ELSEVIER

Toxicology 183 (2003) 39–56

TOXICOLOGY

www.elsevier.com/locate/toxicol

Identification of an abnormal beryllium lymphocyte proliferation test

Edward L. Frome [a,*], Lee S. Newman [b], Donna L. Cragle [c], Shirley P. Colyer [c], Paul F. Wambach [d]

[a] *Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA*
[b] *National Jewish Medical and Research Center, Denver, CO, USA*
[c] *Center for Epidemiologic Research, Oak Ridge Institute for Science and Education, Oak Ridge, TN, USA*
[d] *U.S. Department of Energy, Germantown, MD, USA*

Received 14 May 2002; received in revised form 3 July 2002; accepted 22 July 2002

Abstract

The potential hazards from exposure to beryllium or beryllium compounds in the workplace were first reported in the 1930s. The tritiated thymidine beryllium lymphocyte proliferation test (BeLPT) is an in vitro blood test that is widely used to screen beryllium exposed workers in the nuclear industry for sensitivity to beryllium. The clinical significance of the BeLPT was described and a standard protocol was developed in the late 1980s. Cell proliferation is measured by the incorporation of tritiated thymidine into dividing cells on two culture dates and using three concentrations of beryllium sulfate. Results are expressed as a 'stimulation index' (SI) which is the ratio of the amount of tritiated thymidine (measured by beta counts) in the simulated cells divided by the counts for the unstimulated cells on the same culture day. Several statistical methods for use in the routine analysis of the BeLPT were proposed in the early 1990s. The least absolute values (LAV) method was recommended for routine analysis of the BeLPT. This report further evaluates the LAV method using new data, and proposes a new method for identification of an abnormal or borderline test. This new statistical–biological positive (SBP) method reflects the clinical judgment that: (i) at least two SIs show a 'positive' response to beryllium; and (ii) that the maximum of the six SIs must exceed a cut-point that is determined from a reference data set of normal individuals whose blood has been tested by the same method in the same serum. The new data is from the Y-12 National Security Complex in Oak Ridge (Y-12) and consists of 1080 workers and 33 non-exposed control BeLPTs (all tested in the same serum). Graphical results are presented to explain the statistical method, and the new SBP method is applied to the Y-12 group. The true positive rate and specificity of the new method were estimated to be 86% and 97%, respectively. An electronic notebook that is accessible via the Internet was used in this work and contains background information and details not included in the paper.

Abbreviations: SBP, statistical–biological positive; Be, beryllium; BeLPT, beryllium lymphocyte proliferation test; CBD, chronic beryllium disease; LAV, least absolute values; ORISE, Oak Ridge Institute for Science and Education; q–q, quantile–quantile; ROC, receiver operating characteristic; SI, stimulation index; SLsi, standardized Ln(SI); SE, standard error.

* Corresponding author. Tel.: +1-865-574-3138; fax: +1-865-241-0381
E-mail address: fromeel@ornl.gov (E.L. Frome).

0300-483X/02/$ - see front matter © 2002 Elsevier Science Ireland Ltd. All rights reserved.
PII: S0300-483X(02)00439-0

Keywords: Beryllium; Lymphocyte proliferation test; Least absolute values; Statistical methods; Electronic notebook

1. Introduction

The potential hazards of exposure to beryllium compounds were first reported in the 1930s. The clinical syndrome of chronic beryllium disease (CBD) was first described in 1946 (Hardy and Tabershaw, 1946). Initial speculation on the immunologic basis of CBD (Sterner and Eisenbud, 1951) occurred over 50 years ago. The first in vitro observation of beryllium-specific cell proliferation was demonstrated 20 years later (Hanifin et al., 1970). CBD mainly affects the lung, and occurs in a small percentage of persons exposed to beryllium dusts. Most clinicians rely on evidence of beryllium hypersensitivity as one of several criteria for diagnosis of the disease. In vitro proliferation of lung lymphocytes retrieved by bronchoalveolar lavage and exposed to beryllium is extremely sensitive to and specific for the diagnosis of CBD, but is not suitable for screening since cell retrieval is an invasive procedure (Stokes and Rossman, 1991). A less invasive procedure based on the proliferative response of blood cells to beryllium has been developed and is referred to as the blood beryllium lymphocyte proliferation test (BeLPT). Lymphocytes isolated from a lung lavage or peripheral blood of CBD patients show an in vitro response to beryllium stimulation that distinguishes CBD from other granulomatous lung diseases such as sarcoidosis (Newman, 1996a). The BeLPTwas first used as a clinical diagnostic tool in the 1980s, and the tritiated thymidine BeLPT as used today was further developed in the late 1980s. The historical development and significance of the tritiated thymidine BeLPT in identification of beryllium sensitization and CBD were reviewed at a conference on beryllium-related disease (Newman, 1996b). Epidemiologic and experimental work related to CBD have been reviewed (Kreiss et al., 1994), and it was noted that the laboratories that perform the BeLPT have formed a working group, the Committee to Accredit Beryllium Sensitivity Testing, to resolve technical problems related to the test. Beryllium has been used in the nuclear industry for a number of years. The epidemiology of CBD was examined in a stratified sample of workers at a nuclear weapons plant and the role of the BeLPT in beryllium disease surveillance in the nuclear industry was discussed (Kreiss et al., 1989). The U.S. Department of Energy is operating a screening program for CBD that will eventually include approximately 40 000 current and former beryllium-exposed workers at 20 Department of Energy sites. The use of beryllium in several new economic sectors further emphasizes the need for medical surveillance in the workplace for CBD (Kreiss et al., 1993).

Methods used in the calculation of BeLPT test results are important since it is the primary test that is used in large population screening and surveillance for CBD. The BePLT relies on replicate measurements for each calculated value (12 replicates for control values and four replicates for beryllium-simulated values). In response to concerns over the effect of 'outliers' in the BeLPT data, a new statistical approach was developed and presented at the 1994 conference (Frome et al., 1996). Two outlier-resistant methods were used to estimate the stimulation indices (SIs) and the coefficient of variation. A major advantage of these resistant methods is that they make it unnecessary to identify outlying data values among the replicate well counts. The statistical procedure that was recommended for routine analysis of the BeLPT uses the least absolute values (LAV) method on the Ln of the replicate counts. This new LAV method was developed on a small database from the lymphocyte proliferation test laboratory at Oak Ridge Institute for Science and Education (ORISE). It was presented to Committee to Accredit Beryllium Sensitivity Testing in April 1994, and further evaluation of the LAV method at another laboratory was requested. The LAV method was further 'field tested' using data from the lymphocyte proliferation test laboratory at the National Jewish Medical and Research Center. The results of this evaluation supported the earlier conclusion that the LAV

method is a simple and effective method for routine analysis of the BeLPT (Frome et al., 1997) and was included in the protocol for BeLPT testing that was being developed by the Committee to Accredit Beryllium Sensitivity Testing. In the standard protocol for the BeLPT a SI is calculated for each of three beryllium concentrations on two harvest days. Once SIs have been calculated it is necessary to determine if the results indicate an 'abnormal' response to beryllium. Three methods were considered in previous reports that utilize a 'cut-point' established using a reference data base of BeLPTs (Frome et al., 1994, 1996, 1997). The internal variability in each test was not taken into consideration by any of these methods. The new method proposed here combines clinical judgment and a statistical analysis that utilizes the internal variability to identify an abnormal test or an individual that needs a closer level of monitoring in the future. This new method described in Section 2.3 is referred to as the statistical–biological positive (SBP) method.

A set of BeLPTs performed as part of a study of workers at the Y-12 plant in Oak Ridge is used here to further evaluate the LAV method, and to describe and evaluate the new procedure for identifying an abnormal test. The purpose of the Y-12 study (Cragle and Tankersley, unpublished manuscript) was to examine the workplace characteristics of individuals sensitized to beryllium within a research cohort of 1151 current and retired workers enrolled in a medical examination program. Results presented here are limited to BeLPTs for workers and controls ($N = 1113$) that were done in the same lot of ABi Serum (3040083) as part of the Y-12 study, and only the first test in this serum is included if a worker had more than one test. If a worker's first test was considered abnormal or borderline and a repeat test was done in the same serum, only the first test was included. The serum was used in over 97% ($N = 1080$) of the BeLPTs in the Y-12 study, and 33 non-exposed (control) BeLPTs.

In July of 2000 the U.S. Department of Energy decided that a Specification for the BeLPT was needed to support the worker surveillance programs. A working group was established to write an initial draft version of the BeLPT Specification, and an electronic notebook (Geist and Nachtigal, 2000) was used to document the development, review, and revision of the BeLPT specification—see BeLPT-Notebook (Frome and Cragle, 2001; Frome et al., 2001a,b) for additional details. The final version of the specification was completed in April, 2001 (Wambach, 2001) and is available on page 21 of the BeLPT-Notebook. The data from the Y-12 study was used in this process and the protocol described here is consistent with the Department of Energy Specification.

