An Assessment of Observer Variability in the Identification of Blister Rust Infection in Whitebark Pine

Marsha Huang

I Introduction

Whitebark pine (WBP) occurs in the subalpine zone of western North America, including the Pacific Northwest and Rocky Mountains, it is a highly valuable species ecologically and is a high-energy food source of many wild life species, including red squirrels and threatened grizzly bears. White pine blister rust (caused by the fungus Cronartium ribicola) was introduced into the United States about 1900 and has since spread throughout the range of white pine. Unfortunately, whitebark pine populations are threatened by white pine blister rust. The disease kills the upper, cone-bearing branches of whitebark pine long before the tree dies so that cone production is greatly diminished and subsequent tree regeneration is impossible. Blister rust has had the most devastating effects in climates with coastal influences but has the potential to affect all whitebark pine stands, including those in Yellowstone where the climate is drier and colder.

The following research is based on data provided by Greater Yellowstone Whitebark Pine Monitoring Working Group. For many reasons related to whitebark pines, the interest in blister rust has been increasing, and one of the objective in this study is to estimate the proportion of individual white bark pine trees(>1.4m high) infected with white pine blister rust, and to estimate the rate at which infection of trees is changing over time.

One source of error which has not been addressed by previous studies but may be extremely important is observer differences. Previous studies had showed that observer variability exists when identifying blister rust infection. If the variability is considered to be a fairly large contributor to the standard errors for our estimated parameters, we should add observer variability in our future models. Change in observer differences over time (season) and the relationship among differences by observers (such as observer experience and spatial relation) are also of interest in this study.

II. The Study Area

The Study Area was the Greater Yellowstone Ecosystem which contains 6 national forests and 2 national parks. During 2004, all tree stands sampled were within the Grizzly bear Primary Conservation Area (PCA) because of limitations in the mapped distribution of WBP for the entire study area. Sampling in 2005 was extended beyond the PCA. In the map in Figure 1 the red spots are sampled stands from 2004 and the green spots are sampled stands from 2005. Additional samples outside of PCA were collected in 2005.

III. sampling procedure

The data were collected throughout the Greater Yellowstone Ecosystem in 2004 and 2005 (07/26/2004-09/27/04 and 06/06/2005-08/25/2005). Twenty four transects in 2004 and 2005 were surveyed by multiple observers. More than one observer was present when surveying each transect, and each observer independently recorded the data. It is this data from multiply-recorded observations that forms the basis for the study of observer variability. The investigation is thus focused entirely on the consistency of data recorded by different observers.

To assess the effect of observer differences, independent surveys were conducted by different observers on 6 transects in 2004 and 18 transects in 2005, where one transect was double observed and the remaining 23 transects were triply observed. The observers recorded blister rust infections independently for each tree on the same transect. For each live tree, the presence or absence of indicators of blister rust were also recorded. A tree is classified as infected if either aecia or cankers were present. Ancillary indicators of blister rust include flagging, rodent chewing, oozing sap, roughened bark, and swelling. For a canker to be identified as having blister rust, at least 3 of the ancillary factors needed to be present. The four observers are coded as amy, eks, jjh and kas. In summary, amy observed 6 transects (2004 only) and 63 trees, eks observed 24 transects (2004, 2005) and 441 trees, jjh observed 24 transects (2004, 2005) and 441 trees and kas observed 17 transects (2005 only) and 317 trees.

Table 1 Summary of Multiple-observer data used for analysis of observer variability in 2004

Transect#	Number of trees observed	Observer and number of trees recorded as infected	Observer and number of trees recorded as Aecia present	Date
2602.1	4	Amy (1), eks (1), jjh(1)	Amy (1), eks (1), jjh(1)	07 / 26/04
531.1	13	Amy(1), eks(2), jjh(0)	Amy(1), eks(1), jjh(0)	07 / 27/ 04
4119.1	11	Amy(2), eks (1), jjh(4)	Amy(2), eks(1), jjh(3)	08 / 09/ 04
4299.1	20	Amy(2), eks(3), jjh(3)	Amy(0), eks(1), jjh(1)	08 / 10/ 04
4280.1	5	Amy(0), eks(0), jjh(0)	Amy(0), eks(0), jjh(0)	09 / 13/ 04
1830.1	7	Amy(1), eks(1), jjh(1)	Amy(0), eks(1), jjh(0)	09 / 27/ 04

3 out of the 6 transects observed by the three observers in 2004 have the same record of proportion of infected trees by the 3 observers. 2 out of 6 transects observed by the three observers in 2004 have the same record of proportion of aecia present trees by the 3 observers.

Table 2 Summary of Multiple-observer data used for analysis of observer variability in 2005

