STATISTICAL ANALYSIS

Hard Surface Carrier Test as a Quantitative Test of Disinfection: A Collaborative Study

MARTIN A. HAMILTON and TODD A. DEVRIES
Montana State University, Bozeman, MT 59717
JOSEPH R. RUBINO
Reckitt and Colman, One Philips Parkway, Montvale, NJ 07645

The hard surface carrier test (HSCT) recently was proposed as a qualitative test for disinfectant efficacy. A collaborative study of HSCT led to a suggested performance standard of ≤2 or 3 positive carriers out of 60 tested. Subsequently, it was discovered that HSCT can be used as a quantitative test, because the HSCT protocol requires measurement of inoculum level on some carriers. The data allow estimation of the $\log_{10}$ reduction in number of active bacteria. Producers, consumers, and policymakers will be better able to discuss merits of alternative performance standards if the focus is on log reduction of organisms rather than on number of positive carriers. Data from the collaborative study were reanalyzed from this quantitative viewpoint. If the point estimate of log reduction is LR and the 99% lower confidence limit estimate is LLR, the LR values ranged from 7.0 to 9.0 and the LLR values were greater than 6.0 for all disinfectants except the negative control formulation. The total variance for estimated LR is the sum of interlaboratory and intralaboratory variances. The total variance for LR was 0.095 for *Pseudomonas aeruginosa*, 0.251 for *Staphylococcus aureus*, and 0.118 for *Salmonella choleraesuis*. Percentages of the variance due to interlaboratory variability were 11% for *P. aeruginosa*, 52% for *S. aureus*, and 25% for *S. choleraesuis*. Chances of making false-effective and false-ineffective decisions can be calculated for the quantitative HSCT. The performance standard can be based on LLR.

Rubino et al. (1) collaboratively studied a disinfectant-testing method called the hard surface carrier test (HSCT). HSCT was proposed as an alternative to the use-dilution method (UDM), which the AOAC adopted in 1953 as a test for confirming germicidal activity of disinfectant dilutions (2). In HSCT, a disinfectant is applied to bacteria dried onto a glass penicylinder carrier. Sixty disinfected carriers per test were used in the collaborative study. HSCT is a qualitative test, the recorded outcome being the number of carriers showing presence (positive growth) or absence of active bacteria after incubation in the appropriate medium. Suggested performance standards for effective disinfection are ≤2 positive carriers out of 60 for *Staphylococcus aureus* and *Salmonella choleraesuis* and ≤3 positive carriers for *Pseudomonas aeruginosa*.

The HSCT protocol requires a check on inoculum level. The check eliminates a deficiency that plagued UDM, namely, lack of control over, or measurement of, inoculum level. In HSCT, 66 carriers are inoculated and every 11th is designated a check carrier. The bacterial load on each of 6 check carriers is measured by sonicating the carrier to remove bacteria and then counting bacteria by standard plate count procedures. The test is invalid if the average bacterial count is less than 5×10^5; the test is repeated if there are too many positive carriers and the mean check count is greater than 2×10^6.

While reviewing the results of the HSCT collaborative study, we discovered that check carrier counts, coupled with presence–absence observations for test carriers, provide sufficient information for a quantitative conclusion. The disinfectant-caused reduction in the number of bacteria can be estimated, and it is possible to calculate a lower confidence bound on the reduction. HSCT can now be viewed as a quantitative test.

Let λ denote log reduction, which is the negative logarithm of the fraction of bacteria capable of growth. Specifically, let N_t denote the hypothetical number of active organisms remaining on a test carrier after disinfection and let N_c denote the number of active organisms remaining on an identically inoculated check (nondisinfected) carrier. Then:

$$\lambda = -\log\frac{N_t}{N_c} = \log_{10} N_c - \log_{10} N_t$$

HSCT does not directly provide the counts N_t and N_c but does yield sufficient information for estimating λ.

The purposes of this paper are to introduce methods for quantitating HSCT and to present a reanalysis of data from the HSCT collaborative study (1). For our reanalysis, data from

Received March 29, 1994. Accepted by GL September 1, 1994.

each test are used to calculate estimates of log reduction (LR) and of 99% lower confidence limit (LLR). Graphical and statistical analyses show the extent to which quantitative HSCT results are reproducible. The statistical analysis provides intralaboratory and interlaboratory components of variance. Performance standards and error rates are discussed.

