

Detecting Changes in Field Reliability
Using Data from a Complex Factory Screen

W. DAVID CLARK

Optimal Solutions Consulting, 16 Justice Hill Rd., Sterling MA 01564

RAMON V. LEON

University of Tennessee, Knoxville, TN 37996-0532

New products are sometimes screened in the factory before shipment to the field. We show how to use failure data from one of these screens to predict early field reliability. These reliability predictions are more economical and timely than reliability estimates that wait for data from field tracking studies. Further, these predictions are continuously responsive to the changes in reliability that frequently occur during the manufacture of a new product, and therefore can form the basis of a rapid-feedback system for manufacturing process control.

W. David Clark is a principal consultant at Optimal Solutions Consulting.

Dr. León is Associate Professor in the Department of Statistics of the University of Tennessee, Knoxville. He is a member of ASQC.

Introduction and Summary

Knowledge of early-field reliability is critical in evaluating the performance of a new product. This knowledge is usually obtained from data collected in field tracking studies. Although these studies often do provide good information on early-field reliability they have some inherent drawbacks. They are not only costly but often yield information that is untimely. For example, a product could be in the field several months before a statistically sound measure of early-field reliability is available.

In this paper, we describe a procedure for predicting early-field reliability using data from a complex factory screen. The procedure reduces the need for field tracking. Further, the timeliness of the reliability information provided by the procedure makes it an effective tool for process control, that is, corrective action on the process would be taken if the predictions indicate a sudden decrease in early-field reliability. While we develop the procedure for one particular screen the methods used in deriving the model and in estimating the model parameters is general and can be use for other complex factory screens.

In Figures 3 to 5, the predictions of the first month failure rate using our technique are compared with estimates calculated from field data. Note especially Figure 5, which compares factory screening estimates with field estimates for the 2096A model by quarters. The factory screening estimates predict a sharp rise in first month failures. This prediction was borne out when a reliability audit showed that the reliability was below normal. Further, this problem continued until it was detected by the audit. Had the factory screening estimates been in use the problem would have been detected much earlier.

The Reliability Screening Process

We discuss the system-level screen used by the product. Three tests -- oven, burn-in, and system -- are applied in series to the newly manufactured product. The *oven test* is intended to verify product operation in a worst case ambient temperature, 130° F for 36 hours. As a result, it identifies temperature sensitive devices and infant mortality problems. The *burn-in* test simulates field-use conditions through two days of power cycling (5 hours “On” / 1 hour “Off”) at normal ambient temperature. During these two tests, the product is in self-test mode. The computer-controlled *system test*, conducted after the burn-in test assures that the product functions before shipment from the factory. Product that fails a test is repaired and repeats at least some part of the test during which the failure was observed. See Figure 6 for more detail.

Description of the Data Collection

During the oven and burn-in test the product is in the self-test mode. In this mode, visual failure indications are provided. The exact time of failure is not recorded by the test personnel. Instead, the approach is to divide time into a number of time intervals and record the number of failures in each interval.

The 36-hours oven test, for example, is broken into nine four-hour monitoring intervals. Similarly, the two-day burn-in is broken into eight five-hour intervals that coincide with the “on” portion of the power cycle. Although, the system test takes approximately 15 minutes to conduct, items may be queued in an operating condition for several hours before being tested. The monitoring interval for this test is assumed to be four hours in duration since most items spend at least this time operating in the queue.

Physical Consideration in Modeling the Reliability Screen

The knowledge gained from the reliability screening program for the product gives us reason to expect that the failure data shows the effect of:

- DOAs (Dead-On-Arrivals), e.g., failures related to handling and electrostatic shock damage (ESD)
- Thermally sensitive devices which fail because they cannot tolerate the oven ambient temperature
- Undetected oven and burn-in failures resulting from the limitations of the self-test mode in detecting failures
- Acceleration of the aging process because of the increase in chemical reaction rates at elevated temperatures

DOAs may be the result of damage produced by ESD and physical mishandling and are not related to operating time. These types of damage are likely to occur when the product is physically moved from one test area to another, for example, from the burn-in area to the system test area.

In Figure 1, the estimated failure rate, $\hat{\lambda}(t)$ of the 2224A model is graphed¹.

