# Determining Reliability Requirements and Testing Costs in the Early Stages of Single Use Medical Product Design

Fritz Eubanks, Ph.D.
Battelle Memorial Institute
505 King Avenue
Columbus, OH 43201 USA
eubanksc@battelle.org



**Abstract.** The production of single use medical devices, particularly for home use by patients, continues to grow, and the reliability of these devices is a primary concern for manufacturers and end-users. The systems engineer tasked with the device development needs methods and tools to establish reliability requirements and provide cost estimates for the testing necessary to show compliance with those requirements. This paper examines methods for determining reliability requirements, the cost of reliability testing for single use medical devices in the design input phase of product development, and how the costs of testing and potential errors can be used to perform trade-off analysis between reliability tolerance and confidence level.

## Introduction

Reliability is a product performance parameter, and consequently shares in the three-way balance between product performance, cost, and time to market. Design for high reliability requires varying combinations of high reliability components, functional redundancy, and periodic overhaul/maintenance, all of which make the product more expensive to design, build, and test. On the other hand, disregard for reliability makes products more expensive to operate and maintain, and leads to customer dissatisfaction and loss of sales.

Medical products can range from simple, single use devices, like tongue depressors and syringes, to large, complex systems like MRI systems and multi-assay in vitro diagnostic devices. Likewise, the complexity of the establishing and meeting the device reliability requirements will vary with device complexity.

Product requirements for medical devices are established during the design inputs phase of product development. (QSR 2009) The product requirements define the performance characteristics, safety and reliability requirements, regulatory requirements, applicable product standards, physical characteristics, and packaging and labeling requirements, among other things. (Trautman 1997) At the same time, project managers and systems engineers begin establishing the design and development plan, including major schedule milestones and overall program costs. Chief among design and development costs is product testing to verify compliance with product requirements. In the case of reliability testing, these costs can be substantial due to the large number of items and/or amount of time required to obtain statistically sound data that serve to verify product reliability requirements. Methods to estimate testing lot sizes during the design inputs phase can prove valuable for both cost and

schedule planning, as well as performing trade-off analyses for refining reliability requirements.

# Assessing the Reliability Requirements

## *The Structure of Reliability Requirements*

Reliability is defined as "the probability that an item will perform its intended function under stated conditions over a specified interval." Therefore, the reliability goal must include specifications for the following items:

- Measure of success/failure
  - A probability between 0 and 1, or a percentage between 0% and 100%
  - A mean time to failure (MTTF) or mean time between failure (MTBF)
  - A system availability between 0 and 1, or a percentage between 0% and 100%
- Definition of success/failure
  - Success: No downtime, performance parameters within specification, no lost data
  - Failure: no test result, false positive, insufficient output, complete system failure
- Range of normal operating conditions
  - Temperature, humidity, pressure, vibration, dust/pollution, liquid, power levels
- Interval over which probability of success/failure will be measured
  - Time, cycles, miles
  - Note: this interval is not the same as product life

Examples of good reliability requirements are as follows:

> The system shall have mean time before failure of 1000 hours over a one year period when operating under laboratory conditions where failure is defined as a false positive indication.
>
> The power subsystem shall have a 95% probability of performing in accordance with specifications over 1000 hours in arctic conditions.
>
> The vessel shall remain pressurized at 100±5 psig without operator intervention for 150 hours at 120°F with 99.5% reliability.

## *Collecting Basic Information*

The key reliability issues for any product or system are: (RiAC, 1996)

- What measures of reliability are important to the end-user?
- What levels of reliability are necessary to meet the end-user's needs?
- How will the manufacturer determine if the required levels of reliability have been achieved?

To answer these questions, it may be necessary to engage in a fact-finding effort that may involve a Voice of the Customer (VOC) study, benchmarking, and/or market surveys. Through these activities, the manufacturer should to come to alignment with the end user's needs on the following key reliability questions:

- How often will the product be used?
- How many failures per 1000 attempted uses can be tolerated?
- How much operating time per use is expected?
- Who will be the regular user of the product?
- Where and under what conditions will the product be used?
- How is success/failure of the product defined?
- What is the expected life of the product?
    - For single use products, how long will the product be stored before use and under what conditions?
- Will users be compensated for failed items and, if so, how much?

For repairable systems, additional reliability issues must be considered:

- How many product failures can be tolerated over a 3, 6, or 12 month period?
- How much product downtime for service/repair of failures can be tolerated?
- Who will be tasked with performing service/repair?
- Will there be a warranty period and for how long?
- How much product downtime for routine maintenance can be tolerated?
- Who will be tasked with performing routine maintenance?
- How much will routine maintenance parts cost and who will pay for it?