METHOD

Quantitative HSCT

(**a**) *Laboratory procedure*.—The HSCT method used in the collaborative study was described by Rubino et al. (1). Figures 1 and 2 show the steps involved. The 6 check carriers are inoculated and treated exactly as the 60 test carriers, except for disinfection. The simulated disinfection step shown in Figure 2 was not included in the original protocol of the collaborative study; consequently, collaborating laboratories did not perform that step. Instead, dried carriers were placed directly in sterile medium and sonicated. The simulated disinfection step is included in the schematic, because we now believe it will improve the accuracy of the method.

(**b**) *Calculating LR and LLR*.—Derivation of the formula for LR, description of the method for calculating LLR, and evaluations of the estimators' statistical properties are presented elsewhere (3). Evaluations showed that LR is unbiased and LLR has the correct (99%) level for the range of conditions expected. Estimation depends on 2 assumptions: a test carrier will be positive if and only if it contains 1 or more active organisms after disinfection and the number of active bacteria on a carrier prior to incubation will vary according to a probability distribution that is the same for test carriers when the disinfectant is completely inert as for the check carriers. Let T denote the number of positive carriers among K tested carriers; $K = 60$ for the collaborative study. Let M and S denote the sample mean and sample standard deviation, respectively, of the microbial counts for H check carriers; $H = 6$ for the collaborative study.

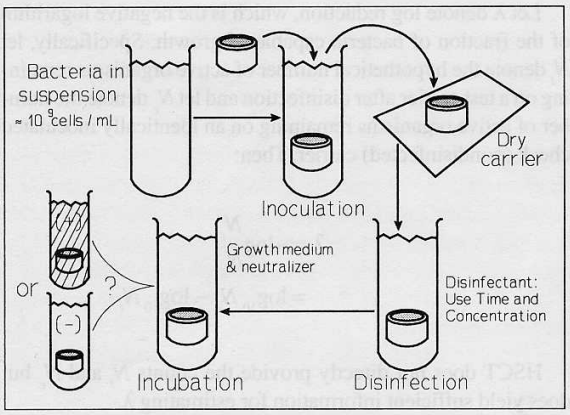

Figure 1. Schematic chart of steps in HSCT protocol for test carriers.

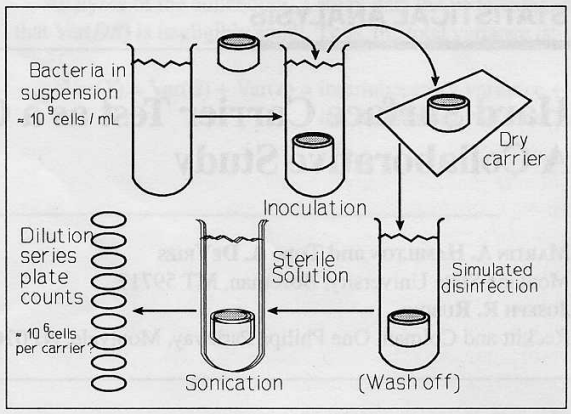

Figure 2. Schematic chart of steps in HSCT protocol for check carriers.

Let C denote the squared coefficient of variation for the check carrier counts; that is, $C = (S/M)^2$. Let

$$R = \left[1 - \frac{T + \frac{1}{2}}{K + 1}\right]^C$$

Then the formula for LR is:

$$LR = \log_{10} \frac{C \times M \times R}{1 - R}$$

If $T = K$, LR should be declared incalculable for 2 reasons. First, although all carriers will be positive ($T = K$) if the disinfectant is weak, inactive, or a bacterial growth medium, LR is always positive, sometimes a large positive value. Second, LLR cannot be calculated because negative values for λ, no matter how large, are not discredited by the observation $T = K$.

The lower confidence limit LLR is calculated by using a bootstrap simulation technique. It does not require normal distribution assumptions or any other assumptions beyond those stated for LR. The calculations are computationally intensive and require a computer.

(**c**) *Examples*.—Eight cases from the collaborative study are shown in Table 1 to illustrate how calculations depend on both check carrier and test carrier data. In the collaborative study, LR estimates ranged from 5.48 (case 1) to 9.01 (case 2). An LR value of 5.48 might seem high for case 1, where 56 of 60 test carriers were positive. However, all bacteria were inactivated in 4 of 60 test carriers for which the mean microbial burden was 3.3×10^6 bacteria. It is reasonable to believe that the formulation was active, even though the number of positive carriers (56) was far above the current performance standard.

