealand Forest Service,
Private Bag, Rotorua, New Zealand

(Received for publication 22 November 1983; revision 23 December 1983)

ABSTRACT

Disease assessment of Dothistroma needle blight in stands of **Pinus radiata** D. Don was evaluated for accuracy. Disease levels were scored visually by estimating the percentage of the normal crown depth infected. Seven independent observers made a tree-by-tree scoring of eight 200-tree transects on the ground and six of the observers assessed the transects from the air; subsequently 11 ground observers assessed two transects of 100 trees and four of the observers re-assessed them 1 week later.

Estimates of variance components for ground assessments showed that appreciable variation was attributable to interactions between observer and tree and the residual error effects. This can be largely eliminated by increasing sample size. The variation caused by change in observer bias in assessments repeated after 1 week was of only minor importance. The effect of observer bias was appreciable but can be reduced by increasing the number of observers. Although the size of the different variance components varied among transects, their relative contribution remained fairly constant. Greater accuracy was achieved by more-experienced observers. Corresponding variance component estimates for aerial assessments were slightly smaller than for ground assessments. Disease ratings of the transects showed good agreement in ranking between ground and aerial assessments.

Accuracy of disease assessments depends on site conditions, observer skill, and sample size and structure. The mean disease level of a 100-tree transect obtained from ground assessments by three observers will have a coefficient of variation of about 10%. An aerial assessment by two observers of a single transect will have a coefficient of variation of about 25% since it involves assigning a single global figure per observer per transect.

INTRODUCTION

Dothistroma pini Hulbary causes needle blight in many exotic pine species in New Zealand (Gilmour 1967). *Pinus radiata* is significantly affected mainly during early growth (Gilmour & Noorderhaven 1973; Gilmour *et al.* 1973; Whyte 1976; van der Pas 1981), but trees normally become resistant by the age of 15 years (Bassett 1972).

Operational control of the disease by aerial application of copper fungicides started in 1966 and proved successful (Gilmour *et al.* 1973). Infected stands are sprayed each year before the main infection periods (Kershaw *et al.* 1982), and the total area that was sprayed in the period from 1966 to 1981 was approximately 640 000 ha (van der Pas *et al.* 1984).

Surveys for disease appraisal are needed for the decisions on control action. The survey of Dothistroma needle blight takes place each year during July and August, thus facilitating the logistics of the spray programme (Kershaw *et al.* 1982). Estimates of disease severity are also needed for evaluating the effect of fungicides and to estimate growth losses. Objective methods such as photogrammetric techniques have not proved practicable for assessing large numbers of trees or stands in a relatively short period. Instead, disease levels are estimated by visual assessment of individual trees on the ground, or of whole stands from the air. The trees or stands are rated by scoring the percentage of green crown that is infected (*see* Kershaw *et al.* 1982).

For satisfactory use the visual scores must be consistent and reproducible (*see* Large 1966), and so this study was initiated to evaluate the scoring techniques that are used in surveys. The main objectives were to examine the reliability and statistical properties of disease scores of both the ground and the aerial assessments.

Annual Survey and Assessment Techniques

Annual surveys to demarcate areas with sprayable levels of Dothistroma needle blight were initially conducted from the ground but this proved too time-consuming for large areas. Subsequently, surveys were conducted from the air with fixed wing aircraft, at between 20 and 100 m above ground level. Since the early 1970s helicopters have been used for the larger forests as this enables observation from low altitudes.

Usually trees at age 2–15 years are assessed as they are susceptible to infection but this rule is not rigid (Kershaw *et al.* 1982). The area of susceptible *P. radiata* stands increased from 140 000 ha in 1971 to 528 000 ha in 1981, and is expected to increase to 645 000 ha in 1991 (Table 1).

Initially disease levels were assessed using the scoring system developed by J. W. Gilmour (Gilmour & Noorderhaven 1971, 1973). This system uses a modified logarithmic scale patterned on the Horsfall-Barrett system for measuring disease in agronomy (Horsfall 1945). The scale is an ascending numerical code from 0 to 7 where scoring is done visually by estimating the percentage of the normal crown infected by *D. pini.* Based on this system, a percentage-in-step method was adopted in 1972 where disease levels are estimated arithmetically in 5% steps. This system, while it requires a high degree of skill and experience, reduces the assessment of disease to the utmost simplicity. To maintain high assessment standards, regular instruction and training sessions are held at the Forest Research Institute in Rotorua for Forest Health Officers.