Transect#	Number of trees observed	Observer and number of trees recorded as infected	Observer and # of trees recorded as Aecia present	Date
4316.1	6	Eks(1),jjh(4),kas(4)	Eks(0), jjh(0), kas(0)	06/06/05
4316.5	5	Eks(2),jjh(2),kas(2)	Eks(0), jjh(1), kas(0)	
1172.1	59,60	Eks(35), jjh(41)	Eks(31), jjh(27)	06/21/05
1345.1	8	Eks(3),jjh(3),kas(3)	Eks(1), jjh(1), kas(1)	06/28/05
1345.2	27	Eks(13),jjh(13),kas(9)	Eks(11),jjh(11),kas(7)	
4175.1	26	Eks(12),jjh(13),kas(15)	Eks(2), jjh(2), kas(2)	07/05/05
11378.1	36	Eks(2), jjh(2), kas(1)	Eks(2), jjh(1), kas(0)	07/18/05
11441.1	71	Eks(17),jjh(18),kas(14)	Eks(16),jjh(18),kas(12)	07/19/05
11418.1	29	Eks(8), jjh(5), kas(5)	Eks(5), jjh(2), kas(4)	07/20/05
11418.5	7	Eks(4), jjh(4), kas(4)	Eks(2), jjh(1), kas(3)	
11390.1	30	Eks(5), jjh(6), kas(3)	Eks(4), jjh(4), kas(3)	07/21/05
11390.4	4	Eks(1), jjh(1), kas(1)	Eks(1), jjh(1), kas(1)	
7336.1	31	Eks(7), jjh(5), kas(8)	Eks(5), jjh(5), kas(5)	07/27/05
5756.3	2	Eks(1), jjh(1), kas(1)	Eks(0), jjh(0), kas(0)	08/22/05
5743.4	15	Eks(2), jjh(2), kas(1)	Eks(2), jjh(2), kas(1)	08/23/05
5741.5	3,2,1	Eks(0), jjh(0), kas(0)	Eks(0), jjh(0), kas(0)	08/24/05
5739.3	8	Eks(7), jjh(7), kas(6)	Eks(0), jjh(0), kas(1)	08/25/05
5739.4	11	Eks(4), jjh(5), kas(2)	Eks(1), jjh(2), kas(1)	

6 out of 17 transects observed by the three observers in 2005 have the same proportion of infected trees. 6 out of 17 transects observed by the three observers in 2005 have the same proportion of aecia present trees.

Table 3 Proportion of Infected trees and Aecia present trees of all multiple observed transects by each observer in year 2004 and 2005

Observer (# trees observed)	Proportion Infected	Proportion with Aecia
1 amy (63)	0.111	0.063
2 eks (441)	0.324	0.202
3 jjh (442)	0.319	0.188
4 kas (317)	0.249	0.129

Table 4 Proportion of Infected Trees and Trees with Aecia recorded by individual observer in 2004 and 2005

Observer	(2004) Proportion Infected	(2004) Proportion with Aecia	(2005) Proportion Infected	(2005) Proportion with Aecia
1 amy	.1228	.0823		
2 eks	.1265	.0994	.3676	.1804
3 jjh	.1511	.0955	.3872	.1803
4 kas			.3209	.1119

In 2004 and 2005 the three observers each has an estimated proportion of infected trees and trees with aecia. In 2004 the proportion of infection ranges from .12 to .15 and the proportion of trees with aecia range from .08 to .10. In 2005, the proportion of infection ranges from .32 to .38 and the proportion of trees with aecia range from .11 to .18.

Table 5 below shows the proportion of agreement by the multiple observers on absence/presence of Infection and Aecia.

Table 5 Proportion on agreement of absence of Infection and Aecia of Multiple Observers

Observer agreement	Inf_P absent	Aec_P absent
1,2 (2004)	.8197 (50/61)	.8852 (54/61)
1,3 (2004)	.8226 (51/62)	.8871 (55/62)
2,3 (both)	.6219 (273/439)	.7585 (333/439)
2,4 (2005)	.6625 (210/317)	.7981 (253/317)
3,4 (2005)	.6467 (205/317)	.7886 (250/317)
1,2,3 (2004)	.7833 (47/60)	.8500 (51/60)
2,3,4 (2005)	.6278 (199/317)	.7666 (243/317)

`

For the agreement of absence of Infection, observers 2 and 3 agree the least (62.19%) on the absence of Infection, while observers 1 and 3 agree the most (82.26%) on the absence of Infection.
For the agreement of absence of Aecia, observers 2 and 3 agree the least (75.85%) on the absence of Aecia while observers 1and 3 agree the most on the absence of Aecia (88.71%).

Table 6 Proportion on agreement on presence of Infection and Aecia of Multiple Observers

Observer (year)	Infection present	Aecia present
1,2 (2004)	.0656 (4/61)	.0328 (2/61)
1,3 (2004)	.0806 (5/62)	.0323 (2/62)
2,3 (2004 and 2005)	.2665 (117/439)	.1481 (65/439)
2,4 (2005)	.1924 (61/317)	.0946 (30/317)
3,4 (2005)	.1830 (58/317)	.0789 (25/317)
1,2,3 (2004)	.0667 (4/60)	.0333 (2/60)
2,3,4 (2005)	.1609 (51/317)	.0726 (23/317)

For the agreement of presence of Infection, observers 1 and 2 agree the least (6.56%) while observers 2 and 3 agree the most (26.65%).
For the agreement of presence of Aecia, observers 1 and 2 agree the least (3.28%) while observers 2 and 3 agree the most (14.81%).

IV. Summary of Observer Agreement

The following statistical procedures were used to study observer variability
1. Kappa coefficient: measure of agreement of multiple raters
2. McNemar test, to test if two observers are equally performing
3. Cochran's test, to test if three observers are equally performing

1. The kappa statistic

Cohen's kappa coefficient was calculated and it is a measure of how strong the agreement is between the two observers relative to random chance.