Consider cases 3 and 4, which had equal check carrier means but different positive test carrier counts of 2 and 0, respectively. LR values were 7.39 and 8.10, properly indicating a higher log reduction for case 4, where all 60 carriers were negative.

Table 1. Quantitative HSCT results for some cases[a] selected from the collaborative study

Case	T = number positive carriers	M = mean burden ($\times 10^6$)	S = std. dev. ($\times 10^6$)	LR = estimated $\log_{10}$ reduction	LLR = 99% lower endpoint
1	56	3.30	3.16	5.48	3.71
2	0	8.36	2.02	9.01	8.11
3	2	1.03	1.48	7.39	6.74
4	0	1.03	2.00	8.10	7.18
5	0	0.78	0.16	7.98	7.12
6	1	0.995	1.55	7.59	6.23
7	2	1.65	0.28	7.60	7.02
8	60	5.70	2.63	NC[b]	NC

[a] Each case involved 60 test carriers and 6 check carriers.
[b] NC = Not calculable.

Consider cases 2 and 5, both of which had 0 positive test carriers but different means for check carrier counts, 8.36×10^6 and 7.8×10^5 organisms, respectively. The associated LR estimates were 9.01 and 7.98, reflecting that case 2 presented a greater challenge to the formulation.

Check carrier counts in HSCT will vary somewhat and can affect LLR. For example, cases 6 and 7 produced almost identical LR values but different LLR values of 6.23 and 7.02, respectively. The reason for this difference is that check carrier counts for case 6 were more variable (standard deviation [SD] $= 1.55 \times 10^6$) than for case 7 (SD $= 0.28 \times 10^6$). The smaller LLR for case 6 reflects the greater uncertainty attached to its check carrier counts.

Case 8 produced data that made LR and LLR incalculable. Because the outcome $T = 60$ would be expected if a formulation is ineffective, the data are consistent with small, even negative, values of λ. No value of λ is too small to be contradicted by the data, so the lower confidence limit does not exist.

Collaborative Study: Data and Statistical Methods

(**a**) *Data.*—In the original analysis of the collaborative study (1), the 10 participating laboratories were assigned code names A, B, …, J. In our reanalysis, we used data from 7 of those laboratories and data from an 11th laboratory, coded K. Laboratory K participated in the collaborative study, but its data were not included in the original analysis. Laboratories B, G, and E were not included in the reanalysis, because their data sheets did not show the check carrier microbial burden data required to calculate LR and LLR. Laboratory D was included only for *P. aeruginosa* and *S. aureus* because it did not submit complete check carrier data for *S. choleraesuis*.

(**b**) *Calculating interlaboratory and intralaboratory reproducibilities.*—The quantities LR and LLR were each submitted to variance component analyses to assess intralaboratory, interlaboratory, and formulation × laboratory interaction variances. Variance components were calculated by PROC VARCOMP and PROC GLM procedures in SAS (4); the 2 procedures gave essentially the same answers. PROC GLM provided F statistics for testing whether interlaboratory variances were significantly different from 0. Although participating laboratories were self-selected into the study, the analysis presumed that they were a random sample, that is, that they were statistically representative of a relevant population of laboratories. This assumption allowed us to analyze LR and LLR values by using a random-effects model, which provided the structure for estimating interlaboratory variances. The exact assumptions and associated statistical model are contained in Appendix A. Because formulations 1 and 6 were duplicates, within-laboratory differences between formulations 1 and 6 values were used to estimate intralaboratory variance. Formulation 2 was the negative

Table 2. Variance component analysis results for LR, the $\log_{10}$ reduction estimate[a]

Species	Source	Variance estimate	% of total	F	df	P value
P. aeruginosa	Interlaboratory	0.0100	11	1.57	7, 28	0.185
	Intralaboratory	0.0852	89	—	—	—
	Total	0.0952	100	—	—	—
S. aureus	Interlaboratory	0.1301	52	6.37	7, 29	<0.0001
	Intralaboratory	0.1212	48	—	—	—
	Total	0.2512	100	—	—	—
S. choleraesuis	Interlaboratory	0.0292	25	2.59	6, 24	0.045
	Intralaboratory	0.0893	75	—	—	—
	Total	0.1185	100	—	—	—

[a] Under the null hypothesis that the variance component attributable to interlaboratory sources is truly zero, the associated test statistic F follows a central F probability distribution with degrees of freedom = df.

control, and it often produced 60 positive carriers; therefore, it was excluded from statistical analyses.