METHODS

Field Procedures

Eight compartments of *P. radiata* with ages ranging from 4 to 6 years were selected in July 1979 in Kaingaroa State Forest for both aerial and ground assessments. The selection was made in order to represent in a realistic way the range of disease levels and different stand conditions. A transect of 350 m was laid down in the middle of

TABLE 1—Areas (000 ha, projected) of 2- to 15-year-old **Pinus radiata** for each planning district with **Dothistroma pini** present

District	Year					
	1971	1976	1981	1986	1991	1996
Northland	5	25	49	60	65	63
Auckland	11	34	64	84	93	90
Rotorua	80	156	224	251	229	186
Gisborne	4	17	31	42	47	46
Taranaki	1	4	7	10	12	13
Wellington	4	13	24	30	34	33
Hawke's Bay	6	21	37	47	54	52
Wairarapa	4	7	11	14	17	19
North Nelson	20	37	51	53	55	58
Marlborough	2	18	20	27	28	28
South Nelson	1	1	2	2	3	4
Westland	3	7	8	9	8	11
TOTAL	146	340	528	629	645	603

Source: New Zealand Forest Service National Planning Model 1969 and 1979.

each compartment along which 200 trees were tagged and numbered. The transects were marked by red painted drums on poles so that they were clearly visible from the air. Aerial assessment of the transects by six observers took place prior to the ground assessment. They were flown over the compartments by helicopter at approximately 50 m above tree level and at a speed of 45 knots. Disease levels were scored as if during a routine survey and all the compartments were assessed in one run. Each transect was then assessed on the ground by seven observers, including the six observers who had done the aerial assessment. In the aerial assessment a percentage rating was given for the whole transect, i.e., the mean value of all 200 trees; in the ground assessments each tree was individually scored. In this way 48 aerial scores and 11 200 individual tree scores were obtained.

In 1980 a test was made to evaluate consistency of personal bias of ground assessments by duplicating scores. Eleven observers, including the seven previously involved, assessed two 100-tree transects with medium to high disease levels in Kaingaroa State Forest. Four of the observers reassessed the same transect 1 week later but in the reverse sequence.

In order to keep assessments independent so as to prevent unwanted bias, the observers were not allowed to discuss results during assessments.

Analysis of Ground Assessment Data

For each transect a scatter plot of tree mean *versus* tree standard deviation was produced. This showed that trees with mean disease levels close to 0% or 100% had lower standard deviations than those with means in the middle of the range. A transformation providing a uniform standard deviation was therefore sought. The arcsin transformation was found to be unsatisfactory. A modified logit function

$$y = \ln ((1 + 0.98 x)/(99 - 0.98 x))$$

where x = assessed value (disease level), and y = transformed value, proved more successful and was used throughout the analysis.

To quantify the different sources of variation within each transect, a linear model was adopted with the following variance components:

(1) σ^2_t Variance between disease levels of trees in the transect

(2) σ^2_o Variance due to across the board observer bias

(3) σ^2_{to} Variance due to observer × tree interaction

(4) $\sigma^2_{r:o}$ Variance due to change in observer bias in repeated assessments

(5) σ^2_e Residual error variance

As there were no assessments repeated during the 1979 exercise, σ^2_{to} and σ^2_e could not be separated and $\sigma^2_{r:o}$ could not be estimated. In 1980, since four of the 11 observers repeated their assessment 1 week after the initial assessment, each component could be estimated using Henderson's method 1 (Henderson 1953) (*see* Table 2 and Appendix 1).

TABLE 2—Estimated variance components of ground assessments

Year	Transect in compartment	σ^{2**}_t	σ^{2NS}_o	$\sigma^{2**}_{r:o}$	$(\sigma^2_o + \sigma^2_{r:o})^{**}$	σ^{2**}_{to}	σ^2_e	$(\sigma^2_{to} + \sigma^2_e)^{**}$
1979	1076B	0.21			0.058			0.29
	1076A	0.28			0.052			0.21
	906B	0.46			0.068			0.25
	906A	0.20			0.050			0.21
	904B	0.66			0.055			0.23
	904A	0.52			0.041			0.21
	903	0.41			0.159			0.23
	902	0.31			0.064			0.18
	Mean	0.38			0.068			0.23
1980	Transect A	1.29	0.033	0.011	0.044	0.091	0.107	0.20
	Transect B	1.39	0.091	0.014	0.105	0.116	0.268	0.38

** Significant in all transects ($p < 0.01$)

NS Not significant

Not surprisingly, not all of the assumptions inherent in the model are fulfilled in practice. In particular, the following shortcomings can be identified.

