Acceptance Criteria for Corrosion Resistance of Medical Devices: Statistical Analysis of Nitinol Pitting in In Vivo Environments

Lawrence E. Eiselstein, Duane Steffey, Andrew Nissan, Nigel Corlett, Roberto Dugnani, Esra Kus, and Sarah G. Stewart

(Submitted October 24, 2008; in revised form February 23, 2009)

ASTM F 2129 test method nor the FDA provides any guidance as to what constitutes an acceptance criterion for the corrosion resistance of implantable medical devices. Neither provide any guidance on how many samples to test or how to handle censored data, i.e. datasets where there are only a few tests that breakdown. The development of both a statistically valid acceptance criterion for corrosion resistance and a method of evaluation would be of significant benefit to the medical device community. This study of 420 nitinol cyclic polarization tests, which builds on previous research that was presented at SMST 2007, investigates the effect of long-term exposure to simulated in vivo environments with differing degrees of aeration. This was accomplished by pre-exposing electropolished (EP) nitinol to phosphate buffered saline (PBS) at 37 °C that had been sparged with either ultra high purity nitrogen or laboratory air. Immersion times ranged from 1 h up to 30 days. A total of 290 EP samples were tested in order to obtain a reasonable number of samples with breakdown, i.e. pitted. In addition, a total of 130 mechanical polished (MP) samples were also analyzed. This data allow us to test our statistical model that was presented at SMST 2007. This model takes into account the probability of breakdown per unit of exposed surface area and, if breakdown occurs, predicts the probability that $E_b - E_r$ is greater than some threshold value. Aerated PBS environments were found to have a large influence on the margin of safety against pitting in vivo. Statistical methods for treating highly right censored pitting data are presented.

1. Introduction

Long-term reliability of implantable medical devices requires testing to assure low corrosion and metal ion release rates. Corrosion can reduce the structural integrity of the device and the release of metal ions has been associated with adverse immunological response (Ref 1-3). ASTM F 2129 is a test method used to assess the corrosion susceptibility of small metallic implants using cyclic potentiodynamic polarization (Ref 4). The United States Food and Drug Administration (FDA) typically requires medical-device manufacturers to provide data that will demonstrate that their device will have sufficient corrosion resistance during the expected in vivo service life which could be greater than 10 years. Frequently, medical-device manufacturers use ASTM F 2129 to satisfy this FDA requirement.

Lawrence E. Eiselstein, Duane Steffey, Andrew Nissan, Nigel Corlett, Roberto Dugnani, Esra Kus, and **Sarah G. Stewart**, Exponent, Menlo Park, CA. Contact e-mail: eisel@exponent.com.

However, neither the ASTM F 2129 test method nor the FDA provides any guidance as to what constitutes an acceptance criterion for the polarization results or how many samples should be tested. ASTM F 2129 suggests comparison with a device already approved and on the market in the United States that is without any postmarket approval incidents of corrosion failures. This approach can, however, be quite expensive as it requires testing twice the number of experimental units (i.e., number of candidate devices and predicate devices); moreover, it presumes the existence of a suitable predicate device, i.e. a device that is implanted in a similar location in the body and manufactured from similar material and that the medical-device company has accurate knowledge regarding the predicate device's in vivo corrosion resistance.

As discussed in our earlier paper on nitinol corrosion, NiTi appears to have very good corrosion resistance based on many patient-years of in vivo experience (Ref 5). Nevertheless, the development of both a statistically valid acceptance criterion for corrosion resistance and a method of evaluation would be of significant benefit to the medical-device community. In addition, there is a need for a protocol by which a medical-device manufacturer can periodically check the corrosion resistance of an approved medical-device to assure that the manufacturing process is still within specification and devices are being manufactured with adequate corrosion resistance, i.e. a need for a manufacturing lot acceptance criterion. This study is an attempt to establish such a methodology and is a continuation of work that was first reported at SMST 2007 in Japan (Ref 5).
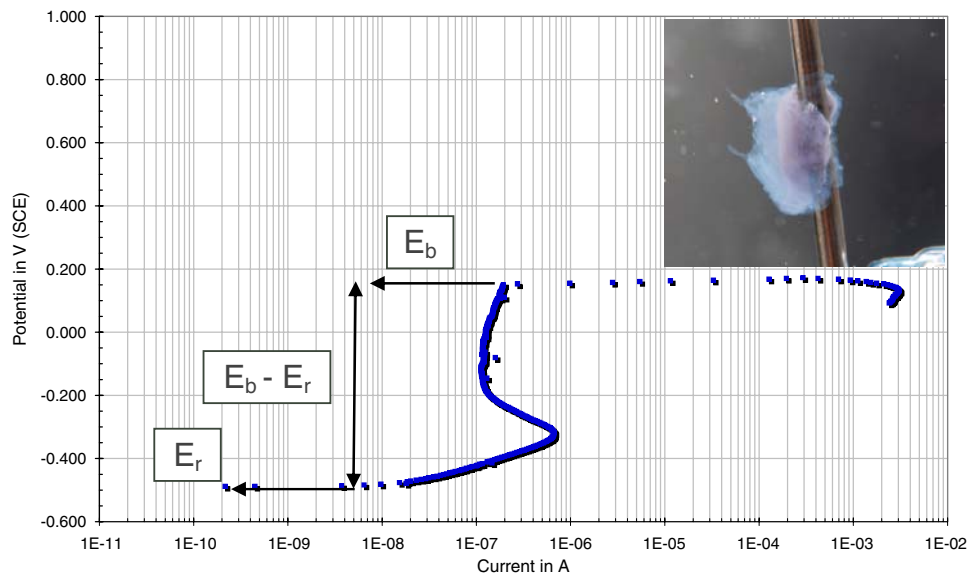
Fig. 1 Definition of electrochemical parameters E_b, E_r, and $E_b - E_r$

Three general methodologies exist as possibilities for an acceptance criterion. The first is that the pitting or breakdown potential (E_b), a measure of resistance to localized corrosion in a specific environment, of an implantable device should have similar or better corrosion resistance as determined by ASTM F 2129 testing than approved devices currently on the market with no known corrosion problems. The electrochemical parameters discussed in this paper are shown in Fig. 1. These exemplar devices should ideally be manufactured from similar materials and be used in the same general location in the body as proposed for the current device to assure a similar in vivo exposure environment, i.e. exposure to blood, bile, urine, saliva, etc. As mentioned above, it is sometimes difficult, and expensive, to find such exemplars and there is little guidance on how many devices should be tested to confirm equivalent or better corrosion resistance. The second methodology that has been proposed is that the breakdown potential (E_b) of the device should be greater than some threshold value, independent of the material used for the implant (Ref 6-8). These implant alloys can have drastically different pitting resistance and it is somewhat arbitrary that a minimum E_b should be specified independent of the material and environment. It is generally agreed that if a NiTi medical-device can be made reliably with an E_b of greater than 0.600 V with respect to a saturated calomel electrode (SCE), then there is little doubt that these materials will be resistant to corrosion in vivo unless other conditions like fretting, galvanic effects, or crevice corrosion take place in vivo. The reality, however, is that many medical devices will not be able to achieve such high values due to material selection, material finish, design, and manufacturing limitations. The third methodology, which we further develop in this paper, is to use the difference between the breakdown potential (E_b) and the open-circuit or rest potential (E_r). This gap ($E_b - E_r$) is considered to be a better parameter to evaluate medical-device margins against pitting in vivo as has previously been discussed by Pound (Ref 9). The basic principle behind this approach is that if the in vivo E_r is much lower than the pitting potential the device will have a large margin of safety against pitting or correspondingly, if $E_b - E_r$ is
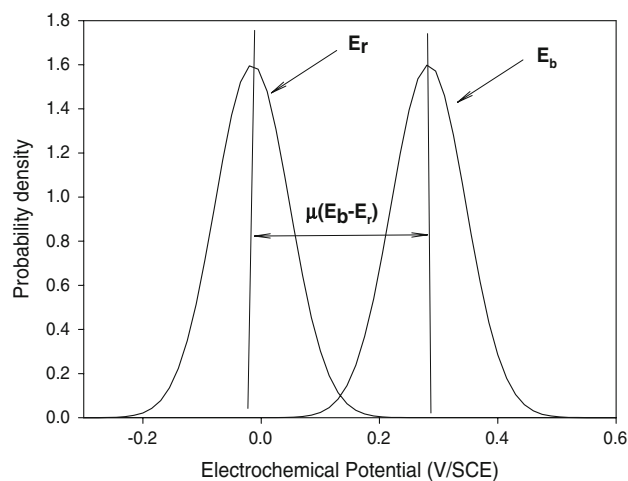

Fig. 2 Schematic illustrating the margin of safety against pitting

nearly zero or negative then the device will most likely pit when exposed to that environment. This is illustrated in Fig. 2. This $E_b - E_r$ parameter has been used frequently in the past to minimize the risk of pitting in many industries (Ref 9-13).