Results

Figure 3 shows LR values for each laboratory and formulation, except formulation 2. All LR values are greater than 7.0. Figure 4 presents LLR values, all of which are greater than 6.0. Figures 3 and 4 provide a visual assessment of HSCT reproducibility. The vertical distance between formulations 1 and 6 within a laboratory is a measure of intralaboratory variability. Table 2 presents variance component analyses for LR. For LR data, the higher intralaboratory variability estimate for *S. aureus* is probably due to laboratories J and K, which had nearly identical LR values for formulations 1 and 6 for *P. aeruginosa* and *S. choleraesuis* but not for *S. aureus*. To visualize interlaboratory variability, approximate each laboratory's typical LR (Figure 3) or typical LLR (Figure 4) across all formulations and observe the spread of those typical values.

For each of the 3 species, the observed formulation × laboratory interaction variance component was 0, indicating that differences among formulations were consistent across laboratories. In the original analysis, the authors found a significant formulation × laboratory interaction variance only for *S. aureus*, but they thought that this interaction was biologically insignificant within the scope and precision of the test (1). We conclude that the formulation × laboratory interaction variance component was negligible for all 3 species and that the total variance is the sum of just 2 components: interlaboratory variance and intralaboratory variance.

For LR, interlaboratory variance estimates were 0.0101 for *P. aeruginosa*, 0.0292 for *S. choleraesuis*, and 0.1301 for *S. aureus*; only the last was clearly statistically significantly different from 0. The percentages of the total variance due to differences among laboratories were 11% for *P. aeruginosa*, 25% for *S. choleraesuis*, and 52% for *S. aureus*. The higher interlaboratory variability for *S. aureus* was probably due to relatively low LR values for *S. aureus* produced by laboratories C and I (Figure 3).

