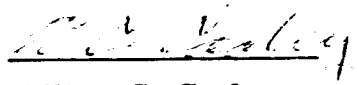
A STOCHASTIC STERILIZATION MODEL

by

Samuel Schalkowsky and Robert Wiederkehr*

Topical Report
Prepared under Contract NASw-1340
for
National Aeronautics and Space Administration
Washington, D. C.

Approved by: 

William C. Cooley
Program Manager

May 6, 1966

Exotech, Inc.
525 School Street, S. W.
Washington, D. C.

Report TR-13

* Consultant, Westat Research Inc.

ABSTRACT

An analytical framework is provided to relate pertinent aspects of the experimental determination of microbial resistance to sterilization on the one hand and to the parameters which enter into the definition of operational sterilization requirements on the other. This analytical model differs from existing models in that it is applicable to any resistance function, e.g., the survivor curve need not be exponential. The model is used to correlate the results from survivor counting tests to sampling tests (end-point) and to evaluate the validity of extrapolating from survivor data to low probabilities of contamination when the latter are specified in a number of alternate ways.

TABLE OF CONTENTS

		Page
1.	INTRODUCTION	1
2.	BASIC ANALYTICAL RELATIONSHIPS	4
3.	PROBABILITY OF SURVIVAL FOR A SINGLE ORGANISM	5
4.	NUMBER OF SURVIVORS IN A FIXED POPULATION	6
5.	STERILITY OF A POPULATION	10
6.	EXTRAPOLATION FROM FRACTIONAL SURVIVORS TO PROBABILITY OF CONTAMINATION	12
7.	EXTRAPOLATION FROM FRACTIONAL SURVIVORS TO PROBABILITY OF ONE SURVIVOR	15
8.	SUMMARY AND CONCLUSIONS	17
	REFERENCES	19

1. Introduction

Under the subject contract a study is in progress to evaluate analytical techniques currently used in the formulation and implementation of spacecraft sterilization requirements, and to evolve new techniques where appropriate. One area found to require considerable clarification and modification is the basic analytical model which relates experimental procedures in the determination of microbial heat resistance to the definition of operational requirements for spacecraft sterilization so as to achieve a specified, low probability of contamination.

Current difficulties arise not from the absence of an analytical model but rather from the fact that the model being used is based upon unproved hypotheses and there is ever increasing evidence that these hypotheses may be wrong. Specifically, it is currently accepted as a "law" that microorganisms exposed to a heat environment lose viability exponentially, i.e. the number of survivors is decreased by one decade in constant intervals of heating time. The validity of this "law" has been questioned ever since it was promulgated and is also being scrutinized in the study under the present contract. However, this is not the subject of the present report. Of interest here is the fact that current analytical models in sterilizations are predicated upon the validity of the exponential "law". But this is quite unnecessary since the model can also be formulated without constraints on the specific form of the survival function. Thus, the relationship between experimental test paramaters and probabilities of sterility (or contamination) of a spacecraft can be evolved on the basis of an undefined survival function. Clearly, such a formulation would be preferred since it is free of any questions concerning the validity of a particular survival function, and can facilitate the study of alternate survival functions. Such a model has been evolved and is described herein. It is referred to as a stochatic model since the unknown survival function is defined in terms of the survival times of each organism in a population and survival time is taken to be a random variable.

The principal use of this, as of any other analytical model, is to provide a framework for specific problem-oriented investigations. As previously noted, one intended application for the model described herein is in the study of alternate survivor functions, i.e. other than the exponential. This effort is currently in progress and will be reported on at a later date. However, a few other applications are considered in the text and include the following:

(a) Experimental Procedures

The resistance of microorganisms to heat sterilization is generally determined in one of two ways. In a counting test, the number of survivors from an initial viable population is determined experimentally, leading to the survivor-time curve. The second method, frequently referred to as the end-point test, utilizes multiple samples of some initially viable populations heated to a range in time where at least some of the samples will show sterility when cultured after heating. A common basis is needed for these two tests and has been provided by Aiba and Toda[1] on the assumption that the survivor curve is exponential. The present model provides more general relationships in that they are applicable to any survivor function. The utility of the present formulation is illustrated by the fact that it explains the lack of correlation between Aibas' and Toda's analytical predictions and their experimental data, see reference (1). However, more generally, the present model can be used to evaluate the relative merits of the two test methods and to define conditions under which one or the other might be more desirable.

