J. Hyg., Camb. (1975), **74**, 149
Printed in Great Britain

Statistical inferences about injury and persistence of environmentally stressed bacteria

BY MARTIN A. HAMILTON
Department of Mathematics
AND GARY K. BISSONNETTE
Department of Microbiology, Montana State University, Bozeman, Montana, U.S.A.

(*Received* 16 *July* 1974)

SUMMARY

A standard technique for ascertaining the survival characteristics of bacteria after being environmentally stressed is to incubate the bacteria on both selective and non-selective media and count the colonies produced. Based on these colony counts, indexes of injury and persistence of the bacteria are calculated. To compare the stress of two different environments, a persistence ratio is calculated. In this paper, methods of statistical inference concerning these indexes and ratios are presented. These statistical methods use well-known procedures for analysis of binomial data and 2×2 table data, and are appropriate when the colony counts follow a Poisson distribution.

1. INTRODUCTION

The survival characteristics of bacteria in aquatic environments are of utmost importance in the determination of bacteriological water quality and its implications concerning public health significance. The detection and enumeration of sanitary indicator organisms or pathogenic organisms from water is dependent upon the efficiency of the specific recovery methods employed.

When detecting particular microorganisms from specific sources, proper consideration must be given to the influence of environmental factors upon detection methods. There have been a number of studies on microbial survival and recovery concerned with the viability of populations after they have been exposed to such stresses as heating (Clark & Ordal, 1969; Ordal, 1970) or freezing (Ray & Speck, 1972, 1973*b*). Attempts to directly enumerate such stressed populations with conventional selective media often indicate that a substantial portion of cells is injured or debilitated to such an extent that they cannot produce detectable colonies on the selective media. It should be emphasized that this environmental stress may merely injure the cell but not necessarily kill or destroy it. Apparently, environmentally injured cells become sensitive to inhibitory agents in selective media in such a way that they are incapable of growth and production of detectable colonies. Only that portion of the total population that is non-injured will have the capability to produce detectable colonies on the selective media.

Table 1. *Data provided by experiment*

Environment	Medium	Replications	Colony counts	Total colonies	Dilution factor
A	Non-selective	n_{AN}	$X_{AN1}, \ldots, X_{ANn_{AN}}$	X_{AN}	D_{AN}
	Selective	n_{AS}	$X_{AS1}, \ldots, X_{ASn_{AS}}$	X_{AS}	D_{AS}
B	Non-selective	n_{BN}	$X_{BN1}, \ldots, X_{BNn_{BN}}$	X_{BN}	D_{BN}
	Selective	n_{BS}	$X_{BS1}, \ldots, X_{BSn_{BS}}$	X_{BS}	D_{BS}

In the study which led to the statistical analysis described below (Bissonnette, 1974), bacteria were exposed to the stresses of two natural aquatic environments of differing chemical and physical characteristics. After exposure to the stresses of these aquatic environments, the surviving organisms were enumerated on both non-selective and selective media. Those organisms producing detectable colonies on the non-selective medium were considered to be the total viable population present, i.e. both injured and non-injured (persistent). The difference between simultaneous counts on the non-selective and selective media was considered to be that portion of the total population which was injured.

2. NOTATION AND DEFINITIONS

The observation recorded is 'number of colonies counted' which is denoted by X. Let X_{ijk} be the colony count on the kth plate observed for environment i using medium j, where $i = A, B$; $j = N, S$ for non-selective and selective media, respectively; $k = 1, \ldots, n_{ij}$. Here n_{ij} is the number of plates incubated for environment i and medium j. It is assumed that the population of bacteria is diluted by a fixed factor before pouring on the n_{ij} plates. The dilution factor D_{ij} is allowed to vary among environment and medium combinations. Let $X_{ij} = \sum_{k=1}^{n_{ij}} X_{ijk}$ = the total number of colonies for the ith environment and jth medium.

The data provided by the experiment are summarized in Table 1.

Let μ_{ij} be the true mean colony count per plate for the population in environment i and medium j when the dilution factor is D_{ij}. The usual measure of *injury* to the bacterial population after being stressed in environment i is (Ray & Speck, 1973*b*):

$$\text{Proportion injury} \equiv I_i \equiv [(\mu_{iN}/D_{iN}) - (\mu_{iS}/D_{iS})]/(\mu_{iN}/D_{iN}).$$

Define *persistence* in environment i as the proportion of non-injury; that is, $P_i = 1 - I_i \equiv (\mu_{iS}/D_{iS})/(\mu_{iN}/D_{iN})$.

