OBSERVER BIAS AND THE DETECTION OF LOW-DENSITY INFESTATIONS: A CASE STUDY WITH THE HEMLOCK WOOLLY ADELGID

Matthew C. Fitzpatrick[1,2], Evan L. Preisser[1], Aaron M. Ellison[2], and Joseph S. Elkinton[3]

[1]University of Rhode Island, Department of Biological Sciences,
100 Flagg Road, Kingston, RI 02881

[2]Harvard University, Harvard Forest,
324 North Main Street, Petersham, MA 01366

[3]University of Massachusetts at Amherst,
Department of Plant, Soil, and Insect Sciences,
270 Stockbridge Road, Amherst, MA 01003

ABSTRACT

Monitoring programs, often comprised of volunteers, increasingly are used to document the spread of forest pests in the hope of detecting and eradicating low-density infestations before they become established. However, interobserver variation in the detection and correct identification of low-density populations of forest pests remains largely unexplored. In this study, we compared the abilities of novice observers and experienced individuals to detect low-density populations of the hemlock woolly adelgid (HWA) and we explore how interobserver variation can bias estimates of the proportion of site infested derived from models. We found that, compared to experienced individuals, novice observers detected HWA infestations at smaller proportion of sites and, on average, failed to detect low-density infestations. In contrast, models suggested that experienced observers had a higher probability of falsely detecting HWA as present than did novice individuals. This latter, unexpected finding can be explained by invoking heterogeneity in detection probabilities associated with variation in population abundance and differences in the ability of observers to detect low-density infestations. Our findings highlight some of the difficulties in sampling for low-density infestations of forest pests in general and for HWA in particular. More broadly, our results caution against the use of different sampling protocols in the same survey and suggest that models that estimate infestation rates should include survey-specific covariates that account for biases in detection probabilities introduced by interobserver variation or survey methods.

INTRODUCTION

The growing threat posed by invasive species has focused increased attention on the importance of documenting the distribution and spread of introduced organisms. Monitoring programs aimed at detecting low-density ‘founder’ populations can play a critical role in slowing or even stopping the spread of harmful invasives by identifying recently established populations that can be targeted for control and/or eradication (e.g., gypsy moth ‘Slow the Spread’ program). These efforts have proven remarkably successful against actively dispersing species, but founding populations of species that disperse passively by means of wind, water, or phoresy often prove far more difficult to locate. Without the ability to attract the organisms to a trapping location, researchers face the often daunting task of repeatedly searching potential habitats for low-density populations of the invading species.

The surveying problems posed by passively dispersing species are exemplified by the hemlock woolly adelgid, an invasive pest of eastern hemlock and Carolina hemlock in the eastern United States. HWA is a minuscule (<1-mm long adult), flightless insect that in the United States is both obligately parthenogenetic and exclusively passively dispersed. The parthenogenetic nature of HWA means that even a single colonizing individual can start a new infestation, producing an initially low-density population that only can be detected by costly and time-consuming surveys. Such surveys are increasingly being met in part by volunteer-based or ‘citizen science’ monitoring programs (e.g., CitSci.org). Although the educational and scientific benefits of volunteer-based invasive species monitoring programs are clear, the reliability of data collected by novice individuals has sometimes been questioned. However, these concerns stem mostly from the lack of studies comparing the quality of volunteer collected data versus professionally collected data rather than from studies demonstrating that volunteers collect unreliable data.

In this study, we first compare the abilities of inexperienced volunteers and experienced observers to detect low-density populations of an actively spreading forest pest, the hemlock woolly adelgid. We then use these data to explore the general question of how interobserver variation can bias estimates of the proportion of sites infested derived from models. We hypothesized that relative to experienced observers, novice individuals should be less likely to detect low-density populations and would be more prone to misidentification of the study species. To explore these hypotheses, we use maximum likelihood methods to select among models that consider differences in the ability of observers to both detect and correctly identify HWA. We parameterize these models using data from a 420-tree survey conducted by nine volunteers and three experienced individuals at Cadwell Memorial Forest in Pelham, MA. Our results support the notion that novice volunteers and experienced observers differ in their ability to detect low-density populations and that such differences in observer ability can bias estimates of the proportion of sites occupied. However, this bias manifests itself in unexpected ways.