(b) Extrapolation of Test Data

Test data in a counting test generally do not cover a range of initial populations and times of heating which incoudes the conditions of the ultimate sterilization process. An extrapolation of test data is therefore necessary to predict the probability of sterility in the actual process. The general validity of such extrapolations has been questioned and recently both Fredrickson[2] and Aiba and Toda[1] have provided an analytical basis for the extrapolation on the

assumption that the survivor function is exponential. The present formulation generalizes the analysis by removing the above assumption. In addition, alternate forms of specifying probability of contamination, e.g. exactly one survivor vs. one or more survivors, are considered and their effect on the accuracy of extrapolation is evaluated.

The following assumptions underly the analytical model developed herein:

(1) A single species of organisms is considered

(2) The deaths of organisms within a population are assumed to be independent events

(3) The intensity of the sterilization environment is assumed to be constant.

Assumption (3) above is noted to clarify the notation to be used. However, it does not constrain the validity of the analytical relationships to constant sterilization environments. Thus, the sterilization time, t, can be replaced by a sterilization dose which is a function of time as well as of a variable sterilization intensity, without invalidating results to be presented here.

2. Basic Analytical Relationships

The principal building block throughout this report will be the well known binomial distribution for repeated trials where the probability of success (or failure) in any trial remains the same. We are thus assuming that the death of one organism has no effect on the probabilities of survival of any other organism, i.e. they are taken to be independent events.

For convenience, we list the following formulas relating to the binomial distribution, e.g. Uspensky[3].

$$P(r,T) = \begin{bmatrix} T \\ r \end{bmatrix} p^r (1-p)^{T-r} = \frac{T!}{r!\,(T-r)!}\, p^r (1-p)^{T-r} \qquad (1)$$

P(r,T) denotes the probability of obtaining r successes in T trials and p is the probability of success in any one trial.

Some limiting cases of interest here are the probability of zero successes in T trials, i.e. r = o, and the probability of a success on every trial, i.e. r = T. From eq. (1) we have

$$P(0,T) = (1-p)^T \qquad (2)$$

$$P(T,T) = p^T \qquad (3)$$

Equation 1 will define a discrete probability function of r when p and T are fixed to particular values. We are interested in the value of r which has the greatest probability of occurring. Using standard nomenclature, we denote this value of r as the expected value of r, E(r), given by:

$$E(r) = T \cdot p \qquad (4)$$

Referring again to equation (1) as a distribution function of r, the variance of r is given by

$$\sigma^2(r) = T p \cdot (1-p) \qquad (5)$$

3. Probability of Survival for a Single Organism

Let τ_d be a random variable, denoting the time at which an organism, subjected to heat for a time τ will die. We then define $f(\tau)d\tau$ as the probability that the organism dies at the time τ, i.e. as the probability that $\tau \leq \tau_d \leq (\tau + d\tau)$. The probability that the organism will die during the time interval zero to t is then given by

$$P_d(t) = p(0 \leq \tau_d \leq t) = \int_0^t f(\tau)d\tau \tag{6}$$

The probability p_s that an organism will survive the time interval t is

$$p_s(t) = 1 - p_d(t) = 1 - \int_0^t f(\tau)d\tau \tag{7}$$

In what follows, we will deal with $p_s(t)$ or $p_d(t)$ without requiring explicit knowledge as to what these functions are. For, as previously noted, one of the major objectives of the model is to facilitate determination of $f(\tau)$ through $p_s(t)$ or $p_d(t)$. If $f(\tau)$ is assumed to be an exponential distribution of survival times, and it is emphasized that we neither require nor accept this assumption, then $p_s(t)$ would be given by

$$p_s(t) = 1 - \int_0^t \mu e^{-\mu t}\, d\tau = 1 - \left[-e^{-\mu\tau}\right]_0^t = e^{-\mu\tau} \tag{8}$$

