

# FRED Reports

REJECTION NUMBER TABLES  
FOR DISEASE SCREENING

by  
Kit Rawson

Number 80



**Alaska Department of Fish & Game**  
Division of Fisheries Rehabilitation,  
Enhancement and Development

REJECTION NUMBER TABLES  
FOR DISEASE SCREENING

by  
Kit Rawson

Number 80

Alaska Department of Fish and Game  
Division of Fisheries Rehabilitation,  
Enhancement and Development

Robert D. Burkett, Chief  
Technology and Development Branch

P. O. Box 3-2000  
Juneau, Alaska 99802-2000

November 1987

TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
ABSTRACT.....	1
INTRODUCTION.....	1
MATERIALS AND METHODS.....	3
Computation of the Rejection Number.....	3
Basic Definitions and Assumptions.....	3
Ossiander and Wedemeyer's Model.....	4
Extension to Rejection Numbers Greater Than One....	6
Approximations to the Hypergeometric Distribution..	8
Binomial Distribution.....	8
Normal Distribution.....	9
RESULTS.....	10
The Rejection Number Tables.....	10
Explanation.....	10
DISCUSSION.....	14
ACKNOWLEDGMENTS.....	16
REFERENCES.....	17

LIST OF TABLES

<u>Table</u>	<u>Page</u>
1. Variable names used in the text.....	5
2. Rejection numbers for different population and sample sizes when the risk level is 10%.....	11
3. Rejection numbers for different population and sample sizes when the risk level is 5%.....	12
4. Rejection numbers for different population and sample sizes when the risk level is 1%.....	13

## ABSTRACT

Ossiander and Wedemeyer (1973) presented a method for determining the minimal sample size necessary to detect a given level of pathogen prevalence in a population of fish. One difficulty with their approach occurs when samples larger than their minimums are screened, because their method mandates rejecting a population if even one disease carrier occurs in a sample of any size. Here Ossiander and Wedemeyer's approach is modified; the number of disease carriers necessary to reject the population (the rejection number) is computed for fixed sample sizes. Tables of rejection numbers for different population sizes, sample sizes, and levels of risk are presented as well.

KEY WORDS: disease, sampling, pathology, rejection number.

## INTRODUCTION

The State of Alaska is currently preparing new regulations to ensure that fish are not transported or transplanted within the state without reasonable assurance that they do not harbor high levels of pathogens. One of the regulatory goals is to prevent fish diseases from causing problems in hatchery fish as well as their transmittal to natural populations. These regulations will also ensure that fish transferred to a hatchery will not jeopardize the health of fish stocks in the facility. Thus many populations will have to be screened to ensure that they do not harbor pathogens at a level above an acceptable threshold.

Acceptable threshold levels for pathogens are established based upon the seriousness of a particular disease, the disease history of the receiving and donor populations of fish, and the economic value of susceptible natural populations. For every situation, however, there is a single sampling problem: determination of

the minimal sample size for ensuring that the risk of not detecting a pathogen is acceptably low. Testing for presence of pathogens requires destructive sampling; because the production and transfer of fish involves great expense, it is imperative that sample numbers are minimally sufficient to detect a given prevalence of pathogens with a reasonable degree of certainty.

The sample-size tables used in North America for disease screening (Piper et al. 1982, 292) are based on work by Ossiander and Wedemeyer (1973). Using a straightforward model for the probability distribution associated with sampling fish, these authors developed a table of the minimal sample sizes necessary to detect pathogens for various levels of pathogen prevalence.

The following hypothetical example will illustrate the use of Ossiander and Wedemeyer's tables. Suppose that the highest acceptable level of prevalence of a particular pathogen in a population of juvenile salmon is 10%, there are 100,000 fry in the population, and an acceptable risk of making a wrong decision about whether or not this population meets the criteria for safe stocking is 5%. According to Ossiander and Wedemeyer, if 27 fish are randomly sampled from this population and if *none* of these is a carrier of the pathogen, then the risk that the population contains 10% or more pathogen carriers will be 5% or less; and this population would be acceptable. If one fish were positive, then the population would be rejected.