Table 7 Agreement of Observers 1 and 2 on Infection of Trees

frequency	observer 2		Total
observer 1	No	Yes	
No	50	4	54
Yes	3	4	7
Total	53	8	61

Consider readings on the trees that are reported as either infection/aecia present or absent by observers 1 and 2. The results are shown in Table 7 in a 2 by 2 format. Diagonal entries represent agreement and off-diagonal entries represent disagreement. The responses of each observer are the marginal totals. The overall proportion of agreement in regard to infection, which we will call Po (stands for observed proportion of agreement) is calculated by combining the diagonal entries. For this data, Po=54/61=88.5%

But Po may give false impression of agreement between the two observers. An alternative to the overall agreement, the absent and present agreement can be estimated separately. This will give an indication of the type of decision on which observers agree on divided by all of the positive readings for both observers. The absent agreement is Pabs= (50+50)/(54+53)=93.5% and the present agreement is Ppre= (4+4)/(8+7)= 53.3% In the example above, although the two observers agree 88.5% of the time overall, they only agree on present 53.3% of the time, whereas they agree on absent 93.5% of the time. The advantage of using Pabs and Ppre is that any imbalance in the proportion of absent and present responses becomes apparent.

The expected proportion of agreement by chance in regard to infection, which we will call Pe is calculated with this equation : Pe=(53*50 + 7 *8)/(61*61). An index called k has been developed as a measure of agreement corrected for chance. The Kappa statistic is defined to be k= (Po-Pe)/ (1-Pe).The equation below shows how to calculate k.

$$\text{kappa} = \frac{\text{Observed agreement} - \text{Expected agreement}}{100\% - \text{Expected agreement}}$$

Table 8 Guidelines for Strength of Agreement Indicated with k Values

K Value	Strength of Agreement Beyond Chance
<0	Poor
0-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

*Landis, JR and Koch, GG. The measurement of observer agreement for categorical data. Biometrics 33:159-174, 1977.

Kappa =0 indicates there is no agreement beyond random chance, this happens when Po equals Pe
Kappa =1 indicates perfect agreement beyond random chance when the observed proportion of agreement beyond random chance reaches its maximum (1-Pe)
The values in between show the various strengths of agreement beyond random chance suggested by Landis and Koch (see table 8). For the observers 1 and 2 data, we have kappa coefficient of 0.4682 which is moderate according to table 8.

Simple Kappa Coefficient

Kappa	0.4682
ASE	0.1709
95% Lower Conf Limit	0.1334
95% Upper Conf Limit	0.8031

Table 9 Agreement by double observers on Infection

Agreement index	Type of agreement	Observers 1,2	Observers 1,3	Observers 2,3	Observers 2,4	Observers 3,4
Po	Overall	0.8853	0.9032	0.8884	0.8549	0.8297
Pabs	Absent (0)	0.9346	0.9444	0.9176	0.9013	0.8836
Ppre	Present (1)	0.5333	0.6250	0.8269	0.7262	0.6824
Pe	Chance	0.7842	0.7747	0.5631	0.6100	0.6068
Kappa (k) coefficient	Chance corrected	0.4682 (moderate)	0.5704 (moderate)	0.7445 (substantial)	0.6280 (substantial)	0.5668 (moderate)
CI for k	95%	.1334-.8031	.2641-.8768	.6774-.8116	.5306-.7253	.4645-.6691

The results of the double observer record are given by Table 9. The kappa values varied between 0.4682 (moderate) and 0.7445 (substantial). Observers 2 and 3 has the highest kappa coefficient (0.7445) with overall proportion of agreement (0.8884) and also has a high proportion of agreement on absence of infection (0.9176) and presence of infection (0.8269). Observers 1 and 2 has the lowest kappa k= 0.4682 with a good overall agreement Po=0.8853 and a strong agreement on absence of infection (0.9346) but with poor agreement on presence of infection (0.5333). This paradoxical result is caused by the high prevalence of 0 cases (absent cases).

Table 10 Agreement of double observers on Aecia

Agreement index	Type of agreement	Observers 1,2	Observers 1,3	Observers 2,3	Observers 2,4	Observers 3,4
Po	Overall	0.9180	0.9194	0.9066	0.8927	0.8675
Pabs	Absent (0)	0.9558	0.9565	0.9420	0.9370	0.9225
Ppre	Present (1)	0.4444	0.4444	0.7602	0.6383	0.5435
Pe	Chance	.8632	0.8652	0.6862	0.7467	0.7514
Kappa (k) coefficient	Chance corrected	0.4008	0.4015	0.7024	0.5765	.4671
CI for k	95%	-.0241-.8257	-.0230-.8261	.6173-.7874	.4496-.7035	.3304-.6037

The results of the double observer record are given by Table 10. The kappa values varied between 0.4008 (moderate) and 0.7024(substantial). Observers 2 and 3 has the highest kappa coefficient (0.7024) with overall proportion of agreement (0.9066) and high proportion of agreement on absence of infection (0.9420) and presence of infection (0.7602). Observer 1, 2 has the lowest kappa k= 0.4008 with strong overall agreement Po=0.9180 and strong agreement on absence of infection (0.9558) but a poor agreement on presence of infection (0.4444). Observers 1 and 3 has the similar results.

2. Mcnemar's Test For Interobserver Variability

McNemar's Test is generally used when the data consist of paired observations of labels. These data cannot be analyzed with a test on binomial proportions because the two samples are not independent. The McNemar test is a test on a 2x2 classification table when the two classification factors are dependent, or when you want to test difference between paired proportions. In this analysis, Mcnemar's test is used to test if the two observers recorded assessments on the same live trees differ randomly or if one of them gives a significantly higher or lower proportion than the other one.

