Journal of Wildlife Diseases, 46(1), 2010, pp. 118–135

A Weighted Surveillance Approach for Detecting Chronic Wasting Disease Foci

Daniel P. Walsh[1,2] and Michael W. Miller[1]

[1] Wildlife Health Program, Colorado Division of Wildlife, Wildlife Research Center, 317 West Prospect Road, Fort Collins, Colorado 80526-2097, USA

[2] Corresponding author (email: dan.walsh@state.co.us)

ABSTRACT: A key component of wildlife disease surveillance is determining the spread and geographic extent of pathogens by monitoring for infected individuals in regions where cases have not been previously detected. A practical challenge of such surveillance is developing reliable, yet cost-effective, approaches that remain sustainable when monitoring needs are prolonged or continuous, or when resources to support these efforts are limited. In order to improve the efficiency of chronic wasting disease (CWD) surveillance in Colorado, United States, we developed a weighted surveillance system exploiting observed differences in CWD prevalence across demographic strata within infected mule deer (*Odocoileus hemionus*) populations. We used field data to estimate sampling weights for individuals from eight demographic strata distinguished by differences in apparent health, sex, and age. In this system, individuals from a sample source with high prevalence and low inclusion probability (e.g., clinical CWD "suspects") received ≥10.3 times more weight than those from a source with low prevalence and high inclusion probability (e.g., apparently healthy, hunter-harvested individuals). We simulated use of this alternative surveillance system for a deer management unit in Colorado and evaluated the potential effects of using biased weights on the probability of failing to detect CWD and on relative surveillance costs. We found that this system should be transparent, cost-effective, and reasonably robust to the inadvertent use of biased weights. By implementing this, or a similar, weighted surveillance system, wildlife agencies should be able to maintain or improve current surveillance standards while, perhaps, collecting and examining fewer samples, thereby increasing the efficiency and cost-effectiveness of ongoing CWD surveillance programs.

Key words: Chronic wasting disease, disease detection, mule deer, *Odocoileus hemionus*, prion, sampling, weighted surveillance.

INTRODUCTION

Surveillance is a key component of effective wildlife disease monitoring and control. Surveillance typically focuses on estimating current levels of disease in areas of known occurrence and on monitoring both peripheral and distant areas to determine the geographic distribution and establishment of new disease foci. For apparently emerging wildlife diseases, determining their broad geographic distribution may be particularly critical to assessing implications and prospects for control. It follows that developing efficient approaches for detecting new disease foci could be valuable to wildlife managers and agencies responsible for surveillance within their jurisdictions.

A recently emerging prion disease of wildlife, chronic wasting disease (CWD), was first recognized in captive mule deer (*Odocoileus hemionus*; Williams and Young, 1980) and elk, and was subsequently diagnosed in free-ranging elk (*Cervus elaphus nelsoni*), mule deer, white-tailed deer (*Odocoileus virginianus*), and moose (*Alces alces*) in scattered foci across North America (Spraker et al., 1997; Williams, 2005; Baeten et al., 2007). Since 2002, considerable resources have been spent by wildlife management and animal health agencies and their partners around the world in conducting surveillance to better define the geographic distribution of CWD. Current CWD surveillance efforts in North America focus mainly on monitoring regions where the disease has as not yet been detected and on estimating prevalence in known infected areas; however, approaches for accomplishing the latter seem more straightforward and efficient than for the former (Samuel et al., 2003). In Colorado, United States, for example, monitoring for

CWD in populations where cases have, thus far, not been detected involves sampling cervids from a variety of sources including vehicle-killed animals, animals culled or recovered dead as CWD "suspects," hunter-killed (or "harvested") animals, and from other sources such as predator-killed or confiscated animals (Miller et al., 2000; Hibler et al., 2003; Krumm et al., 2005). Among submissions from hunters, there are generally numerous individuals from various demographic segments of the state's hunted cervid populations. Past surveillance methods treated these various sources that contributed to the surveillance stream—hereafter referred to as "strata"—separately for purposes of calculating needed sample sizes and did not always combine data from these strata. Consequently, this traditional approach failed to capitalize on available information concerning rather large differences in apparent CWD prevalence among these strata (Miller et al., 2000, 2008; Miller and Conner, 2005; Krumm et al., 2005).

To improve the efficiency of CWD surveillance in Colorado, and perhaps elsewhere, we developed a weighted surveillance system that makes use of all available information on stratum-specific prevalence. Similar approaches have been espoused by Cannon (2002) and have been specifically applied to bovine spongiform encephalopathy monitoring in European Union member states (Wilesmith et al., 2004). For any, user-specified probability of detecting disease (e.g., 95% probability of detecting at least one case where prevalence is ≥1%), weighted surveillance not only provides rigorous estimates of sample sizes to demonstrate that a region is nominally "disease-free" using combined sample sources, but it should also be more cost-effective than traditional approaches. The weighted surveillance system described here is intended to be used in aggregating data from a stratified sample collected from primary sampling units, selected a priori, based on some biologically-relevant spatial sampling scheme. Although we recognize the importance of having a spatial sampling scheme to account for the spatial variability and focal nature of CWD (Samuel et al., 2003; Conner and Miller, 2004; Joly et al., 2006; Nusser et al., 2008), describing the design of such a spatial sampling scheme is beyond the scope of this paper. Consequently, we limit our discussion to the design and application of a weighted sampling system for detecting new CWD foci.

MATERIALS AND METHODS

Derivation of the weighted surveillance system

The first step in developing this weighted surveillance system was to estimate the weights for each specific stratum in the CWD surveillance "stream." To estimate these weights, we made the following assumptions: The number of positive cases at the time of the survey in each of *ith* strata was independently distributed as Poisson (λ_i) random variables, individuals were randomly selected within the *ith* stratum for sampling, and relative prevalence within each stratum was constant across different population prevalence levels. Based on these assumptions, the maximum likelihood estimates for the weights ($\hat{w}_i$) were:

$$\hat{w}_i = \frac{\hat{p}_i}{\hat{p}_0}, \tag{1}$$

where $\hat{p}_i = x_i/n_i$ was the maximum likelihood estimate of the prevalence for the *ith* stratum and $\hat{p}_0$ was the corresponding estimate for the baseline stratum. The weight (w_0) for the "baseline stratum" was defined to be 1; this baseline stratum was considered the reference stratum to which the relative weights for the remaining strata were scaled. Both p_i and p_0 can be empirically estimated from historic data, when available. For an in-depth derivation of the estimated weights and variances, interested readers are referred to Appendix A.

The next step in implementing this surveillance system was to determine when an adequate number of samples had been collected (i.e., to calculate a stopping point for surveillance). To facilitate this, we employed a points system (Cannon, 2002; Wilesmith et al., 2004). Under this system, every sample entering the surveillance stream received a number of weight points ($\hat{w}_i$), based on its respective stratum membership, and sampling continued until the total number of points summed to the target (t). The target, or cumulative number of points needed before enough samples have been collected to cease surveillance, was a function of the desired probability of detecting at least 1 CWD-positive case ($1-\alpha$) and of the specified design prevalence for the baseline stratum. This value was calculated (Dohoo et al., 2003) as

$$t = \frac{-\ln(\alpha)}{p_{design} f}, \quad (2)$$

where p_{design} was the specified design prevalence for p_0 (i.e., the prevalence at which the surveyor wishes to detect at least 1 CWD-positive case in the baseline stratum), $1-\alpha$ was the probability of detecting at least 1 CWD-positive case, and f was the sensitivity of the test. Thus, under weighted surveillance, sample collection from the various strata continues until the following is true:

$$t = \sum_{i=0}^{m} n_i \hat{w}_i, \quad (3)$$

where n_i=the number of samples in the surveillance stream collected from the *ith* stratum and $\hat{w}_i$=the estimated weight for the *ith* stratum.

A potential concern in using weighted surveillance is the effects of biased weights (i.e., what happens when the estimated weights are above or below the unknown true weights). Such bias can lead to an over- or underestimation of the probability of detecting the disease. The amount of increase or reduction in disease detection probability, and the associated number of samples needed to reach the target value, depends on both the bias in the individual weights and the number of samples from each stratum. In Appendix B, we provide equations for estimating the bias in the number of samples required for achieving the target value as well as the bias in disease detection probability when weights are under- or overestimated relative to the true weight. Using simulations, we also examined, in more detail, the effects of varying levels of bias (shown below).