Ossiander and Wedemeyer provided a useful way of balancing two key concerns: (1) the sample must be large enough to detect disease if it is present at a critical level and (2) small enough to avoid the destruction of an undue number of valuable fish. A difficulty with their approach can arise, however, when the number of fish actually sampled is greater than the minimal sample size given in their tables. This occurs because large

samples are likely to contain a few disease carriers even if the pathogen prevalence is very low. To illustrate this, suppose that the sample in the above example contains 50 fish, and one individual is found to be a carrier of the pathogen. Ossiander and Wedemeyer state that if one out of exactly 27 fish is a carrier the population should be rejected; however, they do not discuss what the decision should be if the sample size is greater than 27. In this example, 2% of the sample has been shown to carry the disease. This amount is much less than the highest acceptable pathogen prevalence of 10%, but there is no certainty that the proportion of disease carriers in the population is less than 10%. The purpose of the present report is to show how Ossiander and Wedemeyer's approach can be extended so that decisions about transferring or culturing fish populations on the basis of pathogen prevalence can be made in cases where the sample size may be larger than the Ossiander and Wedemeyer minimum.

## MATERIALS AND METHODS

### Computation of the Rejection Number

#### Basic Definitions and Assumptions:

Most situations concerning fish disease screening include (1) a population comprising some known number of fish and (2) an unknown level of pathogen prevalence; i.e., the fraction of fish in the population that carries the pathogen. It is assumed that a number of individuals can be chosen from the population at random. Furthermore, it is assumed each individual in the sample can be tested for the presence or absence of the pathogen and the test results are always correct. Finally, it is assumed that it is not necessary to combine individual fish in order to complete these laboratory tests. Given these assumptions, the decision procedure will be as follows: each individual fish in the sample

will be tested for the pathogen in the laboratory; there will be some number of fish found to be carriers; if this number is equal to or greater than a predetermined rejection number, the population will be rejected and the proposed transfer or stocking will not be approved; otherwise, the population will not be rejected.

The computation of rejection numbers depends upon several factors: (1) the number of fish in the population, (2) the number of fish in the sample, (3) the maximum acceptable pathogen prevalence, and (4) the acceptable level of risk. The risk level is defined as the probability that a population with a pathogen prevalence beyond the highest acceptable level is not rejected. Table 1 provides variable names and definitions for these factors and for others.

#### Ossiander and Wedemeyer's Model:

The problem of "how large a sample to take in order to detect a given incidence of . . . carrier fish in the population" was considered by Ossiander and Wedemeyer (1983). They fixed the rejection number at  $J=1$ ; and for fixed  $N$ ,  $p$ , and  $r$ , they asked what  $n$  must equal so that the probability of finding one or more carriers in the sample would be less than or equal to  $r$  (Table 1). According to their model, tables of sample sizes are derived from a straightforward consideration of random sampling without replacement. If a population of fish ( $N$ ) includes pathogen carriers ( $M$ ), the probability that the first fish selected for the sample will not be a carrier is  $(N-M)/N$ ; if that occurs, the probability that the second fish sampled will also not be a carrier is  $(N-M-1)/(N-1)$ . The second formula is valid because once the first fish has been selected, the population size is reduced to  $N-1$  fish; and since the first fish was not a carrier, the number of noncarriers remaining is  $N-M-1$ .

Table 1. Variable names used in the text.

Variable name	Meaning
N	Number of fish in the whole population
M	Number of carriers in the population
n	Number of fish in the sample
p	Highest acceptable pathogen prevalence
r	Risk level
J	Rejection number
C	Number of carriers in the sample
s	Standard deviation of the number of carriers in the sample
z	Critical value (one-sided) from a table of the normal distribution

Finally, the probability that *both* of the first two fish are noncarriers is the product of these two quantities, or  $[(N-M)/N] \times [(N-M-1)/(N-1)]$ .