One problem with the $E_b - E_r$ criteria, as discussed by Rosenbloom and Corbett, is that there is little information regarding E_r for in vivo conditions (Ref 8). Another issue is how E_b, E_r, and ultimately $E_b - E_r$ vary during long-term in vivo exposures. The current ASTM F 2129 Standard states that devices are to be tested in their finished state. However, it is unlikely that the device, after exposure to in vivo conditions for several months will have the same resistance to pitting corrosion or even the same rest potential as a result of the surface film modification. Finally, there seems to be a wide variation in E_b values for NiTi depending on surface finish. Some NiTi:

1. will always pit below the minimum potential of 800 mV specified in ASTM F 2129;

2. will never undergo pitting below 800 mV; and
3. will infrequently breakdown.

Normal quality-control methodologies can be used for samples that always pit or never pit, but case 3, samples that only infrequently breakdown (datasets that contain censored data), are more difficult to treat statistically.

In an attempt to answer some of these questions, and to investigate whether long-term in vivo exposure might affect the margin of safety against localized corrosion, we report how long-term immersion in aerated and nonaerated phosphate buffered saline (PBS) will affect E_b, E_r, and ultimately $E_b - E_r$ on a statistically significant number of samples with two different surface conditions: mechanically polished (MP) and electropolished (EP).

Experimental data were collected on the breakdown potential (E_b) of replicate surrogate devices (i.e., wire samples) and on the variation in rest potential (E_r) with immersion time for individual units. Using these data, we conducted simulation studies to investigate the implications for device performance—specifically, the probability that the rest potential would exceed breakdown potential during the life of the device. The simulation results provided statistical support in defining a threshold value that would represent an acceptable margin between breakdown and rest potential ($E_b - E_r$), as well as a correspondingly low probability of corrosion-induced failure during the lifetime of the device.

We then implemented a tolerance interval approach for determining, from standard testing of a candidate device, whether there is sufficient confidence that a high proportion of replicate units will attain the specified threshold value. Finally, we developed guidance for medical-device manufacturers' quality-assurance programs in addressing the key practical question of how to determine the number of replicate units required for testing in order to establish an acceptable lower bound on the value of $E_b - E_r$ for manufactured devices. Developing this lower bound can be quite challenging for NiTi materials with good surface treatments since the vast majority of samples do not exhibit breakdown. We suggest a methodology to treat this type of highly censored data, i.e. few breakdowns with a large number of censored data, i.e. values greater than 800 mV SCE.

2. Materials and Methods

This investigation evaluated the pitting resistance of NiTi wire in two as-received surface conditions: MP and EP. The surface preparation of the material is proprietary to the material supplier and was not provided. The EP and MP NiTi used in the experiments came from different material lots.

The NiTi properties and surface characterization for the MP specimens that were purchased were described 55.9%Ni content superelastic wire with an A_f temperature in the 5-15 °C temperature range with a bright finish described as mechanical polishing that was followed by an argon anneal. The EP specimens that were purchased were described as superelastic straight wire that conforms to ASTM F 2063 Specifications (Ref 14). The EP materials had a nominal composition of 55.8 wt.% Ni. This material had an A_f temperature in the 5-18 °C range. Both the MP and EP wire had a diameter of 1.45 mm (0.057 in).

Auger analysis revealed surface oxide thicknesses of 28.6 ± 3.7 and 5.72 ± 0.79 nm for MP and EP wire, respectively (Ref 15). These film thicknesses were based on four measurements from two different locations on two different MP and EP NiTi wires. The thickness measurement was based on the equivalent sputter rate of SiO_2 under identical conditions used for the NiTi. Oxide thickness measurements were calculated as the thickness where the oxygen intensity dropped to half of its maximum within the profile. Both oxide films on the MP and EP had a 2 wt.% or less Ni concentration at the surface (Ref 15). All electrochemical testing was done with 2.5-cm long wire. Both cut ends of the wire were coated with silicone sealant which resulted in a surface area of 1.15 cm^2.

Figure 3 shows a representation of the size distribution of inclusions from longitudinal and transverse cross sections of the wire, and is similar for both materials used. The longitudinal sections indicate that the inclusions are present as stringers that broke up during the drawing process and are relatively strong and brittle compared to the surrounding matrix, as indicated by the fracturing and voids formed around these inclusions. We were unable to classify the inclusions with energy dispersive spectroscopy (EDS) but were able to observe that they are titanium rich. Infrequently, these inclusions intersect the wire surface as is shown in Fig. 4 for an EP specimen. Locating surface inclusions on the MP wire was much more difficult due to the rough surface and thick oxide layer.

All pre-test immersions and electrochemical testing were performed in PBS (8.0 g L^{-1} NaCl, 0.2 g L^{-1} KCl, 1.15 g L^{-1} Na$_2$H$_2$PO$_4$, and 0.2 g L^{-1} KH$_2$PO$_4$) at 37 ± 1 °C. The pH of the solutions was 7.4 ± 0.1 for all immersions and electrochemical testing. When samples were pre-exposed to aerated PBS, these solutions were aerated with laboratory air (79% N$_2$, 21% O$_2$, and 0.038% CO$_2$). For deaerated tests, the PBS was sparged with ultrahigh-purity nitrogen for 30 mins prior to testing and then continuously throughout the test. The use of deaerated and aerated solutions was chosen to represent the widest range of oxidizing conditions to which an implantable medical-device might be subjected. Obviously, the deaerated PBS has a lower redox potential than would be expected to be found in vivo, and the aerated PBS is expected to provide higher redox potentials than conditions in vivo, except for perhaps implants in the airways or devices that are subjected to peroxides from an immunological response.

All tests were performed using Gamry computer-controlled potentiostats and Gamry DC-105 software. Long-term potential monitoring was performed in accordance with ISO 16429:2004 and the cyclic polarization tests were performed in accordance with ASTM F 2129-06 (Ref 4, 16). All potentials were measured with respect to a SCE. The potentiodynamic scan rate used was 0.167 mV/s and the potential increased until breakdown occurred or 1.1 V (SCE) was reached. All E_r values reported are the corrosion potentials observed at the end of the pre-test immersions, i.e. in aerated or deaerated solutions, not the E_r measured in the deaerated PBS used to determine E_b.

3. Results

We tested a total of 420 nitinol test specimens in order to get sufficient information regarding the probability of pitting in 130 mechanical polished and 290 EP NiTi samples. The pre-test
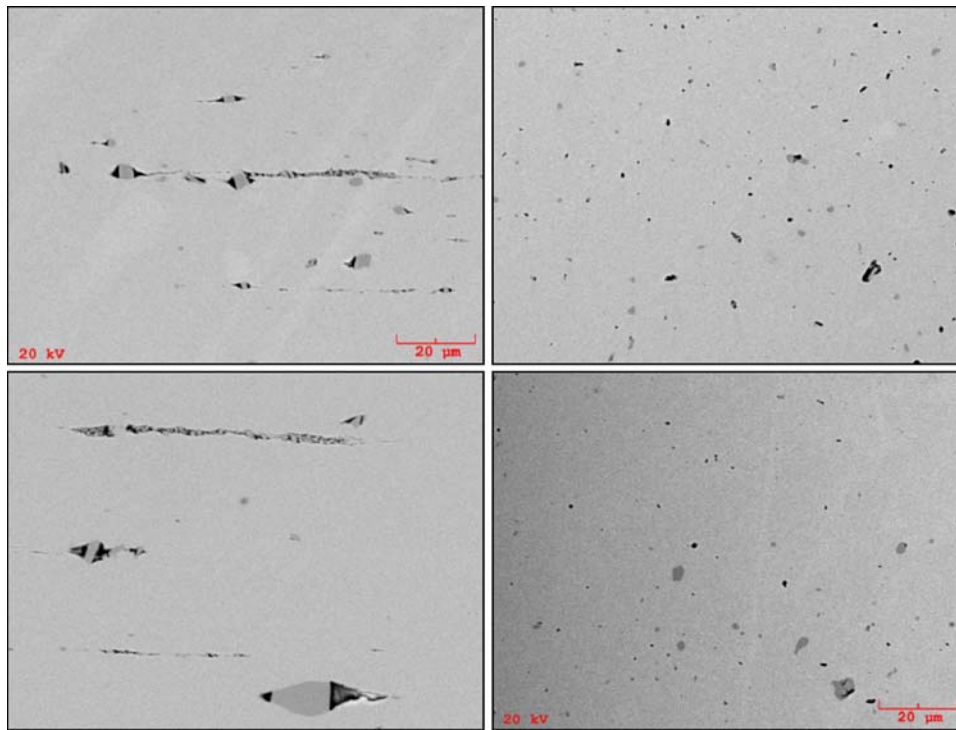
Fig. 3 Inclusions from metallurgical cross sections. Longitudinal (*upper left*) and transverse (*upper right*) of EP wire. The corresponding longitudinal and transverse metallographic sections from the MP wire are shown at the *bottom left* and *right*
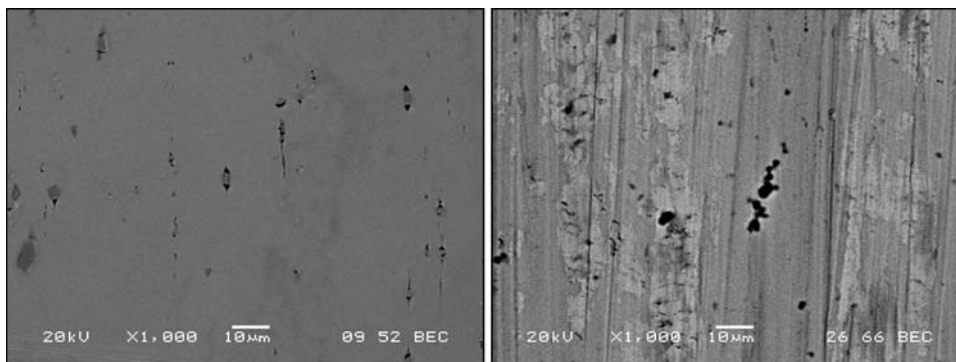