¹ The estimate of instantaneous failure rate, $\hat{\lambda}(t)$ per 1000 hours is in interval i , equal to:

$$\left(\frac{F_i}{N_i \Delta t_i} \right) \times 1000$$

where

F_i = number of failures in interval i

N_i = test population at the beginning of interval i

Δt_i = time length of i^{th} interval

These data come from a reliability screen of approximately 3200 items. Note that the failure rate in the first oven interval is significantly higher than in all other oven intervals. An initially high failure rate is also present in burn-in and in the system test. Some part of these high failure rates is attributable to the presence of DOAs. It is important to separate these failures from the time-related failures.

Some temperature sensitive devices cannot tolerate the oven steady-state temperature. Since this steady-state temperature is reached during the first oven interval these devices fail in this interval. Like DOAs, these failures are not related to operating time.

The self-test, conducted during oven and burn-in, does not completely test all the internal circuitry of the equipment. This is in contrast to the computer-controlled system test that is nearly 100% effective. It is likely, therefore, that the significantly high failure rate at the system test is largely the result of undetected prior failures occurring during the oven and burn-in tests.

Because the equipment operates in the oven at an elevated temperature, acceleration of the aging process must be considered. This effect is usually modeled by the Arrhenius relation (Nelson, 1990). However, this relation is primarily applicable to semiconductor devices. It is not clear how much acceleration is experienced by a system composed not only of devices, but also of printed wire boards, interface connectors, sockets, hard wire, etc. Hence, we are not justified in assuming that oven acceleration of a system satisfies the Arrhenius relation, but instead we should estimate this acceleration from the data.

Screening Process Model

If the power is “On” in a failed product most of the internal circuitry of the product continues to be exercised. Hence, by *operating time* we mean time under power. In particular the model assumes that the equipment:

- does not age or fail during the “Off” part of the power cycle
- ages even after failure as long as it is under power.

So we assume, for example, that after an item goes through a 5 hour “On”/ 1 hour “Off” power cycle it ages exactly five hours even if it failed during the interval.

Let k and h be respectively the number of oven and burn-in intervals.

For $i = 1, 2, \dots, k$ let t_i be the oven operating time by the end of the i^{th} oven interval. For $i = 1, 2, \dots, h$ let s_i be the burn-in operating time by the end of the i^{th} burn-in interval. We note that t_k and s_h are respectively the total operating time of an item in the oven and burn-in. Let r be the duration of the system test².

The model assumes that the high temperature in the oven accelerates the aging process. In particular, it assumes that there is an acceleration factor, A , so that after t_i hours in the oven the equipment is actually At_i hours old. We note that as a consequence of this acceleration, an item after s_i hours of burn-in has an effective age of $At_k + s_i$ hours, and after system test an effective age of $At_k + s_h + r$ hours. (See Figure 2.) Since, in general, the value of the A will not be equal to 1, we distinguish throughout between *age* and *operating time*.

² In our application of the model we have: $k = 9, h = 8, (t_1, t_2, \dots, t_9) = (4, 8, \dots, 36), (s_1, s_2, \dots, s_8) = (5, 10, \dots, 40)$, and $r = 4$.

The model postulates that an item can fail for any one of four independent reasons.

In particular, it assumes that:

- Each item receives a shock in the first oven interval which causes failure with probability p_1 . This shock is associated with DOAs and thermal sensitivities. We note that $\bar{p}_1 = 1 - p_1$ is the probability of surviving the oven shock.
- Each item receives a second shock in the first burn-in interval which causes failure with probability p_2 . This shock is associated with burn-in DOAs. We note that $\bar{p}_2 = 1 - p_2$ is the probability of surviving the burn-in shock.
- Each item receives a third and final shock at the system test which causes failure with probability p_3 . This shock is associated with system test DOAs. We note that $\bar{p}_3 = 1 - p_3$ is the probability of surviving the system test shock.
- Each item is subjected to a continuous stress that causes the item to fail by age t with probability $G(t)$. In our application $G(t)$ is assumed to be a two parameter Weibull, i.e.,

$$G(t) = 1 - \exp\left[-\left(\frac{t^\beta}{\alpha}\right)\right] \text{ for } t \geq 0 \text{ where } \alpha > 0, \text{ and } \beta > 0.$$

We note that $\bar{G}(t) = 1 - G(t)$ is the probability of surviving this stress pass age t .

The model assumes that the system test can detect failures with probability 1. Effectively, we define a failure as a malfunction of the equipment that can be detected by the system test. Any malfunction that cannot be detected by the system test is not treated as a failure in the model. This is not a serious limitation since the system test is very thorough. (The system test checks 93-95% of the circuitry of the equipment.)