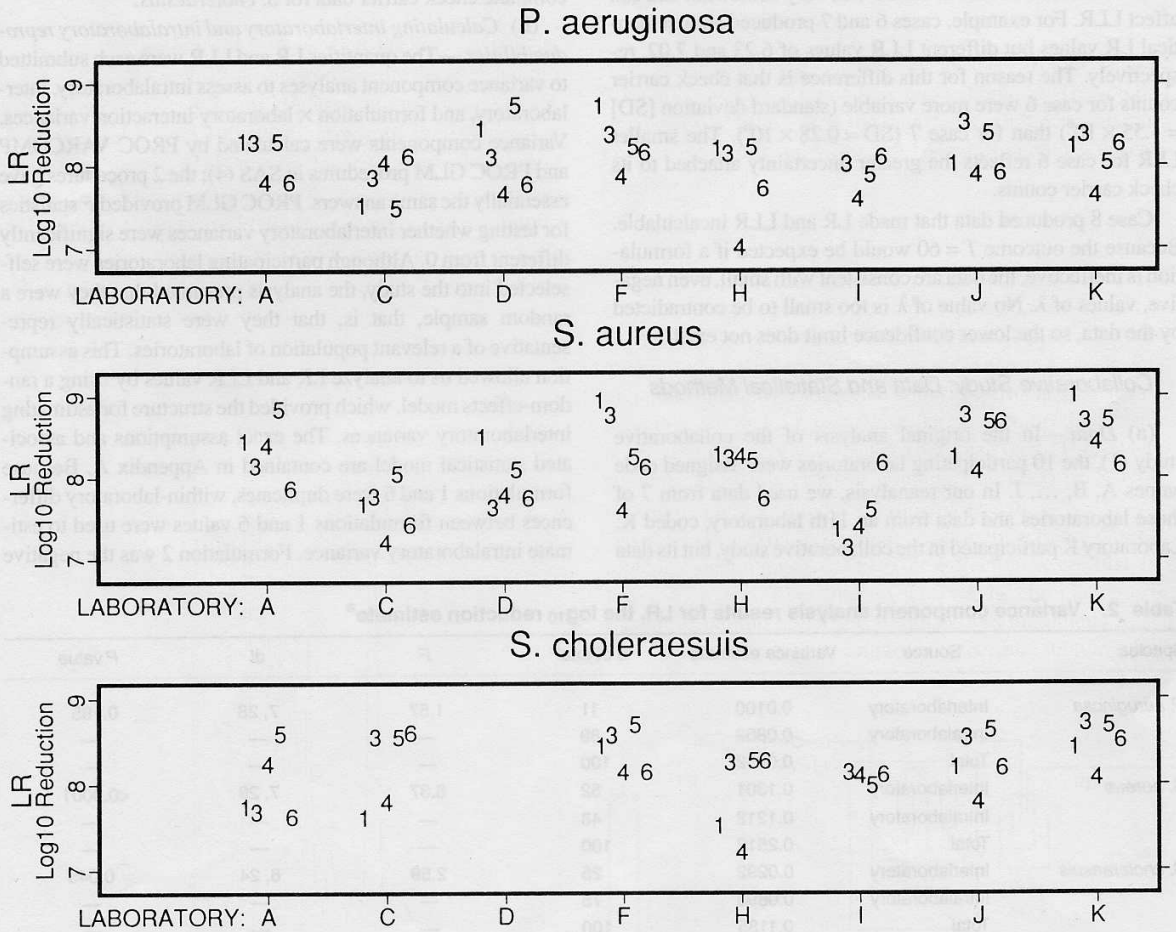

Figure 3. Observed LR for each laboratory. Plot characters are numbers to indicate formulation being tested; 1 and 6 are duplicates of a single formulation.