Fig. 4 Inclusions intersecting the surface. EP (*left*) and MP (*right*) in as-received condition

exposure times to the deaerated and aerated environments were 1, 12, 60, 156, and 720 h for the MP surface finish and 1, 12, and 60 h for the EP samples. Then, 3-99 replicate specimens were tested at each pre-test immersion time. The results from our testing are summarized in Fig. 5-7 and Table 1. The mean and standard deviation for E_b for EP samples can only be calculated for the samples that pitted and therefore underestimate the true resistance of the material to pitting. As can be seen from Table 1, there are very clear differences between the corrosion resistance of the MP and EP samples. All of the MP samples exhibited pitting below 0.8 V (SCE) whereas only 13% of the EP samples pitted. Figure 5 illustrates that increasing oxygen concentration in the PBS increases E_r and that E_r also tends to increase with increasing pre-test exposure time in both deaerated and aerated PBS solutions. E_r increases less rapidly with increasing pre-test exposure time for the EP surface condition. E_b for either the MP or EP surface condition

is about the same and unaffected by oxygen concentration in the PBS as shown in Fig. 6. However, E_b does increase with increasing pre-test exposure time. The safety margin against pitting corrosion, $E_b - E_r$, is lower in aerated environments primarily from the effect of oxygen on raising E_r (Fig. 7).

4. Discussion

All the MP samples pitted, whereas only about 13% of the EP samples pitted indicating a significant effect of surface treatment. The dissolved oxygen concentration also has a very large influence, as expected on the open-circuit potential (E_r). The MP wire had an oxide film about five times as thick as the EP samples and had significantly lower corrosion resistance as seen in Table 1. Clarke et al. showed that NiTi with a thicker
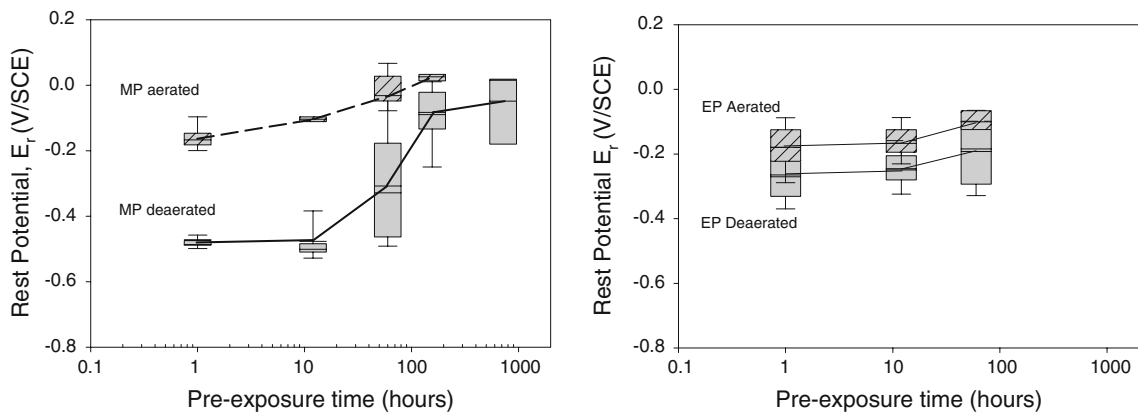
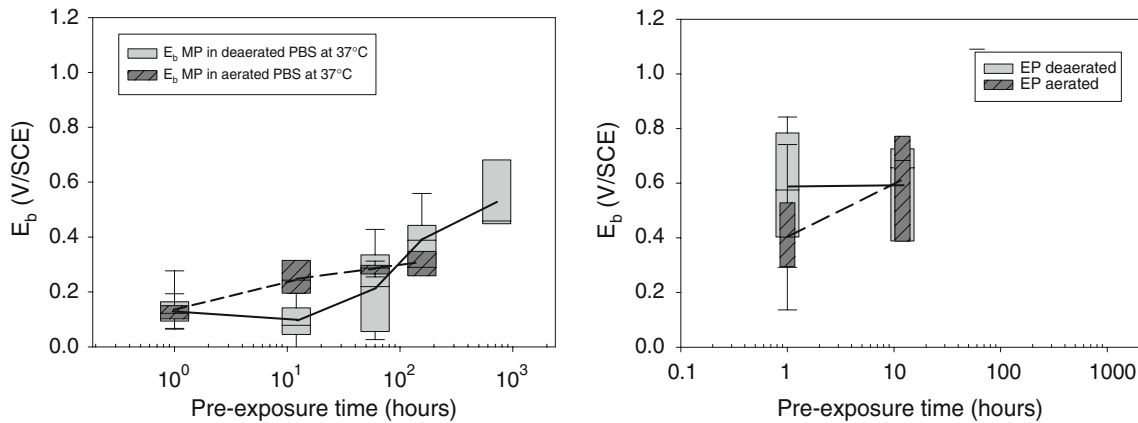**Fig. 5** Effect of pre-test immersion on rest potential E_r

Fig. 6 Effect of pre-test immersion on break down potential E_b

Table 1 Summary of test results

NiTi	Gas	Time, h	No. RPT (a)	E_r Mean	E_r σ	No. CPT (b)	No. pitted	% Pitted	E_b Mean	E_b σ	$E_b - E_r$ Mean	$E_b - E_r$ σ	No. incomplete tests (c)
MP	N_2	1	31	−0.472	0.062	30	30	100	0.130	0.047	0.602	0.078	1
MP	Air	1	13	−0.160	0.035	13	13	100	0.138	0.067	0.298	0.076	0
MP	N_2	12	27	−0.476	0.110	27	27	100	0.095	0.081	0.571	0.095	0
MP	Air	12	6	−0.104	0.009	6	6	100	0.246	0.065	0.350	0.069	0
MP	N_2	60	16	−0.307	0.161	16	16	100	0.212	0.158	0.520	0.153	0
MP	Air	60	13	−0.015	0.049	11	11	100	0.283	0.018	0.310	0.038	2
MP	N_2	156	15	−0.090	0.082	15	15	100	0.393	0.098	0.482	0.168	0
MP	Air	156	6	0.024	0.011	6	6	100	0.306	0.059	0.281	0.062	0
MP	N_2	720	3	−0.049	0.113	3	3	100	0.530	0.131	0.578	0.129	0
MP	Air	720	0	...	...	...	...	...	...	...	...	...	...
EP	N_2	1	98	−0.265	0.082	97	17	18	0.585	0.195	0.832	0.201	1
EP	Air	1	58	−0.180	0.069	57	10	18	0.407	0.184	0.584	0.216	1
EP	N_2	12	58	−0.245	0.060	56	4	7	0.590	0.193	0.769	0.198	2
EP	Air	12	57	−0.159	0.059	55	7	13	0.609	0.259	0.763	0.267	2
EP	N_2	60	13	−0.192	0.093	12	0	0	...	...	...	...	1
EP	Air	60	6	−0.099	0.045	6	1	17	1.090	...	1.199	...	0
EP	N_2	156	0	...	...	...	...	...	...	...	...	...	...
EP	Air	156	0	...	...	...	...	...	...	...	...	...	...

(a) No. RPT: Number of rest potential tests; not all samples had valid cyclic polarization tests
(b) No. CPT: Number of cyclic polarization tests; not all samples had valid cyclic polarization tests
(c) Not all samples had valid cyclic polarization tests

oxide layer has a lower E_b value and a greater nickel release in both long-term and short-term experiments (Ref 1). Clarke et al. also found that after long-term immersion tests breakdown potentials increased for samples that exhibited lower values initially (Ref 1).