The stress on the bacteria by environment B relative to environment A is measured by the *persistence ratio*

$$\psi \equiv P_A/P_B.$$

If ψ is equal to 1, the environments place equal stress on the surviving bacteria; if ψ is less than 1, environment A produces the greater stress; and, if ψ is greater than 1, environment B produces the greater stress.

https://doi.org/10.1017/S0022172400024219 Published online by Cambridge University Press

The statistical problem is to find good procedures for inference about P_A, P_B, and ψ.

The purpose of this paper is to describe some hypothesis-testing procedures and the associated confidence interval estimation procedures for I_i, P_i, and ψ when the colony counts follow a Poisson distribution. An example is provided.

3. INFERENCE ABOUT P_i, I_i, ψ

In the following, it is assumed that $X_{ij1}, \ldots, X_{ijn_{ij}}$ is a random sample from a Poisson distribution having mean μ_{ij}. Information concerning the μ_{ij}'s is consolidated in $\{X_{ij}, i = A, B, j = N, S\}$, the four totals of colonies counted, which are independent Poisson variables. The mean of X_{ij} is $\theta_{ij} = \mu_{ij} n_{ij}$.

3.1. *Inference about P_i and I_i*

Since $\mu_{ij} = \theta_{ij}/n_{ij}$, $P_i = (\theta_{iS}/n_{iS} D_{iS})/(\theta_{iN}/n_{iN} D_{iN})$. The null hypothesis that P_i is equal to a specific constant P_{i0} is equivalent to

$$H_0\colon\ \theta_{iS}/(\theta_{iN}+\theta_{iS}) = P_{i0} n_{iS} D_{iS}/(P_{i0} n_{iS} D_{iS} + n_{iN} D_{iN}).$$

Reformulating the hypothesis in this way allows for a simple analysis because it is well known (Lehmann, 1959) that the test of H_0 may be based on the conditional distribution of X_{iS} given $X_{iN}+X_{iS}$ which is a binomial distribution with parameters $n = X_{iN}+X_{iS}$ and $p = \theta_{iS}/(\theta_{iN}+\theta_{iS})$. Thus the test of H_0 is essentially the usual test on the binomial parameter p and is based on the observed proportion $X_{iS}/(X_{iN}+X_{iS})$.

A confidence interval estimate for p is easily converted into a confidence interval for θ_{iS}/θ_{iN} and then to a confidence interval for P_i. Specifically, if the confidence interval for the binomial parameter p is $p_l < p < p_u$, then the associated confidence interval for P_i is

$$[p_l D_{iN} n_{iN}/(1-p_l) D_{iS} n_{iS} < P_i < p_u D_{iN} n_{iN}/(1-p_u) n_{iS} D_{iS}].$$

Exact confidence intervals for p have been tabled (Diem & Lentner, 1970). If $X_{iN}+X_{iS}$ is large, the normal approximation to the binomial distribution may be used. The formula for an approximate 95% interval is

$$p_l, p_u = X_{iS}/(X_{iN}+X_{iS}) \pm 1\cdot 96[X_{iN} X_{iS}/(X_{iN}+X_{iS})^3]^{\frac{1}{2}}.$$

A point estimate of P_i is $X_{iS} n_{iN} D_{iN}/X_{iN} n_{iS} D_{iS}$. Notice that P_i can equal infinity since $X_{iN} = 0$ with positive probability.

Using the identity, $I_i = 1-P_i$, one can easily adjust these formulas to estimate I_i. The confidence interval estimate for I_i is

$$1-[p_u D_{iN} n_{iN}/(1-p_u) D_{iS} n_{iS}] < I_i < 1-[p_l D_{iN} n_{iN}/(1-p_l) D_{iS} n_{iS}].$$

The point estimate of I_i is

$$(X_{iN} n_{iS} D_{iS} - X_{iS} n_{iN} D_{iN})/X_{iN} n_{iS} D_{iS}.$$

3.2. *Inference about ψ*

Let ψ_θ, ψ_n, ψ_D be cross-product ratios associated with the 2×2 tables having elements $\{\theta_{ij}\}$, $\{n_{ij}\}$, and $\{D_{ij}\}$ in the format:

		Medium: S	Medium: N
Environment	A		
	B		

For example, $\psi_\theta = \theta_{AS}\theta_{BN}/\theta_{BS}\theta_{AN}$.

