SAE International®

SURFACE VEHICLE/ AEROSPACE RECOMMENDED PRACTICE

JA1012 AUG2011

Issued 2002-01
Revised 2011-08

Superseding JA1012 JAN2002

A Guide to the Reliability-Centered Maintenance (RCM) Standard

RATIONALE

The document was updated as a result of the normal 5 year review cycle and to maintain consistency with the most recent revisions to SAE JA1011 – Evaluation Criteria for Reliability-Centered Maintenance (RCM) Processes. Changes were made to clarify the origin of the Reliability Centered Maintenance process and purpose of that document. Additionally, terminology was updated to reflect current usage in the user community and to remove items that might have been considered biased to individual commercial processes. The overall technical process remains unchanged.

FOREWORD

Reliability-Centered Maintenance (RCM) was first documented in a report written by F.S. Nowlan and H.F. Heap of United Airlines and published by the U.S. Department of Defense in 1978. It described the then-current state-of-the-art processes used to develop maintenance programs for commercial aircraft. Since then, the RCM process has been widely used by other industries, and has been extensively refined and developed. These refinements have been incorporated into numerous application documents, published by a variety of organizations around the world. Many of these documents remain faithful to the basic principles of RCM as expounded by Nowlan and Heap.

However in the development of some of these documents, key elements of the RCM process have been omitted or misinterpreted. Due to the growing popularity of RCM, other processes have emerged that have been given the name "RCM" by their proponents, but that are not based on Nowlan and Heap at all. While most of these processes may achieve some of the goals of RCM, a few are actively counterproductive, and some are even dangerous.

As a result, there has been a growing international demand for a standard that sets out the criteria that any process must comply with in order to be called "RCM." SAE JA1011 meets that need. However SAE JA1011 presupposes a high degree of familiarity with the concepts and terminology of RCM. This Guide amplifies, and where necessary clarifies, those key concepts and terms, especially those that are unique to RCM.

Note that this Guide is not intended to be a manual or a procedural guide for performing RCM. Those who wish to apply RCM are strongly encouraged to study the subject in greater detail, and to develop their competency under the guidance of experienced RCM practitioners.

TO PLACE A DOCUMENT ORDER: **Tel: 877-606-7323 (inside USA and Canada)**
Tel: +1 724-776-4970 (outside USA)
Fax: 724-776-0790
Email: CustomerService@sae.org
SAE WEB ADDRESS: http://www.sae.org

SAE values your input. To provide feedback on this Technical Report, please visit http://www.sae.org/technical/standards/JA1012_201108

TABLE OF CONTENTS

1. SCOPE

SAE JA1012 (“A Guide to the Reliability-Centered Maintenance (RCM) Standard”) amplifies and clarifies each of the key criteria listed in SAE JA1011 (“Evaluation Criteria for RCM Processes”), and summarizes additional issues that must be addressed in order to apply RCM successfully.

1.1 Organization of the Guide

Sections 5 to 14, 16, and 17 of this Guide reflect the major sections of SAE JA1011. Section 15 explains in more detail how key elements of the RCM process can be combined to select appropriate policies for managing individual failure modes and their consequences. Section 18 addresses management and resourcing issues essential to the successful performance of RCM.

2. REFERENCES

2.1 Applicable Documents

The following publications form a part of this specification to the extent specified herein. Unless otherwise indicated, the latest issue of SAE publications shall apply.

2.1.1 SAE Publications

Available from SAE International, 400 Commonwealth Drive, Warrendale, PA 15096-0001, Tel: 877-606-7323 (inside USA and Canada) or 724-776-4970 (outside USA), www.sae.org.

SAE JA1011 Evaluation Criteria for Reliability-Centered Maintenance (RCM) Processes

2.2 Related Publications

The following publications are provided for information purposes only and are not a required part of this SAE Technical Report.

2.2.1 U.S. Department of Commerce Publication

Available from NTIS, Port Royal Road, Springfield, VA 22161.

Nowlan, F. Stanley, and Howard F. Heap, “Reliability-Centered Maintenance,” Department of Defense, Washington, D.C. 1978, Report Number AD-A066579

2.2.2 U.S. Department of Defense Publications

Available from DODSSP, Subscription Services Desk, Building 4/Section D, 700 Robbins Avenue, Philadelphia, PA 19111-5098. MIL standards and handbooks may also be obtained from https://assist.daps.dla.mil/quicksearch/.

MIL-HDBK 2173(AS) ”Reliability-Centered Maintenance Requirements for Naval Aircraft, Weapons Systems and Support Equipment,” (U.S. Naval Air Systems Command) (NOTE: canceled without replacement, August 2001.)

NAVAIR 00-25-403 ”Guidelines for the Naval Aviation Reliability-Centered Maintenance Process,” (U.S. Naval Air Systems Command)

MIL-P-24534 ”Planned Maintenance System: Development of Maintenance Requirement Cards, Maintenance Index Pages, and Associated Documentation,” (U.S. Naval Sea Systems Command)

MIL-STD-1629 “Procedures for Performing a Failure Mode, Effects and Criticality Analysis,” Department of Defense, Washington, DC, 1984 (NOTE: Cancelled without replacement, August 1998)

MIL-STD-1843 “Reliability Centered Maintenance for Aircraft, Engines, and Equipment,” United States Air Force (NOTE: Cancelled without Replacement, August 1995)

S9081-AB-GIB-010/MAINT ”Reliability-Centered Maintenance Handbook” (U.S. Naval Sea Systems Command)

2.2.3 U.K. Ministry of Defence Publication

Available from Reliability-centred Maintenance Implementation Team, Ships Support Agency, Ministry of Defence (Navy), Room 22, Block K, Foxhill, Bath, BA1 5AB, United Kingdom.