Below are the 2x2 contingency tables for paired observers and Mcnemar's test results concerning infection of trees.

Table of Infection for observers 1 and 2

frequency	observer 2		Total
observer 1	No	Yes	
No	50	4	54
Yes	3	4	7
Total	53	8	61

McNemar's test	
Statistic (S)	0.1429
Pr > S	0.7055

Table of Infection for observers 1 and 3

frequency	observer 3		Total
observer 1	No	Yes	
No	51	4	55
Yes	2	5	7
Total	53	9	62

McNemar's test	
Statistic (S)	0.6667
Pr > S	0.4142

Table of Infection for observers 2 and 3

frequency	observer 3		Total
observer 2	No	Yes	
No	273	23	296
Yes	26	117	143
Total	299	140	439

McNemar's test	
Statistic (S)	0.1837
Pr > S	0.6682

Table of Infection for observers 2 and 4

frequency observer 2	observer 4 No	Yes	Total
No	210	18	228
Yes	28	61	89
Total	238	79	317

McNemar's test
Statistic (S) 2.1739
Pr > S 0.1404

Table of Infection for observers 3 and 4

frequency observer 3	observer 4 No	Yes	Total
No	205	21	226
Yes	33	58	91
Total	238	79	317

McNemar's test
Statistic (S) 2.6667
Pr > S 0.1025

The tables above show no significant recorded differences between each paired observers at the level of 0.10. The large p-values for comparisons of observers 1 and 2, observers 1 and 3, and observers 2 and 3 suggest that these pairs of observers did not differ systematically on the job of reporting the presence or absence of infection. Although the smaller p-values for the comparisons of observers 2 and 3 with observer 4 (p-value = .1404, .1025) are not statistically significant, they suggest that future assessments with observer 4 should be viewed cautiously.

Below are the 2x2 contingency tables for paired observers on results recorded concerning Aecia of trees.

Table of Aecia for observers 1 and 2

frequency observer 1	observer 2 No	Yes	Total
No	54	3	57
Yes	2	2	4
Total	56	5	61

McNemar's test
Statistic (S) 0.2
Pr > S 1.0

Table of Aecia for observers 1 and 3

frequency observer 1	observer 3 No	Yes	Total
No	55	3	58
Yes	2	2	4

McNemar's test
Statistic (S) 0.2
Pr > S 1.0

Total	57	5	62

Table of Aecia for observers 2 and 3

frequency	observer 3		Total
observer 2	No	Yes	
No	333	17	350
Yes	24	65	89
Total	357	82	439

McNemar's test

Statistic (S) 1.195

Pr > S 0.3489

Table of Aecia for observers 2 and 4

frequency	observer 4		Total
observer 2	No	Yes	
No	253	11	264
Yes	23	30	53
Total	276	41	317

McNemar's test

Statistic (S) 4.235

Pr > S 0.0576

Table of Aecia for observers 3 and 4

frequency	observer 4		Total
observer 3	No	Yes	
No	250	16	266
Yes	26	25	51
Total	276	41	317

McNemar's test

Statistic (S) 2.381

Pr > S 0.1641

The table for observers 2 and 4 has p-value of 0.0576 and it shows a significant recording difference between them at the level of 0.1, which indicates the two observers differ systematically on the job. The rest of the larger p-values in the tables above indicate no significant recording difference between each paired observers at the level of 0.10. The large p-values for comparisons of observers 1 and 2, observers 1 and 3, and observers 2 and 3 suggest that these pairs of observers did not differ systematically on the job of reporting the presence or absence of aecia. Although the smaller p-value for the comparison of observers 3 and 4 (p-value = .1641) is not statistically significant along with the significant observer 2 and 4 difference, suggest that future assessments with observer 4 should be viewed cautiously.

Table 11 Mcnemar's Test on double observers variability

Ho: randomly differ

Ha: do not randomly differ

McNemar's Test	Observer 1,2 p-value	Observer 1,3 p-value	Observer 2,3 p-value	Observer 2,4 p-value	Observer 3,4 p-value
Infection	0.7055	0.4142	0.6682	0.1404	0.1025

Aecia	1.00	1.00	0.3489	0.0576	0.1641

3. Cochran's test for Inter-observer Variability

Cochran's test is an extension of the McNemar test for related samples that provides a method for testing for differences between 3 or more matched sets of frequencies or proportions. Cochran's test will be used for testing if three observers are equally capable of reporting the presence or absence of infection or aecia or there are some differences among the observers beyond random chance.

In general, Cochran's test can be used when matching samples can be based on k characteristics of N individuals that are associated with the response, or alternatively, when N individuals (trees) may be observed under k different treatments or conditions, in this data k is the number of observers.