The self-test, as previously discussed, is not as effective as the system test. The model, therefore, assumes that the self-test detects failures with probability p . Since the model considers the system test to be perfect, those failures that escape detection by the self-test are assumed to be eventually detected by the system test. Of course this detection occurs in the system test interval and not in the period that the failure actually happened.

Probabilities of Failure Detection

We present formulas for the probabilities of an item having a failure *detected* during a given interval. These formulas express the probabilities in terms of the parameters of the model, namely, $p_1, p_2, p_3, p, \alpha, \beta$ and A . First, we need some notation.

For $i=1, 2, \dots, k$ let Q_i be the probability that the failure of the item is detected in the i^{th} oven interval. For $i = k + 1, k + 2, \dots, k + h$, let Q_i be the probability that the failure of the item is detected during the $(i-k)^{\text{th}}$ burn-in interval. For $i=k+h+1$, let Q_i be the probability that the failure of the item is detected in the system test. For $i = k + h + 2$, let Q_i be the probability that the equipment survives the screen. For convenience, we refer to Q_i as the detection probability in the i -th interval.

If $Q_i, i=1, 2, \dots, k+h+2$ is calculated under the assumption that $p_1 = p_2 = p_3 = 0$ and $p = 1$, we write Q_i^S for Q_i . For $i = 1, 2, \dots, k + h + 2$, Q_i^S is the probability of failure (detection) in the i^{th} interval when there are no shocks and all failures are detected when they occur. We call $Q_i^S, i = 1, 2, \dots, k + h + 2$ the time-dependent probabilities.

Result 1, below, expresses the detection probabilities, Q_i , as a function of the time dependent probabilities, Q_i^s , the *shock* probabilities p_1, p_2, p_3 , and the self-test detection probability, p . Result 2 expresses the time dependent probabilities, Q_i^s , as a function of A and G . Recall that G is in turn is a function of the parameters α and β .

Result 1. We have:

$$Q_1 = p \{1 - \bar{p}_1 (1 - Q_1^s)\}$$

$$Q_i = p \bar{p}_1 Q_i^s \text{ for } i = 2, 3, \dots, k$$

$$Q_{k+1} = p \bar{p}_1 \left\{ Q_{k+1}^s + p_2 \sum_{i=k+2}^{k+h+2} Q_i^s \right\}$$

$$Q_{k+i} = p \bar{p}_1 \bar{p}_2 Q_{k+i}^s \text{ for } i = 2, 3, \dots, h$$

$$Q_{k+h+1} = 1 - p + p \bar{p}_1 \bar{p}_2 \{Q_{k+h+1}^s + Q_{k+h+2}^s\} - \bar{p}_1 \bar{p}_2 \bar{p}_3 Q_{k+h+2}^s$$

$$Q_{k+h+2} = \bar{p}_1 \bar{p}_2 \bar{p}_3 Q_{k+h+2}^s$$

Result 2. We have:

$$Q_1^s = 1 - \bar{G}(At_1)$$

$$Q_i^s = \bar{G}(At_{i-1}) - \bar{G}(At_i) \text{ for } i = 2, 3, \dots, k$$

$$Q_{k+i}^s = \bar{G}(At_k + s_{i-1}) - \bar{G}(At_k + s_i) \text{ for } i = 2, 3, \dots, h$$

$$Q_{k+h+1}^s = \bar{G}(At_k + s_h) - \bar{G}(At_k + s_h + r)$$

$$Q_{k+h+2}^s = \bar{G}(At_k + s_h + r)$$

where G is given by

$$G(t) = 1 - \exp\left[-\left(\frac{t^\beta}{\alpha}\right)\right] \text{ for } t \geq 0 \text{ where } \alpha > 0, \text{ and } \beta > 0.$$

The derivation of Results 1 and 2 are straightforward and are left to the reader.

Log-Likelihood Function

Suppose we screen a number of items. Further suppose:

- One detects n_i failures in the i^{th} oven test interval, $i=1, 2, \dots, k$
- One detects n_{k+i} failures in the i^{th} burn-in test interval, $i=1, 2, \dots, h$
- One detects n_{k+h+1} failures in the system test, and
- n_{k+h+2} items survive the screen with no failures detected.

Then the log-likelihood function, $\log L$, for these observations is given by:

$$\log L = \sum_{i=1}^{k+h+2} n_i \log Q_i.$$

The maximum likelihood estimates (MLEs) of the parameters of the model are those values of $\alpha, \beta, A, p_1, p_2, p_3$, and p that maximize the log-likelihood for the

observed data $n_i, i=1, 2, \dots, k+h+2$. (See Crowder, Kimber, Smith, and Sweeting, 1991). We denote the MLEs of $\alpha, \beta, A, p_1, p_2, p_3, p$ by $\hat{\alpha}, \hat{\beta}, \hat{A}, \hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}$.