NES 45 ”Naval Engineering Standard 45, Requirements for the Application of Reliability-Centred Maintenance Techniques to HM Ships, Royal Fleet Auxiliaries and other Naval Auxiliary Vessels” (Restricted-Commercial)

2.3 Other Publications

The following publications were consulted in the course of developing this SAE Technical Report and are not a required part of this document.

Anderson, Ronald T. and Neri, Lewis, “Reliability-Centered Maintenance: Management and Engineering Methods,” Elsevier Applied Science, London and New York, 1990

Blanchard, B.S., D. Verma and Peterson, E.L., “Maintainability: A Key to Effective Serviceability and Maintenance Management,” John Wiley and Sons, New York, 1995

“Dependability Management - Part 3-11: Application guide - Reliability centred maintenance,” International Electrotechnical Commission, Geneva, Doc. No. 56/651/FDIS

Jones, Richard B., “Risk-Based Management: A Reliability-Centered Approach,” Gulf Publishing Co., Houston, TX, 1995

MSG-3, Maintenance Program Development Document,” Air Transport Association, Washington, D.C. Revision 2007.1

Moubry, John, “Reliability Centered Maintenance,” Industrial Press. Inc. New York City, 1997.

Smith, Anthony M., “Reliability Centered Maintenance,” McGraw-Hill, New York. 1993

Zwingelstein, G., “Reliability Centered Maintenance, a practical guide for implementation,” Hermès, Paris. 1996

3. DEFINITIONS

3.1 AGE

A measure of exposure to stress computed from the moment an item enters service or first begins to degrade, either from new or re-entering service after a task designed to restore its initial capability. Age can be measured in terms of calendar time, running time, distance traveled, duty cycles, or units of output or throughput.

3.2 APPROPRIATE TASK

A task that is capable of preventing or mitigating the consequences of failure based on the technical characteristics of that failure

3.3 CONDITIONAL PROBABILITY OF FAILURE

The probability that a failure will occur in a specific period provided that the item concerned has survived to the beginning of that period.

3.4 DESIRED PERFORMANCE

The level of performance desired by the owner or user of a physical asset or system.

3.5 ENVIRONMENTAL CONSEQUENCES

A classification assigned to failure modes, or multiple failures in the case of hidden failure modes, that could result in a breach of any industry or government environmental standard or regulation.

3.6 EVIDENT FAILURE

A failure mode whose effects become apparent to the operator(s)under normal circumstances if the failure mode occurs on its own.

3.7 EVIDENT FUNCTION

A function whose failure on its own becomes apparent to the operator(s)under normal circumstances.

3.8 FAILURE CONSEQUENCES

A classification of the failure effects of failure modes into categories based on evidence of failure, impact on safety, the environment, operational capability, and cost.

3.9 FAILURE EFFECT

What happens when a failure mode occurs.

3.10 FAILURE-FINDING TASK

A scheduled task used to determine whether a specific hidden failure has occurred.

3.11 FAILURE MANAGEMENT POLICY

A generic term that encompasses on-condition tasks, scheduled restoration, scheduled discard, failure-finding, run-to-failure, and one-time changes.

3.12 FAILURE MODE

A single event, which causes a functional failure.

3.13 FUNCTION

What the owner or user of a physical asset or system wants it to do.

3.14 FUNCTIONAL FAILURE

A state in which a physical asset or system is unable to perform a specific function to a desired level of performance.

3.15 HIDDEN FAILURE

A failure mode whose effects do not become apparent to the operator(s) under normal circumstances if the failure mode occurs on its own.

3.16 HIDDEN FUNCTION

A function whose failure on its own does not become apparent to the operator(s) under normal circumstances.

3.17 INITIAL CAPABILITY

The level of performance that a physical asset or system is capable of achieving at the moment it enters service.

3.18 MULTIPLE FAILURE

An event that occurs if a protected function fails while its protective device or protective system is in a failed state.

3.19 NET P-F INTERVAL

The minimum interval likely to elapse between the discovery of a potential failure and the occurrence of the functional failure.

3.20 NON-OPERATIONAL CONSEQUENCES

A classification assigned to failure modes

that do not adversely affect safety, the environment, or operations, but only require repair or replacement of any item(s) that may be affected by the failure.

3.21 ON-CONDITION TASK

A periodic or continuous task used to detect a potential failure.

3.22 ONE-TIME CHANGE

Any action taken to change the physical configuration of an asset or system (redesign or modification), to change the method used by an operator or maintainer to perform a specific task, to change the operational context of the system, or to change the capability of an operator or maintainer (training).

3.23 OPERATING CONTEXT

The circumstances in which a physical asset or system is expected to operate.