(1) It is assumed that all effects are random when this is not strictly true. In particular, there is no drawing from a large population of observers since those who took part in the exercise make up a considerable proportion of the observers engaged in this work. However, this should not greatly affect the results. If anything, it means that the estimate of σ^2_o, based as it is on a nearly complete sample, is more accurate than if it had been based on a sample from an effectively infinite population of observers.

(2) The assumed bias of each observer from the general mean consists of an over-all bias with variance σ^2_o and a random observer × tree interaction with variance σ^2_{to}. To demonstrate that this is an over-simplification, the trees from all transects of the 1979 assessments were divided into four disease classes (0–20%, 21–40%, 41–60%, 61–80%) according to tree mean disease level, and each observer's departure from the over-all class mean was plotted (Fig. 1). This showed that an observer's bias does not necessarily remain constant but can be affected by the disease level of the tree (particularly evident with Observers 6 and 7).

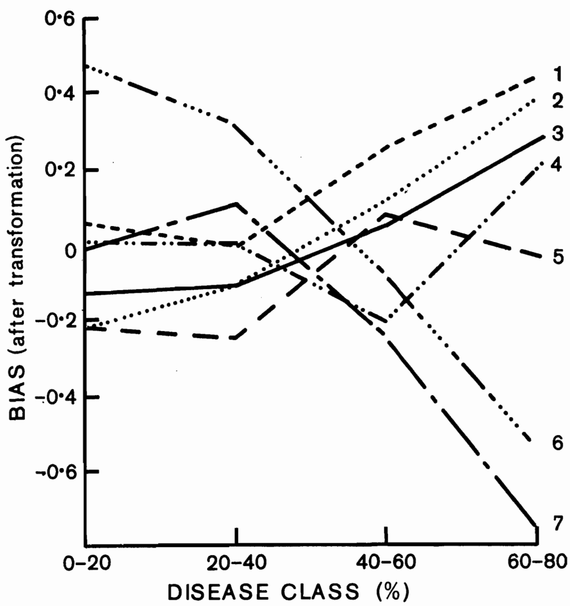

FIG. 1—Bias of seven observers in four disease classes.

(3) The model assumes that σ^2_{to} and σ^2_e are constant for all observers. Bartlett's test for homogeneity of variance was used to test whether the variance of the residuals obtained after removing the tree and observer effects varied between observers. It was significant for all transects in both the 1979 and 1980 assessments indicating that this variance, which is a combination of σ^2_{to} and σ^2_e, is not constant for all observers.

(4) $\sigma^2_{r:o}$ is based on a reassessment after only 1 week. Observer bias may change more over a longer period.

Despite the above deficiencies, it is believed that the analysis gives a good indication of the size and relative importance of the different sources of variation.

A general measure of an observer's accuracy is provided by the mean sum of the squared deviations (MSD) of individual assessments by the observer from assessment means over all observers. These were calculated for each of the 11 observers participating in the 1980 assessments. The MSD were considered in relation to a subjective ranking of the observer's level of experience provided by the senior author prior to the analysis and produced a Kendall's rank correlation coefficient of 0.53 which is significant at the 5% level.

Analysis of Aerial Assessment Data

As with ground assessments, accuracy of aerial assessments varied according to disease level of the transect but this was satisfactorily corrected by use of the same transformation. Applying Bartlett's test in a similar manner to the ground assessments it was shown that differences in precision among the observers were significant. This was caused largely by one observer who had much less experience than the others. When that observer's assessments were omitted from the analysis, there proved to be no significant heterogeneity of precision among the other observers. The three variance components that could be estimated, i.e., variance between transects, variance between observers, and residual variance, are shown in Table 3.

TABLE 3—Estimated variance components of the experienced observers

$\sigma^{2**}_{\text{Transects}}$	$\sigma^{2\text{NS}}_{\text{o}}$	σ^2_{e}
0.79	0.021	0.16

** Significant ($p < 0.01$)

NS Not significant

RESULTS AND DISCUSSION

Before considering the results in detail, it would be well to consider some of their limitations.