2. Materials and methods

2.1. Beryllium lymphocyte proliferation test

The tritiated thymidine BeLPT has followed a standard protocol for laboratory procedure and data collection since the late 1980s. A detailed description of lymphocyte culture methods, quality control measures, and examples of plate maps and printouts of raw data in use at the ORISE BeLPT laboratory has been provided (Frome et al., 1996). This and several alternative assay designs are described in the BeLPT-Notebook—see pages 2–5. The details of the procedure and the equipment used vary at different laboratories and the essential requirements are described in DOE-SPEC-1142-2001 (Wambach, 2001). The ORISE protocol is briefly summarized as follows:

1) A 30 ml blood sample is obtained from each patient using sodium heparin as an anticoagulant. The mononuclear cells are separated using density gradient and centrifugation.
2) Lymphocytes are cultured using standard methods at a final concentration of 2.5×10^5 cells per well in 96-well flat-bottom microtiter plates. For each BeLPT assay 12 replicate control wells, and four replicates for each beryllium concentration (i.e. 1, 10, and 100 μM of $BeSO_4$) are set up in duplicate and harvested after 5 and 7 days of incubation, respectively. Mitogen-stimulated controls are also set up to test the cells ability to grow in culture.

3) Cells are incubated at 37 °C for 5 and 7 days and a pulse of tritiated thymidine is delivered prior to harvest. Cells are harvested on filter paper and counts are measured in a Packard Matrix 96 gas ionization counter. Each filter is counted for 10 min and the results organized as shown in Exhibit A1 in Appendix A.

2.2. Statistical analysis of the BeLPT

As the result of biological variability in the well counts there are different levels of uncertainty present in each BeLPT. This internal variability is described by the standard deviation of the Ln well counts, and is equivalent to the coefficient of variation on the original scale. The 'internal analysis' of the BeLPT is based on estimates of the Ln(SI)s and their standard errors (SE). These estimates are calculated using the LAV method (Frome et al., 1996). This approach only requires the ability to calculate medians and can be done in a spreadsheet (e.g. Excel, see page 14 of the BeLPT-Notebook) or statistical program (e.g. Splus or R). The LAV analysis is based on the assumptions that:

1) the Ln of the well counts follow the normal distribution;
2) standard deviations of Ln counts are constant within harvest days;
3) multiple outliers may be present in the Ln well counts; and
4) if 'responder cells' are present, an increase in cell proliferation relative to the control wells will occur in cultures with beryllium.

The detailed calculations for the LAV analysis are provided in Appendix A with an example. The steps in the analysis are summarized as follows:

1) Calculate the Ln of the well counts.
2) For each 'treatment group' calculate the median of the Ln counts.
3) For each beryllium concentration, calculate the Ln(SI) by subtracting the median of the control well Ln counts from the median of the Be stimulated Ln counts.
4) Calculate the SE of each Ln(SI).
5) Calculate the standardized Ln(SI): SLsi = Ln(SI)/SE[log(SI)].

The results of the LAV analysis for the Y-12 study are summarized graphically using histograms and normal quantile–quantile (q–q) plots (Chambers et al., 1983) of the SIs in original units and in logarithmic units. This is done to verify the assumption that the inter-test variability in SIs is reasonably described by the normal distribution on a logarithmic scale, i.e. a lognormal distribution.

2.3. Identification of abnormal BeLPTS using the SBP method

This new SBP method reflects the clinical judgement that: (i) at least two sets of beryllium stimulated wells should show a positive response; and (ii) the requirement that the maximum SI must exceed a cut-point that is determined from a reference data set of normal individuals. A BeLPT is considered abnormal if both the following statistical and biological criteria are satisfied:

1) Statistical analysis of the BeLPT indicates a positive response to beryllium. A positive response to beryllium occurs if *at least two SLsis are greater than* 2.53. This is referred to as a 'statistical positive' test and has a false positive probability of about 0.001 (see Appendix A for details).
2) The log of the maximum SI—Ln(SImax)—exceeds the 99.9th percentile of a reference data set, indicating a positive response to beryllium, with a false positive probability of 0.001. This is called a 'biological positive' test since it is based on the distribution of the Ln(SImax) values in the reference data set of normal tests. The Ln(SImax) values for the reference data set are assumed to follow the normal distribution with location parameter M and scale parameter S. The 99.9 percentile cut-point is estimated as $\mathrm{Cp}99.9 = M + 3.09S$, where M and S are estimated from the reference data set of normal BeLPTs. An equivalent evaluation is to calculate the metric $\mathrm{Zmax} = [\mathrm{Ln(SImax)} - M]/S$ and determine if

it is greater than 3.09, which is the 99.9 percentile of the standard normal distribution.

If only one of these criteria is met, and the data is otherwise acceptable, then the test is considered to be a 'borderline' test. If neither criteria is met the test is normal. If a patient's first test is not normal a second evaluation is requested, and two repeat BeLPTs are done in different laboratories or in the same laboratory using different sera. If at least two of the three BeLPTs are abnormal the patient is deemed beryllium sensitized. Since the criteria for a single abnormal BeLPT is based on an approximate false positive probability of 0.001, the chance of calling a person a 'sensitized responder' is very small (less than one in 10 000). This is based on the assumption that all non-exposed individuals will show a normal response to beryllium. In practice it is known that some individuals with no known exposure to beryllium will have an abnormal BeLPT. These are referred to as 'biological false positives' (see page 6 of the BeLPT-Notebook for a detailed discussion of a biological as opposed to statistical false positive test and related statistical issues). Consequently, the observed false positive rate will be higher than expected based on statistical considerations alone. A person may be a 'sensitized responder' and not have CBD. If a person is identified as sensitized, then further medical evaluation is available to determine if the worker has CBD (Stokes and Rossman, 1991).

2.4. Receiver operating characteristic (ROC) curve

The second criteria in Section 2.3 is based on results obtained using a single cut-point (c_p) for a biological positive test. The ROC curve is a graphical tool that has been developed to evaluate the accuracy of a diagnostic test when the test result is on a continuous scale, i.e. Ln(SI)s by considering all possible cut-points (Swets and Pickett, 1982; Stokes and Rossman, 1991; Zou and Zhou, 2001). A non-parametric estimate, (Lloyd, 1998) of the ROC curve is obtained by plotting the empirical proportion $\#L_{1i}\text{s} > c_p/n_1$ against $\#L_{0i}\text{s} > c_p/n_0$ for varying c_p. The L_{1i}s are the Ln(SI)s for those individuals that are 'cases' (i.e. they are sensitized and/or have CBD) and L_{0i}s are the Ln(SI)s for the normal individuals. The values on the vertical axis of the ROC curve are estimates of the true positive rate, and the horizontal axis values estimate the false positive rate for each cut-point. The case status of each worker was not known at the time the test was done and was established by following the group of Y-12 workers for 5 years as described in Section 3.4. In ROC analysis the area under the curve is considered as an overall 'index of accuracy' (Swets and Pickett, 1982, Chapter 1) of a test. The partial area under the curve (Pepe, 1998) is an alternate summary of accuracy. It has been argued that a false positive rate above some threshold would not be used in practice and, therefore, the ROC curve is of no interest beyond this point. If c_0 is the largest false positive rate of practical interest, then the partial area under the curve is the area under the ROC curve over the subinterval $(0, c_0)$. In the results $c_0 = 0.05$ is used to calculate a summary measure over a practically relevant range of operating points for the BeLPT. A consistent non-parametric estimation method of the partial area under the curve is used (Pepe, 1998).

3. Results

3.1. Graphical results for mitrogen control BeLPT data

In the ORISE protocol each BeLPT includes two sets of four wells for the mitogen controls—concanavalin-A and phytohemagglutinin. Graphical summaries of the mitogen control SIs for 1113 BeLPTs are shown in Figs. 1 and 2. In a normal q–q plot, if the relation between the empirical quantities (on the vertical axis) and theoretical quantiles (on the horizontal axis) is linear, this indicates that the data are described by a Gaussian (normal) distribution. Each figure displays the data in a histogram (left panel) and a normal q–q plot (right panel). The upper histogram and q–q plot show SIs after a natural log(Ln) transformation and the lower panels are untransformed SIs. In both figures the normal q–q plots for the Ln(SI)s (see top right panels in Figs. 1 and 2)