3.24 OPERATIONAL CONSEQUENCES

A classification assigned to failure modes that adversely affect the operational capability of a physical asset or system (output, product quality, customer service, military capability, or operating costs in addition to the cost of repair).

3.25 OWNER

A person or organization that may either suffer or be held accountable for the consequences of a failure mode by virtue of ownership of the asset or system.

3.26 P-F INTERVAL

The period between the point at which a potential failure becomes detectable and the point at which it degrades into a functional failure.

3.27 POTENTIAL FAILURE

An identifiable condition that indicates that a functional failure is either about to occur or is in the process of occurring.

3.28 PROACTIVE MAINTENANCE

Maintenance undertaken before a failure occurs, in order to prevent the item from getting into a failed state (scheduled restoration, scheduled discard, and on-condition maintenance).

3.29 PROTECTIVE DEVICE OR PROTECTIVE SYSTEM

A device or system which is intended to avoid, eliminate or minimize the consequences of failure of some other system.

3.30 PRIMARY FUNCTION(S)

The function(s) which constitute the main reason(s) why a physical asset or system is acquired by its owner or user.

3.31 RUN-TO-FAILURE

A failure management policy that permits a specific failure mode to occur without any attempt to anticipate or prevent it.

3.32 SAFETY CONSEQUENCES

A classification of failure modes that could injure or kill a human being.

3.33 SCHEDULED DISCARD

A scheduled task that entails replacing an item at or before a specified age limit regardless of its condition at the time.

3.34 SCHEDULED RESTORATION

A scheduled task that restores the capability of an item at or before a specified interval (age limit), regardless of its condition at the time, to a level that provides a acceptable probability of survival to the end of another specified interval.

3.35 SECONDARY FUNCTIONS

Functions which a physical asset or system has to fulfill apart from its primary function(s), such as those needed to fulfill regulatory requirements and those which concern issues such as protection, control, containment, comfort, appearance, energy efficiency and structural integrity.

3.36 USER

A person or organization that operates and/or maintains an asset or system and may either suffer from or be held accountable for the consequences of a failure mode of that system.

4. ACRONYMS

BITE	Built-In Test Equipment
FFI	Failure-finding (task) interval
FMEA	Failure Mode and Effects Analysis
mm	Millimeters
MMF	Mean Time Between Multiple Failures
MTBF	Mean Time Between Failures
MTED	MTBF of the Protected Function
MTIVE	MTBF of the Protective Function
psi	Pounds per Square Inch
RCM	Reliability-Centered Maintenance
RPM	Revolutions Per Minute
UTIVE	Allowed Unavailability of the Protective Function

5. ASSET DEFINITION

“RCM is a specific process used to identify the policies which must be implemented to manage the failure modes which could cause the functional failure of any physical asset in a given operating context.” (SAE JA1011, 1.1)

In order to identify appropriate failure management policies for any physical asset or system, the asset or system must be defined. This entails selecting the asset/system, defining its boundaries, and identifying the most appropriate level of detail at which to carry out the analysis.

SAE JA1011 refers to the processes used to select suitable failure management policies, under the assumption that the asset/system concerned has already been selected and defined. It does not provide criteria for processes to be used for selecting and defining the assets or systems themselves, because such processes tend to be highly dependent on the type of asset/system and where, for what, and by whom it is being (or to be) used. However some general guidance on this topic is provided in Section 18 of this Guide.

6. FUNCTIONS

An RCM process that conforms to SAE JA1011 begins by asking the question, “What are the functions and associated desired standards of performance of the asset in its present operating context (functions)?” This section discusses the following four key concepts concerning functions that are listed in 5.1 of SAE JA1011:

a. Operating context

b. Primary and secondary functions

c. Function statements

d. Performance standards

6.1 Operating Context

"The operating context of the asset shall be defined." (SAE JA1011, 5.1.1)

The functions, failure modes, failure consequences, and failure management policies that will be applied to any asset will depend not only on what the asset is, but also on the exact circumstances under which it is to be used. As a result, these circumstances need to be clearly defined before attempting to answer the question quoted above.

An operating context statement for a physical asset typically includes a brief overall description of how it is to be used, where it is to be used, overall performance criteria governing issues such as output, throughput, safety, environmental integrity, and so on. Specific issues that should be documented in the operating context statement include:

a. Batch versus flow processes: whether the asset is operating in a batch (or intermittent) process or a flow (or continuous) process.

b. Quality standards: overall quality or customer service expectations, in terms of issues such as overall scrap rates, customer satisfaction measurements (such as on-time performance expectations in transportation systems, or rates of warranty claims for manufactured goods), or military preparedness.

c. Environmental standards: what organizational, regional, national, and international environmental standards (if any) apply to the asset.

d. Safety standards: whether any predetermined safety expectations (in terms of overall injury and/or fatality rates) apply to the asset.

e. Theater of operations: characteristics of the location in which equipment is to be operated (arctic versus tropical, desert versus jungle, onshore versus offshore, proximity to sources of supply of parts and/or labor, etc.).

f. Intensity of operations: in the case of manufacturing and mining, whether the process of which the equipment forms a part is to operate 24 hours per day, seven days per week, or at lower intensity. In the case of utilities, whether the equipment operates under peak load or base load conditions. In the case of military equipment, whether the failure management policies are designed for peacetime or wartime operations.

g. Redundancy: whether any redundant or standby capability exists, and if so what form it takes.

h. Work-in-process: the extent to which work-in-process stocks (if any) allow the equipment to stop without affecting total output or throughput.

i. Spares: whether any decisions have been made about the stocking of key spares that might impinge on the subsequent selection of failure management policies.

j. Market demand/raw material supply: whether cyclic fluctuations in market demand and/or the supply of raw materials are likely to impinge on the subsequent selection of failure management policies. (Such fluctuations may occur over the course of a day in the case of an urban transport business, or over the course of a year in the case of a power station, an amusement park, or a food processing business.)