Firstly, we must assume that percentage infection (which is what the observers attempt to assess) is a realistic measure of the disease level. This assumption would apply whether the assessment was provided by a human observer or, for example, a photogrammetric device.

Secondly, there will always be some reservations attached to the assessments owing to the subjective and unpredictable nature of human observers. Visual estimates to quantify disease severity are widely used on many crops in agronomy (*see* Large 1966; Preece 1971; Chester 1950). Horsfall & Cowling (1978) concluded in their study of pathometry that the human eye is in effect a very useful photocell for quantifying disease severity. Nevertheless, the effect of time of day, weather conditions, or tiredness, for example, on an observer's accuracy must remain the subject of some doubt.

Thirdly, we have not considered the effect of sampling error, confining our attention only to observer accuracy. When assessing the disease level of a stand by using a sample of trees or transects, the accuracy of the assessment will be affected by the variability of the disease within the stand, which we have not considered, as well as by observer accuracy. This will not be so when complete sampling is used – for example, in experimental work when all the subject trees are ground assessed, or when assessing a whole compartment from the air to determine whether a spraying operation is necessary. It should be emphasised, however, that the accuracy of an assessment derived from the estimated variance components applies only to the mean of the trees assessed, and not to the tree population from which it was derived.

We will now consider the accuracy of ground assessments. Using MSD as a measure of accuracy, it was demonstrated that considerable differences in skill exist among observers. The square root of MSD is approximately equal to the average deviation from the mean assessment. In Transect A (1980 assessments) this figure was 1.5 times greater for the worst observer than for the best observer. In Transect B, where conditions for assessment were particularly difficult, it was 2.5 times greater – the worst observer was on average 2.5 times further from the general mean for a tree than the best observer. It appears from this that observer skill is more important when conditions for assessment are difficult. The level of skill was found to be related to the experience of the observer, as was apparent in the aerial assessments.

Examination of the variance components (Table 2) shows that, leaving aside the variance between trees, about 80% of the variation is attributable to σ^2_{to} and σ^2_e in all the transects, except in Cpt 903. The effect of these two components on the accuracy of the mean of a sample can be largely eliminated by taking a sufficiently large sample size. The component $\sigma^2_{r:o}$, although statistically significant, is of minor importance and accounts for 3% and 5% of the variance in Transects A and B respectively. It shows that observer bias changed, but only slightly, in the repeat assessment after 1 week. Such changes could be expected to increase when repeat assessments are made after a longer time interval. The non-significance of σ^2_o in part reflects the limited degrees of freedom available to test this component as a result of the small number of observers in the repeat assessment. The variance due to observer bias in a non-repeated assessment of mean disease levels is given by the sum of σ^2_o and $\sigma^2_{r:o}$, which is significant. This source of variation cannot be reduced by taking a larger sample but only by increasing the number of observers.

The size of the variance components varies among transects, reflecting the different levels of difficulty in assessment. Taking the inverse of the square root of the sum of the variance components (excluding σ^2_t) as a measure of relative accuracy the assessments in Transect B are shown to be approximately 40% less accurate than those in Transect A. The trees in Transect B were chosen for contrast since they were much more difficult to assess than those in Transect A. Similarly, the 1979 assessments are on average 11% less accurate than those in Transect A. However, although the accuracy varies among the transects, the relative size of each component within transects remains fairly constant.

Ground assessments are usually taken by more than one observer assessing the mean disease level of one or more trees, sometimes with several assessments during the year. The variance of such an estimate of the mean disease level of the trees assessed will theoretically be: $\sigma^2_o/o + \sigma^2_{to}/to + \sigma^2_{r:o}/ro + \sigma^2_e/tor$ where o = number of observers, t = number of trees, and r = number of repeated assessments. The standard errors of the assessment means have been calculated for different t, o, and r, using the variance components estimated from Transect A (Table 4) and the method of Appendix 2. It is shown that increasing the number of observers contributes most to the accuracy of assessment. For example, two observers each assessing 10 trees will give a more accurate assessment than one observer assessing 100 trees. Repeating the assessment gives little improvement in accuracy.