Since $\psi = \psi_\theta/(\psi_n\psi_D)$, H_0: $\psi = \psi_0$ is equivalent to H_0: $\psi_\theta = \psi_n\psi_D\psi_0$. It can be shown by a simple extension of the argument in Lehmann (1959, p. 141) that the test of H_0 may be based on the conditional distribution of X_{AS} given $(X_{AS}+X_{AN})$, $(X_{AS}+X_{BS})$, and $(X_{BS}+X_{BN})$, which is a non-central hypergeometric distribution. The computational procedure is to place the X_{ij}'s in the 2×2 format and use standard techniques for inference about the cross-product ratio. If $\psi_n\psi_D\psi_0 = 1$, the appropriate test is Fisher's exact test. This procedure was described by Cox & Lewis (1968) and Gart (1974*b*) and was used by Gart (1974*a*) in the analysis of skin cancer incidence data.

Exact confidence intervals based on this 2×2 table approach can be found for ψ_θ by using the computer programme of Thomas (1970). See the paper by Gart & Thomas (1972) for a survey of procedures suggested for calculating confidence limits when all four X_{ij}'s are large. The confidence interval for ψ_θ is converted into a confidence interval for ψ by dividing the endpoints by $\psi_n\psi_D$.

Gart (1971) discusses various methods of point estimation of ψ_θ. The usual estimate is

$$\hat{\psi}_\theta = \frac{X_{AS}X_{BN}}{X_{AN}X_{BS}},$$

which is the cross-product ratio of the total counts when placed in the 2×2 table format. The estimate of ψ based on $\hat{\psi}_\theta$ is $\hat{\psi} = \hat{P}_A/\hat{P}_B = \hat{\psi}_\theta/\psi_n\psi_D$.

4. EXAMPLE

4.1. *Description of the specific experiment*

A suspension of an indicator organism (*Escherichia coli*) at approximately 10^4 organisms/ml was prepared and inoculated into 2 environmental test chambers. The environmental test chambers were identical with those described by McFeters & Stuart (1972). The inoculated environmental test chambers were then immersed in two different stream environments, designated as environments A and B. After 3 days of exposure, 1 ml. samples were withdrawn from the chambers for quantitative analyses.

The samples from each chamber were enumerated simultaneously on a non-selective medium and a selective medium. Surface-overlay plating procedures were

Table 2. *Data collected*

Environment	Medium	Replications	Colony counts	Total colonies	Dilution factor
A	Non-selective	2	57, 42	99	10^{-2}
	Selective	2	50, 55	105	10^{-1}
B	Non-selective	2	14, 9	23	10^{-1}
	Selective	2	1, 2	3	10^{0}

Table 3. *Estimates*

	Environment A			Environment B			
	p_A	P_A	I_A	p_B	P_B	I_B	ψ
Point estimates							
	0·52	0·106	0·89	0·12	0·013	0·99	8·13
Confidence interval estimates							
Confidence							
90%	0·45	0·08	0·86	0·03	0·00	0·96	2·71
	0·57	0·14	0·92	0·27	0·04	1·00	32·6
95%	0·44	0·08	0·86	0·02	0·00	0·96	2·33
	0·59	0·14	0·92	0·30	0·04	1·00	43·3
99%	0·42	0·07	0·84	0·01	0·00	0·94	1·75
	0·61	0·15	0·93	0·36	0·06	1·00	79·9

employed as described by Ray & Speck (1973*a*). The non-selective medium used was Trypticase soy agar supplemented with 0·5% glucose and 0·3% yeast extract. Theoretically, the total viable population (both injured and persistent cells) would produce detectable colonies on this nutritionally rich growth medium. The selective medium employed was deoxycholate lactose agar. Only those cells which were persistent could produce colonies on this selective medium.

In order to obtain few enough colonies per plate to permit accurate counting, it was necessary to dilute the samples. A dilution factor of 10^{-1} represented a 10-fold dilution while a dilution factor of 10^{-2} represented a 100-fold dilution. All platings with both types of media were done in duplicate. Plates were incubated for 24 hr. at 35° C. and enumerated.

4.2. *Data and results*

The data collected are given in Table 2 and the point and confidence interval estimates are shown in Table 3.

4.3. *Calculations*

Consider environment A; $p_A = 105/204$ = proportion of all A colonies which grew in the specific medium. The confidence interval for p_A was calculated using the normal approximation to the binomial distribution. The endpoints of the interval for p_A were converted into endpoints of intervals for P_A and I_A using the formulas of Section 3.1. The calculations for environment B are identical except that exact confidence intervals for p_B were found since $X_{BN}+X_{BS}$ was small.