Several important factors about this weighted surveillance system merit further consideration. First, the weights, as described, are analogous to risk ratios. These types of ratios have been extensively studied (Cox, 1972; Arnanda-Ordaz, 1983; Bedrick et al., 1997). Second, we assumed thereafter that sensitivity of the CWD diagnostic test was one (i.e., f=1). Third, the p_{design} parameter used to calculate the target value is specified by the user and corresponds to the minimum prevalence level at which the user wishes to detect CWD in the baseline stratum (e.g., "a goal of detecting at least one case where CWD prevalence is ≥1% in adult males"). Fourth, the user also needs to choose the baseline population segment to which the specified design prevalence applies and to which other strata are scaled ("adult males" in the previous example). Although any stratum can be selected arbitrarily as the baseline, we recommend that this be a population stratum for which sample sizes are consistently largest, or a stratum that is sensitive to changes in prevalence. Based on the foregoing guidance, we chose harvested adult (≥2 yr old) male mule deer as the baseline stratum in the following example using Colorado field data.

Estimating weights using Colorado mule deer data

To illustrate use of the weighted surveillance approach, we calculated sampling weights for CWD surveillance in mule deer in Colorado, using data collected from 2003–2006, in parts of Colorado where CWD is known to occur. Data (Table 1) were from 20,400 deer from various sources that entered the surveillance stream and were tested for CWD as described elsewhere (Miller et al., 2000; Hibler et al., 2003); samples from live-animal testing, captive facilities, or culling as part of research studies were not included because these represented few samples and were from sources that contributed only sporadically to the surveillance stream. We divided the submitted cases into eight strata distinguished by differences in apparent health, sex, and age of deer included in each stratum: 1) clinical CWD "suspect" females >1 yr old; 2) clinical CWD suspect males >1 yr old; 3) harvested "adult" (≥2 yr old) males (the baseline stratum); 4) harvested adult females; 5) harvested "yearling" (>1 but <2 yr old) males; 6) harvested yearling females; 7) harvested "fawns" (<1 yr old, of either sex); or 8) all "other" dead deer (of both sexes and all ages except fawns). The "other" stratum included individuals recovered as vehicle-kills, predator-kills, and poaching seizures. Each of these strata received a unique weight (Table 1) calculated using equation (1); we estimated an associated SE for each weight using equation (9) in Appendix A.

Simulations to investigate properties of the weighted surveillance system

To examine the potential performance of the weighted surveillance system, we first examined the effects of employing biased weights (i.e., weights that were either greater or less than the true weight) on the probability of detecting CWD where true prevalence= 0.01 in harvested adult male deer, our baseline stratum. Initially, we used the simplest case

TABLE 1. The stratum-specific sample size, sample prevalence, estimated weights, and associated standard errors, based on surveillance results for mule deer in Colorado from 2003–2006, for use in a weighted surveillance system.

Stratum	Prevalence	Sample size	Positives	Total sample size	Total positives	Weights	SE of weights
Suspect—female	0.36	111	40	20,400	595	11.57	1.60
Suspect—male	0.32	125	40	20,400	595	10.27	1.46
Other	0.06	1,300	77	20,400	595	1.90	0.24
Harvest—adult male	0.03	10,046	313	20,400	595	1.00	NA[a]
Harvest—adult female	0.02	5,782	104	20,400	595	0.58	0.06
Harvest—yearling female	0.01	645	9	20,400	595	0.45	0.15
Harvest—yearling male	0.01	1,392	11	20,400	595	0.25	0.08
Harvest—fawn	0.00	999	1	20,400	595	0.03	0.03

[a] NA = Not applicable (Baseline stratum weight is defined to be 1.00).

(i.e., where all samples in the surveillance stream came from one stratum) to illustrate these effects using the stratum weight from Table 1 as the "true" simulation weight. It is important to note that we only assumed that estimated weights from Table 1 were the "true" weight for simulation purposes. We also investigated the effects of these biases on relative total surveillance costs. We repeated this analysis using two different strata, suspect female deer and harvested yearling males, because these strata represented the highest and lowest (aside from fawns) estimated weights. We examined biases of 0–50% in the estimated weights. We used equations (11) and (13) in Appendix B to calculate the bias in $1-\alpha$, sample size, and the relative costs arising from the use of biased weights. To calculate relative cost differences, we assumed an average cost of \$75.00 USD per submission, which was based on the estimated average testing and processing cost for a sample submitted into the Colorado Division of Wildlife's (CDOW) surveillance stream. We did not estimate or include costs of acquiring samples from various strata because CDOW field and laboratory personnel acquire such samples (e.g., vehicle kills, culled clinical suspects, poaching cases, etc.) as part of normal wildlife health monitoring and law enforcement activities. Thus, in Colorado, there are relatively few differences in acquiring samples from different strata.

To further examine the properties and efficacy of this technique using a more realistic scenario, we created a modeled population of mule deer and estimated, via simulation, the probability of detecting at least one CWD-positive individual ($1-\alpha$) with this weighted surveillance system when prevalence was set at 0.01 for the harvested adult male population. For this simulation, we used mule deer population estimates, sex-age ratios, mortality, and harvest data for 2006 from a data analysis unit (DAU) located in southwestern Colorado (DAU D-19; B. Banulis, CDOW, pers. comm.). We chose this DAU because it was apparently CWD-free (Colorado Division of Wildlife, 2009) and relatively good demographic data were available. The population size was estimated at ~40,000 deer. We partitioned this population into the eight strata described above using the 10-yr average of sex and age ratios and the cause-specific mortality estimates from radio-collared deer (CDOW Big Game Harvest Survey, unpubl. data). Because some data were limited for males when stratifying the population into the various demographic strata, cause-specific mortality probabilities for males were based on estimates for females, except for harvest probabilities which were estimated for both sexes. To determine the probability of an individual entering the surveillance stream from the *ith* stratum, we used 2006 statewide estimated harvest rates in conjunction with 2006 CWD test submission rates from DAUs where CWD had been confirmed. We used representative harvest and surveillance data from elsewhere because few samples entered the surveillance stream from DAU D-19 in 2006, thereby precluding reliable estimation of DAU-specific submission parameters. We simulated sampling individuals from the population based on these sampling probabilities (Table 2), with each individual being assigned to one of the eight strata. Once an individual was selected, it was determined to be "positive" if a uniform random variable was less than or equal to the stratum-specific prevalence; the stratum-specific prevalence was based on the prevalences calculated from the Colorado data described above (Table 1), with

TABLE 2. Stratum-specific prevalence and sampling probabilities based on 2006 demographic data from mule deer (*Odocoileus hemionus*) in Data Analysis Unit (DAU) D-19 used in the simulations evaluating the properties of the weighted surveillance system for detecting chronic wasting disease in Colorado mule deer populations.

Manipulation of sampling probability	Stratum identification	Prevalence	Sampling probability
No increase	Suspect—female	0.116	0.010
	Suspect—male	0.103	0.010
	Other	0.019	0.109
	Harvest—adult male	0.010	0.392
	Harvest—yearling male	0.003	0.042
	Harvest—adult female	0.006	0.279
	Harvest—yearling female	0.005	0.016
	Harvest—fawn	0.000	0.142
1% increase	Suspect—female	0.116	0.020
	Suspect—male	0.103	0.020
	Other	0.019	0.119
	Harvest—adult male	0.010	0.386
	Harvest—yearling male	0.003	0.036
	Harvest—adult female	0.006	0.273
	Harvest—yearling female	0.005	0.010
	Harvest—fawn	0.000	0.136
5% increase	Suspect—female	0.116	0.060
	Suspect—male	0.103	0.060
	Other	0.019	0.159
	Harvest—adult male	0.010	0.349
	Harvest—yearling male	0.003	0.032
	Harvest—adult female	0.006	0.236
	Harvest—yearling female	0.005	0.006
	Harvest—fawn	0.000	0.099

prevalence=0.01 assumed in the harvested adult male stratum.