Field Reliability During the First Month

The main objective of the methodology here is to provide an estimate of the average failure rate, R , of the product during the first month of operation. This number is given by:

$$R = R(A, \alpha, \beta) = \frac{\bar{G}(At_k + s_h + r) - \bar{G}(At_k + s_h + r + 730.4)}{\bar{G}(At_k + s_h + r)}$$

where 730.4 is the number of hours in one month.

The MLE of R is $\hat{R} = R(\hat{A}, \hat{\alpha}, \hat{\beta})$. This formula was used to calculate for each quarters the “circle” estimates in Figures 3, 4 and 5.

Using the Delta Method (See Crowder, Kimber, Smith, and Sweeting, 1991) we get that \hat{R} is approximately normal with mean $R(A, \alpha, \beta)$ and approximated variance of

$$Var(\hat{R}) = \left[\begin{array}{l} \left(\frac{\partial R}{\partial \alpha} \right)^2 Var(\hat{\alpha}) + \left(\frac{\partial R}{\partial \beta} \right)^2 Var(\hat{\beta}) + \left(\frac{\partial R}{\partial A} \right)^2 Var(\hat{A}) + \\ 2 \frac{\partial R}{\partial A} \frac{\partial R}{\partial \alpha} cov(\hat{A}, \hat{\alpha}) + 2 \frac{\partial R}{\partial A} \frac{\partial R}{\partial \beta} cov(\hat{A}, \hat{\beta}) + \\ 2 \frac{\partial R}{\partial \alpha} \frac{\partial R}{\partial \beta} cov(\hat{\alpha}, \hat{\beta}) \end{array} \right]_{A=\hat{A}, \alpha=\hat{\alpha}, \beta=\hat{\beta}}$$

This result can be used to give approximate 90% confidence bounds for R using the formula $\hat{R} \pm 1.645\sqrt{Var(\hat{R})}$. This formula was used to calculate for each quarter the confidence bounds around the “circle” estimates in Figures 3, 4, and 5.

Optimization Technique

The MLEs of $\alpha, \beta, A, p_1, p_2, p_3, p$, that is, $\hat{\alpha}, \hat{\beta}, \hat{A}, \hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}$, are those values that maximize the log-likelihood function. Because of the number of parameters in the log-likelihood function a direct attempt to use a Modified Newton-Raphson (MNR) method in which derivatives are calculated numerically failed to find the maximum of the log-likelihood. Test of the MNR Method with simulated data revealed that a large discrepancy often existed between the MNR estimates and the values of the parameters used to drive the simulation. This is a result of the flatness of the log-likelihood function. Moreover, because of this flatness no general optimization algorithm is likely to be successful in finding the maximum of the log-likelihood function. As a result a tailored optimization algorithm had to be developed.

Test of the MNR algorithm with simulated data revealed that it is capable of doing the following:

- Accurate estimation of the stress parameters, α , β , and A when $p = 1$ and p_1, p_2, p_3 are equal to zero. This correspond to the case where the self-test is 100% accurate and there are no shocks.
- Accurate estimation of the shock parameters p_1, p_2, p_3 , and detection probability, p , for fixed values of α , β , and A.

Using these capabilities of the MNR an iterative procedure was developed to optimize the log-likelihood function. This procedure is accurate in finding the values at which the log-likelihood achieves its maximum, that is, the values of the MLEs.

Recall that we let Q_i^s denote the probability of an item failing in the i^{th} interval when $p_1 = p_2 = p_3 = 0$ and $p = 1$. We can use Result 1 to derive the following result:

Result 3. We have:

$$Q_1^s = 1 - \left\{ \left(1 - \frac{Q_1}{p} \right) / \bar{p}_1 \right\}$$

$$Q_i^s = Q_1 / p \bar{p}_1 \text{ for } i = 2, \dots, k$$

$$Q_{k+1}^s = \left\{ Q_{k+1} - p_2 \left(p - \sum_{j=1}^{k+1} Q_j \right) \right\} / p \bar{p}_1$$

$$Q_{k+i}^s = Q_{k+i} / p \bar{p}_1 \bar{p}_2 \text{ for } i = 1, 2, \dots, h$$

$$Q_{k+h+2}^s = Q_{k+h+2} / \bar{p}_1 \bar{p}_2 \bar{p}_3$$

$$Q_{k+h+1}^s = 1 - \sum_{i=1}^{k+h} Q_i^s - Q_{k+h+2}^s.$$

Let $\tilde{P} = (p_1, p_2, p_3, p)$. Then for each value of \tilde{P} the equations in Result 3 define a transformation of the vector $\tilde{Q} = (Q_1, Q_2, \dots, Q_{k+h+2})$ into the vector