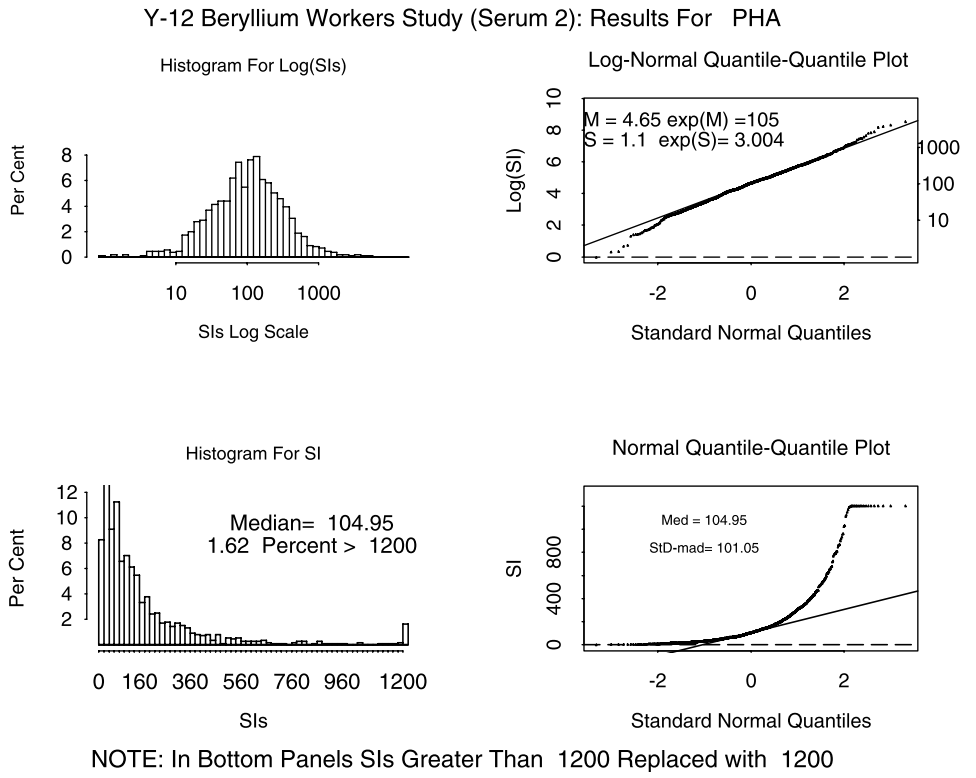

Fig. 1. Histograms amd normal q–q plots phytohemagglutinin (PHA) log and linear scale. The panels on the left show the histograms of the SIs. The top left is for Ln(SI)s and the bottom left is for the SIs. The panels on the right are normal q–q plots. If the data in the histogram (on the left) is normally distributed then the normal q–q plot (on the right) should look like a straight line. These plots clearly show that Ln(SI)s follow the normal distribution, i.e. the SIs follow the lognormal distribution.

strongly support the use of the lognormal distribution to describe the variation in the SIs when the agent is strongly mitotic. The only departure from the lognormal distribution is in the lower tail. This is due to the mitogen-stimulated cultures being well past the peak of their growth curve. If there is a strong mitotic response, and cell overgrowth occurs, the SI may be artificially low. If this occurs the wells have a distinct yellow appearance that indicates the presence of dead cells as the result of depletion of cell nutrient from the growth medium. In over 6000 tests the ORISE lymphocyte proliferation test laboratory has not encountered a single BeLPT in which the mitogen controls failed to show a response.

3.2. Graphical results for beryllium workers and non-exposed BeLPTs

Histograms for the SIs for each harvest day and Be concentration for the BeLPT data are shown in Fig. 3 (data from beryllium workers and non-exposed controls are combined). For the serum supplement used in this study SIs above three were abnormally high, indicating a response to beryllium. For plotting purposes SIs greater than four have been set equal to four. Fig. 4 shows the histograms for the Ln(SI)s for the same data. Comparing the histograms in Figs. 3 and 4 indicates that the SIs are best described by the normal distribution on the log scale. This is further

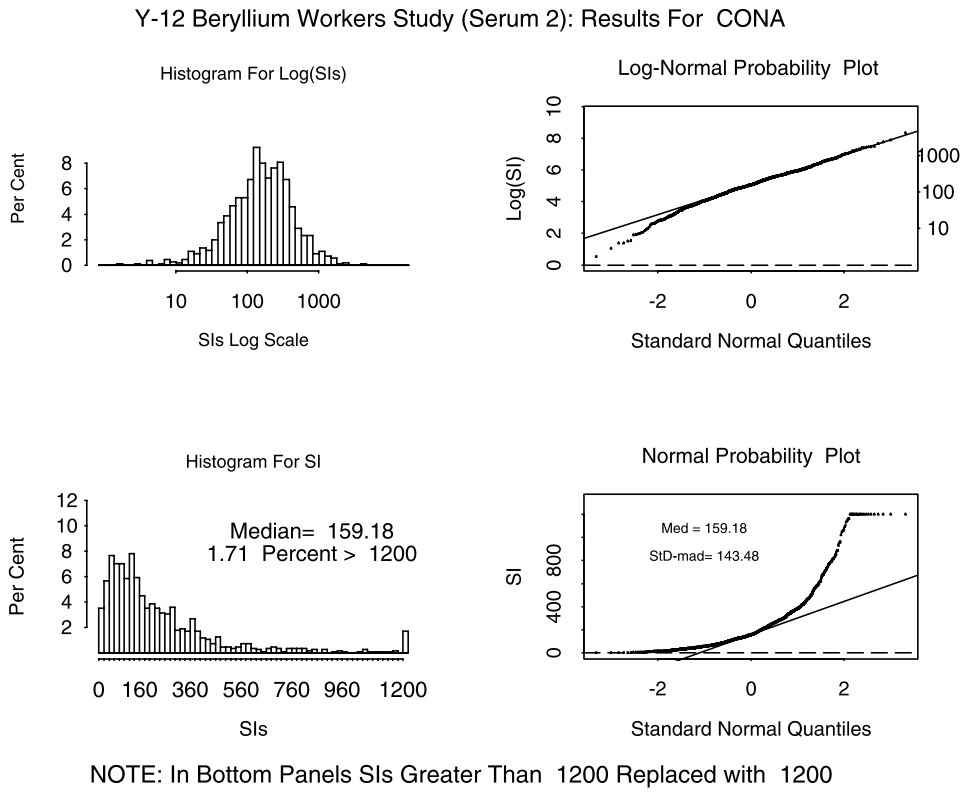

Fig. 2. Histograms and normal q–q plots concanavalin-A (CONA): log and linear scale. The panels on the left show the histograms of the SIs. The top left is for Ln(SI)s and the bottom left is for the SIs. The panels on the right are normal q–q plots. If the data in the histogram (on the left) is normally distributed then the normal q–q plot (on the right) should look like a straight line. These plots clearly show that Ln(SI)s follow the normal distribution, i.e. the Sis follow the lognormal distribution.

supported by Fig. 5 which shows lognormal probability plots for the beryllium workers and non-exposed control SIs for each of the three beryllium concentrations on days 5 and 7. In each of the six plots the data—ordered values of the Ln(SI)s—are shown on the vertical scale on the left, and the quintiles of the standard normal distribution are shown on the horizontal scale. Each plot includes the median (labeled M) and the median absolute deviation estimate of the standard deviation (labeled S) for the Ln(SI)s for the beryllium workers (shown as circles) and the non-exposed Ln(SI)s (shown as triangles). The lines in each plot (solid for non-exposed and dotted for beryllium workers) show the relation that is expected if the Ln(SI) values are from a normal distribution with location parameter M (which determines the intercept) and standard deviation S (which determines the slope).

Fig. 5 reflects the assumptions that most beryllium exposed workers do not show an abnormal response, i.e. they look like the non-exposed group. The relation between the empirical and theoretical quantiles is approximately linear in the center of the distribution indicating that the distribution is Gaussian. For example, consider the plot for day 5 Be-1 in Fig. 5. The Ln(SI)s appear to be approximately normal in the center, for both the non-exposed controls and the beryllium workers. There are several values that are larger than expected (these are the points above the lines). These 'outliers' are SIs that indicates hypersensitivity to beryllium. There are also several points below the line which indicate cell

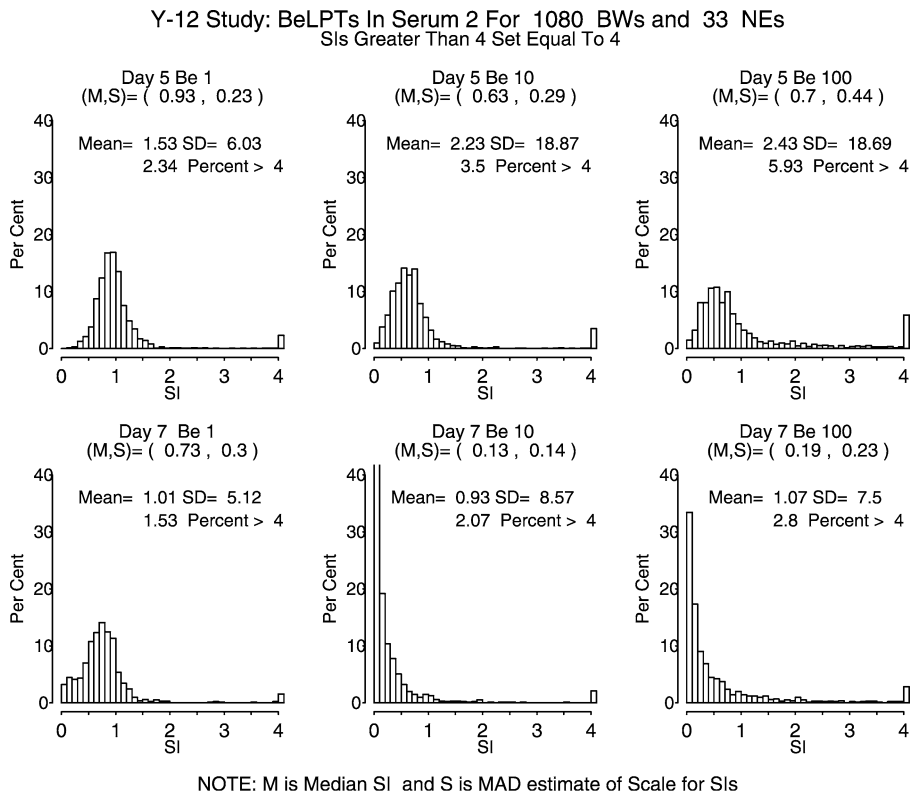

Fig. 3. Histograms of the SIs for the beryllium workers and non-exposed BeLPTs. Numbers in parenthesis are the outlier resistant median (M) estimate of location and S the median absolute deviation estimate of the scale parameter. The mean and standard deviation (SD) for each distribution are also given.

killing. The effect of outliers on these estimates has been minimized since resistant methods were used to estimate the location and scale parameters, M and S, respectively. The results in Fig. 5 are similar to plots for previous ORISE data (Frome et al., 1996), and similar BeLPT test results from the National Jewish Medical and Research Center (Frome et al., 1997, Figs. 5 and 6).