The effect of long-term pre-exposures of MP NiTi to 37 °C PBS clearly increases E_b; however, E_b increases faster for pre-test exposures to aerated solutions compared to deaerated solutions. This can be seen by comparing E_b values for various pre-test exposure immersion times for MP NiTi in Table 1 and Fig. 6. Clarke et al. found a similar trend in E_b for NiTi with a thick oxide (340 nm) exposed to 0.9 wt.% saline for up to 4 months although they do not report how E_r values were affected (Ref 1). Our data show that both E_r and E_b increased with increasing pre-immersion duration in deaerated PBS (Fig. 5 and 6), but the trends were not equitable as $E_b - E_r$ decreased with increasing immersion duration because E_r increased faster than E_b.

The pre-test exposure of MP samples to aerated PBS had a more dramatic effect than the longest exposure to deaerated PBS with the $E_b - E_r$ values being about 300 mV lower which is primarily due to the very large increase of E_r when exposed to aerated PBS.

The statistical distribution of E_b for stainless steels has previously been studied by Shibata and others (Ref 17-19).

Shibata has shown that a normal distribution adequately describes the distribution of E_b for AISI 304 stainless steel tested in 3.5 wt.% NaCl solutions at 30 °C (Ref 17, 18). Nevertheless, Kowaka notes that pitting potentials are actually the lowest breakdown potential that is found only at a small area on a sample under test, and as such, a "weakest link" statistical distribution such as a Gumbel or Weibull distribution is generally expected to fit such data (Ref 18). Kowaka showed that the Weibull distribution also provided a good fit to Shibata's stainless steel E_b data while at the same time being a "weakest link" type of statistical distribution. The E_b data from the MP and EP specimens exposed to deaerated PBS are shown in Fig. 8. Both datasets are well described by a three-parameter Weibull distribution.

5. MP Modeling

Regression analysis of $E_b - E_r$ for MP NiTi data was performed to quantify the effects of aeration and immersion time. The effect of pre-test immersion exposure time was modeled as linear in the logarithm scale and separate regression analyses were performed on the aerated and deaerated data because of the greater variability of the
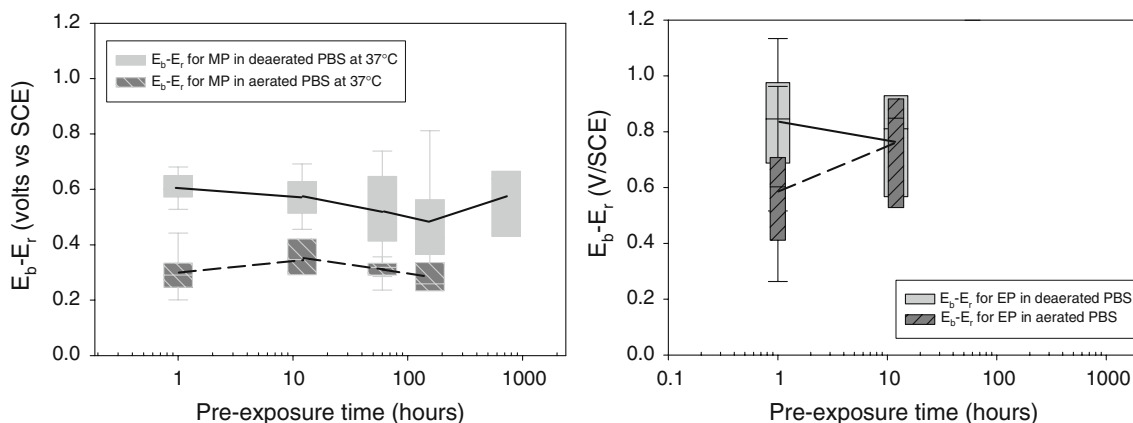

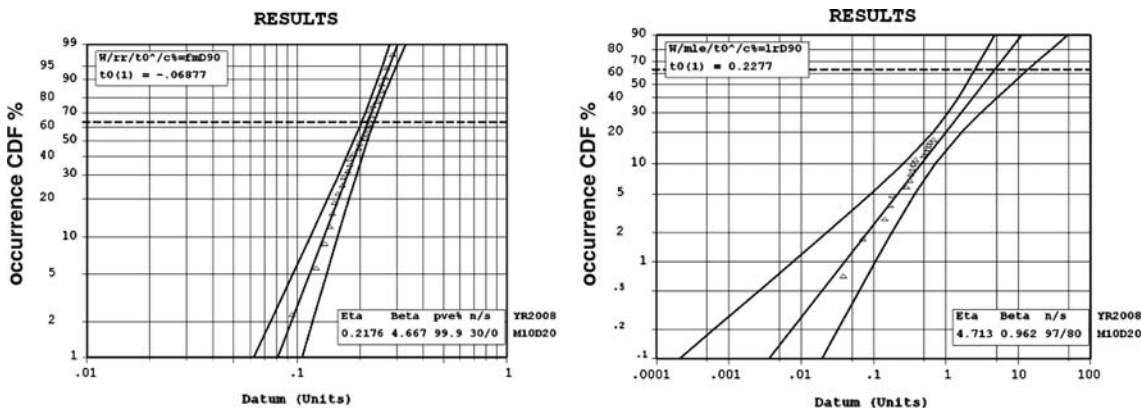**Fig. 7** Effect of pre-test immersion on $E_b - E_r$

Fig. 8 Three-parameter Weibull distribution of E_b for MP (*left*) and EP (*right*) in deaerated PBS for 1 h

Table 2 Regression analysis results of $E_b - E_r$ for MP NiTi

Predictor	Coefficient	SE coefficient	T	P
Deaerated				
Constant	0.6037	0.01962	30.77	<0.001
ln, t	−0.01838	0.006081	−3.02	0.003
$S = 0.1179$	$R^2 = 9.3\%$		R^2 (adjusted) = 8.3%	
Aerated				
Constant	0.3083	0.0172	17.92	<0.001
ln, t	−0.00037	0.00533	−0.07	0.945
$S = 0.0649$	$R^2 = 0.0\%$		R^2 (adjusted) = 0.0%	

deaerated data. The fitted regression equation for the deaerated data is as follows:

$$E_b - E_r(\text{V/SCE}) = 0.604 - 0.0184 \ln[t]$$

Ninety-one data points were used in this regression and the ANOVA results are reported in Table 2.

Similarly for the aerated data the fitted regression equation is as follows:

$$E_b - E_r(\text{V/SCE}) = 0.308 - 0.0053 \ln[t]$$

Thirty-six data points were used in this regression and the ANOVA results are reported in Table 2.

Pre-test exposure to aerated environments has a large effect on the corrosion margin of safety primarily due to the large effect of aeration on E_r as can be seen by the constant term dropping from 604 to 308 mV. Table 2 shows that the $\ln(t)$ factor is statistically significant for deaerated samples but not for aerated samples. However, the low R^2 value for the deaerated samples indicates the presence of large unexplained variability. We suspect that this large variation is a result of the inherent randomness of the pitting process. The average $E_b - E_r$ value for the longest duration pre-test immersion (720 h) appears to violate this decreasing trend, but the average is calculated from only three observations and does not alter the statistical significance of the trend.

For deaerated samples, the baseline for $E_b - E_r$ is defined as material subjected to pre-test exposure for 1 h. The regression equation for deaerated samples can therefore be expressed as:

$$E_b - E_r(\text{V/SCE})$$
$$= \text{baseline} + \text{pre-test immersion duration effect}$$

where baseline = 604 mV (SCE); this is the $E_b - E_r$ from the model for NiTi in the as-received condition (i.e., pre-test exposure to deaerated PBS for 1 h).

Pre-test immersion duration term = $-0.0184 \ln[t]$; the model predicts a decrease in $E_b - E_r$ as the pre-exposure immersion time increases:

12 h = −30 mV
60 h = −96 mV
156 h = −119 mV
720 h = −121 mV
10-years = −209 mV

Sparging with air provides dissolved oxygen levels that are higher than those generally found in vivo and therefore the aeration effect here is exaggerated over that which would be expected when NiTi is exposed to arterial blood. Nevertheless, an aerated PBS solution better represents the in vivo dissolved

oxygen concentration than a deaerated solution. The value provided above of −209 mV for the expected decrease in $E_b - E_r$ after 10-years exposure to deaerated solutions is an extreme extrapolation of the data and therefore may not accurately predict actual in vivo performance. However, if testing is done as we suggest, i.e. pre-exposure to aerated solutions, the inherent variability in $E_b - E_r$ is decreased and there is no need to account for the time the implant will be in service.