Some of the answers to these questions may not be available in the concept/feasibility stage, but need to be considered and, if possible, estimated in order for the manufacturer to decide on how to position the product from reliability and cost perspectives.

## *Reliability Requirement Testing Costs*

The end user needs to know that reliability goes hand in hand with product cost. "Four nines" reliability is great, but may increase the cost of the product to an unacceptable level. Suppose that the end user of the single use syringe demanded 1 failure for every 1000 attempts. Achieving this reliability will likely increase the cost of the device substantially.

The manufacturer needs to define the importance that reliability will have as a performance parameter relative to product cost and time to market. The importance aids in establishing the level of confidence required by the manufacturer when assessing how well the product has met the reliability requirements. This, in turn, allows the systems engineer to make a rough order of magnitude estimate of testing costs, because sample size and test time are driven by the combination of reliability and confidence interval. While testing approaches for more complex reparable and non-reparable system are well studied, simpler single use devices have not receive a lot of attention. O'Connor (2002, 357) recommends that statistical acceptance

sampling methods can be used for such devices. The success/failure nature of single use devices suggests that statistics of population proportions can be applied. (Devore 2008, 306)

Establishing and testing requirements for large reparable systems is well studied and documented. The remainder of this paper will focus on requirements and testing for single-use devices.

### *Notation*

It is important to note that the reliability values expressed in the following development are not the same as are used for time-based reliability calculations. For this work, the device reliability is the ratio of the number of successes to the number of trials, commonly expressed as R = x/n, where x and n are discrete integer values. The notation will be as follows:

| Symbol | Definition |
|---|---|
| R | Required reliability as a population proportion |
| n | Test sample size |
| $\alpha$ | Level of significance; probability of type I error |
| $\beta$ | Probability of type II error |
| R´ | Potential reliability due to type II error |
| $\beta$( R´) | Probability of type II error when R = R´ |

## Single Use Device Reliability Verification

### *Verification Testing Requirements*

Verification testing provides the objective evidence that the product meets performance requirements, and that the product is ready for release to production. Devices used for product verification testing need to be equivalent to the device that will be produced for sale and distribution.

In the language of statistical hypothesis testing, the null hypothesis is that the product performance meets the requirement being tested, while the alternative hypothesis is that the performance falls outside the limits of the requirement. Reliability requirements are generally stated as a minimum; e.g., at least 95% with 95% confidence. Therefore, requirement verification will take the form of a one-tailed hypothesis test with a null hypothesis that the reliability is greater than or equal to 95%. The concern early in the design phase is to plan for a sufficient number of tests to provide the required confidence in the validity of the verification test results.

### *Testing Errors and Sample Size*

The two hypothesis test errors are defined as follows (Devore, 2008, 288):

> A type I error consists of rejecting the null hypothesis when it is true.

A type II error consists of not rejecting the null hypothesis when it is false.

In terms of requirements verification testing, these definitions can be re-written as:

A type I error consists of concluding that the requirement has not be met when it has.
A type II error consists of concluding that the requirement has been met when it hasn't.

While a type I error could result in schedule delays and additional testing cost, a type II error could result in the release of a product that does not meet the reliability requirement. In the context of verification testing, a type II error means that the level of reliability realized in production will be below the level of reliability measured during verification testing.

Statistical confidence is (1 - α), where α is the probability of a type I error. When $nR \geq 10$ and $n(1 - R) \geq 10$, p has approximately a normal distribution, and the lower confidence limit (LCL) for a one-sided, lower bound test of a population proportion can be computed. Therefore, the minimum sample size needed to establish the confidence interval for 95% reliability using the normal approximation is $n = \frac{10}{1-0.95} = 200$. Under the presumption that the reliability of the device will be 95%, and that the desired confidence is 95%, the LCL can be computed as (Devore, 2008, 266):

$$\boldsymbol{LCL = R - z_{\alpha}\sqrt{\frac{R(1-R)}{n}} = 0.95 - 1.645\sqrt{\frac{0.95(1-0.95)}{200}} = 0.925} \quad (1)$$

Increasing the sample size to 400 would make the LCL around 93.2%. Here is the first point where the manufacturer must define the importance of reliability:

Q1: What lower confidence limit of reliability is acceptable at the desired level of statistical confidence?