3.3. Identification of abnormal BeLPTs using the SBP method

The SBP method described in Section 2.3 was used to evaluate each BeLPT. The first step was to calculate the SLsi for each beryllium concentration on days 5 and 7 (see Appendix A for an example). If at least two of the SLsis are greater than 2.53 then the test is a statistical positive. The example in Appendix A has two SLsis greater than 2.53 so it is considered a statistical positive test. The second step requires estimates of the location and scale parameters for the reference data set. *The BeLPTs from the non-exposed controls were used as the reference data set*. The Ln(SImax) values for the non-exposed controls and beryllium workers are shown in box plots (left panel) and normal q–q plots (right panel) of Fig. 7. A detailed example and explanation of boxplots and q–q plots is provided in the BeLPT-Notebook (see 'click here for details' in Item 1 on page 7). The Ln(SImax) for the non-exposed controls appear linear and the Kolmogorov–Smirnov test indicates that lognormal distribution cannot be rejected. The q–q plot for the beryllium workers shows that the most of the test results are described by the same lognormal model, but there are a number of tests that have Ln(SImax) values that are either too large (positive test) or too small (as the result of cell

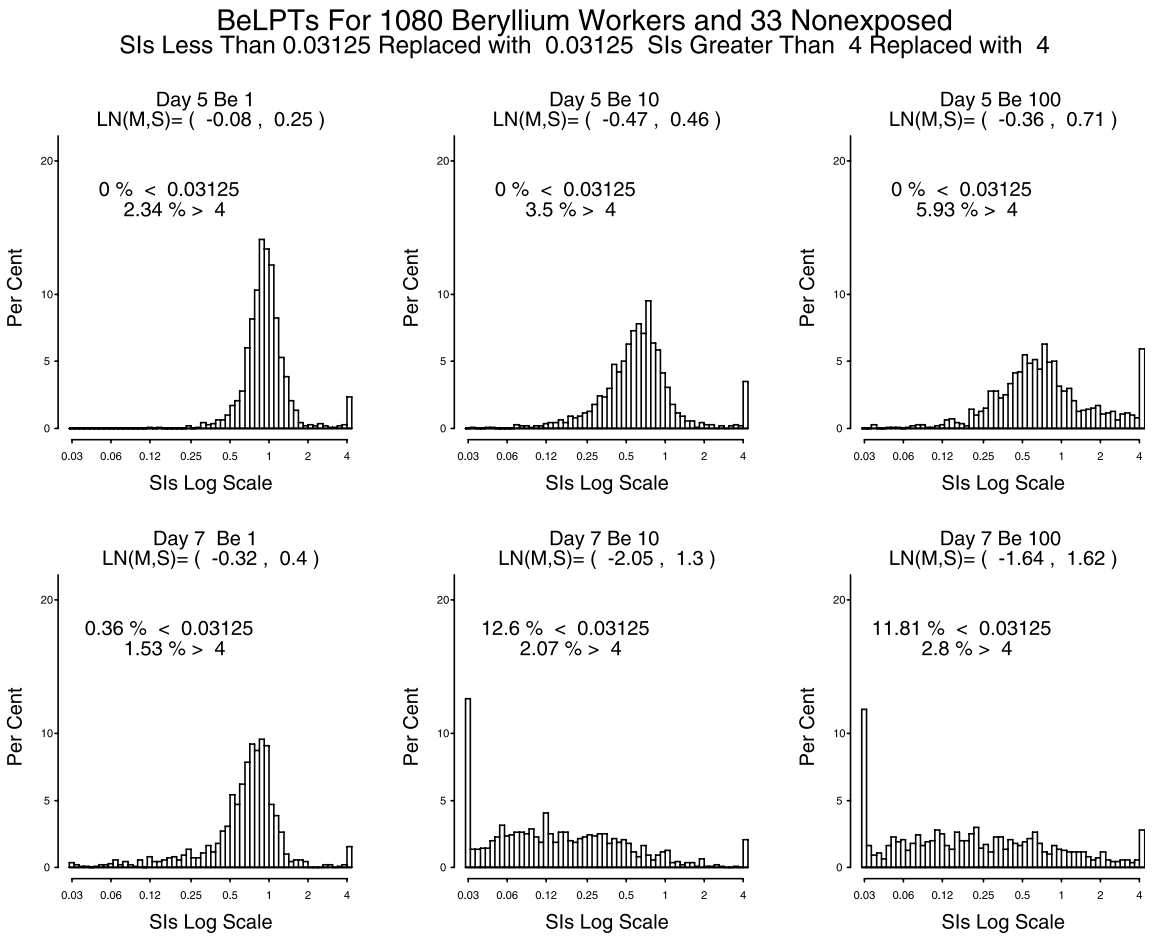

Fig. 4. Histograms of the Ln(SI)s for the beryllium workers and non-exposed BeLPTs. The outlier resistant estimates on the Ln scale of location M (the median) and S the median absolute deviation estimate of the scale parameter for each distribution are given in parenthesis.

killing). This is further supported by the fact that the outliner resistant estimates of the lognormal scale (M) and location (S) parameters for the non-exposed data are almost identical to those for the beryllium workers data. The estimate of M from the reference data set is 0.0812 and the estimate of S is 0.34. A biological positive test occurs (see criterion 2 in Section 2.3) if $Zmax = [Ln(SImax) - M]/S$ is greater than 3.09. For the example in Appendix A $Zmax = [0.98 - 0.0812]/0.34 = 2.64$, indicating that this is not a biological positive test. Consequently, the example is considered a 'borderline' test, and two additional BeLPTs were obtained for this worker. Both of these were abnormal so the worker is considered sensitized to beryllium as described in Section 2.3.

3.4. Identification of cases and ROC curve analysis

All of the BeLPTs in the Y-12 group were done before July, 1996, and all of the workers with a positive test and most of the 944 workers with an initial normal test were followed and retested over the next 5 years. The results for the first test are shown in column 2 of Table 1 and the follow-up results for each worker are shown in Columns 3–7