The probability that an MP sample will have a negative $E_b - E_r$ can be calculated using the probability distribution statistics shown in Table 3 where we have combined both the 1 and 12 h* and 60 and 156 h** deaerated data since there was no statistically significant difference between the 1 and 12 h or the 60 and 156 h data; however, the difference between the 1 and 12 h and the 60 and 156 h data was statistically significant.[†] We also calculated the probability that $E_b - E_r$ will be negative for MP NiTi for the combined 60 and 156 h immersions in deaerated PBS and the probability of negative $E_b - E_r$ for MP NiTi exposed to aerated PBS for all our aerated data (1-156 h data combined) since $E_b - E_r$ versus pre-test exposure time is relative insensitive to exposure duration in aerated environments.

Table 3 shows the results using several different types of distributions. A normal distribution is used since it has been previously used to describe E_b data for AISI 304 stainless steel in 3.5 wt.% saline (Ref 17-19). A Gumbel extreme value distribution has also been used since it theoretically should model the distribution of smallest values for multiple measurements taken in different locations of constant area (Ref 18). Among a set of candidate models that included the two- and three-parameter Weibull and lognormal distributions, the Gumbel extreme value distribution was found to be the best fit to the $E_b - E_r$ data for MP samples pre-exposed to deaerated PBS for 1 and 12 h (Fig. 9). The superiority of the Gumbel versus the normal model is reflected in the larger likelihood value of 65.9 in Table 3. This finding is also reflected in the more reasonable (higher) pitting probability predicted using the Gumbel distribution (6.43×10^{-5} versus 5.14×10^{-12}). A three-parameter Weibull distribution was found to best characterize the longer-term deaerated immersion data (60-156 h) with an estimated pitting probability of zero versus a 6.66×10^{-4} probability if a normal distribution is assumed.

The three-parameter Weibull distribution was also found to best characterize the aerated $E_b - E_r$ data. This is not surprising since the two- and three-parameter Weibull distributions are typically among the probability models providing the best fit to minimum extreme value distributions such as E_b. The three-parameter Weibull model gives an estimated probability of pitting in vivo of zero whenever the threshold parameter—i.e. the minimum observable value of $E_b - E_r$—is estimated to be positive. For example, the $E_b - E_r$ threshold is estimated at 149 mV (SCE) for the samples exposed to aerated solution and at 224 mV (SCE) for the deaerated exposures. Although both distributions would predict zero probability of pitting in vivo, the threshold values suggest that the specimens exposed to aerated PBS have a lower margin of safety against pitting. The fit of three-parameter Weibull distributions to $E_b - E_r$ data for short-term and long-term exposure to deaerated PBS are shown in Fig. 9. Similarly, Fig. 10 shows the three-parameter Weibull fit to the aerated $E_b - E_r$ data (all exposure times included).

---

*Mann-Whitney Rank Sum Test ($P = 0.101$).
**Mann-Whitney Rank Sum Test ($P = 0.385$).
[†]Mann-Whitney Rank Sum Test ($P = 0.001$).

Table 3 Estimated probabilities that $E_b - E_r < 0$ for MP NiTi

| Solution	Distribution type	Immersion time, h	Point estimate (P_E)	95% Upper bound (P_U)	Log-likelihood
Deaerated	Normal	1-12	5.14×10^{-12}	4.96×10^{-9}	58.8
Deaerated	Gumbel (extreme value)	1-12	6.43×10^{-5}	3.60×10^{-4}	65.9
Deaerated	Normal	60-156	6.66×10^{-4}	6.66×10^{-3}	13.6
Deaerated	3-parameter Weibull	60-156	0 (for $E_b - E_r < 224$ mV)	0	16.0
Aerated	Normal	1-156	5.57×10^{-7}	5.07×10^{-5}	48.4
Aerated	3-parameter Weibull	1-156	0 (for $E_b - E_r < 149$ mV)	0	49.5
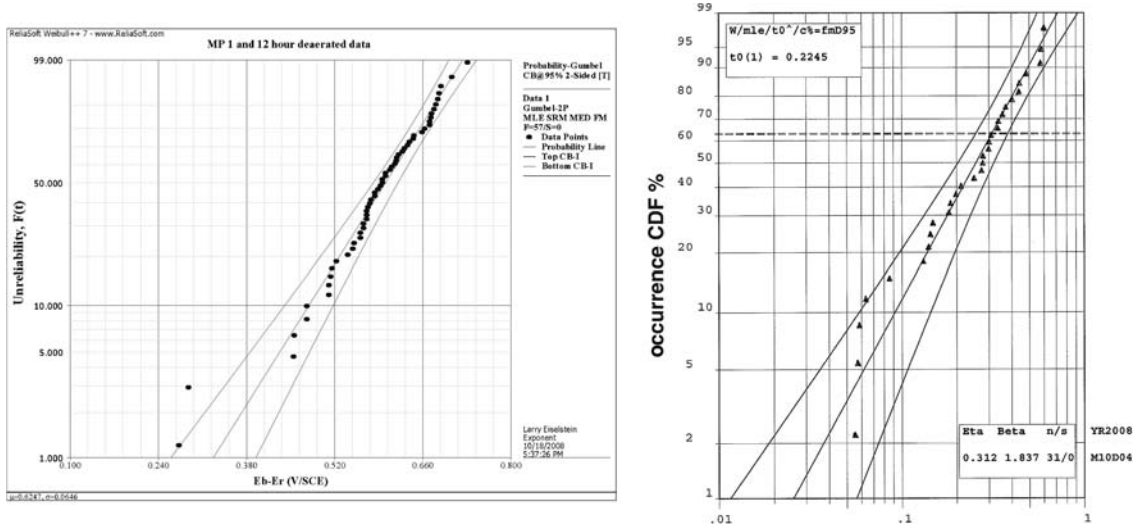
Fig. 9 Distribution of $E_p - E_r$ with confidence bands for MP samples immersed in deaerated PBS for 1 and 12 h as Gumbel (*left*) and immersed in deaerated PBS for 60-156 h as three-parameter Weibull (*right*)

The entries in Table 3 clearly demonstrate that estimated failure probabilities depend very sensitively on the type of probability distribution used to model the observed variation in $E_b - E_r$. When a typical amount of testing is performed, several models may fit the available data and lead to greatly differing inferences regarding extreme values. The normal distribution fits the observed test data reasonably well but, in all cases studied, is not the best distribution to describe the distribution of $E_b - E_r$ data. Specifically, Table 3 shows that, if the normal model is used to calculate the probability of pitting in vivo, it leads the counter intuitive conclusion that pitting is less likely in aerated versus deaerated solutions. This result can be explained by the short tails of the normal distribution and the fact that, although $E_b - E_r$, on average, is greater under deaerated conditions, the associated variability about the mean is also much greater than that under aerated conditions. The Weibull model provides a marginally better fit to the data but effectively precludes the possibility that E_r could exceed E_b. Use of the Gumbel extreme value distribution can be justified by physical considerations and produces conservative estimates of lower-tail probabilities, but the fit of the model to observed data is poor in two of the three cases studied.

6. EP Modeling

A total of 290 EP samples were tested (169 in deaerated PBS and 121 in aerated PBS). Analysis of variance of the EP
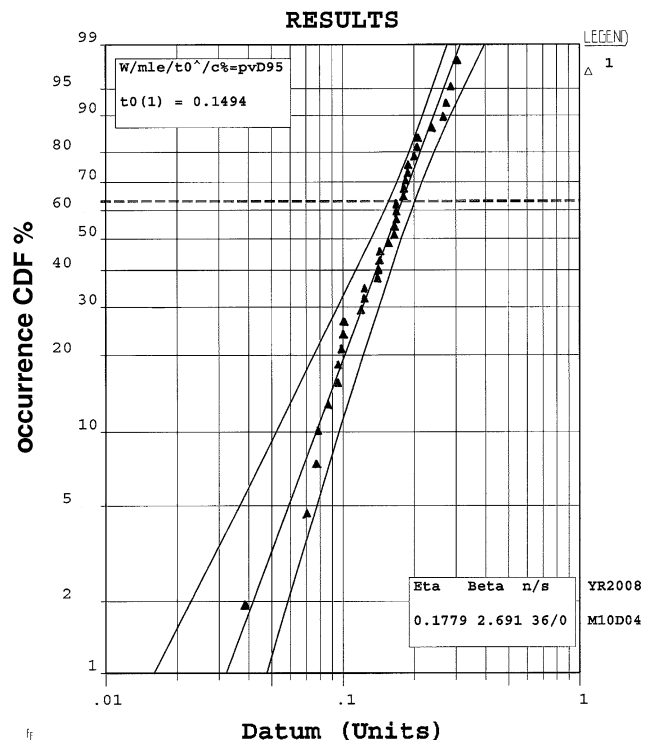

Fig. 10 Distribution of $E_p - E_r$ with confidence bands for MP samples immersed in aerated PBS for 1-156 h (three-parameter Weibull)

rest potential data (see Table 4) indicates that E_r is significantly affected by both the amount of dissolved oxygen in the PBS and the length of time of immersion. The effect of dissolved oxygen on E_r is the combined effect of the dissolved on solution redox potential and the effect of long-term exposure on the passive film (morphology, composition, and thickness).