In the case of very large or very complex systems, it might be sensible to structure the operating context in a hierarchical fashion, if necessary starting with the mission statement of the entire organization that is using the asset.

6.2 List of Functions

"All the functions of the asset/system shall be identified (all primary and secondary functions, including the functions of all protective devices)." (SAE JA1011, 5.1.2)

The objective of the RCM process is to develop a set of policies that preserve the functions of the asset or system under consideration to standards of performance that are acceptable to its owner/user. As a result, the RCM process starts by identifying all the functions of the asset in its operating context.

Functions can be divided into two categories: primary and secondary functions.

6.2.1 Primary Functions

The reason why any organization acquires any asset or system is to fulfill a specific function or functions. These are known as primary functions of the asset. For instance, the main reason why someone acquires a car may be "to transport up to five people at speeds up to 90 km an hour along suitable roads."

6.2.2 Secondary Functions

Most assets are expected to perform other functions, in addition to the primary functions. These are known as their secondary functions. Secondary functions are usually less obvious than primary functions. But the loss of a secondary function can still have serious consequences, sometimes more serious than the loss of a primary function. As a result, secondary functions often need as much if not more attention than primary functions, so they too must be clearly identified.

When identifying secondary functions, care should be taken not to overlook the following:

a. Environmental integrity

b. Safety/structural integrity

c. Control/containment/comfort

d. Appearance

e. Protective devices and systems

f. Economy/efficiency

g. Superfluous

These issues are discussed in more detail as follows.

6.2.2.1 Environmental Integrity

These functions define the extent to which the asset must comply with the corporate, municipal, regional, national, and international environmental standards or regulations that apply to that asset. These standards govern such things as the release of hazardous materials into the environment, and noise.

6.2.2.2 Safety

It is sometimes necessary to write function statements that deal with specific threats to safety that are inherent in the design or operation of the process (as opposed to safety threats that are a result of a functional failure). For example, the function of electrical insulation on a domestic appliance is "to prevent users from touching electrically live components."

6.2.2.3 Structural Integrity

Many assets have a secondary function of providing support for or a secure mount for another item. For example, while the primary function of a wall may be to protect people and equipment from the weather, it might also be expected to support the roof, or to bear the weight of shelves and pictures.

6.2.2.4 Control

In many cases, users not only want assets to fulfill functions to a given standard of performance, but they also want to be able to regulate the performance. This expectation is summarized in separate function statements. For example, a function of a cooling system may be to regulate temperature at will between one specific temperature and another. Indication and feedback form an important subset of the control category of functions.

6.2.2.5 Containment

Systems whose primary function is to store materials must also contain them. Similarly, systems that transfer materials—especially fluids—also have a containment function. These functions must be specified as well.

6.2.2.6 Comfort

Owners and users generally expect that their assets or systems will not cause pain or anxiety to operators or maintainers. These problems should of course be dealt with at the design stage. However deterioration or changing expectations can lead to unacceptable levels of pain or anxiety. The best way to ensure that this does not happen is ensure that the associated function statements are described precisely and that they fully reflect current standards.

6.2.2.7 Appearance

Appearance often constitutes an important secondary function. For example, the primary reason for painting most industrial equipment is to protect it from corrosion. However a bright color may be chosen to enhance its visibility for safety's sake, and this function should also be documented.

6.2.2.8 Protection

Protective functions avoid, eliminate, or minimize the consequences of the failure of some other function. These functions are associated with devices or systems that:

a. Warn operators of abnormal conditions (warning lights or alarms)

b. Shut down equipment in the event of a functional failure (shutdown mechanisms)

c. Eliminate or relieve abnormal conditions caused by a functional failure (relief mechanisms, fire suppression systems, life preservers)

d. Take over from a function that has failed (redundant structural components, stand-by plant)

e. Prevent dangerous situations from arising in the first place (warning signs, protective covers)

A protective function ensures that the failure of the function being protected is much less serious than it would be without the protection. The associated devices are incorporated into systems to reduce risk, so their functions should be documented with special care.

6.2.2.9 Economy/Efficiency

In most organizations, overall cost expectations are expressed in the form of expenditure budgets. However for specific assets, cost expectations can be addressed directly by secondary function statements concerning such things as energy consumption rates and the rate of attrition of process materials.

6.2.2.10 Superfluous Functions

Some systems incorporate items or components that are found to be completely superfluous. This usually happens when equipment or the way in which it is used has been modified over a period of years, or when new equipment has been overspecified.

Although such items have no positive function and are often costly to remove, they can in fact fail and thus reduce overall system reliability. To avoid this, some may require maintenance and so consume resources.