As additional fish are randomly sampled from the population, this process can be extended. The probability that the  $k^{\text{th}}$  fish is a noncarrier equals  $(N-M-k+1)/(N-k+1)$ , and the probability that *all*  $k$  fish are noncarriers will be

$$[(N-M)/N] \times [(N-M-1)/(N-1)] \times \dots \times [(N-M-k+1)/(N-k+1)].$$

Ossiander and Wedemeyer (1973) computed their sample-size tables by repeatedly incrementing the sample size until the product above became less than or equal to  $r$ . The number of terms in the product is, therefore, the sample size necessary to detect a given prevalence of pathogens at the chosen risk level.

#### Extension to Rejection Numbers Greater Than One:

The concept of a rejection number is an extension of Ossiander and Wedemeyer's approach to disease screening. For a given  $N$ ,  $M$ ,  $p$ ,  $r$ , and  $n$ , we can compute a rejection number ( $J$ ). If the number of carriers in the sample ( $C$ ) is equal to or greater than  $J$ , the population is rejected. The computation of  $J$  is a straightforward extension of the previously described model. In general, sampling without replacement is modelled with the hypergeometric probability distribution. Ossiander and Wedemeyer's model is the hypergeometric distribution:  $J$  is fixed and equal to one (1) and  $n$  is allowed to vary. In contrast, to compute the rejection number for a given sample size,  $n$  is fixed and  $J$  is allowed to vary.

The hypergeometric distribution provides a formula for the probability that  $C$  will equal a given value when  $n$  is randomly drawn without replacement from  $N$  that includes  $M$ ; e.g., the

probability that  $C=0$  (written as  $\text{Pr}[C=0]$ ) is given by the formula:

$$\text{Pr}[C=0] = [(N-M)/N] \times [(N-M-1)/(N-1)] \times \dots \times [N-M-n+1)/(N-n+1)].$$

Formulas for  $\text{Pr}[C=1]$ ,  $\text{Pr}[C=k]$ , etc., require more complex notations and will not be reproduced here. The general formula for the hypergeometric distribution can be found in many textbooks (Johnson and Kotz 1969; Cochran 1977). The main concept here is the ability to compute  $\text{Pr}[C=0]$ ,  $\text{Pr}[C=1]$ ,  $\text{Pr}[C=2]$ , . . . ,  $\text{Pr}[C=\min(n,M)]$ . Probabilities for all possible values of  $C$  can be computed, and  $C$  can range from zero to the smaller of  $n$  and  $M$ .

To find the rejection number ( $J$ ), we must first specify  $N$ ,  $M$ , and  $n$ ; e.g., let  $N=100,000$  and determine  $M$  from  $p$ . If  $p=10\%$ , then  $M$  will equal  $10\% \times 100,000 = 10,000$ . Finally, if  $n=50$  then the given values of  $N$ ,  $M$ , and  $n$  can be used to compute hypergeometric probabilities with the aid of a computer. The first few are as follows:

$$\text{Pr}[C=0] = .005 \quad \text{Pr}[C=1] = .029 \quad \text{Pr}[C=2] = .078$$

These numbers can now be applied to disease screening. If an acceptable risk level is  $r=5\%$  (.05), a population will be acceptable for transfer or stocking only if the probability of finding  $C$  in the sample is less than 0.05. For this specific example, the probability of finding  $C=0$  in the sample is .005, much less than the acceptable risk level; therefore, the population would be accepted.

Now suppose that exactly one carrier was found in the sample; the probability of this outcome is .029. In considering the risk of making a wrong decision, however, it is the probability of finding

C or fewer carriers that is meaningful; i.e.,  $\Pr[C=0] + \Pr[C=1] = .005 + .029 = .034$ . This cumulative probability is still less than 5%; so if one carrier was found in a sample of 50 fish, the population would be accepted under the established criteria. This answers the question posed in the hypothetical example provided in the introduction of this report.