METHODS

Twelve observers participated in the sampling effort: three experienced individuals who perform field research on HWA and nine volunteers who had no prior experience sampling for HWA populations. Prior to the sampling, the volunteers were trained for 15 minutes on the sampling methodology (see below) and on identifying HWA infestations. Each person was then assigned to one of four groups (n=3 persons per group). Two of the groups entirely were comprised of volunteers (hereafter referred to as ‘volunteer-only’). The remaining two groups contained one experienced and two volunteer individuals and two experienced and one volunteer individual (hereafter referred to as ‘volunteer/experienced’).

Observers searched all accessible branches for evidence of white woolly masses characteristic of the HWA sistens generation. Each search continued until either HWA was detected or a 2-minute sampling period had expired. To ensure that sampling was independent, no two observers sampled a tree at the same time and observers were instructed not to communicate the infestation status of trees to the other observers. To examine whether there were differences between volunteers and experienced individuals in terms of the density of infestations detected by each type of observer, two experienced individuals returned to all trees where HWA was detected, thoroughly searched all accessible branches, and counted the number of white wooly masses observed on the tree. This count provided an estimate of the number of detectable individuals on the tree. We used a t-test on log-transformed HWA abundance to compare the mean abundance of HWA infestations that were and were not detected by volunteers.

We used differences in detection abilities between volunteer and experienced observers to determine how such differences influence estimates of the proportion of infested hemlock trees. Our models incorporated three parameters: ψ, the proportion of infested hemlock trees, p11, the 'detection probability', the probability of detecting the species, given that the species is actually present at the site, and p10, the 'misclassification probability', the probability of falsely detecting the species at an unoccupied site. We considered four models that make different assumptions regarding p11 and p10. The simplest model assumes false positives are not possible (p10 = 0) and that detection probabilities are constant across observers. The second model again assumes that false positives were not possible, but allows observers to differ in their probability of detecting HWA. The final two models both incorporate the possibility of misclassification (p10 > 0), with the simpler of the two assuming that observers do not differ in their probability of detecting or misclassifying HWA. The more complex of these two models assumes that observers can differ in their probability of detecting and misclassifying HWA.

RESULTS

We found that relative to volunteers, experienced observers (1) detected infestations at a greater proportion of trees; (2) had a higher probability of detecting infestations; and (3) detected smaller infestations. Surprisingly, when compared to volunteers in their group, experienced observers had a higher probability of misclassifying other organisms as HWA. The form of the best-supported model also differed between volunteer-only groups and volunteer/experienced groups. For volunteer-only groups, models where the probability of misidentifying HWA was 0 (p10 = 0) were best supported by the data. In contrast, the best supported model for volunteer/experienced groups assumed misclassification probabilities were greater than 0 and both detection and misclassification probabilities differed between observers. There was little support for models that assumed that experienced and volunteer observers had equal probabilities of detecting HWA infestations.

CONCLUSIONS

We were initially surprised by the apparent result that experienced observers were more likely to misclassify HWA than volunteers. However, when we inspected detection histories we found that, for the team with one experienced and two volunteers, the two inexperienced observers detected HWA on only 1/125 trees when the experienced observer did not. In contrast, the experienced individual detected HWA 23 times when the two volunteers did not. These results suggest a failure by inexperienced observers to detect low-density infestations rather than misidentification by experienced observers. These results also reveal an issue regarding the absence of statistical weighting in the model. When a low-density infestation is detected by one observer, but missed by the remaining two individuals, statistical support tips in favor of misclassification. This finding cautions against the use of different survey protocols (or observers of differing levels of experience) in the same survey and suggest the need to include in models survey-specific covariates that account for biases in detection probabilities introduced by differences in observers or survey methods.

What do our results say about the adequacy of data on the distribution of low-density populations collected by volunteers? We suggest that the answer to this question depends on the ultimate use of the data and on the system under study. HWA, though easy to identify to the trained eye, can be extremely difficult to detect when occurring at low densities; our results suggest field experience can improve the ability to detect such infestations. Taken together, our results underscore the importance of adequate training for novice individuals taking part in monitoring programs and the need to document and account for interobserver variation in analytical estimates of infestation rates.