Results from SAS:

```
1.  Cochran's test for observers 1,2,3 on infection

                Column   total   cases   proportion

                 amy       7       60      0.11667
                 eks       8       60      0.13333
                 jjh       9       60      0.15000

                          Cochran's test
                         chisq    df   pvalue
                         0.667     2   .7165
```

Table 11 Agreement of triple observers 1,2,3 on presence of infection of Blister Rust

Infp123	Frequency	percent
Disagree	9	15
Agree absent	47	78
present	4	7
Cochran's test	Chisqr: 0.667 df: 2 P-value: 0.7165	

```
2.  Cochran's test for observers 2,3,4 on infection

               Column   total   cases   proportion
                 eks      89      317      0.28076
                 jjh      91      317      0.28707
                 kas      79      317      0.24921
                          Cochran's test
                         chisq    df   pvalue
```

3.701 2 .1571

Table 12 Agreement of triple observers 2,3,4 on presence of infection of Blister Rust

Infp234	Frequency	Percent
Disagree	67	21.14
Agree absent present	199 51	62.78 16.09
Cochran's test	Chisqr: 3.701 df: 2 P-value: 0.1571	

The p-values of .7165 and .1571 for the comparisons of observers 1 ,2, and 3 and observers 2,3, and 4 suggest that these three observers did not differ systematically on the job of reporting the presence or absence of infection. Although the smaller p-value for the comparison of observers 2, 3 and 4 (p-value = .1571) is not statistically significant, it is still recommended that future assessments with observer 4 be viewed cautiously.

3. Cochran's test for observers 1,2,3 on aecia

Column	total	cases	proportion
amy	4	60	0.066667
eks	5	60	0.083333
jjh	5	60	0.083333

Cochran's test

chisq	df	pvalue
0.286	2	.8669

Table 13 Agreement of triple observers 1,2,3 on presence of Aecia

Aecp123	Frequency	percent
Disagree 0	7	11. 67
Agree 1 : absent present	51 2	85 3. 33
Cochran's test	Chisqr: 0.286 df: 2 P-value: 0.8669	

3. Cochran's test for observers 2,3,4 on aecia

Column	total	cases	proportion
eks	53	317	0.16719
jjh	51	317	0.16088
kas	41	317	0.12934

Cochran's test

chisq df pvalue

4.863 2 .0879

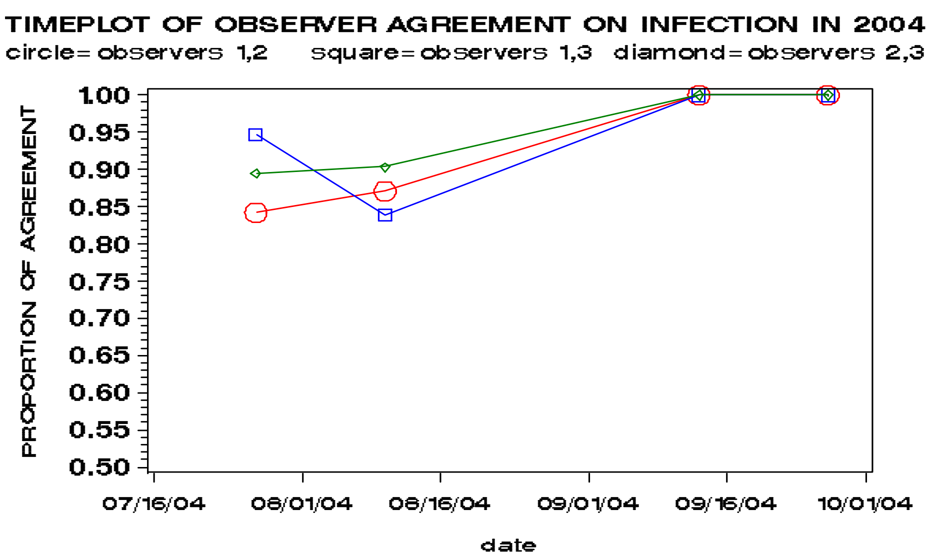
Table 14 Agreement of triple observers 2,3,4 on presence of Aecia

Aecp234	Frequency	Percent
Disagree 0	51	16. 09
Agree 1 : absent present	243 23	76. 66 7. 25
Cochran's test	Chisqr: 4.863 df: 2 P-value: 0.0879	

The table for observers 2,3 and 4 has p-value of 0.0879 and it shows a significant recording difference for aecia among them at the level of 0.1, which indicates there exists a systematic difference on the job of recording presence or absence of aecia among these three observers. The p-value of .8669 for the comparison of observers 1 ,2, and 3 suggests that these three observers did not differ systematically on the job of reporting the presence or absence of aecia.