To evaluate the properties of the weighted surveillance system, we first used the estimated weights from Table 1 as the "true" weights for our simulations, as truth must be established nominally to examine bias effects. Each sampled individual was weighted according to the calculated weight for its assigned stratum. We sampled until the sum of the weights for sampled individuals equaled the target value of 300 (i.e., the cumulative number of weight points needed to assure detection of at least one case with 95% confidence when prevalence is 0.01 among harvested adult males). We ran 500 repetitions of this simulation. We then calculated both the probability of detecting a CWD-positive individual ($1-\alpha$) in the sample and the mean number of samples needed to meet the target. We estimated the cost differential under the weighted surveillance system based on the mean number of samples required to reach the target and on a cost of $75 USD for testing each individual submitted. We also examined the distribution of the "waiting time," which we interpreted to be the cumulative number of samples required before detecting the first positive case, and we used a normal kernel density estimator to provide a smoothed frequency distribution. We then repeated the procedure and used weights with biases ranging from $\pm 10-50\%$ the true simulation weights in order to investigate the performance of the system when weights were biased.

In addition, we conducted simulations incorporating a mixed bias procedure wherein the bias again ranged from $10-50\%$, but was negative (i.e., underestimated the "true" value) for suspect males, suspect females, and the "other" stratum and was positive for the remaining strata. Within the mixed bias simulations, we also evaluated use of a traditional surveillance system that assumed every stratum was equally weighted (i.e., all the weights were one regardless of sample source).

Finally, to study the effects of increased sampling from the higher-prevalence strata, we examined the effects relative to the sampling probabilities, as described above (subsequently classified as "no increase" in the simulation results), by increasing the sampling probability of the suspect male, suspect female, and other strata by 1% and 5% each, while commensurately decreasing sampling probabilities across the remaining strata by a total of 3% and 15% to ensure, for simulation purposes, that the total sampling probability was 1 (Table 2). Subsequently, we examined the statistics previously described. All analyses were performed in SAS 9.2 (SAS Institute Inc., 2008).

Simulations to investigate sample size requirements

For jurisdictions in which CWD has already been detected, responsible management agencies may want to develop their own weighted surveillance system rather than use values estimated for Colorado mule deer. Therefore, we conducted additional simulations to estimate the number of samples that must enter the surveillance stream in order to provide adequate information to estimate strata weights accurately. We used the strata-specific sampling probabilities, calculated as described above, and set the true simulation weights for each stratum equal to our estimated weights for Colorado (Table 1). We set prevalence in the harvested adult male stratum, our baseline stratum, to 0.03, which seemed to be a reasonable estimate for CWD-endemic areas in other jurisdictions. We then created surveillance datasets of 1,000, 5,000, 11,000, and 15,000 samples, calculated the estimated weights for each stratum using equation (1), and repeated this process for 1,000 replications. To examine the effects of small sample size on error in the estimated weights, we calculated the mean absolute error, mean percent error, and the probability that the direction of the bias would be positive across the 1,000 repetitions for each sample size. We used the mean absolute error, instead of mean bias, to look at the mean magnitude of error due to small sample size because the latter would be obscured by averaging positive and negative bias values. We focused on the probability of the bias being positive because positive bias will lead to fewer samples than necessary being collected for surveillance, resulting in a decreased disease detection probability. Also, we expected strata with small weights would tend to be negatively biased.

RESULTS

Estimated weights using Colorado mule deer data

Estimated weights and their associated standard errors for the eight different strata in our weighted surveillance system are in Table 1. The greatest weights were attributed to female and male CWD suspects (about 12 and 10 points, respectively), and the least weight was associated with harvested fawns (0.03), as would be expected given the prevalence and overall sample sizes of these strata. We used these estimated weights as the "true" weights in subsequent simulations.

Simulations to investigate properties of the weighted surveillance system

In the simplest simulation cases, where all samples came from either CWD suspect females (Fig. 1A) or from harvested yearling males (Fig. 1B), $1-\alpha$ was not biased (i.e., equaled 0.95) when unbiased ("true") weights were used. However, using biased weights in simulations influenced both disease detection probability and the costs of associated surveillance; simulated underestimation of true simulation weights increased the probability of detecting at least one case ($1-\alpha$) and overestimation decreased that probability (Fig. 1A, B). Conversely, relative surveillance cost increased with negative bias in the weights, decreased with positive bias, and was most pronounced in the simulation based on harvested yearling males (Fig. 1B).

More detailed simulations, based on sampling a hypothetical deer population where 1% of the adult males were infected, yielded similar results. When the "true" simulation weights were used, $1-\alpha$ was unbiased, but as bias in the weights increased, the bias in the probability of detecting a case also increased (Fig. 2). However, in simulations where sampling probability was not increased and weights were only biased within 30% of true values, the bias in $1-\alpha$ was relatively small, and the probability of detecting a case remained ≥ 0.9 (Fig. 2). Increasing sampling probabilities in the higher prevalence strata did not affect the overall pattern in the simulated effects of using biased weights (Fig. 2), and $1-\alpha$ remained unbiased when the true simulation weights were used, regardless of the

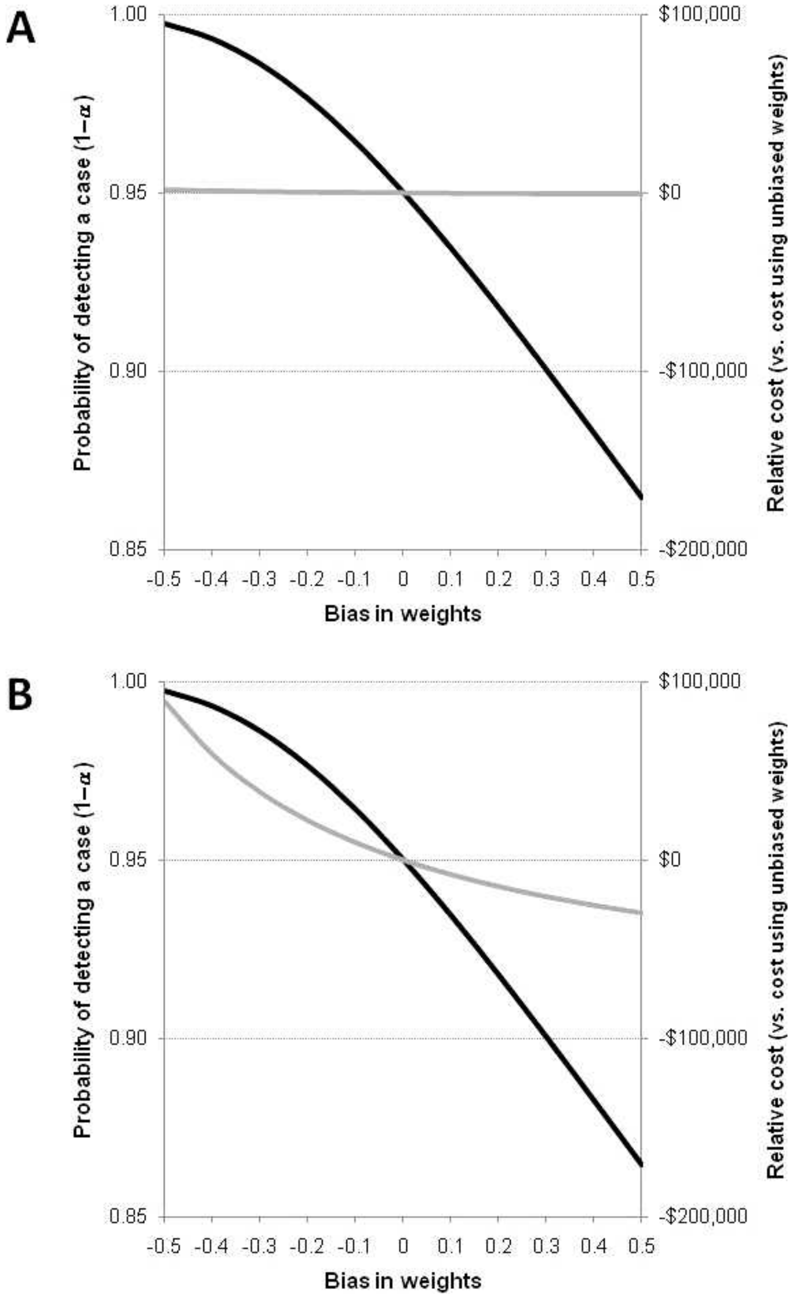

FIGURE 1. Effects of using biased weights on the probability of detecting a positive case ($1-\alpha$; black line) and the associated surveillance cost ("relative cost"; gray line) in a simulated chronic wasting disease (CWD)-infected mule deer population sampled under the proposed weighted surveillance system. For comparison, the influences of bias are illustrated under scenarios where all samples entering the surveillance stream came from either (A) CWD suspect female deer or (B) harvested yearling male deer, representing the highest and second-lowest weighted demographic strata in the proposed system (Table 1). In all simulations, prevalence in the baseline stratum (adult males) was 0.01, and the nominal target detection probability was 0.95.