$\tilde{Q}^s = (Q_1^s, Q_2^s, \dots, Q_{k+h+2}^s)$. Denote this transformation by $T_{\tilde{P}}$ so that

$$\tilde{Q}^s = T_{\tilde{P}}(\tilde{Q}).$$

Let $N = \sum_{i=1}^{k+h+2} n_i$ be the total number of items being screened and

$\hat{Q}_i = n_i/N, i=1,2, \dots, k+h+2$ be the observed proportion of failures detected in the i^{th} interval. Let $\hat{Q} = (\hat{Q}_1, \hat{Q}_2, \dots, \hat{Q}_{k+h+2})$. For each value of $\tilde{P} = (p_1, p_2, p_3, p)$ the vector $T_{\tilde{P}}(\hat{Q})$ is an estimate of the proportions of the failures due to the time dependent causes. Let $\tilde{n} = (n_1, n_2, \dots, n_{k+h+2})$. Define the function $S(\tilde{n}, \tilde{Q})$ by

$$S(\tilde{n}, \tilde{Q}) = \sum_{i=1}^{k+h+2} n_i \log Q_i$$

The algorithm proceed as follows:

Step 1. Make an initial guess about the value of the vector of the parameters

$$\tilde{P} = (p_1, p_2, p_3, p).$$

Step 2. With the value of $\tilde{P} = (p_1, p_2, p_3, p)$ fixed in the previous step find the value of α , β , and A that maximizes the function

$$S(NT_{\tilde{P}}(\hat{Q}), \tilde{Q}^s)$$

Here $\tilde{Q}^s = (Q_1^s, Q_2^s, \dots, Q_{k+h+2}^s)$ is given by Result 2. Use the MNR method for this step.

Step 3. With the values of α , β , and A given by *Step 2* find the value of the parameters $\tilde{P} = (p_1, p_2, p_3, p)$ which maximizes $S(\tilde{n}, \tilde{Q})$ where \tilde{Q} is given by Results 1 and 2. Use the MNR method for this step.

Step 4. Repeat Steps 2 and 3 until convergence.

This algorithm was tested on a number of simulated data items and always converged to the correct parameter values.

Results and Discussion

The predicted first-month failure rate estimates were compared with estimates of this failure rate obtained from field tracking studies. The field tracking study estimate (field estimate) of the first-month failure rate (denoted by $\hat{\lambda}$) was calculated as the number of non-DOA failures in the first 30 days divided by the total observed operating time in the first 30 days.

We assumed an exponential distribution during the first month to calculate the standard error of this estimate: $\hat{\lambda}/\sqrt{r}$ where r is the number of non-DOA failures in the first 30 days. Using these standard errors we calculated the approximate 90% confidence intervals in Figures 3, 4, and 5 (i.e., the intervals with triangles in the middle.)

Further, we calculated the failure rate using a Reliability Block Diagram (RBD) model found in a predecessor of Klinger, D. J., Nakada, Y., and Menendez, M. A. (1990), namely, the Reliability Information Notebook. This model provides first month failure rate estimates for a system by summing the failure rates of the individual components. These RBD predictions are denoted by horizontal lines in figures 3, 4, and 5.

Several observations can be made about the results in figures 3 to 5:

- In all cases, the predicted failure rate is higher than the RBD prediction. We believe that this is to be expected since the RBD model does not adequately take into consideration non-component failures. (About 30 of the failures observed in the screen were not component related.)
- In all but one case the predicted failure rate is higher than the field estimate of the failure rate. This may be the result of a larger proportion of the circuitry being

exercised by the system test than by field operation. Further, there could be an underreporting of the field failures.

- The confidence intervals based on the factory screen are narrower than the corresponding intervals obtained from the field data.

Note especially Figure 5, which compares factory screening estimates with field estimates for the 2096A model by quarters. The fourth quarter factory screening estimate is made using only December data. It predicts a sharp rise in first month failures. This prediction was borne out when a reliability audit showed that the reliability was below normal. Further, this problem continued until it was detected by the audit. Had the factory screening estimates been in use the problem would have been detected much earlier.