Of the 290 EP samples tested, only 39 (13%) pitted below 0.800 V (SCE) the minimum voltage specified in ASTM F 2129 at which the cyclic polarization test can be stopped. Table 5 shows how probability of pitting of EP samples varies with pre-test exposure time and oxygen concentration in the PBS. The results from Fisher's exact test (Table 5) and binary logistic regression analysis using the censored data versus the oxygen and time variable (Table 6) showed that the differences observed were not significant indicating that neither oxygen concentration or duration of immersion in PBS significantly affect the percentage of EP samples that pit below 1.1 V (SCE).

Modeling of the EP test data is more challenging than for the MP test data as most of the data are censored, i.e. most of the samples did not break down. One way to treat such data is through the use of nonparametric statistical methods. However such methods require a large number of samples. For instance, to specify a 99% reliability value with 95% confidence, 299 samples need to be tested. The number of specimens required for a nonparametric approach is much greater than a parametric approach (Ref 20).

One parametric approach is to fit the censored $E_b - E_r$ values to a standard statistical distribution using a methodology that can handle the censoring. We have done this for the aerated and deaerated EP specimens as shown in Fig. 11. These data were best fit by a three-parameter Weibull distribution for both the aerated and deaerated exposures. The threshold parameter (gamma), below which the probability of pitting is zero, is estimated at 512 mV for the deaerated and 286 mV for the aerated exposures. These results suggest that aerated environments are more prone to produce pitting than deaerated solutions, but neither solution is likely to result in breakdown in vivo since the estimated probability of failure is zero for both.

The prediction of a zero failure probability from a fitted three-parameter Weibull or LogNormal distribution, for instance, does not seem physically reasonable, since there is presumably some probability of pitting no matter how small that actual probability. Further, we believe that single-population models lack physical basis in this context, and our investigation has shown that other

Table 4 Analysis of variance of E_r data from EP samples

General linear model: E_r vs. oxygen concentration and time			
Factor	**Type**	**Levels**	**Values**
Oxygen	Fixed	2	Deaerated, aerated
Time	Fixed	3	1, 12, 60 h

Analysis of variance for E_r using adjusted SS for tests						
Source	**DF**	**Seq SS**	**Adj. SS**	**Adj. MS**	**F**	**P**
Oxygen	*1*	*0.531*	*0.230*	*0.230*	*45.55*	*0.000*
Time	*2*	*0.106*	*0.0974*	*0.0487*	*9.65*	*0.000*
Oxygen * Time	2	0.00024	0.00024	0.00012	0.02	0.976
Error	284	1.434	1.434	0.00505		
Total	289	2.071				
$S = 0.0710$		$R^2 = 30.8\%$		R^2 (adj) = 29.6%		

Italics indicate model terms that are statistically significant at 0.05 level

Table 5 EP NiTi—Fisher's exact test indicates probability of breakdown does not vary significantly with oxygen concentration or duration of immersion time in PBS

Oxygen concentration	Immersion time, h	Number of tests	Number pitted	Percent pitted
Deaerated	1	97	17	18
	12	56	4	7
	60	12	0	0
Aerated	1	57	10	18
	12	55	7	13
	60	6	1	17

Fisher's exact test: P values of 1.0 and 0.265 for oxygen concentration and time, respectively

Table 6 EP NiTi—binary logistic regression confirms probability of breakdown does not vary significantly with oxygen concentration or duration of immersion time in PBS

Binary logistic regression: censored vs. oxygen, time		
Variable	**Value**	**Count**
Censored	1	252 (event)
	0	39
	Total	291

Logistic regression table							
Predictor	**Coefficient**	**SE coefficient**	**Z**	**P**	**Odds ratio**	**95% CI lower**	**95% CI upper**
Constant	1.40	0.296	4.73	0.000			
Aeration	−0.282	0.352	−0.80	0.423	0.75	0.38	1.5
Time							
12 h	0.712	0.385	1.85	0.064	2.04	0.96	4.33
60 h	1.305	1.05	1.24	0.214	3.69	0.47	28.9

Log-likelihood = −111.9. Test that all slopes are zero: $G = 5.41$, DF = 3, P value = 0.144
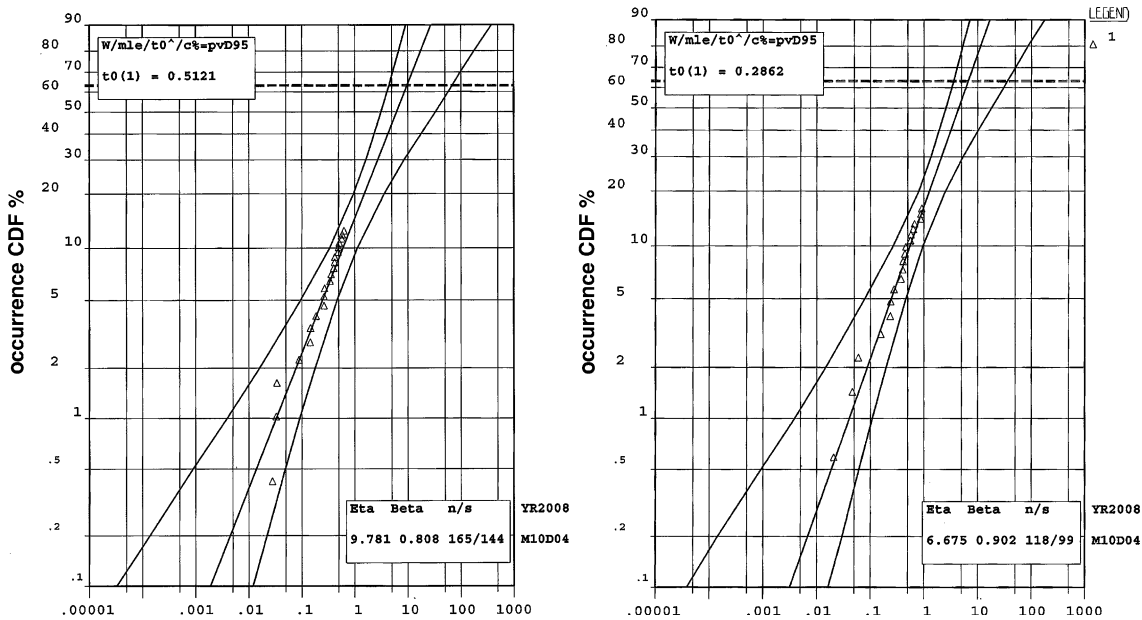
Fig. 11 Censored treatment of $E_b - E_r$ data from EP samples immersed in deaerated (left-two-parameter exponential) and aerated (censored three-parameter Weibull distribution) PBS for 1, 12, and 60 h

probability distributions provide an adequate fit to the observed $E_b - E_r$ data.

To overcome some of the limitations of the parametric approach we have previously proposed (Ref 5), we derive a physically based, mixture model, using a Poisson distribution to describe the fraction of samples that will likely break down. This was done by modeling the occurrence of surface flaws, and then using a normal distribution to model the distribution of breakdown potentials for flawed specimens. The Poisson distribution has been used previously to characterize the number of pits expected (Ref 18). The Poisson probability function $(p(k) = \exp^{-\lambda} \lambda^k / k!)$ actually gives the probability of finding a specimen with k flaws that will pit before 0.8 V is reached. In computing this probability, $\lambda = \theta a$ is proportional to the area (a) of the surface being evaluated (1.15 cm^2) and θ is the expected number of occurrences per unit area.

Our proposed approach to an acceptance criterion is to estimate the probability of corrosion (Π) based on the probability that the sample will pit (Π_1) and, if the specimen will pit, the probability (Π_2) that $E_b - E_r$ is negative:

$$\Pi = \Pi_1 \times \Pi_2,$$

where $\Pi_1 = 1 - \exp(-\lambda)$ is the probability that any individual sample will possess one or more flaws, and will therefore pit, and $\Pi_2 = F(0)$, where $F(\cdot)$ is the cumulative distribution function that describes the variation in $E_b - E_r$. We consider alternative parametric choices for $F(\cdot)$ by considering their suitability for and goodness-of-fit to the values of $E_b - E_r$ observed for the pitted specimens. Using the normal distribution for $F(\cdot)$,

$$\Pi_2 = \left(\frac{1}{2}\right)\left[1 + \mathrm{Erf}\left(\frac{\delta - \mu}{\sigma\sqrt{2}}\right)\right]$$

is the probability that $E_b - E_r < 0$ if the specimen pits. Here, δ is an arbitrary extra margin of safety above $E_b - E_r$ which for this analysis we have assumed to be zero.