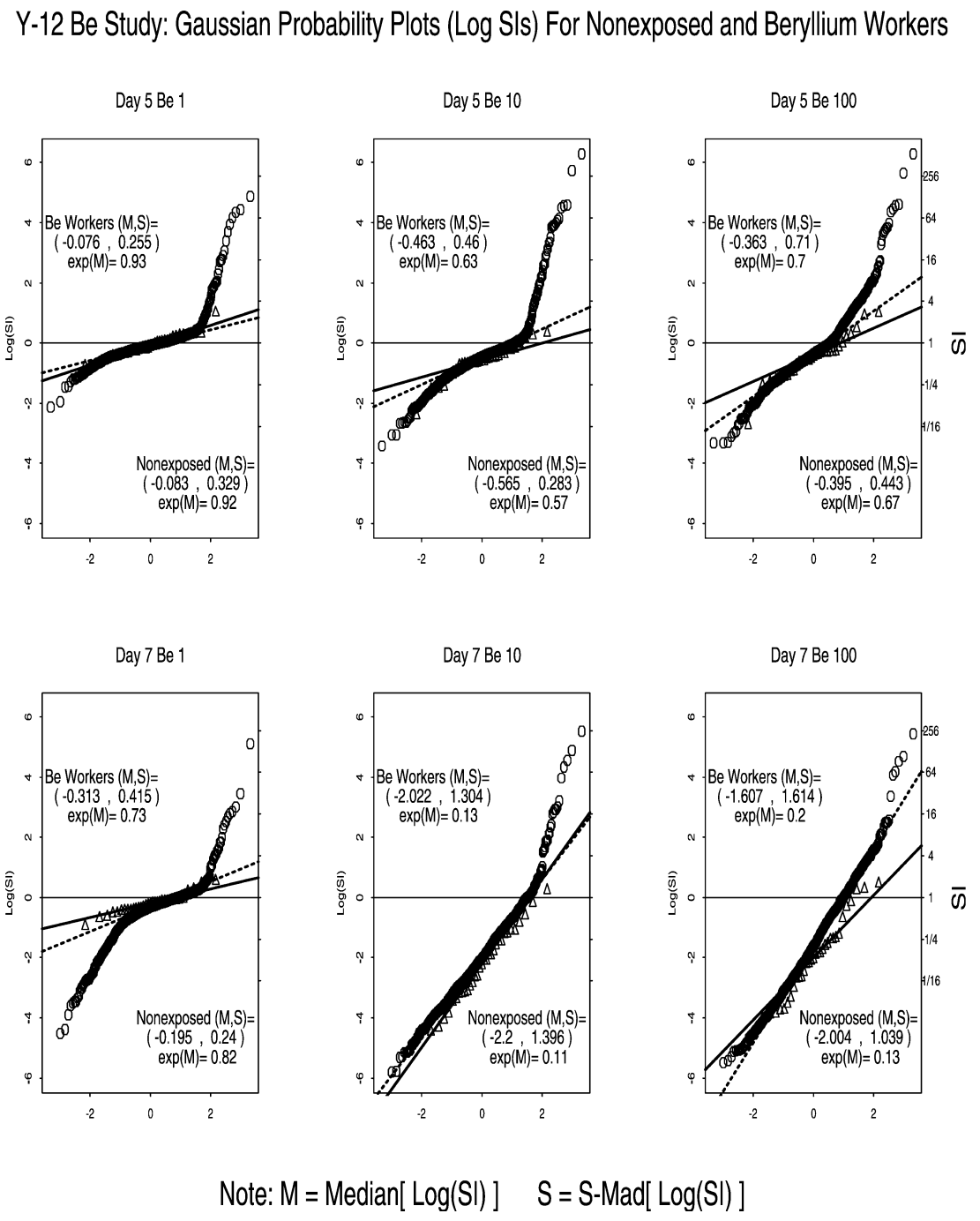

Fig. 5. Normal q–q plots of Ln(SI)s for beryllium concentrations on days 5 and 7 for beryllium workers and non-exposed controls. The data values are shown on the vertical axis. The median (M), median absolute deviation scale estimate (S) of the Ln(SI)s and exp(M) are listed on each plot. Values of M and S for beryllium workers (circles) are in upper left and non-exposed controls (triangles) are in lower right of each panel.

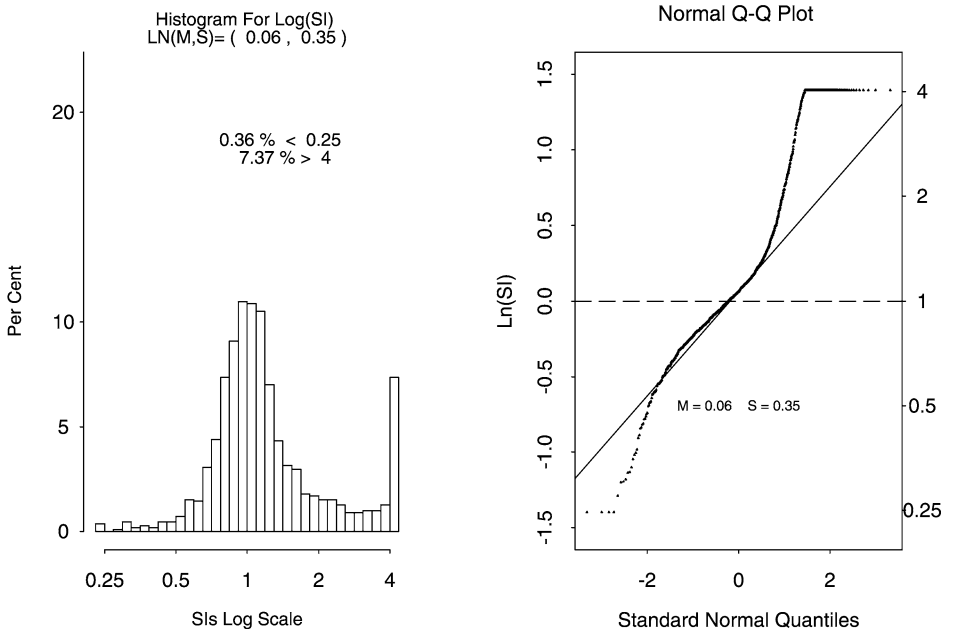

Fig. 6. Histogram and normal q–q plots for Ln(SImax) for beryllium workers and non-exposed combined. The median (M), and median absolute deviation scale estimate (S) of the Ln(SI)s are shown.

of Table 1. A total of 132 BeLPTs had an initial positive test by at least one of the criteria in Section 2.3. There were 80 BeLPTs that were abnormal, 38 tests with Zmax greater than 3.09 (biological positive only), 16 tests with at least two SLsis greater than 2.53 (statistical positive only), and 948 normal tests. These groups are identified in the first column of Table 1. *The classification of individuals in the columns* 3–7 *of* Table 1 *was based on the criteria being used by the ORISE LPT laboratory at the time the tests were done* (not the criteria in Section 2.3). A worker was classified as sensitized if an initial test was repeated twice and at least two of the three results were abnormal. A BeLPT was abnormal if at least two SIs exceeded a cut-point of 2.42. This cut-point was calculated using the SImax for each BeLPT in the REFERENCE DATA SET and is equal to the mean+2(standard deviation). The mean SImax was 1.27 and the standard deviation was 0.576. A test was borderline if only one SI exceeded the cut-point, and the data was otherwise acceptable. If only one BeLPT was done the follow-up status is unknown. If a worker was identified as sensitized, then further medical evaluation was available. If a sensitized worker was evaluated clinically and diagnosed with CBD they are in column 7 of Table 1, otherwise they are in column 6. If a sensitized worker did not have a clinical evaluation their CBD status is not known and they are included in column 6. If a worker was neither abnormal nor normal they are considered borderline and further monitoring is indicated. A worker would be in this classification if, for example, they had an initial abnormal BeLPT, and split tests were borderline and normal.

The results of the SBP method summarized in Table 1 can be used to estimate the true and false positive rates for *a first abnormal BeLPT in a specific serum*. The results in Table 1 were further summarized by assuming that: (i) individuals follow-up status reflects their condition at the

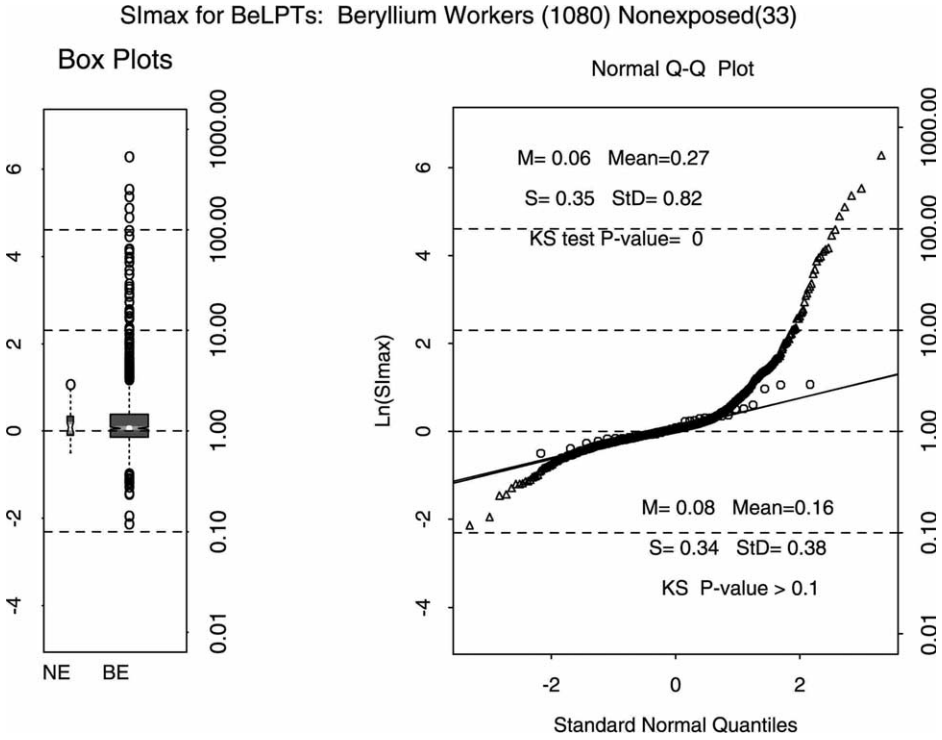

Fig. 7. Boxplots (left panel) and normal q–q plots (right panel) for Ln(SImax). In the right panel summary statistics for non-exposed controls (circles) are shown in lower right, and for beryllium workers (triangles) in upper left of q–q plot. A small P value for Kolmogorov–Smirnov (KS) goodness-of-fit test indicates departure from normal distribution for Ln(SImax).

time the first test was done; (ii) individuals with unknown status were normal (these are mostly retired workers with a normal first test that are asymptomatic); (iii) individuals that have CBD are sensitized; and (iv) individuals that were not sensitized to beryllium are normal. The true positive rate of the first BeLPT in Serum 3040083 is 48/56 or 85.7% 1- and the specificity (false positive rate) is 992/1024 or 96.9%. The ORISE LPT laboratory identified abnormal BeLPTs using the methods and criteria in place at the time that each test was done. Using the

Table 1
Summary follow-up data for Y-12 group

Group	Initial results[a]	Follow-up results				
		B[b]	N	UN	SEN	CBD[c]
Abnormal test	80	7	21	4	27	21
Biological positive	36	6	22	6	1	1
Statistical positive	16	0	10	4	2	0
Normal	948	6	629	309	3	1
Total	1080	19	682	323	33	23

[a] Results of SBP method for first test in serum 3040083.
[b] B, borderline; N, normal; UN, unknown; SEN, sensitized.
[c] CBD see Section 2.3 for explanation.

information from the ORISE historical data base the true positive rate was 78.6% and the specificity was 98.3%. If individuals with unknown status (see ii above) are not included in the calculations, then the specificity for the SBP method is 96.0%, and the specificity for ORISE historical method is 92.7%.