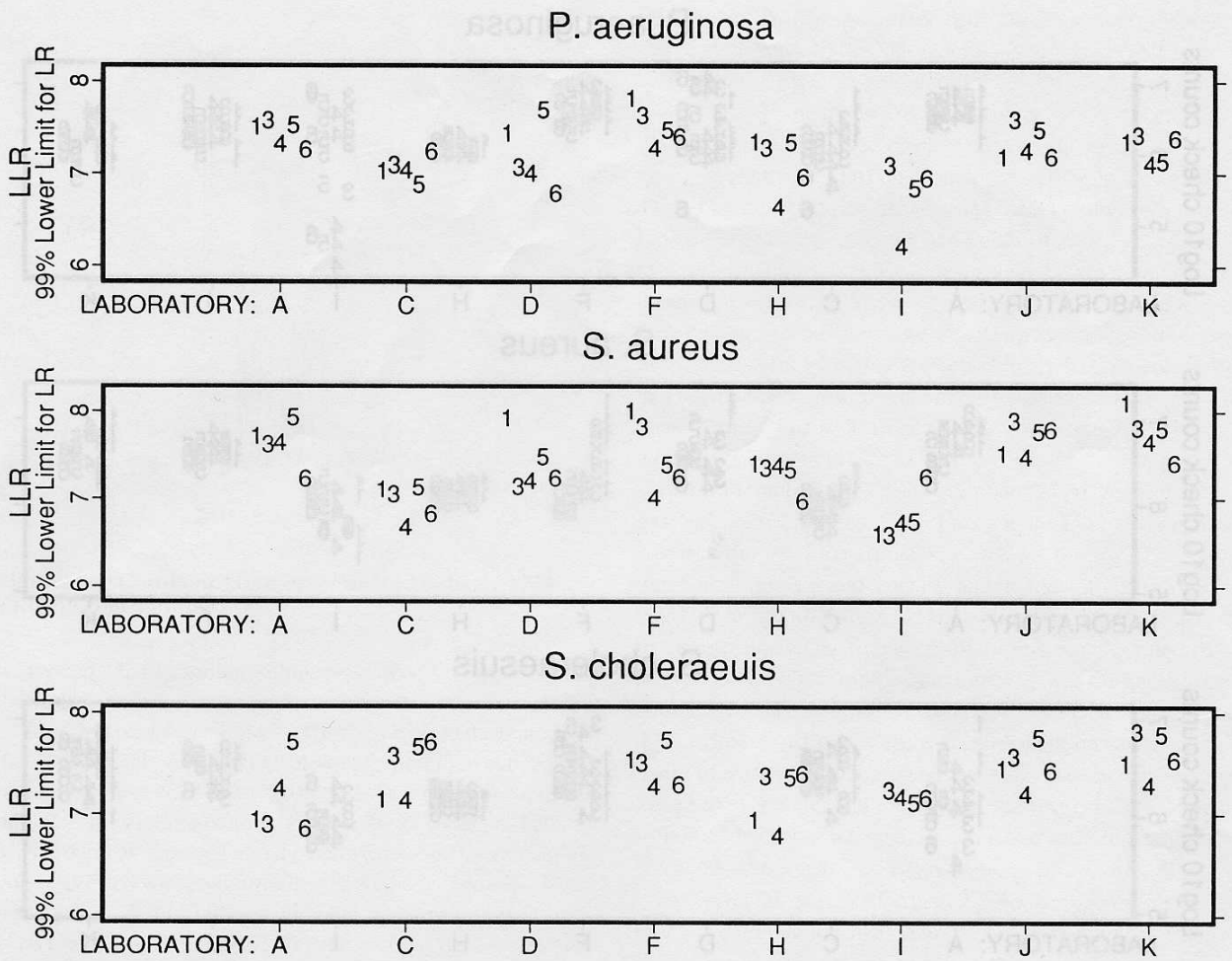

Figure 4. Observed LLR for each laboratory. Plot characters are numbers to indicate formulation being tested.

Table 3 presents the variance component analysis for LLR. Comparison of Table 3 with Table 2 shows that intralaboratory variability and total variability were less for LLR than for LR. This relationship is evident also from Figures 3 and 4, where each within-laboratory vertical distance between formulations 1 and 6 was at least as large for LR as for LLR. An explanation is presented in Appendix B. For LLR, interlaboratory variance was about half the total variance and was statistically significantly different from 0.

Table 3. Variance component results for LLR, the 99% confidence lower endpoint for $\log_{10}$ reduction[a]

Species	Source	Variance estimate	% of total	*F*	df	*P* value
P. aeruginosa	Interlaboratory	0.0410	48	5.42	7, 28	0.0005
	Intralaboratory	0.0451	52	—	—	—
	Total	0.0861	100	—	—	—
S. aureus	Interlaboratory	0.1086	60	8.58	7, 29	<0.0001
	Intralaboratory	0.0716	40	—	—	—
	Total	0.1802	100	—	—	—
S. choleraesuis	Interlaboratory	0.0272	42	4.47	6, 24	0.004
	Intralaboratory	0.0380	58	—	—	—
	Total	0.0652	100	—	—	—

[a] Under the null hypothesis that the variance component attributable to interlaboratory sources is truly zero, the associated test statistic *F* follows a central *F* probability distribution with degrees of freedom = df.