If they are removed, the associated failure modes and costs will also be removed. However, before their removal can be recommended with confidence, their functions need to be clearly identified and understood.

6.2.2.11 “Reliability” Functions

There is often a temptation to write “reliability” function statements such as “to operate seven days a week, 24 hours per day.” In fact, reliability is not a function in its own right. It is a performance expectation that pervades all the other functions. Overall reliability/availability goals can be documented in the context statement. The reliability of a specific asset is in fact managed by dealing appropriately with each of the failure modes that could cause each loss of function.

6.3 Describing Functions

“All function statements shall contain a verb, an object, and a performance standard (quantified in every case where this can be done).” (SAE JA1011, 5.1.3)

For example, Figure 1 shows a pump pumping water from one tank into another. The nominal capacity of the pump is 1000 liters per minute, and water is withdrawn from the tank at a maximum rate of 800 liters per minute. The primary function of this pump would be described as: “to pump water from tank X to tank Y at not less than 800 liters per minute.” Here the verb is “to pump,” the object is “water,” and the performance standard is “from tank X to tank Y at not less than 800 liters per minute.”

Protective function statements need special handling. These functions act by exception—in other words, when something else goes wrong—so the function statement should reflect this fact. This is usually done by incorporating the words “if” or “in the event of,” followed by a very brief summary of the circumstances or event that would activate the protection. For example, the function of a pressure safety valve may be described as follows: “To be capable of relieving the pressure in the boiler if it exceeds 250 psi.”

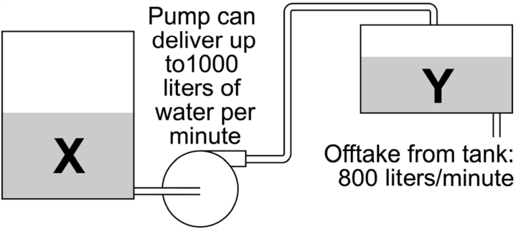
Pump can deliver up to1000 liters of water per minute
X
Y
Offtake from tank: 800 liters/minute

FIGURE 1 - FUNCTION OF A PUMP

6.4 Performance Standards

"Performance standards incorporated in function statements shall be the level of performance desired by the owner or user of the asset/system in its operating context." (SAE JA1011, 5.1.4)

Any organized system exposed to the real world will deteriorate—to total disorganization (also known as "chaos" or "entropy")—unless steps are taken to deal with whatever process is causing the system to deteriorate.

For example, centrifugal pumps are subject to impeller wear. This happens whether a pump moves acid or lubricating oil, and whether the impeller is made of titanium or mild steel. The only question is how fast the impeller will wear to the point that it can no longer pump fluid at the minimum required flow rate.

Once the performance of an asset drops below the minimum that is acceptable to the user, that asset has failed. Conversely, if the performance of the asset is maintained above this minimum, it continues to function to a level that is satisfactory to the user. As used in this Guide, "users" include the owners of the assets, the users of the assets—usually the operators—and society as a whole. Owners are satisfied if their assets generate a satisfactory return on the investment made to acquire them (usually financial return for commercial operations, or other measures for non-commercial operations). Users are satisfied if each asset continues to do whatever they want it to do to a standard of performance that they—the users—consider satisfactory. Finally, society as a whole is satisfied if assets do not fail in ways that threaten public safety and the environment.

In essence, this means that if we are seeking to cause an asset to continue to function to a level that is satisfactory to the user, then the objective of maintenance is to ensure that assets continue to perform above the minimum level that is acceptable to those users. If it were possible to build an asset that could deliver the minimum performance without deteriorating in any way, then it would be able to run continuously, with no need for maintenance.

However, deterioration is inevitable, so it must be allowed for. This means that when any asset is put into service, it must be able to deliver more than the minimum standard of performance desired by the user. What the asset is able to deliver at this point in time is known as its initial capability. Figure 2 shows the correct relationship between this capability and desired performance.

This means that performance can be defined in two ways:

a. Desired performance (what the user wants the asset to do)

b. Built-in capability (what it can do)

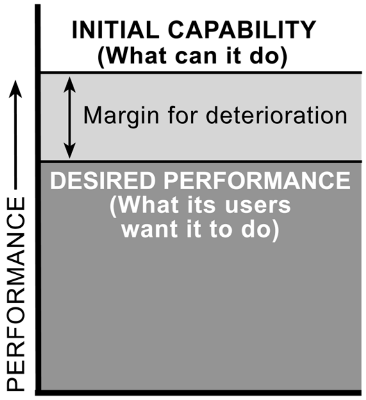

FIGURE 2 - ALLOWING FOR DETERIORATION

The margin for deterioration must be large enough to allow for a reasonable amount of use before the component degrades to functional failure, but not so large that the system is "over-designed" and hence too expensive. In practice, the margin is adequate in the case of most components, so it is usually possible to develop maintenance programs accordingly.

However if the desired performance is higher than built-in capability, no amount of maintenance can deliver the desired performance, in which case the asset is not maintainable.

All this means that, in order to ascertain whether an asset can be maintained, we need to know both kinds of performance: the built-in capability of the asset, and the minimum performance that the user is prepared to accept in the context in which the asset is being used. This minimum performance is the performance standard that must be incorporated in the function statement.