As we have seen when new products are screened in the factory the failure data collected in the screen can be used to predict changes in early field reliability. These predictions are more timely than estimates obtained from field tracking studies and thus can form the basis of a rapid-feedback process control system. The statistical and optimization techniques we employed - though developed for a particular factory screen - can be adapted to other complex screens.

References

Crowder, M. J., Kimber, A. C, Smith, R. L., and Sweeting, T. J. (1991) *Statistical Analysis of Reliability Data*. Chapman and Hall, London.

Klinger, D. J., Nakada, Y., and Menendez, M. A. (1990) *AT&T Reliability Manual*.

Van Nostrand Reinhold, New York.

Nelson, W (1990). *Accelerated Testing*. Wiley, New York, NY.

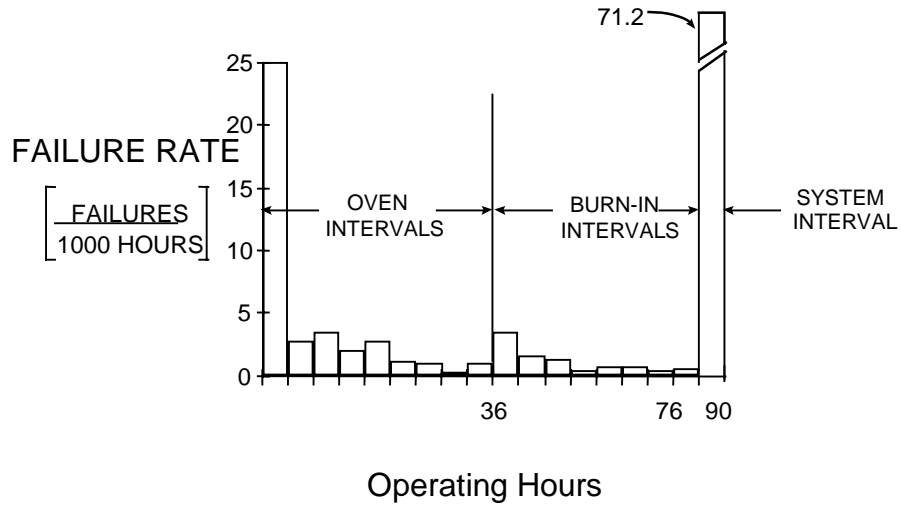


FIGURE 1. Average failure rate for the 2024A product during reliability screening.