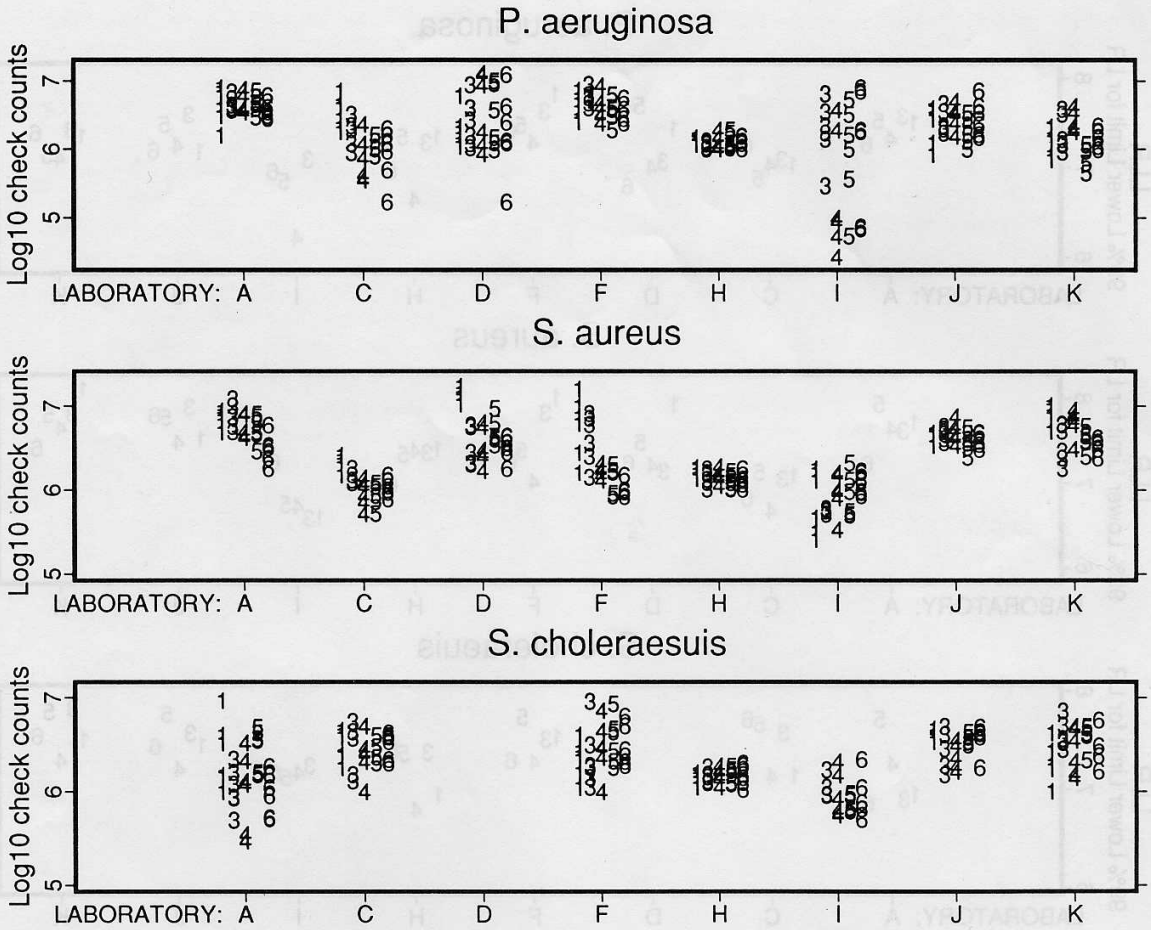

Figure 5. Observed counts of bacteria on check carriers within each laboratory. Plot characters are numbers to indicate formulation being tested.

Counts of bacteria on check carriers play a major role in quantitation of HSCT. Figure 5 shows check carrier counts that entered the analyses. Differences among laboratories are evident in Figure 5. For the *P. aeruginosa* data, carriers in laboratory A produced higher counts than those in laboratory C. Counts for laboratory H exhibited remarkably small variability and, for each of the 3 species, were almost exactly centered on the target of 10^6 bacteria. For some laboratories (e.g., A and C), the typical check carrier count seemed to decrease steadily as the formulation number increased. Assuming that formulations were tested in numerical order, the microbial burden per carrier may have been affected by a time trend. Because LR and LLR calculations take into account the microbial burden, they are not invalidated by such a trend.

Discussion

In this collaborative study, the 99% confidence lower endpoint for log reduction was greater than 6.0 for all formulations, except formulation 2, which was the study's negative control. The estimated interlaboratory variance was about equal to the intralaboratory variance, a fact of practical significance. The method we used for calculating LLR was based on assessment of intralaboratory variability only. If interlaboratory variability must be accounted for, then a different, more conservative statistical procedure for estimating a lower limit on LR will be necessary.

It is anticipated that HSCT will be the basis for a decision concerning the tested formulation's effectiveness. Standardized tests usually are accompanied by a performance standard, which is a set of potential test outcomes for which the decision is effective. A suitable performance standard for HSCT is to decide that a formulation is effective if and only if the lower confidence limit is greater than a designated log reduction value, called the performance reliability value (PRV). PRV could be chosen by policymakers, in consultation with consumers, producers, and independent experts. Some implications of alternative choices for PRV can be ascertained from this collaborative study. For example, if the PRV were 6, then all formulations, other than formulation 2, would have been de-

clared effective by every test in the collaborative study. However, if the PRV were 7, the majority of tests would have concluded that the formulations were ineffective (Figure 4). These statements are not meant to imply that 6 or 7 is an appropriate PRV value.

Decisions based on quantitative HSCT will not be perfect. Two types of errors can be made: a false-effective decision, where a formulation having a λ less than PRV is found by the test to be effective, and a false-ineffective decision, where HSCT leads one to decide that a formulation with a sufficiently large λ is ineffective. By stating performance standard in terms of the confidence endpoint for LR, the risk of a false-effective decision can be controlled. Specifically, for a performance standard based on LLR, the risk of a false-effective decision is 1.0% at most. If the 95% confidence endpoint were used, then the risk would be 5% at most. The appropriate risk level should be chosen by policymakers. In this paper, we used a 99% confidence level for illustrative purposes only. The risk of a false-ineffective decision can be calculated, but because those calculations depend on the policymaker's performance standard and chosen target for λ, we give no examples here.