4. Number of Survivors in a Fixed Population

Consider a population of N_o organisms subjected to heat sterilization for a time t and let N(t) be the number of organisms which have survived up to the time t. Since we consider the deaths of organisms to be independent events, and looking at a particular time interval t, we can view each organism as being subjected to a trial. The total number of trials therefore equals N_o. If we call survival up to time t a success, then $p_S(t)$ is the probability of success in any one of these repeated trials. We can therefore apply equation (1) with $T = N_o$ and $r = N(t)$. Thus

$$P\left[N(t), N_o\right] = \frac{N_o!}{N(t)!\left[N_o - N(t)\right]!} \; p_S(t)^{N(t)} \left[1 - p_S(t)\right]^{N_o - N(t)} \qquad (9)$$

Figure 1 illustrates the results of a hypothetical experiment in which we take n samples, each containing an initial population of N_o viable organisms, apply heat for a fixed time t_c and note the number of samples $n(N_1)$ having a particular number of survivors N_1. We might attempt to fit a discrete distribution to Figure 1 using equation (9) with $p_S(t_c)$ as a parameter. Having obtained a suitable fit, we would have one point of $p_S(t)$. Repeating this procedure for different heating times, we could define the manner in which $p_S(t)$ varies with time.

Needless to say, the above would not be an efficient experimental procedure. We note, however, that the expected value of N(t) is, according to equation (4)

$$E\left[N(t)\right] = N_o \, p_S(t) \qquad (10)$$

Since we generally expect $p_S(t)$ to be a monotonic function of time, we draw a smooth (not necessarily straight) curve thru the measured values of N. Points on this curve can then be thought of as averaged values of N, which we will denote as N^A, and the curve itself represents an estimated $p_S(t)$ which will be denoted as $\hat{p}_S(t)$. Thus,

$$\hat{p}_S(t) = N^A(t)/N_o \qquad (11)$$

as shown in Figure 2.