Finally, suppose that two carriers were found in the sample. The cumulative probability that C is less than or equal to two is 0.112, as can be seen from summing the above probabilities. This number is greater than the 5% risk level; therefore, the population would be rejected for transfer or stocking. Thus the rejection number is  $J=2$ ; i.e., the population will be rejected if the sample has two or more carriers. A similar procedure could be followed to compute rejection numbers for any given values of N, p (from which we can compute M), n, and r.

#### Approximations to the Hypergeometric Distribution:

Binomial Distribution. Even with the aid of a computer program, the computation of rejection numbers using the hypergeometric distribution can be time-consuming. As the population size increases, the hypergeometric distribution will be closely approximated by the binomial distribution. As the sample size increases, either the Poisson or the normal distribution will provide a good approximation to the binomial. These approximations are discussed in detail by Johnson and Kotz (1969); i.e., guidelines for applying these well known approximations to computing rejection numbers.

First, consider a population of 1,000 fish, highest acceptable pathogen prevalence of 20%, and a risk level of 1%; the sample size is 200. With the given parameters, the rejection number

computes to 29. If the population size is increased to 5,000 and all other parameters remain the same, the rejection number decreases by 1 to 28. At a population of 10,000, the rejection number becomes 27; and it will never be smaller than this, no matter how large the population becomes.

This value  $r(27)$  can also be computed using the binomial distribution instead of the hypergeometric distribution. The binomial distribution has sample size and pathogen prevalence as parameters, but the population size is not included in the probability formulas. Mathematically as the population size becomes larger and larger, the binomial distribution gives the same rejection number as the hypergeometric distribution. Of course, the population can never really contain an infinite number of fish, so this really means that the binomial distribution will give the right answer whenever the population is bigger than some number. For the hypothetical example used here, the binomial distribution will give the right answer if the population is greater than 10,000. Even for a population as small as 1,000 individuals, the binomial approximation (27) is close to the rejection number computed using the hypergeometric distribution (29).

Johnson and Kotz (1969) suggest that the binomial distribution is an adequate approximation to the hypergeometric whenever  $n < N/10$ . For the ranges of parameters commonly encountered in disease-screening work, following this rule will give rejection numbers in error by, at most, one fish. Since it is often easier to compute binomial probabilities than hypergeometric ones, this rule should save time in computing tables of rejection numbers.

Normal Distribution. Where the binomial distribution provides an appropriate approximation to the hypergeometric (i.e., when  $n < N/10$ ) and where  $n > 100$  and  $p > 0.1$ ), rejection numbers may be even more rapidly computed using the normal distribution. The expected number of carriers in the sample must first be computed as  $np$  and the variance of the number of carriers as

$np(1-p)$ . The square root of the variance is the standard deviation of the number of carriers (s). The rejection number is computed as  $np - zs$ , where z is the appropriate value from a table of the cumulative normal distribution for the chosen risk level. In tables giving values for one- and two-sided tests, the values should be taken from the column for one-sided tests (e.g., 10% risk,  $z=1.282$ ; 5% risk,  $z=1.645$ ; 1% risk,  $z=2.326$ ).

As an example of applying the normal approximation, assume  $p=0.20$ ,  $n=100$ , and  $r=0.05$ . Then  $s =$  the square root of  $(100)(0.2)(0.8) = 4$ , and  $J = (100)(0.2) - (1.645)(4) = 13.4$ . Since J must be a whole number, we take the next whole number bigger than this, so  $J=14$ . This is the same value computed using the hypergeometric distribution (for  $N=10,000$ ), and it is much easier to compute.

## RESULTS

### The Rejection Number Tables

#### Explanation:

Tables 2, 3, and 4 contain rejection numbers for different values of risk, population size, carrier rate, and sample size. These tables were computed using a program written in Turbo Pascal<sup>®</sup> for the IBM/PC<sup>®</sup> microcomputer. The hypergeometric distribution was used for all finite population sizes, and the binomial distribution was used for the infinite population.

---

<sup>®</sup> Mention of commercial products or trade names does not constitute endorsement by ADF&G, FRED Division.