The statistical properties of HSCT depend on the numbers of carriers used: K test carriers and H check carriers. By increasing K and H, the risks of false-effective and false-ineffective decisions can be reduced. The authors of the HSCT protocol used $K = 60$, because 60 carriers were recommended for UDM. They used $H = 6$, because it seemed intuitively reasonable to have a check carrier for every 10 test carriers. After policymakers establish target values for λ and values for acceptable risks of decision errors, it will be possible to determine the appropriate numbers of carriers K and H.

LR and LLR calculations for this collaborative study might be affected by a positive bias. When the collaborative study was conducted, the protocol for check carriers did not include a wash-off step to simulate possible removal of bacteria from the carrier during disinfection. Thus check carrier counts may be too high, which would lead to overstated LR and LLR values. A second collaborative study is being planned, mainly to assess the reliability of HSCT in the presence of organic soil and hard water. That study will include a wash-off step in the check carrier protocol. In this step, it is important to use an inert solution that washes off the same proportion of active organisms as does the disinfectant formulation.

Compared with the qualitative form of HSCT, the quantitative approach is easier to interpret. We believe that policymakers, producers, and consumers will be better able to discuss the merits of alternative performance standards if the focus is on log reduction instead of number of positive carriers. False-effective results for qualitative HSCT could occur because of a small bacterial burden per carrier; the associated risk of a false-effective result is probably impossible to calculate. The risks of false-effective and false-ineffective results, however, can be calculated for quantitative HSCT.

Acknowledgments

The statistical methodology development and analyses were partially supported by Cooperative Agreement ECD-8907039 between the National Science Foundation and Montana State University and by Cooperative Agreement CR818324-02 between the Environmental Protection Agency and Montana State University. The biological work that led to HSCT and the collaborative study was conducted independently by the investigators and laboratories listed in reference 1. We thank anonymous referees for suggestions and corrections that improved the paper.

References

(1) Rubino, J.R., Bauer, J.M., Clarke, P.H., Woodward, B.B., Porter, F.C., & Hilton, H.G. (1992) *J. AOAC Int.* **75**, 635–645

(2) Stuart, L.S., Ortenzio, L.F., & Friedl, J.L. (1953) *J. Assoc. Off. Agric. Chem.* **36**, 466–480

(3) Hamilton, M.A., & DeVries, T. (1995) submitted for publication to *Biometrics*

(4) SAS Institute, Inc. (1988) *SAS/STAT User's Guide, Release 6.03 Edition*, SAS Institute, Inc., Cary, NC

Appendix A

The statistical model for variance component analyses of LR and LLR has a deterministic component for formulation and 3 independent stochastic components for laboratory, formulation × laboratory interaction, and inherent chance error. Let Y denote either LR or LLR, depending on the context. Let D_i denote the deterministic mean Y for the ith formulation. Let the stochastic part of the model be denoted by $B_j + DB_{(ij)} + e_{ijk}$, where B_j is the random effect due to the jth laboratory, $DB_{(ij)}$ is the random effect due to the interaction between the ith formulation and the jth laboratory, and e_{ijk} is the inherent chance error due to intralaboratory sources from the kth replicate of the ith formulation and jth laboratory combination. The true mean of each random component is 0. The true, unknown variances of the random components are denoted by Var(B), Var(DB), and Var(e), respectively. Let Y_{ijk} denote the Y value for the ijkth formulation × laboratory × replicate combination.

Formulations 1 and 6 are duplicates and were both coded as 1 for the analysis. Formulation 2 is not used, and thus the formulation subscript values are 1, 3, 4, and 5. The number of replicates for any formulation × laboratory combination is either 1 or 2, depending on formulation subscript. If the formulation subscript is 1, then there are 2 replicates (i.e., values from formulation 1 and 6), but for formulation subscripts 3, 4, and 5, there is only 1 replicate. The statistical model is:

$$Y_{ijk} = D_i + B_j + DB_{(ij)} + e_{ijk}$$

where: $i = 1, 3, 4, 5$; $j = 1,\ldots,8$; $k = 1, 2$, if $i = 1$, and $k = 1$, if $i = 3, 4$, or 5.

If subscripts are suppressed, the model implies that the total variance for Y is:

$$\mathrm{Var}(Y) = \mathrm{Var}(B) + \mathrm{Var}(DB) + \mathrm{Var}(e)$$

Analyses of the collaborative study data led us to conclude that Var(DB) is negligibly small. Thus, the total variance is:

$$\mathrm{Var}(Y) = \mathrm{Var}(B) + \mathrm{Var}(e) = \text{interlaboratory variance} + \text{intralaboratory variance}$$