

SOME PROBABILITIES ASSOCIATED WITH SAMPLING FOR DISEASES IN BIGHORN SHEEP

JOHN D. WEHAUSEN, University of California, White Mountain Research Station, Bishop, CA 93514

Desert Bighorn Counc. Trans. 31: 8-10.

Abstract: I calculated probabilities for obtaining no positive evidence of disease organism in a population given different sample sizes, population sizes, and proportions of the population that test positive. Widespread sampling of populations to determine distribution of certain disease organisms; i.e., presence or absence in each population, is an impractical goal. Large samples are needed to state with sufficient confidence that the prevalence of a disease organism in any population is $\leq 10\%$. Some alternative classifications of disease prevalence in populations are: presence versus undetermined; high versus low, where low includes absence; and results expressed with confidence limits.

Much of the research concerning diseases of bighorn sheep (*Ovis canadensis*) has been concerned with studying potentially fatal disease processes in individual sheep. Only recently have population surveys been initiated to investigate exposure to various disease organisms, especially viruses (Parks and England 1974, Chilelli et al. 1982, DeForge et al. 1982, Turner and Payson 1982, Clark et al. 1985). Such surveys are generally concerned with establishing the distribution of diseases among populations rather than a measure of the percent of any population infected or showing evidence of exposure to a particular disease organism. However, establishing the presence or absence of a disease organism is not just a question of finding or not finding evidence of it because sampling probabilities are involved. The purpose of my paper is to present some probability distributions pertinent to the question of the absence of disease organisms in a population. The results are intended to be useful in the planning and interpretation of disease sampling.

I appreciate comments from D. A. Jessup on this manuscript, and fruitful discussions with M. C. Hansen during derivation of the equation used.

METHODS

Any finding of positive serologic or other evidence of the organism is simple to interpret: the disease organism is present. More difficult to interpret is a negative finding from any sampling. Consequently, the probability distributions generated in this study concern the latter. Probabilities calculated were those for obtaining no positive evidence of a particular disease organism, given (1) a sample size, (2) a total population size, and (3) a proportion of the total population that tests positive. Because recaptured individuals are usually recognizable due to marking, this is a probability problem involving sampling without replacement. The equation derived was:

$$P(O) = [(qN)! \times (N-n)!] / [N! \times (qN-n)!] \quad (1)$$

where $P(O)$ is the probability of obtaining no positive samples for the disease organism, N is the size of the total population, n is the number sampled (randomly) from that population, and q is the proportion of the population that will test negative under the particular laboratory technique used ($q = 1-p$, where p is the proportion that will test positive). For each case, calculations were carried out for increasing sample sizes only until the probability dropped to 0.05 (the 5% level). Higher confidence levels can be calculated using equation (1).

RESULTS

Results for 6 population sizes (20–200) and 7 levels of infection in the population (10–70%; $p = 0.1-0.7$) are presented in Table 1. Because population sizes are seldom accurately known, Table 1 can safely be used by choosing the closest population size. The probabilities calculated for a population of 200 can be used for larger populations because values do not change significantly for larger populations. The use of this Table 1 can best be explained by example.

Suppose that 4 sheep were sampled from a population of about 50, and none showed significant titers to parainfluenza-3 (PI-3). Find 4 along the top row of Table 1, and drop down that column until you reach the section for $N = 50$. Your confidence levels will in each case be $1 -$ the probabilities in Table 1. Thus, with a 35% confidence ($1 - 0.65$) the infection rate in that population would be $\leq 10\%$ ($p = 0.1$), with 60% confidence that it was $\leq 20\%$.



Table 1. Probabilities (P) of obtaining no positive samples for a particular disease organism for different sample sizes, given population size (N) and proportion of the population that would test positive if sampled (p). To interpret samplings with no positive results, find the sample size along the top row, then drop down that column to the section with N closest to your population size. For each row in that section, you can state with confidence 1-P that the infection level in your population is $\leq p$. Where P values are missing, the confidence level is $> 1-P$ for the last P value in that row. Probabilities not listed in this table can be calculated using equation (1).