For example, the initial capability of the pump in Figure 1 is 1000 liters per minute, and the rate at which the water is being withdrawn from the tank (offtake) is 800 liters per minute. In this context, the pump is fulfilling the expectations of its users as long as it continues to pump water into the tank faster than the water is being withdrawn. As a result, the primary function of the pump was described as being "to pump water from tank X to tank Y at not less than 800 liters per minute," and not "to pump at 1000 liters per minute."

Note that if the same pump were to be used in a situation where the offtake from the tank was (say) 900 liters per minute, then the primary function would read, "to pump water from tank X to tank Y at not less than 900 liters per minute," and the maintenance program would be changed to reflect this new performance expectation.

Note that users and maintainers often have significantly different views about what constitutes acceptable performance. As a result, in order to avoid misunderstandings about what constitutes "functional failure," the minimum standards of acceptable performance must be clearly understood and accepted by the users and maintainers of the asset, together with anyone else who has a legitimate interest in the behavior of the asset.

Performance standards must be quantified where possible, because quantitative standards are clearer and more precise than qualitative standards. Occasionally it is only possible to use qualitative standards, for example when dealing with functions relating to appearance. In such cases, special care must be taken to ensure that the qualitative standard is understood and accepted by users and maintainers of the asset.

7. FUNCTIONAL FAILURES

An RCM process that conforms to SAE JA1011 then asks the question, "In what ways can it fail to fulfill its functions (functional failures)?" In order to answer this question satisfactorily, SAE JA1011, 5.2 states that "all the failed states associated with each function shall be identified."

Section 6 explained that an asset is failed if it is incapable of doing what its users want it to do. It also explained that what the asset must do is defined as a function, and that every asset has more than one (and often several) different functions. Since it is possible for each one of these functions to fail, it follows that any asset can suffer from a variety of failed states.

For example, the primary function of the pump in Figure 1 was "to pump water from tank X to tank Y at not less than 800 liters per minute," while a secondary function is "to contain the water in the pump." It is possible for such a pump to be capable of pumping the required amount of water (not failed in terms of its primary function) while leaking excessively (failed in terms of its secondary function). Conversely, it is equally possible for the pump to deteriorate to the point where it cannot pump the required amount (failed in terms of its primary function) while it still contains the required liquid (not failed in terms of its secondary function).

For this reason, it is more accurate to define failure in terms of the loss of specific functions, rather than failure of an asset as a whole. The previous example also shows why the RCM process uses the term "functional failure" to describe failed states, rather than "failure" on its own. (Note that RCM distinguishes between a functional failure, or failed state, and a "failure mode," which is an event that causes a failed state.)

Two further points that need to be considered when defining functional failures are: partial and total failure, and upper and lower limits.

7.1 Partial and Total Failure

Functional failures that represent total failure of the function are relatively easy to identify. For example, it is clear that the pump mentioned in 6.3, will have suffered a functional failure if it fails to pump any water at all ("total failure"). However the pump will also have suffered a functional failure if it can pump water but the rate at which it does so is less than 800 liters per minute.

The second failed state in this example is known as a "partial failure." Partial failures need to be identified separately because they are nearly always caused by different failure modes from total failures, and because the consequences are also nearly always different.

Bear in mind that partial failure is not the same as deterioration below initial capability. Everything deteriorates below initial capability after some use, and such deterioration can be tolerated as long as it does not reach the point that is unacceptable to the user of the asset, as shown in Figure 2. Deterioration only becomes functional failure (partial or total) when the performance drops below the minimum level required by the user.

7.2 Upper and Lower Limits

The performance standards associated with some functions incorporate upper and lower limits. These limits mean that the asset has failed if it performs above the upper limit or below the lower limit. In these cases, the breach of the upper limit needs to be documented separately from the breach of the lower limit. This is because the failure modes and/or consequences associated with exceeding the upper limit are usually different from those associated with going below the lower limit.

For example, the primary function of a grinding machine may be listed as: "To grind bearing journals in a cycle time of 3.00 min ± 0.03 min, to a diameter of 75 mm ± 0.1 mm, with a surface finish of no greater than Ra 0.2." This machine has failed if:

a. It stops altogether

b. It grinds workpiece in a cycle time longer than 3.03 min

c. It grinds workpiece in a cycle time less than 2.97 min

d. Diameter exceeds 75.1 mm

e. Diameter is below 74.9 mm

f. Surface finish too rough (exceeds Ra 0.2)

8. FAILURE MODES

An RCM process that conforms to SAE JA1011 then asks the question, "What causes each functional failure (failure modes)?" This section discusses the following five key concepts concerning failure modes that are listed in 5.3 of SAE JA1011:

a. Identifying failure modes

b. Establishing what is meant by "reasonably likely"

c. Levels of causation

d. Sources of information

e. Types of failure modes

8.1 Identifying Failure Modes

“All failure modes reasonably likely to cause each functional failure shall be identified.” (SAE JA1011, 5.3.1)

Section 7 of this Guide mentioned that RCM distinguishes between the failed state of the asset (functional failure) and the events that cause the failed states (failure modes). Because it is impossible to define causes of failure until we have established exactly what we mean by “failed,” the RCM process identifies functional failures before failure modes. The way in which this is usually documented is shown in Figure 3, for the primary function of the pump illustrated in Figure 1. Figure 3, which lists an asset’s functions, functional failures, and failure modes, shows nearly all elements of a Failure Mode and Effects Analysis (FMEA). The “effects” of each failure mode are listed in a subsequent step (see Section 9 of this Guide).