Other choices for $F(\cdot)$ can be considered; for example, using the three-parameter Weibull distribution,

$$\Pi_2 = 1 - \exp[-(-\gamma/\alpha)^\beta],$$

where α and β are the (positive) scale and shape parameters, respectively, and γ is the (negative) threshold parameter. Alternatively, one could develop probability models for E_b and E_r separately. Although we do not pursue such an approach in this paper, some justification can be found in the observed lack of correlation ($r = 0.024$, $P = 0.882$) between the E_b and E_r values for pitted specimens.

The probability (Π) that $E_b - E_r$ is negative can be approximated by:

$$\Pi \approx P_E = P_1 \times P_2$$

where P_1 is the observed proportion of breakdowns and P_2 is computed by substituting data-derived estimates of μ and σ in the expression for Π_2, as shown in Fig. 12 for the deaerated exposures and Fig. 13 for the aerated exposures. The next step is then to determine the upper bound P_U for Π. The lot of material can then be accepted or rejected if P_U is less than a specified limit on the probability for pitting and subsequent breakdown.

For our data, 39 out of 290 specimens broke down. With λ representing the expected number of flaws per 1.15 cm^2 of surface area, this rate can be estimated by

$$\lambda = -\ln(1 - P_1) = -\ln(1 - 39/290) = 0.144$$

or about one flaw per seven specimens.

Restricting attention to the 39 pitted specimens, we can use analysis of variance to see if E_b is affected by oxygen concentration or duration of exposure to PBS. The result of this analysis is shown in Table 7 and confirms that E_b is not significantly affected by oxygen concentration but is influenced by the duration of exposure. Similarly, an analysis of variance of $E_b - E_r$ (see Table 8) indicates that the margin of safety
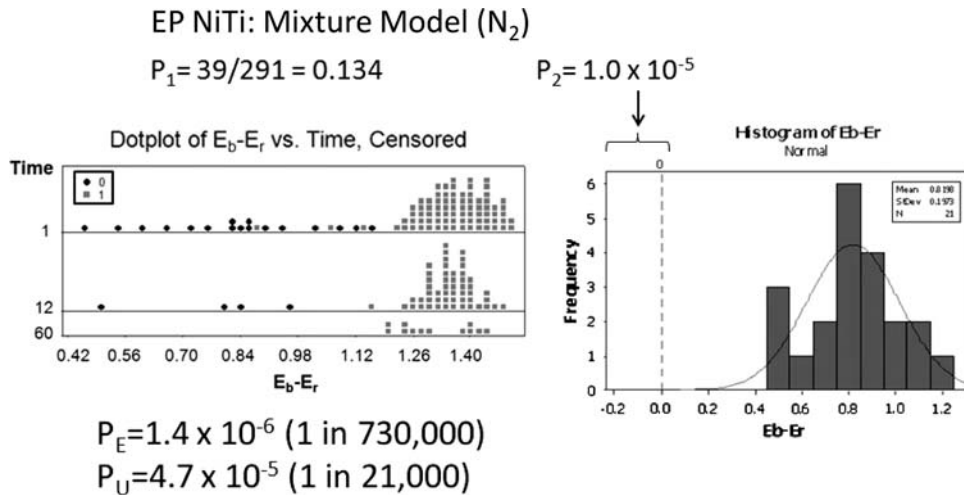
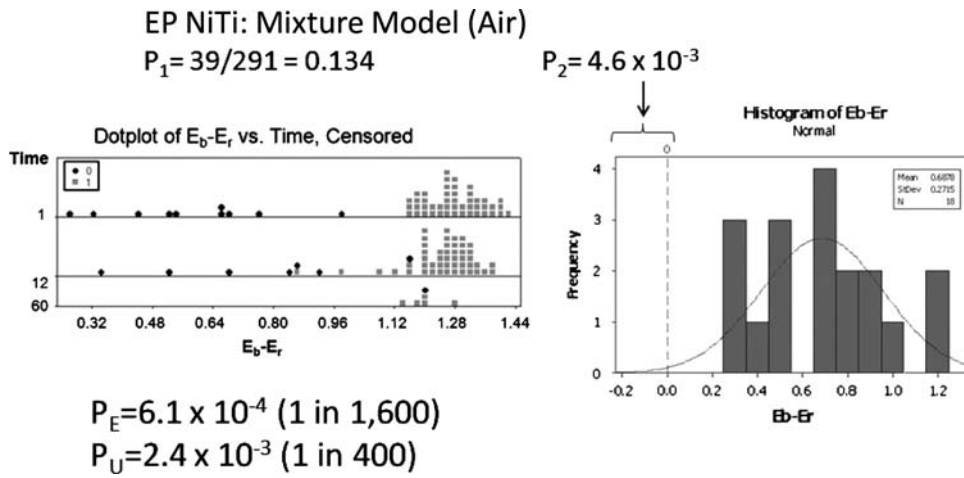**Fig. 12** EP mixed probability model for deaerated PBS

Fig. 13 EP mixed probability model for aerated PBS

Table 7 Analysis of variance of E_b data from EP samples

General linear model: E_b vs. oxygen concentration and time

Factor	Type	Levels	Values
Oxygen	Fixed	2	Deaerated, aerated
Time	Fixed	3	1, 12, 60 h

Analysis of variance for E_b using adjusted SS for tests

Source	DF	Seq SS	Adj. SS	Adj. MS	F	P
Oxygen	1	0.0381	0.131	0.131	3.04	0.090
Time	2	0.437	0.437	0.219	5.10	0.011
Error	35	1.50	1.50	0.0429		
Total	38	1.98				
S = 0.207		R^2 = 24.1%		R^2 (adj) = 17.6%		

Italics indicate model terms that are statistically significant at 0.05 level

Table 8 Analysis of variance of $E_b - E_r$ data from EP samples

General linear model: $E_b - E_r$ vs. oxygen concentration and time

Factor	Type	Levels	Values
Oxygen	Fixed	2	Deaerated, aerated
Time	Fixed	3	1, 12, 60 h

Analysis of variance for $E_b - E_r$ using adjusted SS for tests

Source	DF	Seq SS	Adj. SS	Adj. MS	F	P
Oxygen	*1*	*0.169*	*0.280*	*0.280*	*5.72*	*0.022*
Time	2	0.315	0.315	0.158	3.21	0.052
Error	35	1.72	1.72	0.0490		
Total	38	2.20				
S = 0.221		R^2 = 22.0%		R^2 (adj) = 15.3%		

Italics indicate model terms that are statistically significant at 0.05 level

against pitting $E_b - E_r$ is not significantly affected by duration of exposure but is significantly affected by the dissolved oxygen concentration in the PBS. Thus, from this analysis we see that for EP samples:

1. E_r is greater for aerated solutions
2. E_b is similar for aerated and deaerated solutions
3. Both E_b and E_r increase with exposure time
4. $E_b - E_r$ is lower in aerated solutions
5. $E_b - E_r$ does not vary significantly with immersion time for aerated or deaerated solutions

Because aeration significantly affected $E_b - E_r$ and immersion time did not, estimates of P_2 are developed by fitting probability distributions to the 21 pitted specimens tested under deaerated conditions and to the 18 pitted specimens tested under aerated conditions. Compared to a single-population approach, analysis using this physically based mixture model gives a more realistic result of 1 in 1600 for the probability of pitting for EP exposed to aerated PBS versus 1 in 730,000 for specimens exposed to deaerated solutions. This can be compared to the probability for pitting of MP samples of 1 in 1500 in deaerated PBS (60-156 h) and 1 in 1,800,000 for MP exposed to aerated PBS (1-156 h).

It appears that the MP samples have a very high surface density of pit-initiation sites and that the breakdown potentials for these sites are reasonably well characterized by a normal distribution of voltages with a small standard deviation. The breakdown is most likely associated with defects in the oxide coating from mechanical polishing. The EP samples, however, appear to have a relatively low surface-defect density that results in pitting only once in 10 specimens. For the samples that pit, the distribution of the breakdown potentials are slightly higher than those observed for the MP samples. It is possible that once the defective oxide coating has been removed by electropolishing, the only defects that remain are where inclusions and cracks surrounding the inclusions intersect the surface. Shabalovskaya et al. investigated typical inclusions in NiTi and found that the inclusions are generally titanium rich as we observed (Ref 21, 22). They indicate that these titanium rich inclusions are generally Ti_2NiO, TiO_2, TiC, or oxidized TiC. The probability of inclusions, associated crevices, and voids intersecting the surface is, hopefully, a rare event and may explain why only one in seven of the EP samples failed. The intersection of inclusions and voids and cracks surrounding such inclusions has been used in the past to characterize the probability of fatigue failure (Ref 23) and we believe a similar process can be used to explain the pitting probabilities for NiTi once the more damaging/frequent oxide defects have been removed by electropolishing.