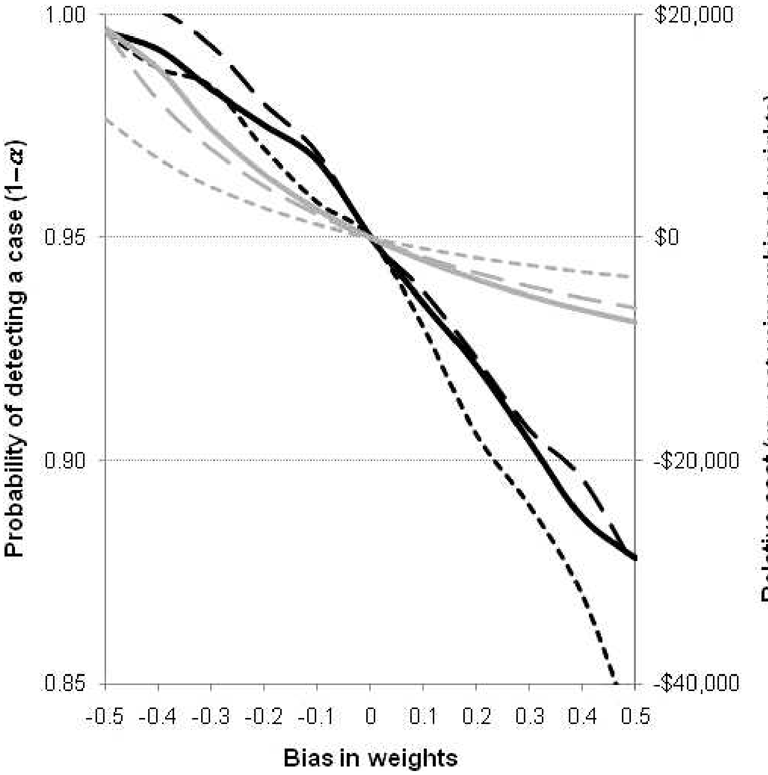

FIGURE 2. Effects of using weights biased in a constant direction across strata (*x* axis) on the probability of detecting a positive case ($1-\alpha$; black lines) and on the associated surveillance cost ("relative cost"; gray lines) in a simulated chronic wasting disease (CWD)-infected mule deer population sampled under the proposed weighted surveillance system. The three line styles represent different levels of emphasis placed on sampling from "high-prevalence" strata (CWD suspect males, CWD suspect females, and "other"): No increase in sampling probabilities of high-prevalence strata (solid lines), sampling probabilities of these strata increased by 1% (dashed lines), or sampling probabilities of these strata increased by 5% (dotted lines); Table 2 lists the stratum-specific sampling probabilities used in simulations under these three scenarios. In all simulations, prevalence in the baseline stratum (adult males) was 0.01, samples entered the surveillance stream from multiple sources, and the nominal target detection probability was 0.95.

sampling probabilities (Fig. 2). When positively biased weights were employed, the resulting bias in $1-\alpha$ tended to be slightly greater in simulations with increased sampling of high-prevalence strata (Fig. 2). The increasing relative surveillance costs associated with negatively biased weights were somewhat offset in simulations where sampling probabilities were increased for high-prevalence strata (Fig. 2).

Comparable patterns occurred in simulations where the direction of the bias was mixed among the strata (Fig. 3). The overall bias in $1-\alpha$ was relatively negligible, in simulations where weights were negatively biased for the suspect and other strata, and were positively biased for the rest of the strata (Fig. 3). The effects on relative costs were also diminished in simulations with mixed bias in the weights (Fig. 3).

The traditional surveillance approach (i.e., all weights being equal) represents an extreme case where the direction of the bias in weights is mixed across strata. Comparing our weighted surveillance system, using unbiased weights, to the traditional system revealed that implementing the weighted system in a large Colorado DAU would cost $315 USD more than the current system. However, as the probability

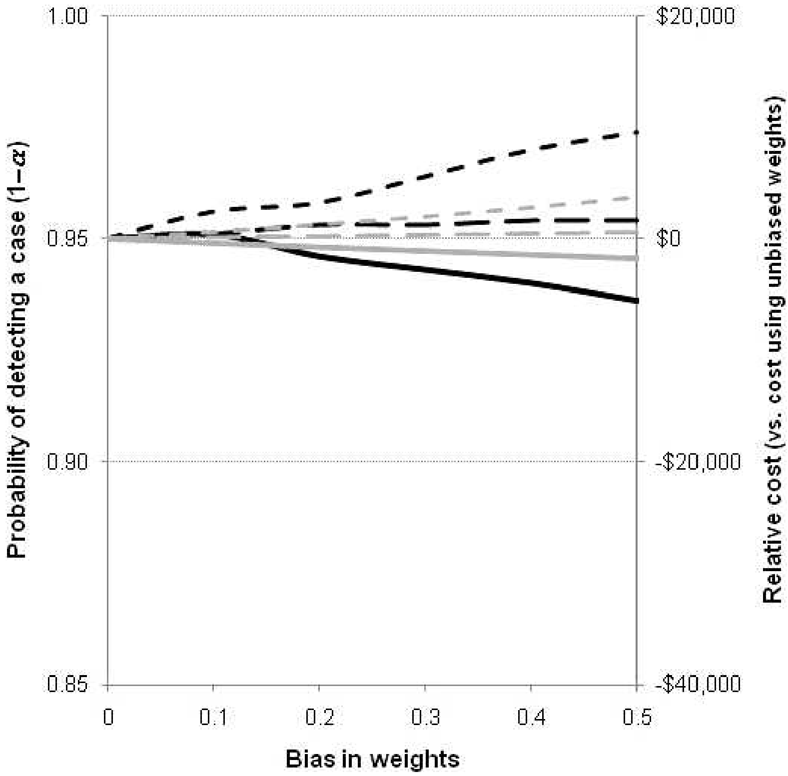

FIGURE 3. Effects of using weights biased in mixed direction across strata (x axis) on the probability of detecting a positive case ($1-\alpha$; black lines) and on the associated surveillance cost ("relative cost"; gray lines) in a simulated chronic wasting disease (CWD)-infected mule deer population sampled under the proposed weighted surveillance system. Weights were biased negatively for the three "high-prevalence" strata (CWD suspect males, CWD suspect females, and "other") and were biased positively for the remaining five demographic strata in the proposed weighted surveillance system. The three line styles represent different levels of emphasis placed on sampling from high-prevalence strata: No increase in sampling probabilities of high-prevalence strata (solid lines), sampling probabilities of these strata increased by 1% (dashed lines), or sampling probabilities of these strata increased by 5% (dotted lines); Table 2 lists the stratum-specific sampling probabilities used in simulations under these three scenarios. In all simulations, prevalence in the baseline stratum (adult males) was 0.01, samples entered the surveillance stream from multiple sources, and the nominal target detection probability was 0.95.

of sampling higher-prevalence strata was minimally increased by 1% and 5%—reflecting a shift in emphasis toward collecting clinical CWD suspect and "other" dead deer—weighted surveillance became considerably more cost-effective, with projected savings of $3,863 and $11,682 USD, respectively. Simulations also demonstrated that fewer samples were required before detecting the first positive case when sampling effort focused on higher-prevalence strata: The distribution of "waiting times" shifted closer to zero, with increased emphasis on sampling higher-prevalence strata; mean values were 79, 69, and 49 for no increase, and a 1% and 5% increase in sampling probabilities, respectively (Fig. 4).