Table 1 is based on results obtained using a single cut-point (c_p) for a biological positive test as described in Section 2.3. The ROC curves for each harvest day and beryllium concentration are shown in Fig. 8. The area under the curve and the partial area under the curve over the interval (0, 0.05) are given for each curve (see Section 2.4).

4. Discussion

The graphical results in Figs. 1–7 of Section 3 provide empirical evidence that the assumptions described in Section 2.2 are reasonable. The results in Table 1 indicate that the SBP method, using the LAV approach to estimate the SIs, is at least as good as current methods for evaluating the

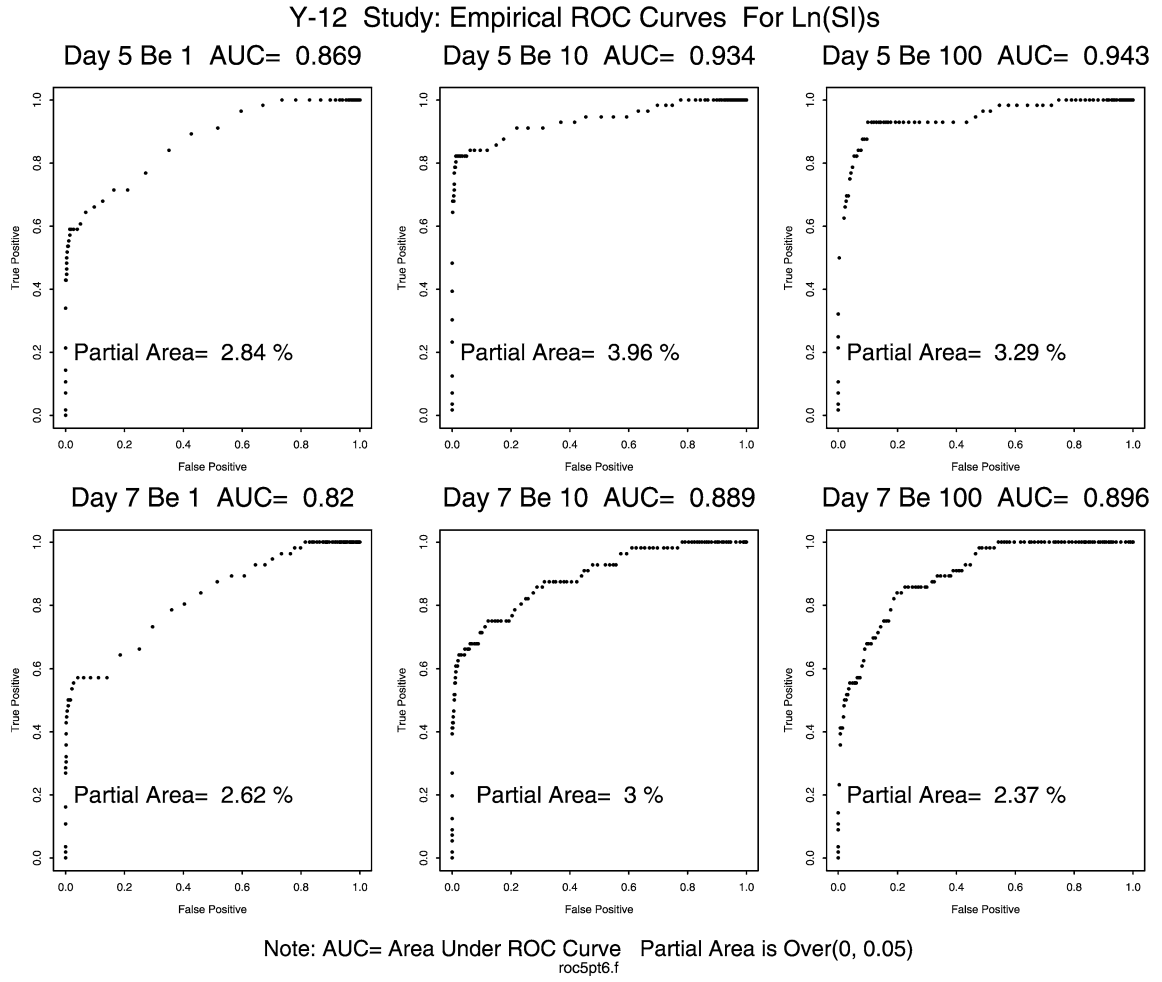

Fig. 8. Empirical ROC curves for Ln(SI)s for each beryllium concentration on days 5 and 7. AUC is the area under the curve. The partial AUC shown in each plot is based on a non-parametric estimate of the area under the ROC curve from 0 to 0.05 on the x-axis (i.e. maximum false positive rate of practical interest is 0.05).

BeLPT. The 'outlier rejection method' that is used by some laboratories has no logical statistical basis (see page 16 of the BeLPT-Notebook for further discussion). Further evaluation of the SBP approach is currently underway using results from ORISE obtained in several different sera after 1996, and using data from at least two additional laboratories. The results of this work and any additional information related to the tritiated thymidine BeLPT will be added to the BeLPT-Notebook.

The ROC analysis in Fig. 8 indicates that results on day 5 are generally more accurate than day 7 and that the 10 μM $BeSO_4$ challenge provides the best results on both days. A possible verification bias occurs since all workers with a normal first test do not receive additional tests during follow-up. This problem primarily occurs in the group of retired workers with a normal first test that are asymptomatic. Active workers in the beryllium surveillance program were retested on a regular basis. For the ROC analysis it was assumed that retired workers that were asymptomatic would be normal in subsequent testing. The 'gold standard' used to identify 'cases' is therefore imperfect, since a worker is considered 'sensitized' to beryllium if they have at least two abnormal BeLPTs, i.e. clinical verification of CBD status is optional (Zou and Zhou, 2001).

A new method for the measurement of lymphocyte proliferation (the Immuno-BeLPT) that uses a flow cytometric assay has been developed (Farris et al., 2000). The Immuno-BeLPT provides additional information about the type of lymphocytes (CD4+ and CD8+ T cells) that are responding. Genetic testing has shown that certain allelic variations of the HLA-DPB1 gene occur in individuals having beryllium hypersensitivity, but no symptoms of CBD. These results (Wang et al., 2001) suggest that the combination of the Immuno-BeLPT and HLA genotyping may be useful in diagnosing CBD, and in the evaluation of the risk of developing CBD for sensitized individuals without disease. This proposed association between Immuno-BeLPT results and different HLA-DPB1 genotypes and the risk for the development of CBD is currently being evaluated.

Acknowledgements

This research was support by the Offices of Occupational Medicine, U.S. Department of Energy, Environment, Safety and Health Studies contract DE-AC05-00OR22725 with UT-Battelle, LLC. The authors thank ORISE, Center for Epidemiologic Research for help with data collection and Carole Holbrook, Oak Ridge National Laboratory, Computer Science and Mathematics Division for assistance in preparing this report. The work has been authored by a contractor of the U.S. Government. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this work, or to allow others to do so for U.S. Government purposes.

Appendix A: Least absolute values analysis for BeLPT

The main results required for the LAV method (Frome et al., 1996) are summarized here. Let y_{jk} denote the well count (see Exhibit A1) for the kth replicate of the jth set of culture conditions (see Column 1, Exhibit A1). The data in Exhibit A1 are the raw counts for worker ID-271 and are used here to demonstrate the calculations.