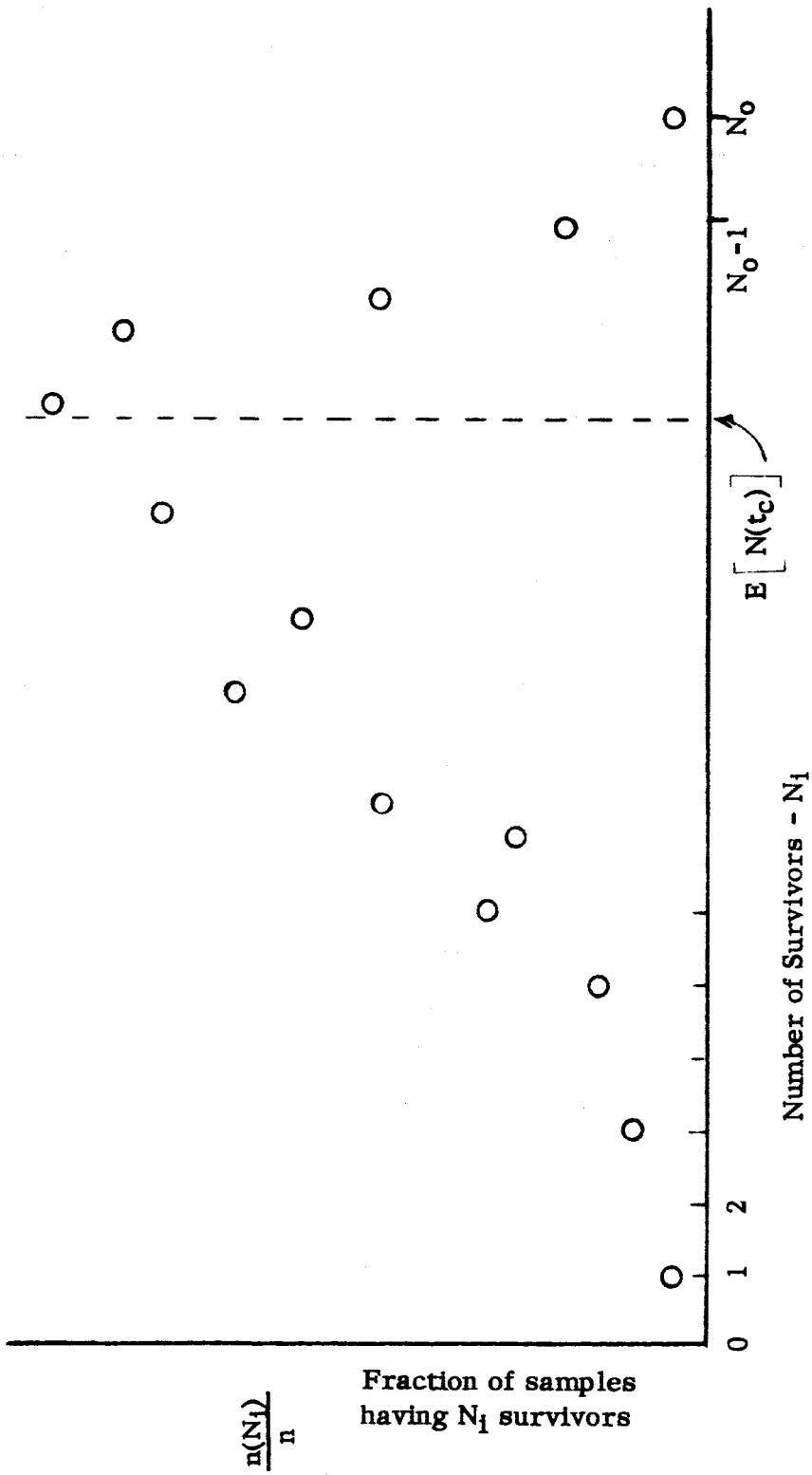


Figure 1. Hypothetical Survivor Frequencies

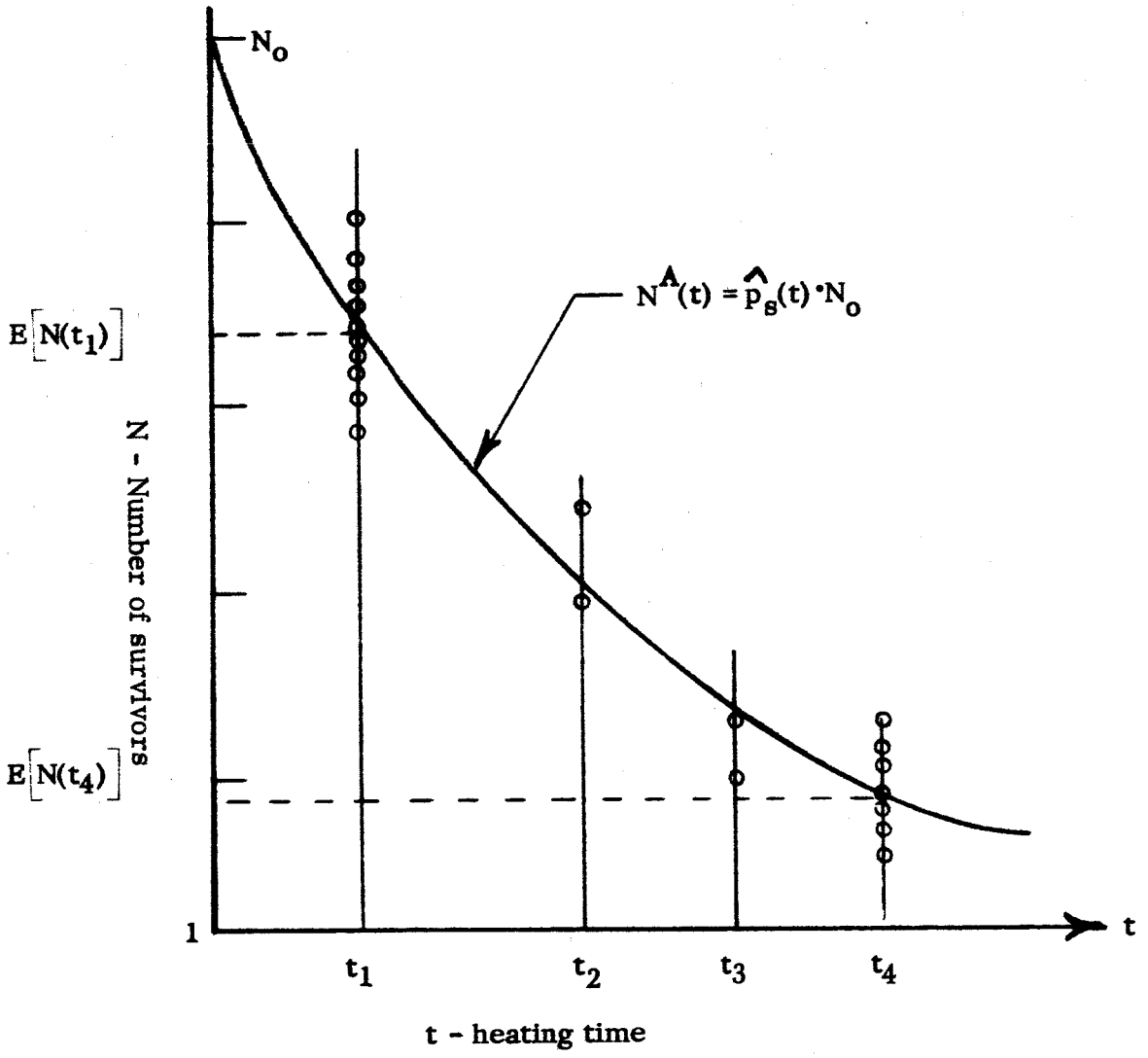


Figure 2. Illustrative Counting-Test Data

We can thus obtain $\hat{p}_S(t)$ from a counting experiment in which the averaged number of survivors, $N^{\blacktriangle}$, are normalized by dividing by N_o.

From eq. (5) and eq. (11) it is readily shown that the variance of $\hat{p}_S$ is given by

$$\sigma^2(\hat{p}_S) = \frac{\hat{p}_S(1 - \hat{p}_S)}{N_o} \tag{13}$$

Thus, by choosing a sufficiently large N_o, the variance can be made small.

5. Sterility of a Population

Consider n_o samples, each containing initially N_o viable organisms, and let $n_s(t)$ be the number of samples which are sterile after time t. Similarly, $n_c(t)$ will denote the number of contaminated samples at time t. Clearly

$$n_s(t) + n_c(t) = n_o \tag{14}$$

We can view each sample as one of n_o repeated trials and define "success" in any one trial to mean that the sample remained contaminated. This probability of "success" is therefore $P_c(t)$, i.e. the probability of one or more survivors. Applying the binomial distribution to define the probability of $n_c(t)$ contaminated samples, we obtain

$$P\left[n_c(t), n_o\right] = \frac{n_o!}{n_c!\,(n_o-n_c)!} \quad P_c^{n_o}\,(1 - P_c)^{n_o-n_c} \tag{15}$$