Simulations to investigate sample size requirements

Our simulations revealed that, with a sample size of at least 5,000 CWD test results, it appears that the mean percent error for strata with true simulation weights >1 will be ≤20% (Table 3). Strata with larger mean percent error values were harvested fawns, harvested yearling males, and harvested yearling females; although

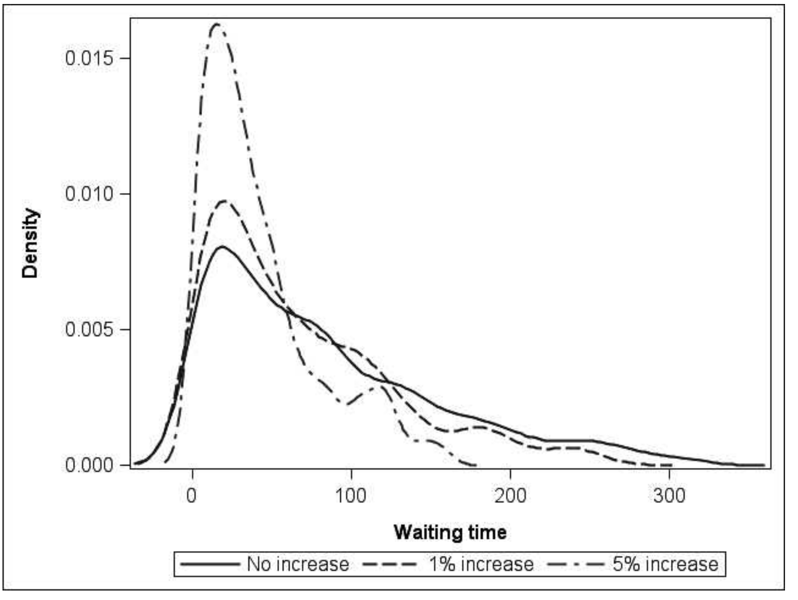

FIGURE 4. The distribution of "waiting time" (expressed as cumulative number of samples) until the first positive deer was detected in a simulated chronic wasting disease (CWD)-infected mule deer population sampled under the proposed weighted surveillance system. The three line styles represent different levels of emphasis placed on sampling from "high-prevalence" strata (CWD suspect males, CWD suspect females, and "other"): No increase in sampling probabilities of high-prevalence strata (solid lines), sampling probabilities of these strata increased by 1% (dashed lines), or sampling probabilities of these strata increased by 5% (dot-dashed lines); Table 2 lists the stratum-specific sampling probabilities used in simulations under these three scenarios. In all simulations, unbiased weights (Table 1) were used, prevalence in the baseline stratum (adult males) was 0.01, samples entered the surveillance stream from multiple sources, and the nominal target detection probability was 0.95.

these strata had much larger mean percent error values (≥0.68), their mean absolute errors were small (i.e., ≤0.33) at a sample size of 5,000 CWD test results, most likely because prevalence in these groups was low. Thus, for strata with small weights, a much-larger number of samples must be acquired before error drops significantly. Simulation results also demonstrated that, across most sample sizes and strata weights, on average results tend to be negatively biased, although the probability of a positively biased weight for most strata and sample sizes is nearly 50% (Table 3). This was expected because equation (1), which was used to estimate the weights, is an asymptotically unbiased estimator.

DISCUSSION

The weighted CWD surveillance system and simulation analyses described here provide statistical justification for the intuitive weighting of individuals from various demographic strata, based on the estimated apparent prevalence and sampling probabilities of those strata as they enter the surveillance stream and are tested. In our system, demographic strata of deer that have a higher prevalence and lower probability of being sampled, relative to the baseline stratum, receive more weight than deer from strata with lower prevalence and from which samples are more common (Table 1). The probability of detecting at least one case of CWD among mule deer from strata where infection is relatively rare is lower than the probability of detecting an infected individual from strata where infection is more likely. Therefore, larger numbers of deer must be sampled from "low-prevalence" strata in order to achieve the high probability of detection typically ascribed

TABLE 3. Stratum-specific mean absolute error, mean percent error, probability of estimating positively biased weights, and true weights based on 2006 demographic data from mule deer (*Odocoileus hemionus*) in Data Analysis Unit (DAU) D-19 used in the simulations evaluating the effects of sample size on estimating the weights for a weighted surveillance system.

Stratum identification	Number of samples	True weight	Mean absolute error	Mean percent error	Probability of positive bias
Suspect—female	1,000	11.57	5.60	0.48	0.51
Suspect—male		10.27	5.32	0.52	0.48
Other		1.90	0.79	0.42	0.48
Harvest—adult female		0.58	0.28	0.48	0.47
Harvest—yearling female		0.45	0.80	1.78	0.21
Harvest—yearling male		0.25	0.38	1.52	0.28
Harvest—fawn		0.03	0.06	1.94	0.13
Suspect—female	5,000	11.57	2.29	0.20	0.50
Suspect—male		10.27	2.10	0.20	0.52
Other		1.90	0.31	0.17	0.50
Harvest—adult female		0.58	0.11	0.19	0.48
Harvest—yearling female		0.45	0.33	0.73	0.43
Harvest—yearling male		0.25	0.17	0.68	0.46
Harvest—fawn		0.03	0.03	1.09	0.49
Suspect—female	11,000	11.57	1.51	0.13	0.49
Suspect—male		10.27	1.43	0.14	0.50
Other		1.90	0.22	0.12	0.48
Harvest—adult female		0.58	0.07	0.13	0.48
Harvest—yearling female		0.45	0.23	0.52	0.45
Harvest—yearling male		0.25	0.11	0.43	0.46
Harvest—fawn		0.03	0.02	0.65	0.40
Suspect—female	15,000	11.57	1.26	0.11	0.53
Suspect—male		10.27	1.19	0.12	0.50
Other		1.90	0.19	0.10	0.52
Harvest—adult female		0.58	0.06	0.11	0.50
Harvest—yearling female		0.45	0.20	0.44	0.44
Harvest—yearling male		0.25	0.09	0.36	0.47
Harvest—fawn		0.03	0.02	0.55	0.50

to such surveillance efforts (Fig. 5). Our weighted approach provides a framework for combining data from various sample sources in order to more transparently evaluate and compare the overall efficacy of surveillance activities.

Our simulations demonstrated the utility of maximizing the collection and submission of mule deer from demographic strata with a higher weight to increase the probability of detecting disease and to minimize the overall economic commitment toward CWD surveillance efforts: On average, surveys based on collecting and examining clinical CWD suspects would only be expected to encounter a case by examining about one tenth the number of submissions needed if only harvested animals were collected and examined. By design, the weighted surveillance system promotes sampling from higher-prevalence (and thus higher-risk) strata because more points are assigned to such strata, thereby motivating users to reach the target value with fewer samples and, thus, reduce overall surveillance effort. The shift in the distribution of cumulative cases examined before detecting the first CWD case (Fig. 4) further emphasizes the benefits of sampling more heavily from higher-prevalence strata. The cost savings associated with this system could be substantial because the number of samples required for testing decreases significantly as sampling is focused less on harvested (and typically healthy) animals and more on collection of “suspect” or

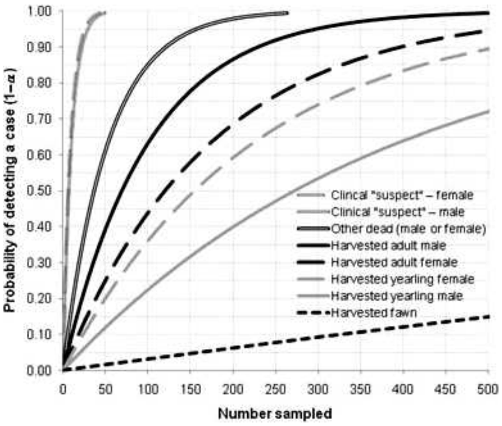

FIGURE 5. The potential contributions of samples from different demographic strata in detecting chronic wasting disease (CWD) are reflected in the relationships between sample size and the probability of detecting at least one CWD case across the eight (apparent health×sex×age) strata developed for mule deer sampled in Colorado, United States. The eight strata are arrayed from upper left to lower right in order of descending estimated weights (Table 1), calculated as described in the text. The solid black line is for harvested adult (≥2 yr old) male mule deer, the baseline stratum to which other strata weights were referenced; assumed prevalence among harvested adult male mule deer was 0.01.