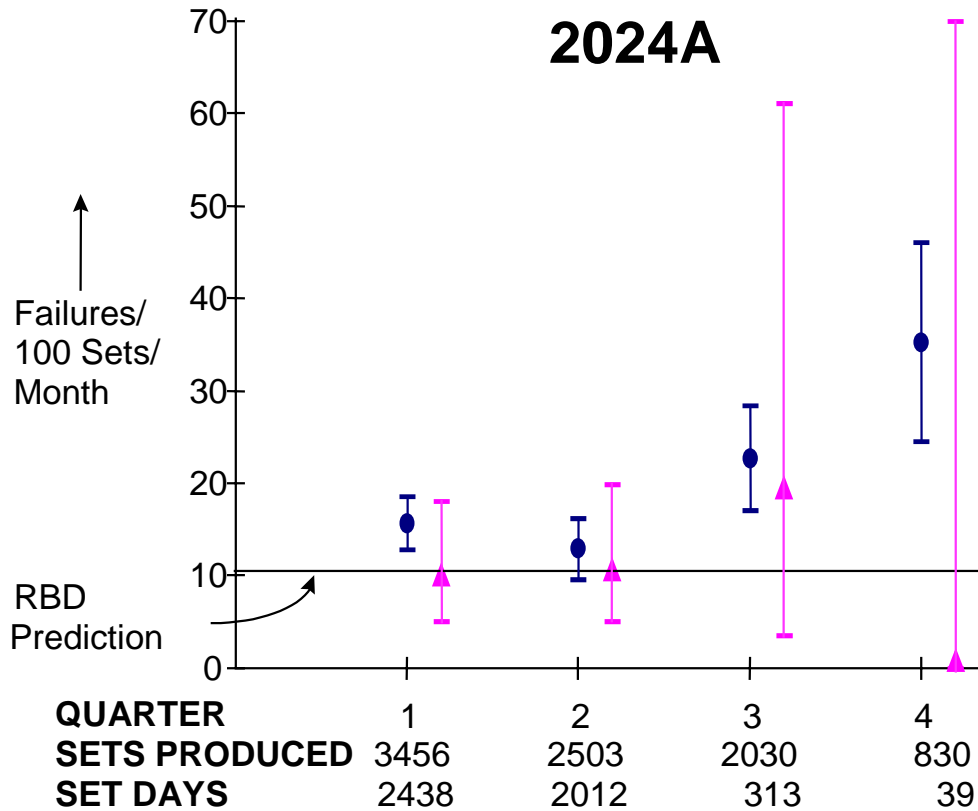


FIGURE 3. Point estimates of the first month failure rate with 90% confidence bounds are shown for model 2024A. The circles correspond to estimates from factory data and the triangles to estimates from field data. A theoretical Reliability Block Diagram (RBD) prediction based on part data is also provided. The numbers along the bottom indicate the size of the factory and field samples, and explain the extreme width of the confidence bands for the third and fourth quarter field estimates.

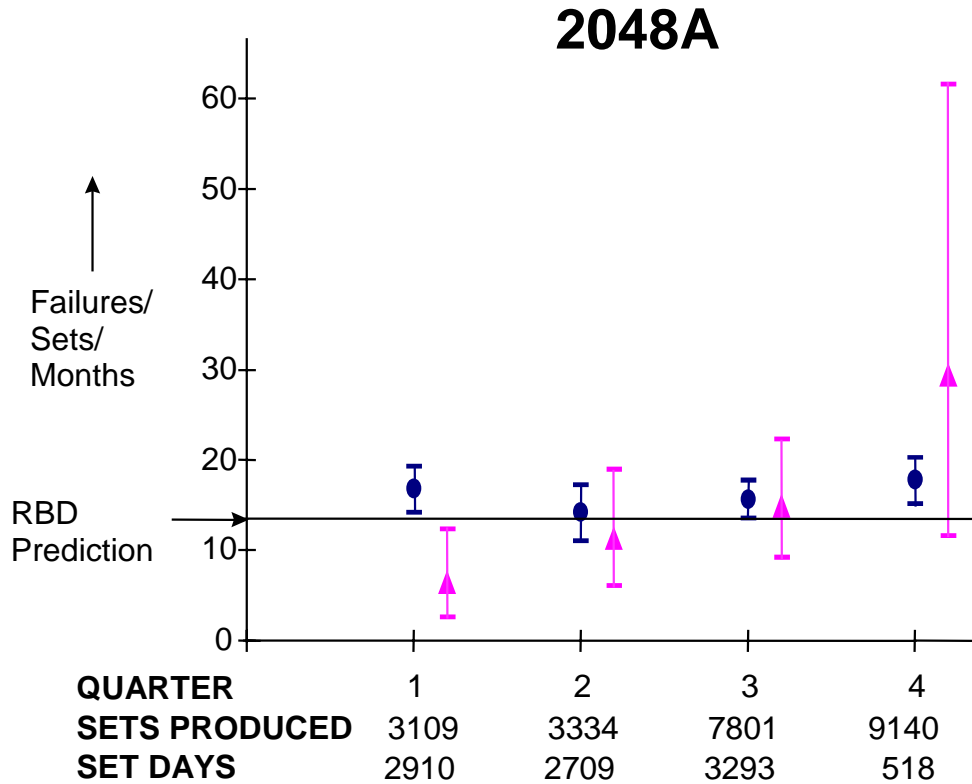


FIGURE 4. Point estimates of the first month failure rate with 90% confidence bounds are shown for model 2048A. The circles correspond to estimates from factory data and the triangles to estimates from field data. A theoretical Reliability Block Diagram (RBD) prediction based on part data is also provided. The numbers along the bottom indicate the size of the factory and field samples.