The expected value of $n_c(t)$ is, by equation (4)

$$E\left[n_c(t)\right] = n_o \cdot P_c(t) \tag{16}$$

or

$$P_c(t) = E\left[\frac{n_c(t)}{n_o}\right] \tag{17}$$

In a manner similar to that discussed for survivor curves, an estimated value of P_c can be obtained by plotting a smooth curvethru the data points given by $n_c(t)/n_o$. Denoting again by n_c^A the points on this curve, we have

$$\hat{P}_c(t) = \frac{n_c^A(t)}{n_o} \tag{18}$$

It is of interest to relate P_c to p_s since we may wish to obtain p_s not by the counting methods previously described, but to infer it from sampling test in which the measured parameter is $\hat{P}_c$. Towards this end, we first define P_s as the probability that a sample of N_o initially viable organisms will be sterile as a function of exposure time t. Since this implies no survivors, application of equation (9) with N = 0 yields

$$P_s = P(0,N_o) = \left[1 - p_s(t)\right]^{N_o} \tag{19}$$

Since

$$P_c = 1 - P_s$$

$$P_c = 1 - \left[1 - p_s(t)\right]^{N_o} \tag{20}$$

Let $\hat{p}_s^s$ denote the value of p_s inferred from a sampling experiment, i.e. it is obtained from $\hat{P}_c$. Then, from equation 20

$$\hat{p}_s^s = 1 - \left[1 - \hat{P}_c\right]^{1/N_o} \tag{21}$$

Equation 21 provides a means for converting results from sample sterility experiments to estimates of the organism survival probability function.