"other" deer that inherently tend to be unhealthy. The mean cost savings shown in our simulation results are for one DAU in Colorado; however, there are 33 of 55 DAUs across the state where CWD has not been detected (Colorado Division of Wildlife, 2009) and where this weighted survey system could be applied; fiscal savings could be even more dramatic in other jurisdictions with greater numbers of cervid populations of unknown status that require sampling. However, it is important to consider that these cost savings estimates are based on the surveillance system established in Colorado and do not incorporate differences across strata for sample collection. This is reasonable, in our system, because field personnel are responsible for collecting samples from these strata as part of their normal duties (i.e., culling suspect deer, collecting vehicle-kills, etc.) and, therefore, there is little added cost for collecting samples from these sources as compared to hunter-submitted samples. Other jurisdictions with limited field personnel, or other restrictions, may have marked differences in sample collection costs and, therefore, cost savings as reported here may vary. On balance, however, it seems likely that using a weighted surveillance approach would provide some cost savings and would also provide a method for managers to exploit all available information on CWD epidemiology when conducting surveillance.

Our simulations emphasized several important considerations related to CWD (and other wildlife disease) surveillance. First, there is a clear relationship in CWD surveillance between the probability of detecting at least one positive case and the cost associated with surveillance: Regardless of sampling scheme, increasing CWD detection probability comes at an increased cost and vice versa; however, exploiting differences in sampling and prevalence rates should lessen the relative costs associated with increasing the likelihood of detecting new foci. If biased weights are inadvertently used in the weighted surveillance system, then costs will increase or decrease depending on the direction of the bias, as will the disease detection probability. Fortunately, it appears that the weighted surveillance system is relatively robust to modest bias in the weights: In our simulations, the probability of detecting at least one case was not less than 0.9 until bias in the weights was 40% or greater.

As with any surveillance system, we recognize that violating sampling design assumptions will likely diminish the ability to detect CWD at the specified $1-\alpha$. This consequence was evident in simulations where bias was introduced into the weights. The assumption that CWD cases were distributed as a Poisson random variable within each stratum seemed reasonable, given the low probability of an individual being infected, but we did not examine the effects of violating this assumption. The assumption that individuals were randomly selected within a stratum for sampling is undoubtedly violated in practice (Otis et al., 1978), but this problem of individual heterogeneity plagues all practiced CWD surveillance approaches of which we are aware. The assumption that relative prevalence within each stratum is constant across different population prevalence levels may also likely be violated due to factors such as transmission probabilities, spatial heterogeneity, density-dependent mechanisms and, perhaps, other factors (Miller et al., 2000; Miller and Conner, 2005), although the difference in prevalence between sexes appears remarkably robust across a wide range of prevalence (Miller et al., 2008). Our system performed reasonably well in simulations examining the use of biased weights that equated to a situation where prevalence for each stratum differed from the true simulation prevalence, suggesting it is robust to minor violations of this assumption. Despite these potential limitations, we believe that the assumptions associated with our, or similar, weighted surveillance approaches are more defensible than the assumption that every individual sample entering the surveillance stream comes from a uniform population and is of equal detection value, an assumption that is central to harvest-based CWD surveillance approaches (Samuel et al., 2003), yet clearly violated (Fig. 5).

One suggested advantage of the harvest-based CWD surveillance approach is that sampling apparently healthy harvested animals is "conservative" compared to using weighted surveillance. In other words, because in the weighted surveillance system 300 individual samples may not be tested (i.e., the target value may be reached before the total number of samples tested equals 300), systems currently in use will have a higher probability of detecting CWD. Examined in the same framework as our weighted surveillance system, however, the contemporary CWD surveillance approach in wide use represents a case of extreme bias in stratum-specific weights most comparable to simulations incorporating mixed bias in the weights—the bias in harvest-based surveys tends to be negative for high-prevalence strata (e.g., unthrifty individuals) and positive for low-prevalence strata. It follows that such an approach may result either in increased probability of disease detection or in decreased probability of disease detection—the direction of the bias in this probability depends on sample composition. For example, if most samples actually come from high-prevalence stratum, then α will be negatively biased and the probability of detection will approach 1.0; alternatively, if the majority of the samples in the surveillance stream come from low-prevalence strata like yearling males (which are abundant in many heavily harvested North American deer populations), then α will be positively biased and the true probability of detecting at least 1 positive case will fall below that believed to be assured, based on the original survey design (Fig. 5). Under either scenario, the contemporary harvest-based approach will be inefficient relative to the weighted surveillance system, because either costs will be higher than necessary or disease detection probability will be below the nominal "advertised" level. Moreover, if the majority of samples come from low-prevalence strata, such as yearling males, then the weighted surveillance system would actually require sampling more individuals than prescribed by the contemporary approach because the weights for this stratum are less than one. The notion that traditional approaches for CWD surveillance are somehow superior to a weighted approach seems largely based on the premise that a greater number of samples will always be "better" when, in fact, our results illustrate that both the number of samples and the source of those samples can influence the probability of disease detection (Fig. 5). By design, weighted surveillance encourages sampling from demographic strata with the highest probability of infection in order to maximize speed of disease detection and minimize cost—we see no clear advantage to using approaches that ignore information available on differences in prevalence and probability of CWD infection across demographic strata within a deer population.

Another concern among prospective users is whether adequate data are available to create a jurisdiction-specific weighted system, as well as how to estimate the required weights. Based on our sample size simulations, it appears that estimating weights from at least 5,000 samples will yield errors $\leq$20% for strata with weights $>$1, which our bias simulations suggest should not markedly affect disease detection. These sample size simulations also demonstrate that, if weights are biased, the biases tend to be negative and thereby provide conservative estimates of the number of samples needed (i.e., $1-\alpha$ will be higher than the level prescribed by the user). Moreover, the strata most affected by small sample size are those with low weights and, thus, these biases will have minimal effects on overall disease detection probability.

The information provided here should be sufficient to allow other jurisdictions to develop weighted surveillance systems as the need arises. However, as a first step in doing so, we recommend constructing a demographic model and conducting sample size simulations, as described herein, based on strata-specific sampling probabilities and prevalence estimates from other regions of interest. If sufficient data are available, but jurisdictions also want to incorporate the data from Colorado included here, we suggest using a Bayesian approach to estimate weights. Such an approach should be relatively straightforward, given the likelihoods described in Appendix A and using weights from Table 1 as prior values; as more local data are acquired over time to estimate the weights, these prior values will be overwhelmed by local data. We encourage other jurisdictions with adequate data to consider estimating weights independently, using the maximum likelihood framework provided herein for comparison to the weights we have reported to help elucidate potential regional or species-specific differences that could be important in further refining CWD surveillance approaches.

For jurisdictions lacking appropriate or adequate data (e.g., a state, province, or nation where CWD has not been detected), an alternative to the traditional "random sampling" approach is to simply use the weights directly from Table 1. This assumes that probabilities of samples from the various strata entering the surveillance stream from other jurisdictions are similar to those we reported and that the effects and epidemiologic patterns of CWD are relatively constant between regions and across species, which seem reasonable, based on published observations (Miller et al., 2000; Miller and Wild, 2004; Miller and Conner, 2005; Williams, 2005; Joly et al., 2006; Grear et al., 2006) and on similarities between the weight values independently estimated for mule deer (Table 1) and for elk (D. Walsh, unpubl. data) using data from Colorado. Given the relatively robust nature of these estimated weights, as demonstrated by our simulations, it seems unlikely that a CWD surveillance approach based on our estimated weights for mule deer in northern Colorado would be any less reliable than alternatives that incorrectly assume that the probability of CWD infection across all demographic strata is equal.