Table 2. Rejection numbers for different population and sample sizes when the risk level is 10%.

Population Size = 1,000

Pathogen prevalence	Sample Size						
	30	60	100	120	200	300	500
0.20	3	8	15	19	34	53	92
0.10	1	3	6	8	15	24	44
0.05		1	2	3	7	11	21
0.01						1	3

Population Size = 5,000

Pathogen prevalence	Sample Size						
	30	60	100	120	200	300	500
0.20	3	8	15	19	33	51	89
0.10	1	3	6	8	15	24	42
0.05		1	2	3	6	10	19
0.01						1	2

Population Size = 10,000

Pathogen prevalence	Sample Size						
	30	60	100	120	200	300	500
0.20	3	8	15	18	33	51	89
0.10	1	3	6	8	15	24	42
0.05		1	2	3	6	10	19
0.01						1	2

Population Size = infinite

Pathogen prevalence	Sample Size						
	30	60	100	120	200	300	500
0.20	3	8	15	18	33	51	89
0.10	1	3	6	8	15	23	41
0.05		1	2	3	6	10	19
0.01						1	2

Table 3. Rejection numbers for different population and sample sizes when the risk level is 5%.

Population Size = 1,000

Pathogen prevalence	Sample Size						
	30	60	100	120	200	300	500
0.20	3	7	14	17	32	51	90
0.10	1	3	6	7	14	23	42
0.05		1	2	3	6	10	19
0.01						1	2

Population Size = 5,000

Pathogen prevalence	Sample Size						
	30	60	100	120	200	300	500
0.20	3	7	14	17	31	49	86
0.10	1	2	5	7	13	22	40
0.05		1	2	2	5	9	18
0.01						1	2

Population Size = 10,000

Pathogen prevalence	Sample Size						
	30	60	100	120	200	300	500
0.20	3	7	14	17	31	49	86
0.10	1	2	5	7	13	22	39
0.05		1	2	2	5	9	17
0.01						1	2

Population Size = infinite

Pathogen prevalence	Sample Size						
	30	60	100	120	200	300	500
0.20	3	7	14	17	31	49	85
0.10	1	2	5	7	13	22	39
0.05		1	2	2	5	9	17
0.01						1	2

Table 4. Rejection numbers for different population and sample sizes when the risk level is 1%.

Population Size = 1,000

Pathogen prevalence	Sample Size						
	30	60	100	120	200	300	500
0.20	2	5	12	15	29	47	85
0.10		1	4	5	12	20	39
0.05			1	1	4	8	17
0.01							1

Population Size = 5,000

Pathogen prevalence	Sample Size						
	30	60	100	120	200	300	500
0.20	1	5	11	14	28	45	81
0.10		1	4	5	11	19	36
0.05			1	1	4	7	15
0.01							1

Population Size = 10,000

Pathogen prevalence	Sample Size						
	30	60	100	120	200	300	500
0.20	1	5	11	14	27	45	80
0.10		1	4	5	11	19	35
0.05			1	1	4	7	15
0.01							1

Population Size = infinite

Pathogen prevalence	Sample Size						
	30	60	100	120	200	300	500
0.20	1	5	11	14	27	44	80
0.10		1	4	5	11	19	35
0.05			1	1	4	7	14
0.01							1

The application of these tables and subtables will depend on the acceptable risk level and the size of the population under consideration, respectively. In the subtable, find the maximally acceptable carrier rate in the left-hand column and the sample size in the top row. The number at the intersection of the sample-size column and the carrier-rate row is the rejection number. If the number of carriers in the sample is greater than or equal to the rejection number, the proposed transfer or introduction will not be allowed.

Notice that there is very little change in the tables between a population size of 1,000 and an infinite population. This indicates that the binomial approximation to the hypergeometric distribution is appropriate for most of the situations encountered in fish disease screening work.