TABLE 4—Standard errors (%) of ground estimates of disease - the figures are based on the variance components of Transect A

No. of trees	No. of observers	No. of assessments	Estimated disease level 10%	20%	40%
1	1	1	4.8	8.2	12.1
1	1	2	4.2	7.1	10.5
1	2	1	3.4	5.8	8.5
1	2	2	3.0	5.0	7.4
1	3	1	2.8	4.7	7.0
1	3	2	2.4	4.1	6.1
10	1	1	2.5	4.2	6.2
10	1	2	2.3	3.8	5.6
10	2	1	1.8	3.0	4.4
10	2	2	1.6	2.7	4.0
10	3	1	1.4	2.4	3.6
10	3	2	1.3	2.2	3.3
25	1	1	2.2	3.8	5.6
25	1	2	2.1	3.5	5.2
25	2	1	1.6	2.7	4.0
25	2	2	1.5	2.5	3.6
25	3	1	1.3	2.2	3.2
25	3	2	1.2	2.0	3.0
100	1	1	2.1	3.6	5.3
100	1	2	2.0	3.3	4.9
100	2	1	1.5	2.5	3.7
100	2	2	1.4	2.4	3.5
100	3	1	1.2	2.1	3.0
100	3	2	1.1	1.9	2.8

The variance components for aerial assessment of the transects (Table 3) are slightly smaller than, but of a similar order to the components for ground assessments of trees. This may imply that the type of skill required for the two kinds of assessments is similar. Table 5 shows the standard errors of assessments for various disease levels, numbers of observers, and transects. A comparison between means for ground and aerial assessments (Table 6) indicates that the aerial assessments gave higher figures for low disease levels and lower figures for high disease levels than ground assessments. However, the two methods give a very similar ranking to the transects (Kendalls' rank correlation coefficient = 0.93) indicating that the discrepancy between methods is consistent. This could, if necessary, be corrected for during training sessions prior to annual aerial surveys.

TABLE 5—Standard errors (%) of aerial estimates of disease level

No. of transects	No. of observers	Estimated disease level		
		10%	20%	40%
1	1	4.2	7.1	10.4
1	2	3.0	5.0	7.4
1	4	2.1	3.5	5.2
2	1	3.1	5.3	7.8
2	2	2.0	3.3	4.9
2	4	1.5	2.6	3.9

TABLE 6—Comparison of ground with aerial assessments

Transect in compartment	Mean ground assessment (%)	Mean aerial assessment (%)	Difference (%)
1076B	55.0	47.5	7.5
1076A	61.6	54.2	7.4
906	10.8	19.2	- 8.4
904B	11.7	19.2	- 7.5
904A	9.3	10.0	- 0.7
903	10.2	13.3	- 3.1
902	16.5	22.5	- 6.0
1183	10.9	21.7	- 10.8

In practice the reliability and economy of sampling a stand or forest for disease severity are diametrically opposed. Therefore neither can be increased except at the expense of the other and usually a compromise will be sought. Both aerial and ground assessments of Dothistroma needle blight should have a practical degree of accuracy depending on the purpose of the appraisal, e.g., management survey or research trial. The results, however, should be comparable from one worker, location, or season to another. This can be accomplished by the use of standardised visual estimates and appropriate sample size and number of observers. The periodic training and "calibrating" of the observers is a necessity. The above results indicate that the accuracy of disease assessments is dependent mainly on site conditions, skill, and sample size. For research trials, an estimate with a coefficient of variation of less than 10% is thought to be suitable for most practical purposes. This figure is somewhat arbitrary and is given only to suggest a desirable order of magnitude of the permissible error. The coefficient of variation of a sample of 100 trees ground-assessed once by three observers (Table 4) is just within these limits. Aerial assessments of stands for sprayability require less accuracy. Usually such assessments are made by two observers and will therefore have a coefficient of variation of about 25% (Table 5) which probably represents an adequate level of accuracy.

ACKNOWLEDGMENTS

The authors are grateful to the Forest Health Officers who participated in the exercise. We would like to thank Dr R. D. Burdon, Dr P. D. Gadgil, and Mr I. A. Andrew for critical comments.

REFERENCES

BASSETT, C. 1972: The Dothistroma situation. **Farm Forestry 14:** 47-52.

CHESTER, K. S. 1950: Plant disease losses: their appraisal and interpretation. **Plant Disease Reporter, Supplement 193:** 189-362.