The need for a weighted surveillance system that incorporates all available information regarding stratum-specific prevalence and sampling probabilities has become apparent in Colorado because public interest and economic support for CWD surveillance has begun to wane. As interest and funding have declined, it has become necessary to streamline our surveillance to refocus on sampling individuals with the greatest probability of being infected in order to continue detecting changes in the geographic distribution of CWD in a timely manner. Additionally, the inability to collect the "standard 300 samples" from high-risk management units due to lack of harvest, hunter participation, or other constraints (Colorado Division of Wildlife, 2009) has compelled a shift in emphasis from harvest submissions to those that can be collected by agency personnel. Using this weighted surveillance system can reduce the total number of samples that agency personnel will need to acquire from a population in order to achieve a nominal "disease-free" determination at the same probability of disease detection as afforded under harvest-based surveillance approaches, by exploiting variability in stratum-specific prevalence and sampling probabilities and by allowing all surveillance submissions to count toward reaching local surveillance goals. Moreover, the clearly defined scoring system and target should make this an intuitive system for agency personnel to use and track. Although not described further here, our weighted surveillance system can also potentially use data from multiple cervid species in situations where agencies are attempting to detect CWD in a region rather than in a particular species. Under such an approach, samples from other susceptible host species (elk, white-tailed deer, and moose) could be included in the weighting scheme and thereby contribute to the overall confidence and probability of detecting CWD in a particular geographic area (D. Walsh, unpubl. data).

In the face of shrinking budgets and dwindling public participation, we believe that our weighted surveillance system will be a useful tool for CDOW and, perhaps, for other wildlife management agencies charged with monitoring cervid populations to detect new CWD foci. Weighted surveillance is intended to encourage local wildlife managers to remain actively involved in CWD surveillance by encouraging sample submissions from demographic sources where CWD is more prevalent; the greatest strengths of this approach are the intuitive assignment of different values to samples from different sources and the ability to combine contributions from multiple sample sources toward reaching a quota of survey points. As interest shifts from CWD to other wildlife diseases in Colorado and elsewhere, surveillance systems will need to continue to evolve and become more efficient in order to be sustainable. We believe the weighted surveillance system described here represents a step forward in this surveillance process.

ACKNOWLEDGMENTS

The data analyzed here are the cumulative product of numerous laboratory and field efforts to study CWD in Colorado, efforts that now span over a decade. We extend special thanks to K. Larsen, L. Baeten, I. Levan, K. Griffin, and many, many others in the CDOW for submitting and sampling cervids over the years; and to E. Williams, P. Jaeger, and others at the University of Wyoming and to B. Powers, C. Hibler, T. Spraker, and others at the Colorado State University Diagnostic Laboratory for diagnostic support. We also thank L. Creekmore and A. Scott for their assistance. Numerous hunters facilitated this work by submitting harvested mule deer for CWD testing. We thank B. Banulis for providing the DAU D-19 deer population data used in the simulations. We thank P. Lukacs, D. Heisey, M. Samuel, and two anonymous reviewers for helpful comments on earlier drafts of our manuscript. Our study was funded by the Colorado Division of Wildlife and the Federal Aid in Wildlife Restoration Project W-153-R, with partial support provided from a cooperative agreement with the US Department of Agriculture, Animal and Plant Health Inspection Services, Veterinary Services.

LITERATURE CITED

ARNANDA-ORDAZ, F. J. 1983. An extension of the proportional-hazards model for grouped data. Biometrics 39: 109–117.

BAETEN, L. A., B. E. POWERS, J. E. JEWELL, T. R. SPRAKER, AND M. W. MILLER. 2007. A natural case of chronic wasting disease in a free-ranging moose (*Alces alces shirasi*). Journal of Wildlife Diseases 43: 309–314.

BEDRICK, E. J., R. CHRISTENSEN, AND W. JOHNSON. 1997. Bayesian binomial regression: Predicting survival at a trauma center. American Statistician 51: 211–218.

CANNON, R. M. 2002. Demonstrating disease freedom—Combining confidence levels. Preventive Veterinary Medicine 52: 227–249.

COLORADO DIVISION OF WILDLIFE. 2009. Colorado Division of Wildlife. http://www.wildlife.state.co.us/. Accessed: 2/07/2009.

CONNER, M. M., AND M. W. MILLER. 2004. Movement patterns and spatial epidemiology of a prion disease in mule deer population units. Ecological Applications 14: 1870–1881.

COX, D. R. 1972. Regression models and life-tables. Journal of the Royal Statistical Society, Series B (Methodological) 34: 187–220.

DOHOO, I., W. MARTIN, AND H. STRYHNN. 2003. Veterinary epidemiologic research. AVC Inc., Charlottetown, Canada, 706 pp.

GREAR, D. A., M. D. SAMUEL, J. A. LANGENBERG, AND D. KEANE. 2006. Demographic patterns and harvest vulnerability of chronic wasting disease infected white-tailed deer in Wisconsin. Journal of Wildlife Management 70: 546–553.

HIBLER, C. P., K. L. WILSON, T. R. SPRAKER, M. W. MILLER, R. R. ZINK, L. L. DEBUSE, E. ANDERSEN, D. SCHWEITZER, J. A. KENNEDY, L. A. BAETEN, J. F. SMELTZER, M. D. SALMAN, AND B. E. POWERS. 2003. Field validation and assessment of an enzyme-linked immunosorbent assay for detecting chronic wasting disease in mule deer (*Odocoileus hemionus*), white-tailed deer (*Odocoileus virginianus*), and Rocky Mountain elk (*Cervus elaphus nelsoni*). Journal of Veterinary Diagnostic Investigation 15: 311–319.

JOLY, D. O., M. D. SAMUEL, J. A. LANGENBERG, J. A. BLANCHONG, C. A. BATHA, R. E. ROLLEY, D. P. KEANE, AND C. A. RIBIC. 2006. Spatial epidemiology of chronic wasting disease in Wisconsin white-tailed deer. Journal of Wildlife Diseases 42: 578–588.

KRUMM, C. E., M. M. CONNER, AND M. W. MILLER. 2005. Relative vulnerability of chronic wasting disease infected mule deer to vehicle collisions. Journal of Wildlife Diseases 41: 503–511.

MILLER, M. W., AND M. M. CONNER. 2005. Epidemiology of chronic wasting disease in free-ranging mule deer: Spatial, temporal, and demographic influences on observed prevalence patterns. Journal of Wildlife Diseases 41: 275–290.

———, AND M. A. WILD. 2004. Epidemiology of chronic wasting disease in captive white-tailed and mule deer. Journal of Wildlife Diseases 40: 320–327.

———, E. S. WILLIAMS, C. W. MCCARTY, T. R. SPRAKER, T. J. KREEGER, C. T. LARSEN, AND E. T. THORNE. 2000. Epizootiology of chronic wasting disease in free-ranging cervids in Colorado and Wyoming. Journal of Wildlife Diseases 36: 676–690.

———, H. M. SWANSON, L. L. WOLFE, F. G. QUARTARONE, S. L. HUWER, C. H. SOUTHWICK, AND P. M. LUKACS. 2008. Lions and prions and deer demise. PLoS ONE 3: e4019. doi:10.1371/journal.pone.0004019.

NUSSER, S. M., W. R. CLARK, D. L. OTIS, AND L. HUANG. 2008. Sampling considerations for disease surveillance in wildlife populations. Journal of Wildlife Management 72: 52–60.

OTIS, D. L., K. P. BURNHAM, G. C. WHITE, AND D. R. ANDERSON. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs 62: 3–135.

SAMUEL, M. D., D. O. JOLY, M. A. WILD, S. D. WRIGHT, D. L. OTIS, R. W. WERGE, AND M. W. MILLER. 2003. Surveillance strategies for detecting chronic wasting disease in free-ranging deer and elk: Results of a CWD surveillance workshop, 10–12 December 2002. United States Geological Survey National Wildlife Health Center, Madison, Wisconsin, 41 pp.

SAS INSTITUTE INC. 2008. SAS 9.2. SAS Institute Inc., Cary, North Carolina.