#### DISCUSSION

This report provides an extension of the tables provided by Ossiander and Wedemeyer (1973) for screening fish populations for disease. The present approach is appropriate when it is important to ensure that pathogen prevalence in a population is below some predetermined level. While Ossiander and Wedemeyer's approach mandates that populations be rejected for transfer or stocking if any pathogen carriers are found in a sample, the approach presented here uses a rejection number. A population is rejected only if the sample contains a number of disease carriers equal to or greater than the rejection number, which may be greater than one.

The approach to disease screening presented here relies upon several assumptions: (1) fish can be randomly sampled, (2) the laboratory test for pathogen presence is 100% accurate, and (3) sample fish do not have to be pooled in the laboratory. If the

assumptions cannot be met, these methods should be modified. Worlund and Taylor (1983) discuss estimating the rate of disease incidence when the last of these assumptions is relaxed.

The rejection number can be computed for any given values of population size, sample size, pathogen prevalence, and risk level. The exact computation using the hypergeometric distribution is time-consuming; however, the normal approximation discussed in this report is applicable to many situations encountered in fish-disease screening (see guidelines above). With this approximation, rejection numbers can be computed using only a hand calculator.

In summary, this report presents a way of determining whether the level of pathogen prevalence in a population of fish is at or below some threshold level. This approach is appropriate in situations where a small level of pathogen prevalence is acceptable, as long as there is assurance that this level is below a previously determined threshold. If it is not acceptable to transfer or culture a population of fish with any evidence of pathogen carriers, Ossiander and Wedemeyer's (1973) approach to disease screening should be followed.

## ACKNOWLEDGMENTS

R. Saft, T. Meyers, and H. Geiger provided technical reviews; remaining errors are my own. I have incorporated many suggestions of W. Hauser and T. Meyers for improving the presentation of this material. S. Cartwright provided expert typing assistance. K. Leon and S. Morgan edited the final draft. The sampling problems addressed here were brought out in discussions of the Pathology Practices Committee whose members included P. Krasnowski, J. Cochran, T. Meyers, and R. Saft of ADF&G, FRED Division; D. Amend of the Southern Southeast Regional Aquaculture Association; and B. Bachen of the Northern Southeast Regional Aquaculture Association.

## REFERENCES

- Cochran, W. G. 1977. Sampling techniques, 3rd ed. John Wiley & Sons, New York.
- Johnson, N. L. and S. Kotz. 1969. Discrete distributions. John Wiley & Sons, New York.
- Ossiander, F. J. and G. Wedemeyer. 1973. Computer program for sample sizes required to determine disease incidence in fish populations. J. Fish. Res. Bd. Can. 30:1383-1384.
- Piper, R. G., I. B. McElwain, L. E. Orme, J. P. McCaren, L. G. Fowler, and J. R. Leonard. 1982. Fish hatchery management. U.S. Dept. of the Interior, Fish and Wildlife Service, Washington, DC.
- Worlund, D.D. and G. Taylor. 1983. Estimation of disease incidence in fish populations. Can. J. Fish. Aquat. Sci. 40:2194-2197.

The Alaska Department of Fish and Game administers all programs and activities free from discrimination based on race, color, national origin, age, sex, religion, marital status, pregnancy, parenthood, or disability. The department administers all programs and activities in compliance with Title VI of the Civil Rights Act of 1964, Section 504 of the Rehabilitation Act of 1973, Title II of the Americans with Disabilities Act of 1990, the Age Discrimination Act of 1975, and Title IX of the Education Amendments of 1972.

If you believe you have been discriminated against in any program, activity, or facility, or if you desire further information please write to ADF&G, P.O. Box 25526, Juneau, AK 99802-5526; U.S. Fish and Wildlife Service, 4040 N. Fairfax Drive, Suite 300 Webb, Arlington, VA 22203 or O.E.O., U.S. Department of the Interior, Washington DC 20240.

For information on alternative formats for this and other department publications, please contact the department ADA Coordinator at (voice) 907-465-6077, (TDD) 907-465-3646, or (FAX) 907-465-6078.