7. Conclusion

1. $E_b - E_r$ as determined from the current version of ASTM F 2129 which does not include pre-test exposures to aerated solutions and can over estimate the margin of safety of NiTi to pitting corrosion in vivo. This indicates that current acceptance testing of medical device implants (such as ASTM F 2129) should be modified or extended to include at least several days pre-test exposure of implant material to aerated physiological solutions prior to cyclic polarization testing using the E_r values at the end of the pre-test immersion and the E_b from the testing in deaerated PBS.

2. Acceptance criterion should be based on a calculation of the probability that the device will pit in vivo is sufficiently small. This can be done by cyclic polarization testing a sufficient number of devices that have been pre-exposed to aerated PBS for sufficient times. This data should then be used to calculate the probability of pitting in vivo, i.e. the probability that $E_b - E_r$ is negative.

3. The margin of safety $(E_b - E_r)$ of MP NiTi decreased significantly after exposure to deaerated PBS at 37 °C. Modeling showed that with a pre-test exposure of 1 h E_b E_r decreased approximately 296 mV when exposed to aerated rather than deaerated solutions. This change is primarily a result of the increase in E_r rather than from a change in E_b. If testing is done with pre-exposure to aerated solutions, the inherent variability in $E_b - E_r$ is decreased and there is no need to account for the time the implant will be in service.

4. To achieve a high degree of confidence that the probability of in vivo pitting is acceptably small, the amount of testing required will depend on the population variation in breakdown and rest potentials. Small-scale pilot testing should be performed when this information cannot be obtained from other sources.

5. The pitting resistance $(E_b - E_r)$ of EP NiTi is higher than MP NiTi. Only five of the EP NiTi samples broke down whereas all of the MP samples broke down. Using only the samples that broke down, the average $E_b - E_r$ was 0.774 V SCE for the EP NiTi and 0.586 V SCE for the MP NiTi. This, however, does not take into account the fact that over 90% of the EP samples did not break down.

6. MP NiTi breakdown potentials can be treated with a normal quality-assurance type of statistics and sampling theory. $E_b - E_r$ distributions can be used to predict the probability that $E_b - E_r$ from the sample will be negative or some pre-determined threshold value. The use of a threshold value can be used to account for the effect of drift in E_b and E_r with long-term immersion, or testing can be done on samples subjected to long-term immersion in aerated solutions to verify that $E_b - E_r$ has stabilized.

7. Quality-assurance testing and statistical analysis of EP NiTi, and other medical-device materials, with breakdown potentials frequently greater than 0.8 V SCE, is much more challenging. Here we propose a combined model in which the censored data (no break down data) is used to model the probability that any given sample may, or may not break down as a Poisson process, then model the breakdown potentials of the samples that have failed as a normal distribution.

8. The probability that NiTi will pit in vivo $(E_b - E_r < 0)$ was estimated as:

    (a) 1 in 15,500 for MP exposed to deaerated PBS for 1 and 12 h (Gumbel)
    (b) 0 for MP exposed to aerated for 1-156 h (three-parameter Weibull)
    (c) 1 in 730,000 for EP exposed to deaerated for 1 h (mixture model)
    (d) 1 in 1,600 for EP exposed to aerated solutions for 1 h (mixture model)

9. The estimated probability of pitting in vivo is higher for EP samples pre-exposed to aerated PBS than for deaerated PBS primarily as a result of the increase in dissolved oxygen raising E_r.

References

1. B. Clarke, et al., Influence of Nitinol Wire Surface Treatment on Oxide Thickness and Composition and Its Subsequent Effect on Corrosion Resistance and Nickel Ion Release, *J. Biomed. Mater. Res. A*, 2006, **79**(1), p 61–70

2. L.E. Eiselstein, D.M. Proctor, and T.C. Flowers, Trivalent and Hexavalent Chromium Issues in Medical Implants, *Mater. Sci. Forum*, 2007, **539–543**, p 698–703

3. J.J. Jacobs, J.L. Gilbert, and R.M. Urban, Current Concepts Review - Corrosion of Metal Orthopaedic Implants, *J. Bone Joint Surg.*, 1998, **80**, p 268–282

4. "Standard Test Method for Conducting Cyclic Potentiodynamic Polarization Measurements to Determine the Corrosion Susceptibility of Small Implant Devices," *ASTM F 2129-06, ASTM International Standard F 2129-06*, Vol. 13.01, ASTM International, West Conshohocken, PA, 2006

5. L. Eiselstein, et al., Toward an Acceptance Criterion for the Corrosion Resistance of Medical Devices: A Statistical Study of the Pitting Susceptibility of Nitinol, *Proceedings of SMST-2007 the International Conference on Shape Memory and Superelastic Technologies*, 2007, ASM International, Tsukuba City, Japan

6. R.A. Corbett, Laboratory Corrosion Testing of Medical Implants, *Proceedings of Materials and Processes for Medical Devices Conference*, S. Shrivastava, Ed., ASM International, Materials Park, OH, 2004, p 166-171

7. R.A. Corbett and S.N. Rosenbloom, An Assessment of ASTM F 2129 Test Results Comparing Nitinol to Other Implant Alloys, *SMST 2006: Proceedings of the International Conference on Shape Memory and Superelastic Technologies*, 2006

8. S.N. Rosenbloom and R. Corbett, "An Assessment of ASTM F 2129 Electrochemical Testing of Small Medical Implants—Lessons Learned," Paper no. 07674, NACE 2007, Houston, TX, 2007

9. B.G. Pound, Susceptibility of Nitinol to Localized Corrosion, *J. Biomed. Mater. Res. A*, 2006, **77**(1), p 185–191

10. G. Frankel, Pitting Corrosion, *ASM Handbook*, Vol 13A, S.D. Cramer and B.S. Covino Jr., Eds. ASM International, Materials Park, OH, 2003, p 236

11. J.R. Scully and M.J. Kelly, Methods for Determining Aqueous Corrosion Reaction Rates, *ASM Handbook*, Vol 13A, S.D. Cramer and B.S. Covino Jr., Eds., ASM International, Materials Park, OH, 2003, p 68

12. D. Silverman, "Tutorial on Cyclic Potentiodynamic Polarization Technique," Paper no 299, CORROSION/98, NACE International, Houston, TX, 1998

13. D. Silverman, Practical Corrosion Prediction Using Electrochemical Techniques, in *Uhlig's Corrosion Handbook*, 2nd ed., R.W. Revie, Ed., Wiley, New York, 2000, p 1179–1225

14. "Standard Specification for Wrought Nickel-Titanium Shape Memory Alloys for Medical Devices and Surgical Implants," ASTM F 2063-05, Vol. 13.01, ASTM International, West Conshohocken, PA, 2005

15. B.J. Burrow and J. Moskito, "Auger Electron Spectroscopy (AES) Surface Analysis Report," Job Number: C07S6006, Evans Analytical Group, Sunnyvale, CA 94086, USA, 2007

16. "Implants for Surgery—Measurements of Open-Circuit Potential to Assess Corrosion Behaviour of Metallic Implantable Materials and Medical Devices Over Extended Time Periods," ISO 16429-04, Geneva, Switzerland. 2004

17. T. Shibata and W.R. Whitney, Award Lecture: Statistical and Stochastic Approaches to Localized Corrosio, *Corrosion*, 1996, **52**(11), p 813–830

18. M. Kowaka, *Introduction to Life Prediction of Industrial Plant Materials*, Allerton Press, Inc., New York, 1994

19. T. Shibata, Corrosion Probability and Statistical Evaluation of Corrosion Data, in *Uhlig's Corrosion Handbook*, R.W. Revie, Ed., Wiley, New York, 2000

20. W.J. Conover, *Practical Nonparametric Statistics*, 2nd ed., Wiley Series in Probability and Mathematical Statistics, Wiley, New York, 1980

21. S. Shabalovskaya, J. Anderegg, and J.V. Humbeeck, Recent Observations of Particulates in Nitinol, *Mater. Sci. Eng. A*, 2008, **481–482**, p 431–436

22. S. Shabalovskaya, et al., The Effect of Surface Particulates on the Corrosion Resistance of Nitinol Wire, *SMST-2003: The International Conference on Shape Memory and Superelastic Technologies*, 2003, p 399–408

23. J. Schaffer, A Probabilistic Approach to Modeling Microstructural Variability and Fatigue Behavior in ASTM F562 Medical Grade Wire, *Presented May 19th, 2006 at the 9th International Fatigue Congress*, Atlanta, USA. 2006