Figure 3 also shows that at the very least, a description of a failure mode should consist of a noun and a verb. The description should contain enough detail for it to be possible to select an appropriate failure management policy, but not so much detail that excessive amounts of time are wasted on the analysis process itself.

In particular, the verbs used to describe failure modes should be chosen with care, because they strongly influence the subsequent failure management policy selection process. For instance, verbs such as “fails” or “breaks” or “malfunctions” should be used sparingly, because they give little or no indication as to what might be an appropriate way of managing the failure mode. The use of more specific verbs makes it possible to select from the full range of failure management options.

For example, failure mode 1A4 in Figure 3 might have been phrased, “coupling fails.” However such a phrase would provide no clue as to what might be done to anticipate or prevent the failure mode. If we say “coupling bolts come loose” or “coupling hub shears due to fatigue,” then it becomes much easier to identify a possible proactive task.

ASSET: Pumping System					
FUNCTION		**FUNCTIONAL FAILURE (Loss of Function)**		**FAILURE MODE (Cause of Failure)**	
1	Transfer water from tank X to tank Y at not less than 800 liters/minute	A	Unable to transfer any water at all	1	Bearing seizes
				2	Motor burns out
				3	Impeller comes adrift
				4	Coupling hub shears due to fatigue
				5	Inlet valve jams closed
				6	Impeller jammed by foreign object … etc
		B	Transfers less than 800 liters/minute	1	Impeller worn
				2	Partially blocked suction line … etc

FIGURE 3 - FAILURE MODES OF A PUMP

For valves, switches, and similar devices, the failure mode description should indicate whether the loss of function is caused by the item failing in the open or closed position. “Valve jams closed” says more than “valve jams.” Furthermore, the purpose of identifying failure modes is to identify the cause of the functional failure so that a way may be found to anticipate or prevent it. As a result, it may sometimes be necessary to take this one step further, such as: “valve jams closed due to rust on lead screw.” In this context, the use of the word “rust” suggests that it may be appropriate to focus failure management efforts on detecting or controlling rust.

8.2 Establishing What Is Meant By "Reasonably Likely"

"The method used to decide what constitutes a 'reasonably likely' failure mode shall be acceptable to the owner or user of the asset." (SAE JA1011, 5.3.2)

Section 8.1 mentioned that all failure modes reasonably likely to cause each functional failure shall be identified. "Reasonable likelihood" means just that: a likelihood that meets the test of reasonableness, when applied by trained and knowledgeable people. (A term often used instead of "reasonable" in this context is the term "credible.") If people who are trained to use RCM, and who are knowledgeable about the asset in its operating context, agree that the probability that a specific failure mode could occur is sufficiently high to warrant further analysis, then that failure mode should be listed.

In practice, it is sometimes very difficult to decide whether a failure mode should or should not be listed. This issue is related to both probability of occurrence and the level of detail used to describe failure modes. Too few failure modes, and/or too little detail, leads to superficial and sometimes dangerous analysis. Too many failure modes, and/or too much detail, causes the entire RCM process to take much longer than it needs to. In extreme cases, this can cause the process to take two or even three times longer than necessary (a phenomenon known as "analysis paralysis"), and can also lead to excessively cumbersome maintenance programs.

In situations where there may be doubt or disagreement as to what constitutes the threshold of "reasonableness," the final decision must be taken by the organization that owns or uses the asset, because that organization will ultimately be held accountable for the consequences if the failure mode does occur.

Note that the decision to list a failure mode should be tempered by a consideration of its consequences. If the consequences are likely to be very severe indeed, then less likely failure modes should be listed and subjected to further analysis.

For example, if the pump described in Figure 3 was installed in a food factory or a vehicle assembly plant, the failure mode "casing smashed by an object falling from the sky" would be dismissed immediately as being laughably unlikely. However if the same pump were a primary coolant pump in a nuclear reactor at a commercial power plant, this failure mode is more likely to be taken seriously—even though it is still highly improbable. (Appropriate failure management policies might be either to ban aircraft from flying over the facility, or to design a roof that can withstand a crashing aircraft. This is not simply speculation, of course—both policies are routinely considered at nuclear power stations.)

8.3 Levels of Causation

"Failure modes shall be identified at a level of causation that makes it possible to identify an appropriate failure management policy." (SAE JA1011, 5.3.3)

Previous sections of this Guide stated that failure modes should be described in enough detail for it to be possible to select an appropriate failure management policy, but not in so much detail that excessive amounts of time are wasted on the analysis process itself.

The extent to which failure modes can be described at different levels of detail is illustrated in Figure 4, which is based on the pump whose functions and functional failures were described in Figure 3. Figure 4 lists some of the failure modes that might cause the functional failure "unable to transfer any water at all." In this example, these failure modes are considered in seven levels of detail, beginning with the failure of the pump set as a whole.

The first point to emerge from this example is the connection between the level of detail and the number of failure modes listed. The example shows that the further one "drills down" in an FMEA, the larger the number of failure modes that can be listed. For example, there are 3 failure modes listed for the pump at level 3 in Figure 4, but 20 at level 6.