GILMOUR, J. W. 1967: Distribution and significance of the needle blight of pines caused by **Dothistroma pini** in New Zealand. **Plant Disease Reporter 51:** 727-30.

GILMOUR, J. W.; NOORDERHAVEN, A. 1971: Influence of time of application of cuprous oxide on control of **Dothistroma** needle blight. **New Zealand Journal of Forestry Science 1:** 160-6.

——— 1973: Control of **Dothistroma** needle blight by low volume aerial application of copper fungicides. **New Zealand Journal of Forestry Science 3:** 120-36.

GILMOUR, J. W.; LEGGAT, G. J.; FITZPATRICK, F. 1973: Operational control of **Dothistroma** needle blight in radiata pine plantations in New Zealand 1966-73. **Proceedings 26th New Zealand Weed and Pest Control Conference:** 130-8.

HORSFALL, J. G. 1945: "Fungicides and Their Action". Chronica Botanica Co., Waltham, Massachusetts. 239 p.

HORSFALL, J. G.; COWLING, E. B. 1978: Pathometry: The measurement of plant disease. Pp. 119-36 **in** Horsfall, J. G.; Cowling, E. B. (Ed.) "Plant Disease: An Advanced Treatise", Vol. 1. Academic Press, New York.

HENDERSON, C. R. 1953: Estimation of variance and covariance components. **Biometrics 9:** 226-52.

KENDALL, M. G.; STUART, A. 1963: "The Advanced Theory of Statistics", Vol. 1 (2nd ed.). Charles Griffen and Company.

KERSHAW, D. J.; GADGIL, P. D.; LEGGAT, G. J.; RAY, J. W.; VAN DER PAS, J. B. 1982: Assessment and control of **Dothistroma** needle blight (rev. ed.). **New Zealand Forest Service, FRI Bulletin No. 18.**

LARGE, E. C. 1966: Measuring plant disease. **Annual Review of Phytopathology 4:** 9-28.

PREECE, T. F. 1971: Disease assessment. Pp. 8-20 **in** Western, J. H. (Ed.) "Disease of Crop Plants". Wiley, New York.

van der PAS, J. B. 1981: Reduced early growth rates of **Pinus radiata** caused by **Dothistroma pini**. **New Zealand Journal of Forestry Science 11:** 210-20.

van der PAS, J. B.; BULMAN, L.; HORGAN, G. P. 1984: Disease control by aerial spraying of **Dothistroma pini** in tended stands of **Pinus radiata** in New Zealand. **New Zealand Journal of Forestry Science 14:** 23-40.

WHYTE, A. G. D. 1976: Spraying pine plantations with fungicides - the manager's dilemma. **Forest Ecology and Management 1:** 7-19.

APPENDIX 1

EXPECTED MEAN SQUARES FOR THE 1980 ASSESSMENTS

Source	d.f.	Expected mean squares
Tree	99	$\sigma^2_e + 1.347\ \sigma^2_{to} + 15\ \sigma^2_t$
Observer	10	$\sigma^2_e + 100\ \sigma^2_{r:o} + 1.347\ \sigma^2_{to} + 1.347\ \sigma^2_o$
Tree × observer	990	$\sigma^2_e + 1.347\ \sigma^2_{to}$
Re-assessment : observer	4	$\sigma^2_e + 100\ \sigma^2_{r:o}$
Residual	396	σ^2_e
TOTAL	1499	

APPENDIX 2

CALCULATION OF THE VARIANCE OF AN ASSESSMENT GIVEN THE VARIANCE OF THE TRANSFORMED ASSESSMENT

For a given disease level x, we have used the transformation:

$$y = \ln\left[\frac{1 + 0.98x}{99 - 0.98x}\right] \qquad (1)$$

Thus,

$$x = \frac{99\ \exp(y) - 1}{0.98\ (1 + \exp(y))}$$

Now, var(x) $\left[\frac{dx^2}{dy}\right]$ var. (y) evaluated at the expected value of y (e.g., Kendall & Stuart 1963, p. 232)

$$= \frac{100\ \exp(y)}{0.98\ (1 + \exp(y))^2}\ \text{var}(y) \qquad (2)$$

Thus, to find the variance of x for a particular mean disease level, we calculate y for this disease level using (1) and substitute it into (2).