SPRAKER, T. R., M. W. MILLER, E. S. WILLIAMS, D. M. GETZY, W. J. ADRIAN, G. G. SCHOONVELD, R. A. SPOWART, K. I. OROURKE, J. M. MILLER, AND P. A. MERZ. 1997. Spongiform encephalopathy in free-ranging mule deer (*Odocoileus hemionus*), white-tailed deer (*Odocoileus virginianus*) and Rocky Mountain elk (*Cervus elaphus nelsoni*) in northcentral Colorado. Journal of Wildlife Diseases 33: 1–6.

WILESMITH, J., R. MORRIS, M. STEVENSON, R. CANNON, D. PRATTLEY, AND H. BENARD. 2004. Development of a method for evaluation of national surveillance data and optimization of national surveillance strategies for bovine spongiform encephalopathy. European Union TSE Community Reference Laboratory, Weybridge, UK, 34 pp.

WILLIAMS, E. S. 2005. Review article: Chronic wasting disease. Veterinary Pathology 42: 530–549.

———, AND S. YOUNG. 1980. Chronic wasting disease of captive mule deer: A spongiform encephalopathy. Journal of Wildlife Diseases 16: 89–98.

Received for publication 10 April 2008.

APPENDIX A: DERIVATION OF WEIGHTED SURVEILLANCE SYSTEM

The first step in developing the weighted surveillance system is to determine the method for calculating the number of samples needed across the various strata to achieve some user-specified disease detection probability ($1-\alpha$) at a given prevalence (p_{design}). To facilitate this, we will rely on the assumptions previously stated (see Materials and Methods). Based on these assumptions, the joint probability density function can be formulated as follows:

$$P(X_0, \ldots, X_m \mid \lambda_i) = \prod_{i=0}^{m} \exp\left(-\lambda_i\right) \frac{\lambda_i^{X_i}}{X_i!}, \quad (1)$$

where X_i=the total number of positives from the *ith* stratum in the surveillance stream, and the expected number of CWD-positive cases in the *ith* stratum is $\lambda_i = N_i \times \delta_i \times p_i \times f = n_i \times p_i \times f$, where N_i=population size of the ith stratum, δ_i=sampling probability of the ith stratum (i.e., probability an individual enters the surveillance stream), n_i = the realized number of samples in the surveillance stream from *ith* stratum, p_i is the prevalence for the *ith* stratum, p_0 is the prevalence for the baseline stratum (i.e., the stratum with weight (w_0) equal to 1), and f=the sensitivity of the test. Based on equation (1), the probability of failing to detect a positive case during surveillance (α) is:

$$prob\left(\sum_{i=0}^{m} X_i = 0\right) = \exp\left(-\sum_{i=0}^{m} n_i p_i f\right) = \alpha,$$
$$\sum_{i=0}^{m} n_i p_i = \frac{-\ln(\alpha)}{f}. \quad (2)$$

The probability of detecting ≥1 CWD-positive case can be calculated as $1-\alpha$. Using our assumption that relative prevalence within each stratum is constant across different population prevalence levels, we let $p_i = w_i p_{design}$, where w_i = weight for the *ith* stratum and p_{design}=the specified design prevalence for p_0 (i.e., this is the prevalence at which the practitioner wishes to detect at least 1 CWD-positive case with probability $1-\alpha$ in the baseline stratum; commonly 0.95 is used). Assuming f=1, equation (2) can be rewritten as follows:

$$\sum_{i=0}^{m} n_i w_i p_{design} = -\ln(\alpha). \quad (3)$$

Thus, the number of samples needed from each stratum, or from a combination of strata, to achieve a disease detection probability of $1-\alpha$ is:

$$\sum_{i=0}^{m} n_i w_i = \frac{-\ln(\alpha)}{p_{design}} = t, \quad (4)$$

where t is the target value, as described earlier. It is clear from this equation that current surveillance systems (i.e., $w_i \equiv 1$ for all i) are special cases of the weighted surveillance system. This equation provides the basis for determining how many samples from each stratum or combination of strata are required to reach a desired disease detection probability of $(1-\alpha)$.

The next step is to estimate weights (w_i) for the various strata in equation (4). Once again, using the assumption that relative prevalence within each stratum is constant across different population prevalence levels, we let $p_i = w_i p_0$, with p_0 representing the prevalence in the user-specified base-line stratum. Thus, prevalence for each stratum is scaled relative to the baseline stratum. Earlier we had set $p_0 = p_{design}$, but to estimate the weights we will use an estimate of p_0 derived from our surveillance data. Then, given the data vector x, the likelihood function for w_i is as follows:

$$L(w_i \mid x_0, \ldots, x_m, n_0, \ldots, n_m, p_0) = \prod_{i=0}^{m} \exp(-\lambda_i)\frac{\lambda_i^{x_i}}{x_i!} = \prod_{i=0}^{m} \exp(-w_i p_0 n_i)\frac{(w_i p_0 n_i)^{x_i}}{x_i!}. \quad (5)$$

Based on our assumption, if the number of positive cases at the time of the survey in each of *ith* strata is independently distributed as Poisson (λ_i) random variables and the realization of $\lambda_i = n_i \times p_i = n_i \times w_i \times p_o$, then the joint likelihood function can be expressed as:

$$L(w_i, p_0 \mid x_0, \ldots, x_m, n_0, \ldots, n_m) = \exp(-p_0 n_0)\frac{(p_0 n_0)^{x_0}}{x_0!} \times \prod_{i=1}^{m} \exp(-w_i p_0 n_i)\frac{(w_i p_0 n_i)^{x_i}}{x_i!}, \quad (6)$$

from which we generate maximum likelihood estimates for w_i as:

$$\hat{w}_i = \frac{x_i}{n_i \hat{p}_0} = \frac{\hat{p}_i}{\hat{p}_0}, \quad (7)$$

and we also generate $\hat{p}_0 = x_0/n_0$, the maximum likelihood estimate of prevalence for the baseline stratum. It is clear from equation (7) that the weight for the baseline stratum is ≡1.

The weighted surveillance system is then employed by collecting samples from the various strata until:

$$\sum_{i=0}^{m} n_i \hat{w}_i = t, \quad (8)$$

where t has been calculated before the onset of surveillance from equation (4) using a user-specified α and p_{design}.

Using the delta method, the estimate of the variance for the individual weights can be calculated as follows:

$$\hat{var}(\hat{w}_i) = \frac{\hat{p}_i(1-\hat{p}_i)}{n_i(\hat{p}_0)^2} + \frac{(\hat{p}_i)^2 \hat{p}_0(1-\hat{p}_0)}{n_0(\hat{p}_0)^4}. \quad (9)$$

APPENDIX B: BIAS EQUATIONS

If weights are biased, the target value is reached too early, or too late, depending on the direction of the bias. The difference in the total number of samples needed to reach the target value, compared to the actual number needed for each stratum, can be derived. Let q_i=difference in number of samples from the true number of samples (n_i) needed in the *ith* stratum to reach the target (t), when there exists a bias (b_i) in the estimated weights relative to the true weights (w_i). Then, for stratum i, the following is true:

$$(n_i+q_i)\times(w_i+b_i)=n_i w_i=t, \qquad (10)$$

Using simple algebra, the difference in the number of samples collected based on biased weights from the true number of samples needed is:

$$\mathrm{bias}(n_i)=q_i=\frac{-n_i b_i}{(w_i+b_i)}. \qquad (11)$$

In addition to affecting the number of samples needed to reach the target value, biased weights will also increase or decrease the disease detection probability $(1-\alpha)$ beyond the intended level, depending on the direction of the bias of the weights. For simplicity, we present the formula for calculating this change in α due to biased weights for 1 stratum here; the extension to all strata follows logically. Using equation (1) and based on the true weights,

$$bias(\alpha)=\alpha_b-\alpha_{true}=\exp\left(-(n_i+q_i)\times w_i p_{design}\right)-\exp\left(-n_i w_i p_{design}\right), \qquad (12)$$

where $1-\alpha_b$=the actual disease-detection probability based on using biased weights, and $1-\alpha_{true}$=the actual disease-detection probability based on using the true weights. Then it can be shown that:

$$bias(\alpha)=\alpha\left(\exp\left(-w_i\frac{n_i b_i}{(w_i+b_i)}\times p_{design}\right)-1\right). \qquad (13)$$