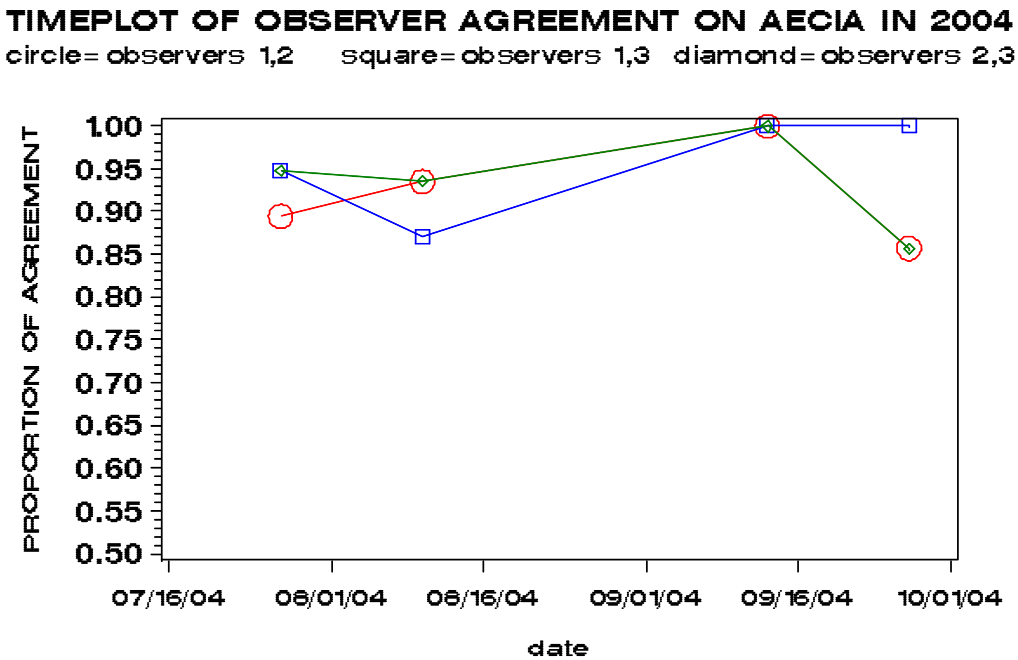
V. Seasonal effects

To study potential seasonal effects (such as learning curve for inexperienced observers), plots of the proportion of observer agreement were generated across time on days when multiple observers recorded measurements on the same transect. In the following plots, a circle represents observers 1 and 2 , a square represents observers 1 and 3, and a diamond represents observers 2 and 3. Data tables summarizing the plotted values are also given.



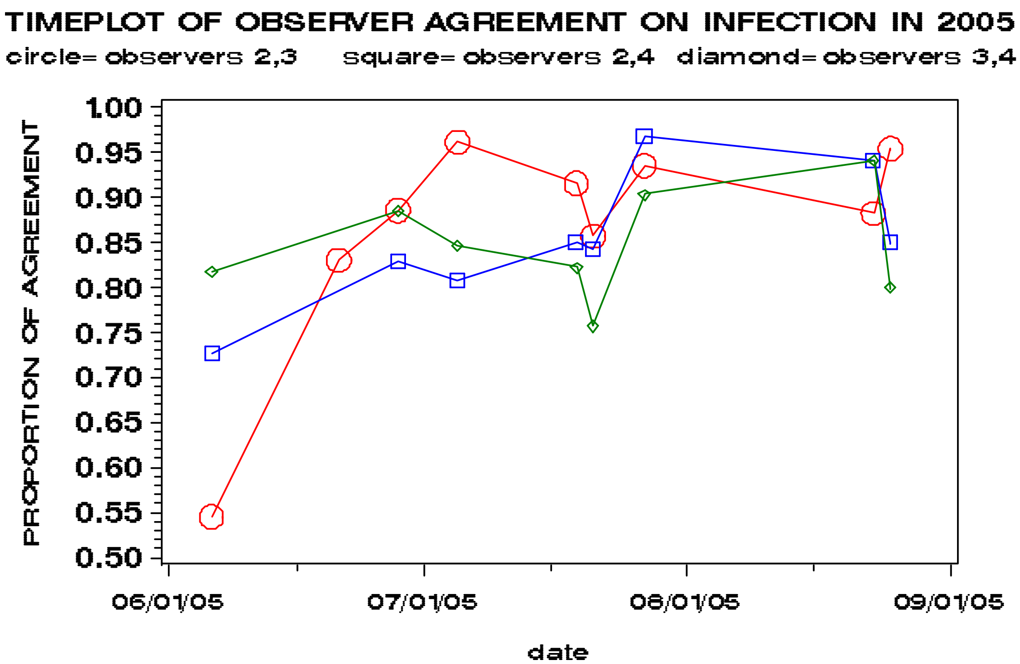
	circle (1,2)	square(1,3)	diamond(2,3)
07/26 07/27 (17)	0.9000	0.9667	0.9334
08/09 08/10 (31)	0.8795	0.8349	0.8636
09/13 (5)	1.0000	1.0000	1.0000

09/27 (7)	1.0000	1.0000	1.0000



	circle(1,2)	square(1,3)	diamond(2,3)
07/26 07/27 (17)	0.9335	0.9667	0.9667
08/09 08/10 (31)	0.9295	0.8386	0.9091
09/13 (5)	1.0000	1.0000	1.0000
09/27 (7)	0.8571	1.0000	0.8571

The agreement on infection and on aecia in 2004 is fairly good with values larger than .8349. There is a slight increasing tendency in the proportion of agreement which may indicate a learning effect.