6. Extrapolation from Fractional Survivors to Probability of Contamination

As shown in references (1) and (2), the probability of contamination can be obtained with little error by an extrapolation of the survivor curve to values of $N < 1$. In the above references, only the case where $f(\tau)$ is an exponential distribution is considered. We can readily generalize the extrapolation to any distribution $f(\tau)$ (and hence for any shape of a survival curve $p_S(t)$) by expanding equation 20 into a Taylor series. Omitting the functional notation (t) for the sake of brevity, combining 11 and 20 and then expanding the exponential term

$$\left[1 - \frac{E(N)}{N_o}\right]^{N_o}$$

into a Taylor series with a remainder R_c, we obtain

$$P_c = 1 - \left[1 - \frac{E(N)}{N_o}\right]^{N_o} = E(N) - R_c \tag{22}$$

$$R_c = \frac{[E(N)]^2}{2}\left(1 - \frac{1}{N_o}\right)\left(1 - \theta\frac{E(N)}{N_o}\right)^{N_o-2}; \; 0 < \theta < 1 \tag{23}$$

Since

$$(1 - 1/N_o) < 1 \text{ and } \left(1 - \theta\frac{E(N)}{N_o}\right)^{N_o-2} < 1 \tag{24}$$

$$R_c < \frac{[E(N)]^2}{2}$$

Since R_c is always positive, $P_c < E(N)$. Thus, as shown in Figure 3, the curve of P_c must always be below the extrapolated E(N). Since, by definition, $P_c \leq 1$, the extrapolation can only have meaning for values of $N < 1$.

Although the extrapolation has been discussed in terms of P_c and the expected values E(N), it would be more proper to define it in terms of $\hat{P}_c$ and N^A. In practice, curves such as the one shown in Figure 3, would be obtained by fitting an analytical relationship $N^A(t)$ to the data points over the experimental range of N. The extrapolation would then consist of extending the analytical curve by calculations, using values of $N^A < 1$.