If the manufacturer desires 95% reliability with 95% confidence and LCL of 93%, then the necessary sample size can be estimated as (Devore, 2008, 267):

$$\boldsymbol{n = R(1-R)\left(\frac{z_{\alpha}}{(R-LCL)}\right)^2 = 0.95(0.05)\left(\frac{1.645}{0.02}\right)^2 = 322} \quad (2)$$

Statistical power is (1 - β), where beta is the probability of a type II error. Unlike α, there is not a single value for β. There will be a different β for each value of p contained within the bounds of the alternative hypothesis. For example, if a test of 400 units shows that the reliability is 95% with 95% confidence, there is a 19% probability that the actual population reliability is 92%. In other words, there is an almost 1 in 5 chance that the actual device reliability in production will be below the 95% one-sided lower confidence limit. Here is the second point where the manufacturer must define the importance of reliability:

Q2: What tolerance for type II error (combination of actual reliability in production and probability of realizing that reliability) is acceptable at the desired level of statistical confidence?

The calculation of sample size necessary to properly control both type I and type II errors in reliability verification testing contains 4 variables: the required reliability ($R_0$), the confidence level (1 - $\alpha$), the probability of a type II error ($\beta$), and the lower bound at which $\beta$ applies ($R'$). Continuing our example, the manufacturer desires at least 95% reliability with 95% one-sided confidence ($\alpha = 0.05$). In addition, the manufacturer feels they can only tolerate a 10% chance ($\beta = 0.10$) that the actual reliability in production is as low as 93%. The sample size can be estimated using (Devore, 2008, 308):

$$\boldsymbol{n = \left[\frac{z_\alpha\sqrt{R_0(1-R_0)}+z_\beta\sqrt{R'(1-R')}}{R'-R_0}\right]^2 = \left[\frac{1.645\sqrt{0.95(1-0.95)}+1.282\sqrt{0.93(1-0.93)}}{0.93-0.95}\right]^2 = 1176} \quad (3)$$

Note that the increased sample size also reduces the confidence interval, resulting in a LCL of 94%. In this case, the desire for a low probability of type II error has driven the sample size to a level that provides 95% confidence that the LCL will be within 1% of the required reliability.

## *Sample Size Tradeoff Analysis*

Considering that pre-production samples for testing can cost anywhere from $50 to $500 each, the cost of parts alone for a sample size of 1176 starts at $58,800 and goes up from there, not to mention the time and effort required to manufacture the preproduction parts for testing. The manufacturer may want to examine options for possibly reducing the lot size for testing. Using the previous development, sample size can be calculated for combinations of acceptable limits of type I and type II errors. However, working with probabilities and sample sizes alone can be a little too abstract for making tradeoff decisions. What the manufacturer really wants at this stage is a rough order of magnitude estimate of the total cost of the reliability testing.

Using experience from previous programs and judicious estimation, the systems engineer can collect some basic parameters used to estimate the costs of conducting verification testing. These values can be used to calculate a simple estimate of testing costs for various sample sizes as follows:

$$\boldsymbol{Testing\ Cost = [(Sample\ Size) * (Part\ Cost)] + \left[\frac{(Sample\ Size)}{(Testing\ Rate)} * (Labor\ Rate + Facility\ Rate)\right] + (Fixture\ Cost)} \quad (4)$$

Testing costs are driven by sample size, and in this context, lower is better. However, lower sample size results in a higher probability of type II error, and thus a better chance that the production reliability will be lower than anticipated. The impact of lower reliability in production will be felt as a loss to the manufacturer due to warranty returns, customer dissatisfaction, and potential claims for property damage or personal injury. If the losses can be roughly estimated for each incremental shortfall in reliability, it can provide the basis for a tradeoff against testing costs.

In some cases, the cost of device failures may have been computed as part of the business case used to justify the decision to proceed with design. Otherwise, a rough estimate can be obtained by summing up the estimated probability and severity of each potential outcome of a device failure. Potential outcomes and estimates of severity and probability can be generated from previous experience with similar devices, or from high level risk assessments. For outcomes involving injury or property loss, Ayyub (2003) and Wilson and Crouch (2001) can be used to estimate costs. Expressing the severity in terms of cost to the manufacturer, the general expression would be:

$$\boldsymbol{C(failure) = \sum_{i=1}^{m} P\,(outcome)_i \times C(outcome)_i} \quad (5)$$

where: C(failure) = cost of a device failure
$P(outcome)_i$ = probability of potential outcome i occurring
$C(outcome)_i$ = cost of potential outcome i to manufacturer
m = number of potential outcomes identified

Recall that there will be a different probability of type II error for each value of R´ < R. For one-sided hypotheses, the probability is calculated as:

$$\boldsymbol{\beta(R') = 1 - \Phi\left[\frac{R - R' - z_\alpha\sqrt{\frac{R(1-R)}{n}}}{\sqrt{\frac{R'(1-R')}{n}}}\right]} \quad (6)$$

Therefore, the cost of potential type II errors can be expressed as the sum over potential values of R´ of the probability of type II error multiplied by the cost associated with products having reliability R´ instead of $R_o$. This calculation is not as intractable as it seems. For moderate values of sample size ($n \geq 400$) with R = 95%, β(90%) is less than 1%.