N	P	SAMPLE SIZE																									
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
200	0.1	0.90	0.81	0.73	0.65	0.59	0.53	0.47	0.42	0.38	0.34	0.30	0.27	0.24	0.22	0.19	0.17	0.15	0.14	0.12	0.11	0.10	0.09	0.08	0.07	0.06	0.05
200	0.2	0.80	0.64	0.51	0.41	0.32	0.26	0.20	0.16	0.13	0.10	0.08	0.06	0.05													
200	0.3	0.70	0.49	0.34	0.24	0.16	0.11	0.08	0.05																		
200	0.4	0.60	0.36	0.21	0.13	0.08	0.04																				
200	0.5	0.50	0.25	0.12	0.06	0.03																					
200	0.6	0.40	0.16	0.06	0.02																						
200	0.7	0.30	0.09	0.03																							
100	0.1	0.90	0.81	0.73	0.65	0.58	0.52	0.47	0.42	0.37	0.33	0.29	0.26	0.23	0.20	0.18	0.16	0.14	0.12	0.11	0.10	0.08	0.07	0.06	0.06	0.05	
100	0.2	0.80	0.64	0.51	0.40	0.32	0.25	0.20	0.16	0.12	0.10	0.07	0.06	0.04													
100	0.3	0.70	0.49	0.34	0.23	0.16	0.11	0.07	0.05																		
100	0.4	0.60	0.36	0.21	0.12	0.07	0.04																				
100	0.5	0.50	0.25	0.12	0.06	0.03																					
100	0.6	0.40	0.16	0.06	0.02																						
100	0.7	0.30	0.09	0.03																							
80	0.1	0.90	0.81	0.73	0.65	0.58	0.52	0.46	0.41	0.37	0.33	0.29	0.26	0.23	0.20	0.17	0.15	0.13	0.12	0.10	0.09	0.08	0.07	0.06	0.06	0.05	
80	0.2	0.80	0.64	0.51	0.40	0.32	0.25	0.20	0.15	0.12	0.09	0.07	0.05														
80	0.3	0.70	0.49	0.34	0.23	0.16	0.11	0.07	0.05																		
80	0.4	0.60	0.36	0.21	0.12	0.07	0.04																				
80	0.5	0.50	0.25	0.12	0.06	0.03																					
80	0.6	0.40	0.16	0.06	0.02																						
80	0.7	0.30	0.09	0.02																							
50	0.1	0.90	0.81	0.73	0.65	0.58	0.51	0.45	0.40	0.35	0.31	0.27	0.24	0.21	0.18	0.15	0.13	0.11	0.10	0.08	0.07	0.06	0.06	0.05			
50	0.2	0.80	0.64	0.50	0.40	0.31	0.24	0.19	0.14	0.11	0.08	0.06	0.05														
50	0.3	0.70	0.49	0.33	0.23	0.15	0.10	0.07	0.04																		
50	0.4	0.60	0.36	0.21	0.12	0.07	0.04																				
50	0.5	0.50	0.24	0.12	0.05																						
50	0.6	0.40	0.16	0.06	0.02																						
50	0.7	0.30	0.09	0.02																							
30	0.1	0.90	0.81	0.72	0.64	0.57	0.50	0.44	0.38	0.33	0.28	0.24	0.20	0.17	0.14	0.11	0.09	0.07	0.05								
30	0.2	0.80	0.63	0.50	0.39	0.30	0.23	0.17	0.13	0.09	0.07	0.05															
30	0.3	0.70	0.48	0.33	0.22	0.14	0.09	0.06	0.03																		
30	0.4	0.60	0.35	0.20	0.11	0.06	0.03																				
30	0.5	0.50	0.24	0.11	0.05																						
30	0.6	0.40	0.15	0.05																							
30	0.7	0.30	0.08	0.02																							
20	0.1	0.90	0.81	0.72	0.63	0.55	0.48	0.41	0.35	0.29	0.24	0.19	0.15	0.11	0.08	0.05											
20	0.2	0.80	0.63	0.49	0.38	0.28	0.21	0.15	0.10	0.07	0.04																
20	0.3	0.70	0.48	0.32	0.21	0.13	0.08	0.04																			
20	0.4	0.60	0.35	0.19	0.10	0.05																					
20	0.5	0.50	0.24	0.11	0.04																						
20	0.6	0.40	0.15	0.05																							
20	0.7	0.30	0.08	0.02																							

DISCUSSION

The capture of free-ranging bighorn sheep for disease sampling is generally expensive and sample sizes rarely exceed 10. For the range of population sizes listed, samples < 10 have little resolution relative to the question of presence or absence of a disease organism (Table 1). A sample size of 10 with no positive results allows only a statement with 90% confidence that infection level is $\leq 20\%$, or > 95% confidence that it is $\leq 30\%$ for the larger population sizes. Dropping the population size to 30 for a sample of 10 only raises the confidence level from 90 to 93% that the infection rate is $\leq 20\%$. Thus, even a small sample of 10 has limited resolution relative to the question of presence or absence of disease organisms in populations.