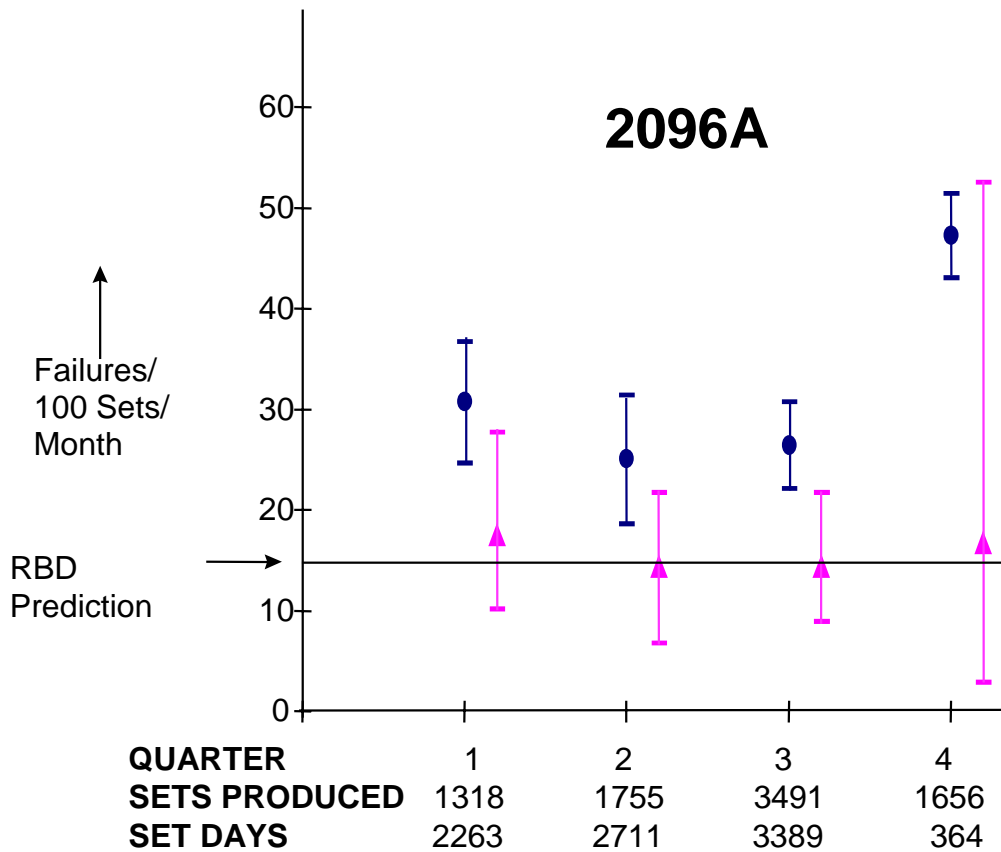


FIGURE 5. Point estimates of the first month failure rate with 90% confidence bounds are shown for model 2096A. The circles correspond to estimates from factory data and the triangles to estimates from field data. A theoretical Reliability Block Diagram (RBD) prediction based on part data is also provided. The numbers along the bottom indicate the size of the field samples. The fourth quarter factory estimate is based solely on December data as data for October and November were unavailable.

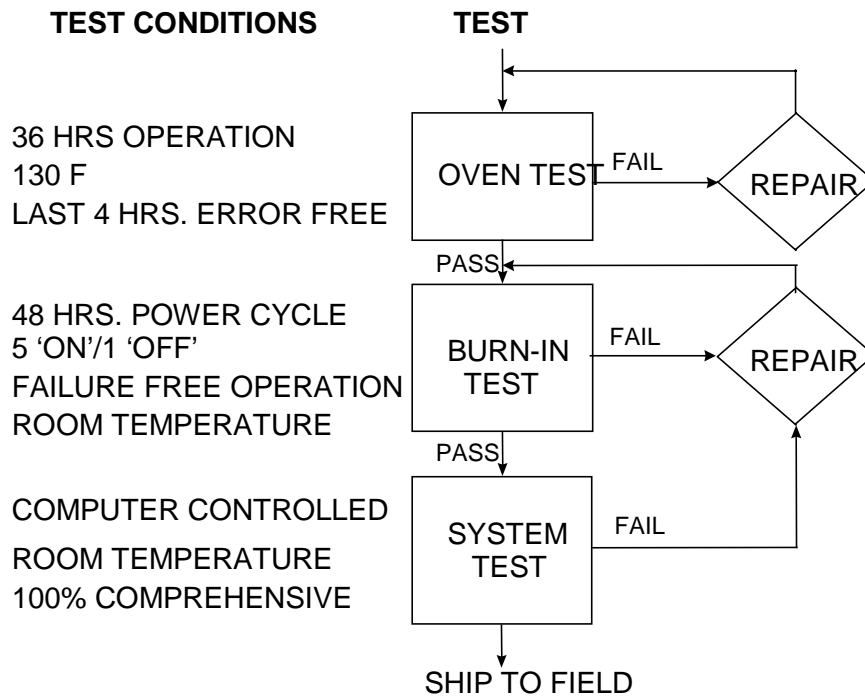


FIGURE 6. Reliability Screening Process and Test Conditions. (Sets with intermittent failure symptoms are returned to the oven after repair. The entire sequence is then repeated.)