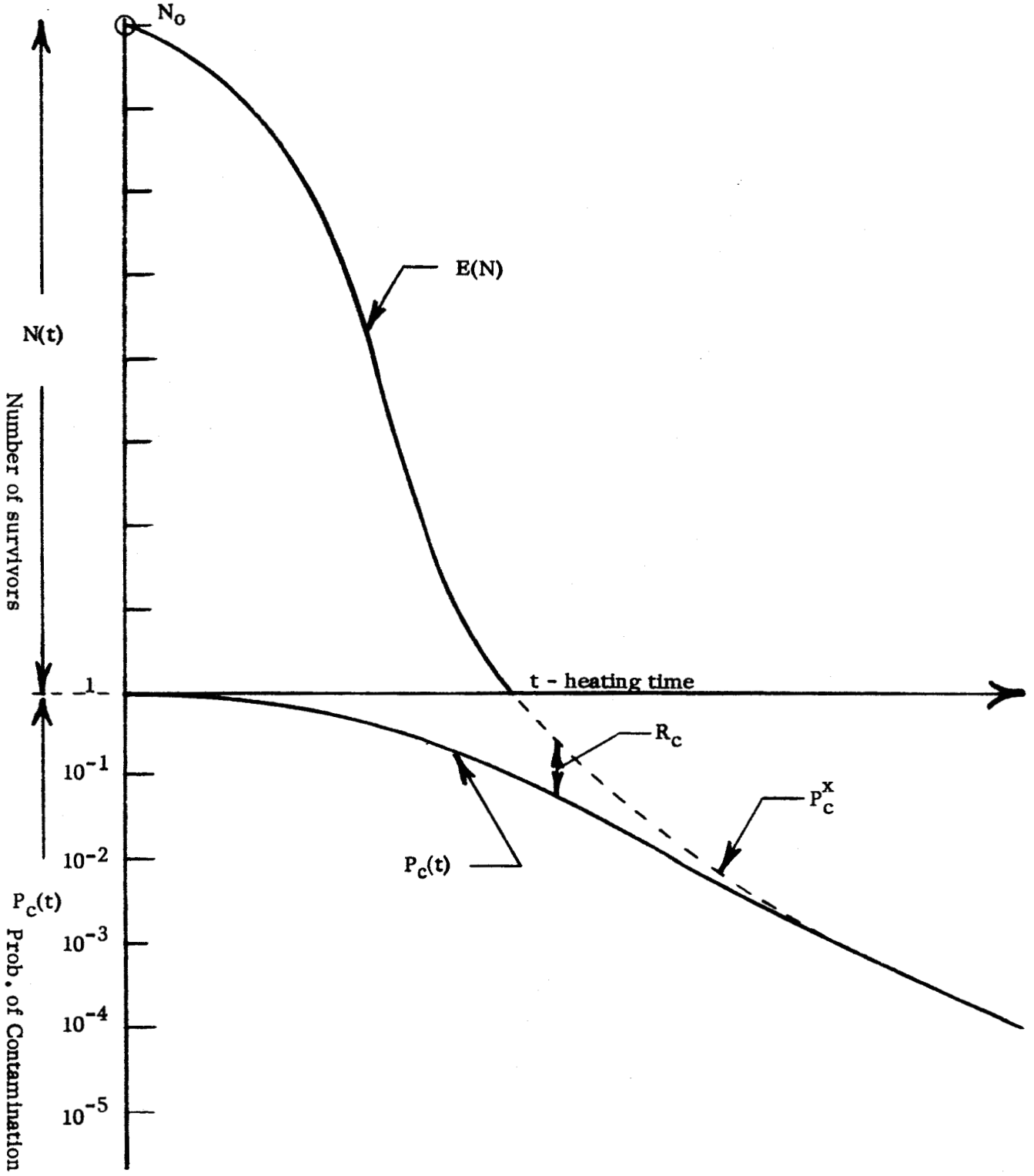


Figure 3. Extrapolation from Survivor Curve to Probabilities of Contamination

It is to be noted that P_c always refers to a probability of contamination for a given initial population. Hence, the extrapolated P_c, which we denote as P_c^x, relates to the N_o of the curve of N^A from which it was extrapolated. This is not a restriction since P_c^x for any other population is readily obtained from the fact that curves of $N^A(t)$, when plotted on semi-log paper, can be translated up or down to match different N_o's. In general, from eq. 11 for the same $p_s(t)$

$$\frac{N^A{}_1}{N_{o1}} = \frac{N_2{}^A}{N_{o2}} \tag{25}$$

The maximum error associated with the extrapolation is seen to be small for values of P_c less than about 0.1. Some numerical values are illustrated below.

P_c^x	Max. error in P_c
10^{-1}	+5%
10^{-2}	+0.5%
10^{-3}	+0.05%
10^{-4}	+0.005%

These errors indicate that the extrapolation is quite accurate. Furthermore, as previously noted, the extrapolated values will always be larger than P_c. The extrapolation therefore produces slightly conservative values of the probability of contamination P_c.

7. Extrapolation from Fractional Survivors to Probability of One Survivor

Currently, planetary quarantine requirements for sterilized landing vehicles are specified not in terms of the probability that the lander will be contaminated but rather in terms of a probability that it will have one viable organism. The latter has been denoted as p_N and would be given by equation (9) as

$$p_N = P(1, N_o) = N_o p_s (1 - p_s)^{N_o - 1} \tag{26}$$

It would also be convenient to obtain p_N by extrapolating curves of E(N) as was done for P_c. To evaluate such an extrapolation we again expand the right hand side of 26 into a Taylor series with a remainder denoted by R_1. This yields

$$p_N = N_o p_s (1 - R_1) \tag{27}$$

$$R_1 = N_o p_s (1 - 1/N_o)(1 - \theta p_s)^{N_o - 2} \tag{28}$$

Since

$$N_o p_s = E(N) = p_N^x$$

i.e. p_N^x is the value of p_N obtained by extrapolating E(N), we can write

$$P_N^x - p_N = \Delta p_N = p_N^x \cdot R_1 \tag{29}$$

To compare the extrapolations of P_c and p_N, equations 22 and 29 are slightly rearranged below:

$$P_c = P_c^x - R_c \tag{22a}$$

$$p_N = p_N^x - \Delta p_N = p_N^x - R_1 \cdot p_N^x \tag{29a}$$

Since $p_N^x \cdot R_1$ is always positive, p_N^x will also be a conservative value of p_N. By reasoning similar to that applied for R_c,

$$R_1 < p_N^x \tag{30}$$

Hence

$$\Delta p_N < (p_N^x)^2$$

Figure 3 can be applied to this extrapolation by replacing P_c with p_N and R_c with Δp_N. Maximum values of Δp_N are illustrated below

P_N^x	Max. error in p_N
10^{-1}	10%
10^{-2}	1%
10^{-3}	0.1%
10^{-4}	0.01%

Again, the inaccuracy in the extrapolation is on the side of conservative estimates of p_N and the magnitudes of the errors, although twice as large as that for P_c^x, are sufficiently small for $p_N^x \leq 10^{-2}$ to make the extrapolation of p_N quite accurate.