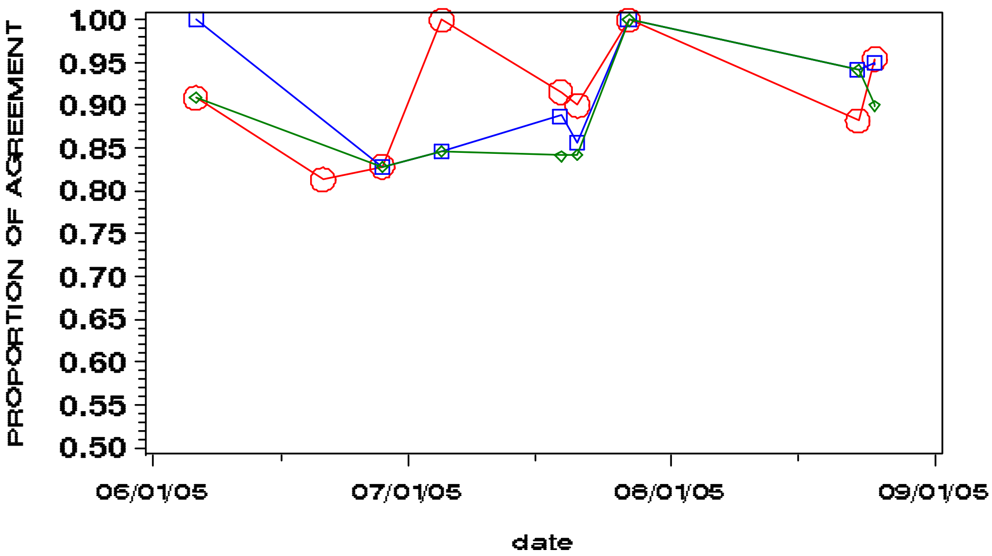


	circle(2,3)	square(2,4)	diamond(3,4)
06/06 (11)	0.5455	0.7273	0.8182
06/21 (59)	0.8305		
06/28 (35)	0.8857	0.8286	0.8857
07/05 (26)	0.9615	0.8077	0.8462
07/18 07/19 (107)	0.9159	0.8505	0.8224
07/20 07/21 (70)	0.8571	0.8429	0.7571
07/27 (31)	0.9355	0.9677	0.9032
08/22 08/23 (15)	0.8824	0.9412	0.9412
08/24 08/25 (20)	0.9546	0.8500	0.8000

There is a large variability among the observers in 2005 with values ranged from .5455 to .9677. Again there is an increasing trend in the agreement which may indicate a learning effect.

TIMEPLOT OF OBSERVER AGREEMENT ON AECIA IN 2005

circle= observers 2,3 square= observers 2,4 diamond= observers 3,4



	red (2,3)	blue(2,4)	green(3,4)
06/06	0.9091	1.0000	0.9091
06/21	0.8136		
06/28	0.8286	0.8286	0.8286
07/05	1.0000	0.8462	0.8462
07/18 07/19	0.9159	0.8879	0.8411
07/20 07/21	0.9000	0.8571	0.8429
07/27	1.0000	1.0000	1.0000
08/22 08/23	0.8824	0.9412	0.9412
08/24 08/25	0.9546	0.9500	0.9000

Average:

The agreement still has some variability but with a smaller range of .8136 to 1.0000. The agreement is fairly good and there seems to be a random fluctuation trend in the plot.

Conclusion:

At the transect level, consistency among observers in estimating the proportions of trees infected or aecia present on each transect was, in general, only moderate. Because of the variability among the observer assessments, managers should be very careful when reporting estimates of the proportion of infected trees and estimates of the proportion of trees with aecia given potential observer-specific effects.

Although the overall proportions of agreement for the presence or absence of infection or aecia may seem relatively high (between 82% and 92%), this is not the case when the separate cases of agreement for presence and agreement for absence are studied separately. The majority of cases involve absence of infection or aecia, and agreement among observers remains high (between 88% and 96%). However, for the minority of cases that involve presence of infection or aecia, the agreement among observers is substantially lower (between 44% and 83%). Thus, it will be misleading to only consider overall proportions of agreement. This is further supported by the kappa statistics results. In addition, the McNemar's Test and Cochran's Test results indicate that the level of disagreement is highest when observer 4 is one of the observers, while there is much less concern for comparisons among observers 1, 2 and 3.

At this point in time, management needs to address the issues concerning observer-to-observer variability and its effect on the quality and reliability of estimation. Two ways of handling these concerns may be to either (1) delete points associated with disagreement between observer assessments of the presence or absence of infection or aecia or (2) when observers disagree, only use the recorded assessment of the more experienced observers. Conducting further experimental designs to better understand and, hopefully, lead to a reduction and control of observer variability are recommended.

The following output is good for plot the time plots of observer overall agreement in % of Inf and Aec.

agreement by day 2004

---------------------------------- year=1 ----------------------------------

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
day	6	24.6666667	26.1966919	1.0000000	64.0000000
Date	6	24.6666667	26.1966919	1.0000000	64.0000000
TYPE	6	0	0	0	0
FREQ	6	10.8333333	6.4316924	4.0000000	20.0000000
Inf_P1	6	0.1228626	0.0886578	0	0.2500000
Inf_P2	6	0.1264610	0.0816618	0	0.2500000
Inf_P3	6	0.1510823	0.1419045	0	0.3636364
Inf_P4	0	.	.	.	.
Aec_P1	6	0.0823864	0.1083873	0	0.2500000
Aec_P2	6	0.0993777	0.0875481	0	0.2500000
Aec_P3	6	0.0954545	0.1301620	0	0.2727273
Aec_P4	0	.	.	.	.
aecp12	6	0.9304834	0.0631745	0.8571429	1.0000000
aecp13	6	0.9351010	0.1058804	0.7272727	1.0000000
aecp23	6	0.9347763	0.0804807	0.8181818	1.0000000
aecp24	0	.	.	.	.
aecp34	0	.	.	.	.
aecp123	6	0.8985931	0.1058266	0.7272727	1.0000000
aecp234	0	.	.	.	.
infp12	6	0.9265152	0.0875949	0.8000000	1.0000000
infp13	6	0.9335859	0.0818990	0.8181818	1.0000000
infp23	6	0.9323232	0.1137340	0.7272727	1.0000000
infp24	0	.	.	.	.
infp34	0	.	.	.	.
infp123	6	0.8938312	0.1226113	0.7272727	1.0000000
infp234	0	.	.	.	.
na12	6	10.5000000	6.4109282	4.0000000	20.0000000