Treatment	Well counts			
Day 5 controls	1220	2391	1774	947
Day 5 controls	1499	1568	1410	1131
Day 5 controls	969	2265	1743	728
Day 5 Be1	1777	1890	1702	1885
Day 5 Bel0	3368	7221	1473	3097
Day 5 Bel00	3631	3655	2452	1634
Day 7 controls	3616	17 410	3989	3144
Day 7 controls	669	1257	1497	4460
Day 7 controls	2897	4174	1366	1152
Day 7 Be1	1670	2186	629	1264
Day 7 Be10	330	598	254	264
Day 7 Be100	3611	4436	14 452	14 892
PHA	102 160	44 223	59 344	51 088
CONA	115 673	104 146	252 237	159 421

Exhibit A1. Well counts for BeLPT assay 271

The expected count in each well can be represented by a log–linear regression function:

$$E(y_{jk}) = \lambda_j = \exp(\boldsymbol{X}_j\boldsymbol{\beta}), \quad (1)$$

where $j = 1, \ldots, 10$ and $k = 1, \ldots, 12$ for the controls and $k = 1, 2, 3, 4$ for the beryllium stimulated cells and the positive controls (see column 1 of Exhibit A1). In Eq. (1), $\boldsymbol{X}_j$ is a row vector of indicator variables and $\boldsymbol{\beta}$ is the vector of regression parameters (see below). It is further assumed that the variance of the well counts is proportional to the square of the expected count, i.e. the standard deviation is proportional to the mean:

$$\mathrm{Var}(y_{jk}) = (\phi\lambda_j)^2. \quad (2)$$

Eqs. (1) and (2) together are referred to as a generalized linear model with constant coefficient of variation ϕ (McCullagh and Nelder, 1989).

(1) *The first step in the LAV analysis is to take the Ln of the counts in Exhibit A*1 (see Columns 2–4 of Exhibit A2). This is the variance-stabilizing transformation and leads to a linear model in say $z_{jk} = \ln(y_{jk})$ with $\mathrm{Var}(z_{jk}) \simeq \phi^2$. The Ln of the counts are shown in columns 2–5 of Exhibit A2. If outliers are not present, applying ordinary least-squares to the transformed data will yield consistent estimates for the Ln(SI) parameters (McCullagh and Nelder, 1989). The effect of outliers is minimized by using LAV (or some other robust method) on z_{jk}.

(2) *The second step is to calculate the median of the Ln counts for each Treatment Group*. The median of a data set is the middle value when all the data values are put in ranked order. When the number of data values is even, the median is the average of the middle two values. For example, for day 5, Be100 (Exhibit A2, Line 6) the median is $(7.8047 + 8.1973)/2 = 8.001$. Let $\tilde{z}_j$ denote the median for the jth beryllium concentration and $\hat{z}_o$, denote the median of the Ln well counts for the corresponding control wells (see Column 6 of Exhibit A2).

(3) *Step three is to calculate the LAV estimate of the jth Ln(SI)*, $\tilde{\boldsymbol{\beta}}_j = \tilde{z}_j - \tilde{z}_o$. For example, on day 5 Be100 the $\mathrm{Ln(SI)} = 8.0010 - 7.2819 = 0.7191$, and the estimate of the SI is $\exp(0.7191) = 2.05$.

Treatment group	Ln(well counts)				Median
Day 5 controls	7.1066	7.7795	7.4810	6.8533	7.2819
Day 5 controls	7.3126	7.3576	7.2513	7.0309	7.2819
Day 5 controls	6.8763	7.7253	7.4634	6.5903	7.2819
Day 5 Be1	7.4827	7.5443	7.4396	7.5417	7.5122
Day 5 Be10	8.1221	8.8847	7.2951	8.0382	8.0801
Day 5 Be100	8.1973	8.2039	7.8047	7.3988	8.0010
Day 7 control	8.1931	9.7648	8.2913	8.0533	8.0123
Day 7 control	6.5058	7.1368	7.3112	8.4029	8.0123
Day 7 control	7.9714	8.3366	7.2196	7.0493	8.0123
Day 7 Be1	7.4206	7.6898	6.4441	7.1420	7.2813
Day 7 Be10	5.7991	6.3936	5.5373	5.5759	5.6875
Day 7 Be100	8.1917	8.3975	9.5789	9.6086	8.9880
PHA	11.5343	10.6970	10.9911	10.8413	10.96162
CONA	11.6585	11.5535	12.4381	11.9793	11.8189

Exhibit A2. Ln of well counts for BeLPT assay ORISE ID = 271

(4) *The fourth step is to calculate the SE of each Ln(SI)*. This requires an estimate of ϕ the standard deviation of the Ln counts (corresponds to coefficient of variation on original scale). An outlier resistant estimate (Frome et al., 1996) 'phitilde' of ϕ is $\tilde{\phi} = 1.48 \times \sqrt{n/(n-p)} \times \mathrm{median}\{|z_{jk} - \tilde{z}_j|\}$. Estimates of ϕ are calculated for days 5 and 7, since it has been observed that there is generally more variability on day 7. The estimated standard error of Ln(SI) is $\mathrm{SE} = \tilde{\phi}\sqrt{\mathrm{diag}(V)}$, where $V = (\pi/2)(X'X)^{-1}$. Using day 5 as an example, the residuals $u_{jk} = z_{jk} - \tilde{z}_j$ are shown in Exhibit A3, and $\tilde{\phi} = 1.48 \times \sqrt{24/20} \times (0.1963) = 0.3183$ where median $|z_{jk} - \tilde{z}_j| = 0.1963$. The SE of the day 5 Ln(SI)'s is $\tilde{\phi}\sqrt{\mathrm{diag}(V)} = 0.3183 \times \sqrt{1.571 \times 0.25} = 0.230$.

(5) *Divide the Ln(SI) by its SE to obtain the standardized Ln(SI)*. Dividing each Ln(SI) by its SE results in a statistic that is in 'standard measure', having mean 0 and standard deviation 1 (Kotz and Johnson, 1988). The LAV estimate of the standardized Ln(SI) for day 5 Be100 is $\mathrm{SLsi} = \mathrm{Ln(SI)/SE[Ln(SI)]} = 0.7197/0.230 = 3.13$. If outliers are not present, and the z_{jks} are normally distributed, then the least-squares estimate of the jth Ln(SI) is $\hat{\boldsymbol{\beta}}_j = \hat{z}_j - \hat{z}_o$, i.e. the treatment effects on the Ln scale. Here $\hat{z}_j$ denotes the mean for the jth beryllium concentration and $\hat{z}_o$ denotes the mean of the Ln well counts for the corresponding

control wells, i.e. these are the 'least-squares' estimates of location and scale. Under the null hypothesis of no treatment effect, H_0:$\boldsymbol{\beta}_j = 0$ dividing the Ln(SI) by its SE results in a standardized Ln(SI) that will follow Student's t-distribution with $n-p$ degrees of freedom. The LAV estimates are asymptotically Gaussian with covariance matrix ϕV (Basset and Koenker, 1978). In large samples the outlier SLsi will follow the standard normal distribution. The small sample distribution of the outlier resistant SLsi is not know. For this application the appropriate t-distribution (with degrees of freedom = 20) is used as a reference distribution to identify a large value for this test statistic. The critical value or 'cut-point' of 2.53 is selected so that under H_0 the probability that the SLsi exceeds 2.53 is approximately 1%. For a BeLPT assay to be called 'abnormal' *at least two SLsis must exceed this value*, so the statistical false positive probability that a normal test is called abnormal is about 0.1%.

	$u[j,k]$			
Day 5 controls	−0.1753	0.4976	0.1991	−0.4286
Day 5 controls	0.0307	0.0757	−0.0306	−0.2510
Day 5 controls	−0.4056	0.4434	0.1815	−0.6916
Day 5 Be1	−0.0295	0.0321	−0.0726	0.0295
Day Be10	0.0420	0.8046	−0.7850	−0.0419
Day 5 Be100	0.1963	0.2029	−0.1963	−0.6022

Exhibit A3. Residuals from day 5 in Exhibit A2

The results of all the calculations and additional statistics are combined into single laboratory LAV report (see Exhibit A4). Note that 'coefficient of variation-MAD' (this is the median absolute deviation estimate $\tilde{\phi}$ for each treatment group) and the residuals in the Panel I of Exhibit A4 have been multiplied by 100, i.e. they are in Log percent units (L%) (Tornquist et al., 1985). For example, the first residual for the day 5 control wells is $100 \times \text{Ln}(1220/1453.8) = -18L\%$. Panel II of Exhibit A4 lists the LAV estimates of the SIs, Ln(SI)s and Slsis for each beryllium concentration on days 5 and 7 and the positive controls. The estimates of $\tilde{\phi}$ for the control wells and treated wells for days 5 and 7 with and without 'pooling' are provided in the Panel III of Exhibit A4. An 'overall' pooled estimate is also provided. These values are used to calculate the SE of the [Ln(SI)] and to evaluate the amount of variation within control and treated groups on days 5 and 7.