Sampling populations with the intent of classifying them as to the presence or absence of disease organisms is not a workable goal. Absence of evidence of a disease can be stated confidently only after the entire population has been sampled; but even to arrive at a statement that the prevalence is $\leq 10\%$ at a reasonable confidence level requires a large sample size. This was illustrated in the biglong sheep population in the Marble Mountains, California. No significant titers for PI-3 were found in 21 blood samples obtained during 2 bighorn sheep captures in 1983. However, after another 15 blood samples were obtained the following year, positive evidence of PI-3 was found (Clark et al. 1985).

DeForge et al. (1982:79) interpreted their 1981 finding of substantial serologic evidence of PI-3 in the Santa Rosa Mountains, California, compared with the negative evidence from 10 sheep sampled by Turner and Payson (1982) in 1977, as "indicating recent introduction of the disease organism". However, one can only state with reasonable confidence that the infection level was $\leq 30\%$ when Turner and Payson (1982) sampled the population. While introduction of the virus is a possibility, a more likely interpretation is that the virus became more widespread. It would be a reasonable hypothesis that PI-3 is present in all bighorn sheep populations in North America where domestic sheep or cattle grazing has occurred. This hypothesis cannot be practically tested because of our inability to establish unambiguous absence in any population without sampling every bighorn sheep. In other words, the hypothesis is not falsifiable, and thus falls outside of the realm of empirical science according to Popper's (1968) criterion of demarcation.

Populations whose samples produce ambiguous results relative to absence of a disease organism are also ambiguous relative to its presence. Consequently, it will not be possible to arrive at the proportion of populations sampled that harbor the organism, except for the case in which positive results are obtained from every population sampled. Given this problem, it is necessary to ask what a meaningful goal of such disease samplings might be. One such goal would be determination of the minimum number of populations infected, wherein the presence/absence dichotomy is replaced by presence versus undetermined. Another approach might be a classification of high versus low disease activity with 30% prevalence being the cutoff between the 2 categories. With sample sizes ≥ 10 , it should often be possible to make statements that disease prevalence is \leq or $\geq 30\%$ with high confidence levels. However, with this classification a third category will be necessary to include the many sampling results that will not statistically fit into either category with confidence. In some regards, the problem lies with trying to pigeonhole populations into categories. Perhaps the most meaningful approach would be to calculate confidence limits around any sampling result in terms of percent of population testing positive for the disease organism in question.

LITERATURE CITED

- Chilelli, C., M. Marshall, and J. G. Songer. 1982. Antileptospiral agglutinins in sera of desert bighorn sheep. *Desert Bighorn Council Trans.* 26:15-18.
- Clark, R. K., D. A. Jessup, M. D. Koch, and R. A. Weaver. 1985. Survey of desert bighorn sheep in California for exposure to selected infectious diseases. *J. Am. Vet. Med. Assoc.* 187:1175-1179.
- DeForge, J. R., D. A. Jessup, C. W. Jenner, and J. E. Scott. 1982. Disease investigations into high lamb mortality of desert bighorn in the Santa Rosa Mountains, California. *Desert Bighorn Council Trans.* 26:76-81.
- Parks, J. and J. England. 1974. A serological survey for selected viral infections of Rocky Mountain bighorn sheep. *J. Wildl. Dis.* 10: 107-110.
- Popper, K. R. 1968. *The logic of scientific discovery.* Harper and Row, New York, N.Y. 480pp.
- Turner, J. C. and J. B. Payson. 1982. Prevalence of antibodies of selected infectious disease agents in the peninsular desert bighorn sheep (*Ovis canadensis cremnobates*) of the Santa Rosa Mountains, California. *J. Wildl. Dis.* 18:243-245.