The observed data were placed in the 2×2 tables:

Environment	Medium S	Medium N
A	105	99
B	3	23

Total colonies

Environment	Medium S	Medium N
A	2	2
B	2	2

Replications

Environment	Medium S	Medium N
A	10^{-1}	10^{-2}
B	10^{0}	10^{-1}

Dilution factors

The corresponding cross-product ratios are

$$\hat{\psi}_\theta = \frac{(105)\,(23)}{(3)\,(99)} = 8{\cdot}13; \quad \psi_n = \frac{(2)\,(2)}{(2)\,(2)} = 1; \quad \psi_D = \frac{(10^{-1})\,(10^{-1})}{(10^{0})\,(10^{-2})} = 1.$$

The point estimate of ψ is $\hat{\psi}_\theta/(\psi_n\psi_D) = 8{\cdot}13$. The programme of Thomas (1970) was used to place the confidence interval around ψ_θ and the endpoints of the interval were divided by $\psi_n\psi_D$ to yield the interval for ψ.

4.4. *Discussion*

This example illustrates a method of statistical inference concerning injury, persistence and ratios of persistence for bacterial populations which have been placed under stress. The analysis relies on known methods of inference about the binomial parameter and about the true cross-product ratio in a 2×2 table.

The suggested procedures were derived for experiments where the colony counts follow a Poisson distribution. Examples exist in the literature where colony count data do not conform to the Poisson distribution. As a check on the appropriateness of the Poisson distribution for data of the example, 80 replicates of the experiment using the same (but unstressed) bacterial population incubated in the non-selective medium were run. There was no obvious departure of the data from the Poisson distribution. The ratio of the sample variance to the sample mean was 0·95; the χ^2-test was not significant.

REFERENCES

BISSONNETTE, G. K. (1974). Recovery characteristics of bacteria injured in the natural aquatic environment. Ph.D. Dissertation, unpublished.

COX, D. R. & LEWIS, P. A. W. (1968). *The Statistical Analysis of Series of Events*, p. 226. Methuen & Co.

CLARK, C. W. & ORDAL, Z. J. (1969). Thermal injury and recovery of *Salmonella typhimurium* and its effect on enumeration procedures. *Applied Microbiology* **18**, 332–6.

DIEM, K. & LENTNER, C. (eds) (1970). *Scientific Tables*. Ardsley, N.Y.: Geigy Pharmaceuticals.

GART, J. J. (1971). The comparison of proportions: a review of significance tests, confidence intervals and adjustments for stratification. *Reviews of the International Statistical Institute* **39**, 148–69.

GART, J. J. (1974*a*). The analysis of ratios and cross products of ratios of Poisson variates with applications to incidence rates. Paper presented at the meeting of the American Statistical Association, St Louis, Mo., Aug. 1974.

GART, J. J. (1974*b*). The Poisson distribution: the theory and application of some conditional methods. Presented to the International Conference on Characterizations of Statistical Distributions, Calgary, Alberta, Canada, Aug. 1974.

Gart, J. J. & Thomas, D. G. (1972). Numerical results on approximate confidence limits for the odds ratio. *Journal of The Royal Statistical Society*, Series B, **34**, 441–7.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses*, p. 141. John Wiley & Sons.

McFeters, G. A. & Stuart, D. G. (1972). Survival of coliform bacteria in natural waters: field and laboratory studies with membrane filter chambers. *Applied Microbiology* **24**, 805–11.

Ordal, Z. J. (1970). Current developments in the detection of microorganisms in foods: influence of environmental factors on detection methods. *Journal of Milk and Food Technology* **33**, 1–5.

Ray, B. & Speck, M. L. (1972). Repair of injury induced by freezing *Escherichia coli* as influenced by recovery medium. *Applied Microbiology* **24**, 258–63.

Ray, B. & Speck, M. L. (1973*a*). Discrepancies in the enumeration of *Escherichia coli*. *Applied Microbiology* **25**, 494–8.

Ray, B. & Speck, M. L. (1973*b*). Enumeration of *Escherichia coli* in frozen samples after recovery from injury. *Applied Microbiology* **25**, 499–503.

Thomas, D. G. (1970). Exact confidence limits for the odds ratio in a 2 × 2 table. *Applied Statistics* **20**, 105–10.