agreement by day 2005

---------------------------------- year=2 ----------------------------------

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
day	9	40.8888889	26.6947769	1.0000000	80.0000000
Date	9	40.8888889	26.6947769	1.0000000	80.0000000
TYPE	9	0	0	0	0
FREQ	9	42.1111111	31.1144839	11.0000000	107.0000000
Inf_P1	0	.	.	.	.
Inf_P2	9	0.3675622	0.1994101	0.1764706	0.7796610
Inf_P3	9	0.3871815	0.1987834	0.1612903	0.6833333
Inf_P4	8	0.3208559	0.1768389	0.1176471	0.5769231
Aec_P1	0	.	.	.	.
Aec_P2	9	0.1803927	0.1625481	0	0.5254237
Aec_P3	9	0.1802657	0.1296550	0.0769231	0.4500000
Aec_P4	8	0.1118626	0.0705107	0	0.2285714
aecp12	0	.	.	.	.
aecp13	0	.	.	.	.
aecp23	9	0.9115564	0.0661379	0.8135593	1.0000000
aecp24	8	0.9138619	0.0682894	0.8285714	1.0000000
aecp34	8	0.8886214	0.0603274	0.8285714	1.0000000
aecp123	0	.	.	.	.
aecp234	8	0.8628606	0.0783488	0.7428571	1.0000000
infp12	0	.	.	.	.
infp13	0	.	.	.	.
infp23	9	0.8631810	0.1269630	0.5454545	0.9615385
infp24	8	0.8519724	0.0751261	0.7272727	0.9677419
infp34	8	0.8467531	0.0600560	0.7571429	0.9411765
infp123	0	.	.	.	.
infp234	8	0.7827112	0.1102156	0.5454545	0.9032258
na12	9	0	0	0	0

agreement by day 2004

Obs	year	day	Date	_TYPE_	_FREQ_	aecp12	aecp13	aecp23	aecp24	aecp34
1	1	07/26/04	1	0	4	1.00000	1.00000	1.00000	.	.
2	1	07/27/04	2	0	17	0.86667	0.93333	0.93333	.	.
3	1	08/09/04	15	0	11	0.90909	0.72727	0.81818	.	.
4	1	08/10/04	16	0	20	0.95000	0.95000	1.00000	.	.
5	1	09/13/04	50	0	6	1.00000	1.00000	1.00000	.	.
6	1	09/27/04	64	0	7	0.85714	1.00000	0.85714	.	.

Obs	date	aecp123	aecp234	infp12	infp13	infp23	infp24	infp34	infp123	infp234	na12
1	07/26/04	1.00000	.	1.00000	1.00000	1.00000	.	.	1.00000	.	4
2	07/27/04	0.85714	.	0.80000	0.93333	0.86667	.	.	0.78571	.	15
3	08/09/04	0.72727	.	0.90909	0.81818	0.72727	.	.	0.72727	.	11
4	08/10/04	0.95000	.	0.85000	0.85000	1.00000	.	.	0.85000	.	20
5	09/13/04	1.00000	.	1.00000	1.00000	1.00000	.	.	1.00000	.	4
6	09/27/04	0.85714	.	1.00000	1.00000	1.00000	.	.	1.00000	.	7

agreement by day 2005

Obs	year	day	Date	_TYPE_	_FREQ_	aecp12	aecp13	aecp23	aecp24	aecp34
1	2	06/06/05	1	0	11	.	.	0.90909	1.00000	0.90909
2	2	06/21/05	16	0	60	.	.	0.81356	.	.
3	2	06/28/05	23	0	35	.	.	0.82857	0.82857	0.82857
4	2	07/05/05	30	0	26	.	.	1.00000	0.84615	0.84615
5	2	07/18-19/05	43	0	107	.	.	0.91589	0.88785	0.84112
6	2	07/20-21/05	45	0	70	.	.	0.90000	0.85714	0.84286
7	2	07/27/05	52	0	31	.	.	1.00000	1.00000	1.00000
8	2	08/22-23/05	78	0	17	.	.	0.88235	0.94118	0.94118
9	2	08/24-25/05	80	0	22	.	.	0.95455	0.95000	0.90000

Obs	date	aecp123	aecp234	infp12	infp13	infp23	infp24	infp34	infp123	infp234	na12
1	06/06/05	.	0.90909	.	.	0.54545	0.72727	0.81818	.	0.54545	0
2	06/21/05	.	.	.	.	0.83051	.	.	.	.	0
3	06/28/05	.	0.74286	.	.	0.88571	0.82857	0.88571	.	0.80000	0
4	07/05/05	.	0.84615	.	.	0.96154	0.80769	0.84615	.	0.80769	0
5	07/18-19/05	.	0.82243	.	.	0.91589	0.85047	0.82243	.	0.79439	0
6	07/20-21/05	.	0.80000	.	.	0.85714	0.84286	0.75714	.	0.72857	0
7	07/27/05	.	1.00000	.	.	0.93548	0.96774	0.90323	.	0.90323	0
8	08/22-23/05	.	0.88235	.	.	0.88235	0.94118	0.94118	.	0.88235	0
9	08/24-25/05	.	0.90000	.	.	0.95455	0.85000	0.80000	.	0.80000	0