8. Summary and Conclusions

The principal function of interest in characterizing the resistance of a given strain of organisms to a sterilizing environment is the probability density of survival times, $f(\tau)$. It is, however, more convenient to use the survival function $p_s(t)$, i.e. the probability of survival in the interval of time up to t, since the two are interrelated thru the integ ration

$$p_s(t) = 1 - \int_0^t f(\tau)d\tau \tag{7}$$

$p_s(t)$ can be determined experimentally in one of two ways. In a counting experiment, $\hat{p}_s(t)$ is given by a plot of N^A/N_o as a function of the heating time t. Such a plot is independent of N_o. Hence a plot of the log of N^A versus t for a specific value of log N_o can be translated up or down to other values of log N_o without change in the shape of the curve and regardless of the shape, i.e. it need not be a straight line on a semi-log plot of N(t).

An alternate approach to the evaluation of $p_s(t)$ is to measure the number of samples out of an initial n_o samples which are not sterile after exposure up to a time t. If each sample had an initial population N_o^s, the measurement is that of the probability $\hat{P}_c$ that a population N_o^s will have one or more survivors. However $\hat{P}_c$ thus obtained can be related to $\hat{p}_s$ through

$$\hat{p}_s^{\,s} = 1 - (1 - \hat{P}_c)^{1/N_o^s} \tag{21}$$

Furthermore, if a sufficiently large number of samples is used, i.e. 50 or more, to yield some non-sterile samples at long heating times, p_s can be found to an accuracy of better than 1% from

$$\hat{p}_s^{\,s} \approx \frac{\hat{P}_c}{N_o^s}$$

Since either a counting or a sampling experiment can be used to obtain sections of the curve of $\hat{p}_s(t)$, the question arises as to which of these methods would be more accurate. It can be shown that if N_o^c, the population in the counting experiment, is the same as $N_o^s \cdot n_o$, the total population in the sampling experiment, the variance in $\hat{p}_s^{\,c}$ is essentially the same as the variance

in $\hat{p}_s^{\,c}$, i.e. the estimates of p_s would be of equal accuracy in both cases. This, however, is only valid if the initial populations $N_o^{\,c}$ and $N_o^{\,s}$ are both assumed to be known exactly. Since this is not generally true, the question as to which experimental method is to be preferred remains to be answered and should be examined including the uncertainties in the measurements of N_o.

Having established $p_s(t)$ in a functional form, various probabilities of organism survival for any initial population can be established. Although the desirability of defining planetary quarantine requirements in terms of a probability of exactly one survivor (p_N) can be questioned, there is little difference between such a requirement and the alternate specification for a probability of one or more survivors (P_c). In either case, provided the magnitudes of P_c or p_N are 10^{-2} or less, they can be obtained from

$$P_c \approx p_N \approx N_o p_s(t) \tag{31}$$

In general, $p_s(t)$ would be obtained experimentally over a range of time t much smaller than the time needed to achieve the desired p_N (or P_c). Thus, if $p_N = 10^{-3}$ and $N_o = 10^8$, $p_s(t) = 10^{-11}$. To measure this value of $p_s(t)$ with any kind of accuracy would, in the case of a counting test, require a population of 10^{12} and heating times which would reduce the population to 10 or more. Since, generally, population sizes in experimental determinations of $p_s(t)$ are much less than 10^{12}, it is essential that an appropriate functional relationship for $p_s(t)$ be available so as to minimize errors due the analytical extension of the curve to the range of unmeasured values.

References

1. Shuichi Aiba and Kiyoshi Toda, "An Analysis of Thermal Death Rate of Bacterial Spores". Progress Report No. 30, Biochemical Engineering Laboratory, Institute of Applied Microbiology, University of Tokyo, Tokyo, Japan, July 30, 1964.

2. Fredrickson, A.G., "Stochastic Models for Batch and Continuous Sterilization". To be published in J. Applied Microbiology.

3. Uspensky, J.V. Introduction to Mathematical Probability, McGraw-Hill, 1937.