The cost estimation and trade-off process is best illustrated through the following example.

## Single Use Medical Device Example

A pharmaceutical manufacturer developed a drug for treating a chronic pain condition. The drug requires intramuscular injection on a daily basis, and the manufacturer wanted to develop a one-button, home use solution for making the injection. Answers to the salient questions are as follows:

- How often will the product be used? Once
- How many failures per 1000 attempted uses can be tolerated? 50
- How much operating time per use is expected? No more than 5 seconds
- Who will be the regular user of the product? Adults, 25-80 years old, no physical disabilities
- Where and under what conditions will the product be used? Home use, weekly or monthly, US, Canada, EU

- How is success/failure of the product defined? Success = proper dose delivered to patient's thigh muscle within 5 seconds of activation
- For single use products, how long will the product be stored before use and under what conditions? 2 years at 5°C
- Will users be compensated for failed items and, if so, how much? The cost of the device plus shipping.

Using the information above, the systems engineer can establish the following product reliability requirement:

> The product shall deliver the proper dose to the patient within 5 seconds of actuation with a probability of at least 95% when used in an environmentally controlled interior space with temperature of 15-35°C, humidity of 10-95% RH, and atmospheric pressure of 14.7-10.3 psia following storage at 5°C for no more than 2 years.

The manufacturer believes that a single use device with 95% reliability provides a good balance between performance and cost. Production volumes are estimated at 50,000 devices per year. The trade-off process starts by determining required sample size based on lower confidence limit and the level of confidence in achieving that limit in accordance with Equation (2). Table 1 provides the trade-offs between confidence, LCL, and sample size for R=95%.

Table 1. Lot Size for Testing, $R_0$=95%

| | LCL | | | | |
|---|---|---|---|---|---|
| Confidence | 94% | 93% | 92% | 91% | 90% |
| 95% | 1286 | 322 | 143 | 81 | 52 |
| 90% | 781 | 196 | 87 | 49 | 32 |
| 85% | 511 | 128 | 57 | 32 | 21 |
| 80% | 337 | 85 | 38 | 22 | 14 |

Note that using sample sizes below 200 will require different treatment due to the restriction that $n(1-R) \geq 10$ in order for the normal distribution assumption of R to apply. The estimated costs for reliability testing as a function of sample size calculated using Equation 4 are shown in Table 2.

Table 2. Estimated Cost of Reliability Testing

| | | |
|---|---|---|
| Part Cost = | $200 | each |
| Testing Rate = | 6 | /hour |
| Labor Rate = | $100 | /hour |
| Fixture Cost = | $5,000 | |
| Facility Rate = | $75 | /hour |

| | LCL | | | | |
|---|---|---|---|---|---|
| Confidence | 94% | 93% | 92% | 91% | 90% |
| 95% | $299,708 | $78,792 | $37,771 | $23,563 | $16,917 |
| 90% | $183,979 | $49,917 | $24,938 | $16,229 | $12,333 |
| 85% | $122,104 | $34,333 | $18,063 | $12,333 | $9,813 |
| 80% | $82,229 | $24,479 | $13,708 | $10,042 | $8,208 |

Assume a preliminary selection of the 90% confidence level. Based on the confidence level, the additional cost of development must be weighed against the possible additional cost of operation due to a higher than expected failure rate, as measured by the probability of a Type II error. Calculated probabilities for type II error for each level of combination of R′ and LCL in accordance with Equation 6 are shown in Table 3.