The LAV report is used for quality control (e.g. to identify unacceptable tests) and to help in the interpretation of acceptable BeLPTs that are not confirmed abnormals or normals. The BeLPT in Exhibit A4 has two large SLsis (D5be10 and D5be100) indicating a statistical positive test (see Item 1, Section 2.3). The standardized Ln(SImax) for this test, (see Section 2.3, Item 2) is $\text{Zmax} = (0.98-M)/S = (0.98-0.081)/0.34 = 2.64$. This is below the cut-point of 3.1 (see Section 2.3) for the reference data set. The values of $\tilde{\phi}$ are high on day 7 and the SLsi for D7be10 is −3.98, indicating cell killing in at least one well. Based on the criteria in Section 2.3 this is a borderline test and was also interpreted as borderline by the ORISE LPT laboratory using the criteria described in Section

I. BeLPT analysis using LAV ID = 0271

Treatment GRP	Well counts				FIT*	CV-MAD	Residuals			
Day 5 controls	1220	2391	1774	947	1453.8	34.9	−18	50	20	−43
Day 5 controls	1499	1568	1410	1131	1453.8	34.9	3	8	−3	−25
Day 5 controls	969	2265	1743	728	1453.8	34.9	−41	44	18	−69
Day 5 Be1	1777	1890	1702	1885	1830.2	5.3	−3	3	−7	3
Day 5 Be10	3368	7221	1473	3097	3229.7	70.8	4	80	−79	−4
Day 5 Be100	3631	3655	2452	1634	2983.8	34.2	20	20	−20	−60
Day 7 controls	3616	17410	3989	3144	3018.0	84.5	18	175	28	4
Day 7 controls	669	1257	1497	4460	3018.0	84.5	−151	−88	−70	39

Table (*Continued*)

Treatment GRP	Well counts				FIT*	CV-MAD	Residuals			
Day 7 controls	2897	4174	1366	1152	3018.0	84.5	−4	35	−79	−96
Day 7 Be1	1670	2186	629	1264	1452.9	46.9	14	41	−84	−14
Day 7 Be10	330	598	254	264	295.2	22.4	11	71	−15	−11
Day 7 Be100	3611	4436	14 452	14 892	8006.8	103.7	−80	−59	−59	62
PHA	102 156	44 221	59 346	51 090	55063.5	25.2	62	−22	7	−7
CONA	115 673	104 146	252 237	159 421	135796.6	36.4	−16	−27	62	16

*FIT, Fitted Value−exp[median(log z)] for each treatment group. Coefficient of variation-mad (CV−MAD) and residuals are in L% units.

3.4. Subsequent testing found this patient to be sensitized.

II. Stimulation indices (SI)

	Day 5			Day 7			Positive PHA	Controls CONA
	D5be1	D5be10	D5be100	D7be1	D7be10	D7be100		
SI	1.26	2.22	2.05	0.48	0.10	2.65	37.88	93.41
Log SI	0.26	0.80	0.72	−0.73	−2.32	0.98	3.63	4.54
Slsi	1.00	3.48	3.13	−1.25	−3.98	1.67	15.83	19.76

SLsi = Log SI/SE is standardized Log SI.
Large POSITIVE values (GT 2.5) indicate POSITIVE response.
Large NEGATIVE values (LT −2.5) indicate cell killing.

*III. Summary statistics Phitlide (coefficient of variation)**

Overall	0.385				
Day 5 control	0.349	Day 5 treated	0.23	Day 5 pooled	0.319
Day 7 control	0.845	Day 7 treated	0.855	Day 7 pooled	0.811

*Phitilde is median absolute deviation estimate of the standard deviation on log scale (corresponds to coefficient of variation on original scale).

Exhibit A4. LAV report for BeLPT assay 271

References

Basset, G., Jr., Koenker, R., 1978. Asymptotic theory of least absolute error regression. Journal of the American Statistical Association 73, 618–622.

Chambers, J.M., Cleveland, W.S., Kleiner, B., Tukey, P.A., 1983. Graphical Methods for Data Analysis. Duxbury Press, Boston.

Farris, G.M., Newman, L.S., Frome, E.L., Shou, Y., Barker, E., Habbersett, R.C., Mair, L., Smith, H.N., Marrone, B.L., et al., 2000. Detection of beryllium sensitivity using a flow cytometric lymphocyte proliferation test: the IMMUNO-BE-LPT. Toxicology 143, 125–140.

Frome, E.L., Cragle, D.L., 2001. BeLPT-Notebook: Notes Concerning the Tritiated Thymidine Beryllium Lymphocyte Proliferation Test at Oak Ridge National Laboratory, Oak Ridge, TN (http://www.csm.ornl.gov/ ~ frome/BE/tno-te.html).

Frome, E.L., Smith, M.H., Littlefield, L.G., Neubert, R.L., Colyer, S.P., 1996. Statistical methods for the blood beryllium lymphocyte proliferation test. Environmental Health Perspectives Supplement 104, 957–968.

Frome, E.L., Cragle, D.L., Colyer, S.P., Wambach, P.F., 2001a. Development and evaluation of a screening test for beryllium sensitization. Proceedings of the Statistics in Epidemiology Section, American Statistical Association, Alexandria, VA.

Frome, E.L., Cragle, D.L., Littlefield, L.G., Colyer, S.P., 2001b. Results of the Analysis of the Blood Beryllium Lymphocyte Proliferation Test Data from the Oak Ridge Y-12 Study. Oak Ridge National Laboratory, Technical Report ORNL/TM-2001/79.

Frome, E.L., Newman, L.S., Morz, M.M., 1997. Results of the Analysis of the Blood Lymphocyte Proliferation Test Data from the National Jewish Center. Oak Ridge National Laboratory, Technical Report ORNL/TM-13338.

Frome, E.L., Smith, M.H., Littlefield, L.G., Neubert, R.L., Colyer, S.P., 1994. Statistical Methods for the Analysis of a Screening Test for Chronic Beryllium Disease. Oak Ridge National Laboratory, ORNL/TM-6818.

Geist, A., Nachtigal, N., 2000. Oak Ridge National Laboratory Electronic Notebook Project at Oak Ridge National Laboratory, Oak Ridge, TN (http://www.csm.ornl.gov/ ~ geist/java/applets/enote/).

Hanifin, J.M., Epstein, W.L., Cline, M.J., 1970. In vitro studies of granulomatous hypersensitivity to beryllium. Journal of Investigative Dermatology 55, 284–288.

Hardy, H.L., Tabershaw, J.R., 1946. Delayed chemical pneumonitis occurring in workers exposed to beryllium compounds. Indust. Hygine Toxicol. 28, 197–211.

Kotz, S., Johnson, N.L., 1988. Encyclopedia of Statistical Science, vol. 8. John Wiley & Sons, New York, NY, p. 631.

Kreiss, K., Miller, F., Newman, L.S., Ojo-Amaize, A., Rossman, M., Saltini, C., 1994. Chronic beryllium disease—from the workplace to cellular immunology, molecular

immunogenetics, and back. Clinical Immunology and Immunopathology 71, 123–129.

Kreiss, K., Mroz, M.M., Zhen, B., Marrtny, J.W., Newman, L.S., 1993. Epidemiology of beryllium sensitization and disease in nuclear workers. American Review of Respiratory Disease 148, 985–991.

Kreiss, K., Newman, L.S., Mroz, M.M., Campbell, P.A., 1989. Screening blood test identifies subclinical beryllium disease. Journal of Occupational Medicine 31, 603–608.

Lloyd, C.J., 1998. Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems. Journal of the American Statistical Association 93, 1356–1364.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Chapman and Hall, London, pp. 285–286.

Newman, L.S., 1996a. Beryllium disease and sarcoidosis: clinical and laboratory links. Sarcoidosis 12, 7–19.

Newman, L.S., 1996b. Significance of the blood beryllium lymphocyte proliferation test (BeLPT). Environmental Health Perspectives Supplement 104, 953–956.

Pepe, M.S., 1998. Three approaches to regression analysis of receiver operating characteristic curves for continuous test results. Biometrics 54, 124–135.

Sterner, J., Eisenbud, M., 1951. Epidemiology of beryllium intoxication. Arch. Ind. Hyg. Occup. Med. 4, 123–157.

Stokes, R.F., Rossman, M.D., 1991. Blood cell proliferation response to beryllium: analysis by receiver-operating characteristics. Journal of Occupational Medicine 33, 23–28.

Swets, J.A., Pickett, R.M., 1982. Evaluation of Diagnostic Systems: Methods from Signal Detection Theory. Academic Press, New York, NY, pp. 31–32.

Tornquist, L., Vartia, P., Vartia, Y., 1985. How should relative changes be measured? The American Statistician 39, 43–46.

Wambach, P., 2001. Beryllium Lymphocyte Proliferation Testing (BeLPT), at Document Number DOE-SPEC-1142-2001. National Technical Information Service, Springfield, VA 22161 (http://tis.eh.doe.gov/techstds/standard/spec1142/SPEC11422001.pdf).

Wang, Z., Farris, G.M., Newman, L.S., Shou, Y., Mair, L.A., Smith, H.N., Marrone, B.L., 2001. Beryllium sensitivity is linked to HLA-DP genotype. Toxicology 165, 27–38.

Zhou, H.K., Zhou, X.A., 2001. Receiver Operating Characteristic (ROC) Analysis, Statistics in Epidemiology Newsletter. American Statistical Association, Alexandria, VA, 1–6 (http://www.csm.ornl.gov/asasie/newsletter/SIEv2001n1.pdf).