Table 3. Probability of Type II Error

| R = | 95% |
|---|---|
| Confidence = | 90% |

| | LCL | | | | |
|---|---|---|---|---|---|
| R′ | 94% | 93% | 92% | 91% | 90% |
| 0.94 | 0.18 | 0.55 | 0.69 | 0.74 | 0.78 |
| 0.93 | 0.00 | 0.20 | 0.42 | 0.55 | 0.63 |
| 0.92 | 0.00 | 0.04 | 0.21 | 0.37 | 0.47 |
| 0.91 | 0.00 | 0.01 | 0.09 | 0.23 | 0.34 |
| 0.90 | 0.00 | 0.00 | 0.03 | 0.13 | 0.24 |

Assume that the manufacturer has performed a rough cost assessment of potential failure outcomes as follows:

| Potential outcome | Probability | Cost |
|---|---|---|
| Serious injury | 0.0001 | $500,000 |
| Moderate injury | 0.005 | $45,000 |
| Minor injury | 0.05 | $6000 |
| No injury – returned item | 0.94 | $500 |

Using Equation 5, the cost per device failure is estimated to be $1,045. For a population of 500,000 devices, an additional failure rate of 1% represent 5000 devices, for a potential annual loss of $5,225,000. Calculated potential losses for each level of combination of probability and magnitude of type II error and their totals are shown in Table 4.

Table 4. Potential Loss due to Type II Error

| R = 95%<br>Confidence = 90% | | | | | |
|---|---|---|---|---|---|
| | LCL | | | | |
| R′ | 94% | 93% | 92% | 91% | 90% |
| 0.94 | $9,593 | $28,971 | $35,813 | $38,866 | $40,554 |
| 0.93 | $290 | $20,956 | $44,143 | $57,452 | $65,315 |
| 0.92 | $1 | $6,995 | $33,625 | $57,618 | $74,303 |
| 0.91 | $0 | $1,399 | $19,352 | $47,183 | $71,426 |
| 0.90 | $0 | $191 | $9,129 | $33,734 | $61,682 |
| Total Cost | $9,884 | $58,512 | $142,062 | $234,854 | $313,281 |

Our rough calculations indicate that a reliability test program that exhibits 90% confidence in a lower reliability bound of 93% is a reasonable trade-off of testing cost versus potential loss due to reliability uncertainty. A sample size for verification testing of 1176 is calculated using Equation 3. Note that the increased sample size brings the total estimated testing costs to around $94,000, but still represents a good trade when compared to the potential cost of lowering the acceptable LCL to 92%.

## Conclusions and Future Work

The production of single use medical devices, particularly for home use by patients, continues to grow, and reliability of these devices is a primary concern. The systems engineer tasked with the device development needs methods and tools to establish reliability requirements and provide cost estimates for the testing necessary to show compliance with those requirements. This paper presented a set of basic questions for determining reliability requirements during the design input stage. We also demonstrated that the cost of reliability testing for single use medical devices can be estimated during the design input stage, and the results used to perform trade-off analysis of between required tolerance, confidence level, and cost. We will continue to develop and refine the questions we ask to determine the proper reliability requirements, and the cost models for providing rough order of magnitude cost estimates as we apply them to future product development projects.

## References

Ayyub, B. 2003. *Risk Analysis in Engineering and Economics.* Boca Raton, FL: Chapman & Hall/CRC Press LLC.

Devore, J. 2008. *Probability and Statistics for Engineers and Scientists, 7th Edition.* Belmont,

CA: Duxbury.

ISO 2007. ANSI/AAMI/ISO 14971:2007 – Medical devices – Application of risk management to medical devices. Arlington, VA: Association for the Advancement of Medical Instrumentation.

O'Connor, P. 2002. *Practical Reliability Engineering, Fourth Edition.* West Sussex, England: John Wiley and Sons, Ltd.

QSR 2009. Code of Federal Regulations Title 21 – Food and Drugs, Chapter I – Food and Drug Administration, Department of Health and Human Services, Subchapter H – Medical Devices, Part 820 – Quality System Regulation. Revised April 1, 2009.

RiAC. 1996. Blueprints for Product Reliability Part 1 – Defining Reliability Programs. Reliability Information Analysis Center. Published May 15, 1996. http://theriac.org/DeskReference/viewDocument.php?id=280&Scope=reg.

Trautman, K. 1997. *The FDA and Worldwide Quality System Requirements Guidebook for Medical Devices.* Milwaukee, WI: ASQ.

Wilson, R. and Crouch, E. 2001. *Risk Benefit Analysis.* Cambridge, MA: Harvard University Press.

## Biography

Fritz Eubanks is a Systems Engineer with Battelle's Health and Life Science's Medical Device Solutions group, where he is involved in system-level design and analysis of both medical and commercial products, and serves as Safety Risk Management lead engineer. His 19 years of experience includes 8 years civil service in quality engineering with the Air Force Logistics Command, and 11 years in medical and commercial product development at Battelle. He received a B.S. in Mechanical Engineering from Kansas State University in 1982, and M.S. and Ph.D. from Ohio State University in 1992 and 1996, respectively. He is a member of INCOSE and is an ASQ Certified Reliability Engineer.