One other point that arises from Figure 4 is "root causes." It is discussed as follows.

Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
Pump set fails	Pump fails	Impeller fails	Impeller comes adrift	Mounting nut undone	Nut not tightened correctly	Assembly error
				Mounting nut worn away	Nut eroded/corroded away	
					Nut made of wrong material	Wrong material specified
						Wrong material supplied
				Impeller nut cracked	Impeller nut overtightened	Assembly error
					Nut made of wrong material	Wrong material specified
						Wrong material supplied
				Impeller key sheared	Wrong key steel specified	Design error
					Wrong key steel supplied	Procurement error
						Storekeeping error
						Requisitioning error
			Object smashes impeller	Part in system after maintenance	Assembly error	See (human error)
				Foreign object enters system	Suction strainer not installed	Assembly error
					Strainer holed by corrosion	
		Casing ruptured	Casing bolts come loose	Casing bolts undertightened	Assembly error	See (human error)
				Bolts loosened by vibration		
				Casing bolts corroded away		
				Bolts fail due to fatigue		
			Casing joint fails	Joints incorrectly fitted	Assembly error	See (human error)
				Joint fails due to fretting		
			Casing smashed	Casing smashed by vehicle	Operating error	See (human error)
				Smashed by object from sky	Casing hit by meteorite	
					Casing hit by part of aircraft	
		Pump seal fails	Normal wear and tear	Seal abraded		
			Pump runs dry	*See "water supply fails" below*		
			Seal misaligned	Assembly error	See (human error)	
			Seal faces dirty	Assembly error	See (human error)	
			Wrong seal fitted	Wrong seal supplied	Procurement error	See (human error)
					Storekeeping error	See (human error)
				Wrong seal specified	Design error	See (human error)
			Damaged seal installed	Pump seal dropped in stores	Storekeeping error	See (human error)
				Pump seal damaged in transit	Procurement error	See (human error)
	Motor fails	Etc				
	Driveline fails	Etc				
	Valve closed	Etc				
	Power fails	Etc				

FIGURE 4 - FAILURE MODES AT DIFFERENT LEVELS OF DETAIL

8.3.1 Root Causes

The term “root cause” is often used in connection with the analysis of failures. It implies that it is possible to arrive at a final and absolute level of causation, if only one drills down far enough. In fact, this is not only very difficult to do but it is also usually unnecessary.

For instance, in Figure the failure mode “impeller nut comes adrift” is listed at level 3, which in turn is caused by “impeller nut cracked” at level 4. If we were to go down one level further, this might have been caused by “impeller nut overtightened” (level 5), which might have been caused in turn by “assembly error” (level 6). The assembly error might have occurred because the “technician was distracted” (level 7). He might have been distracted because his “child was ill” (level 8). This failure mode might have occurred because the “child ate bad food in restaurant” (level 9).

Clearly, this process of drilling down could go on almost forever—way beyond the point at which the organization responsible for operating and maintaining the asset has any control over the failure modes. This is why SAE JA1011 requires an RCM process to identify failure modes at a level of causation that makes it possible to identify an appropriate failure management policy. This level will vary for different failure modes. Some failure modes might be identified at level 3, others at level 5, and the rest at other levels.

Note that some of the failure modes shown in Figure 4 might not be considered reasonably likely in a context different from that used to develop Figure 4. In that case, there would be no reason to list them at all. Conversely, other failure modes that are not shown in Figure 4 but that are considered to be reasonably likely in that context might be added to the list. Note also that the failure modes listed in Figure 4 only apply to the functional failure, “unable to transfer water at all.” Figure 4 does not show failure modes that would cause other functional failures, such as loss of containment or loss of protection.

8.4 Sources of Information about Failure Modes

“Lists of failure modes shall include failure modes that have happened before, failure modes that are currently being prevented by existing maintenance programs and failure modes that have not yet happened but that are thought to be reasonably likely (credible) in the operating context.” (SAE JA1011, 5.3.4)

Failure modes that have occurred before on the same or similar assets are the most obvious candidates for inclusion in the list of failure modes, unless something has been changed in such a way that the failure mode cannot occur again. Sources of information about these failure modes include people who know the asset well (operators, maintainers, equipment vendors, or other users of the same equipment), technical history records, and data banks.

Failure modes that are the subject of existing proactive maintenance routines should also be incorporated in the list of failure modes. One way to ensure that none of these failure modes has been overlooked is to study existing maintenance schedules for identical or very similar assets and ask, “what failure mode would occur if this task was not performed?” However existing schedules should only be reviewed as a final check after the rest of the RCM analysis has been completed, in order to reduce the possibility of perpetuating the status quo.

Finally, the list of failure modes should include failure modes that have not yet occurred but that are considered to be real possibilities in the context under consideration. Identifying and deciding how to deal with failure modes that have not happened yet is an essential feature of proactive management in general, and of risk management in particular. It is also one of the most challenging aspects of the RCM prospect, because it calls for a high degree of judgment applied by skilled and knowledgeable people.

8.5 Types of Failure Modes

"Lists of failure modes should include any event or process that is likely to cause a functional failure, including deterioration, design defects, and human error whether caused by operators or maintainers (unless human error is being actively addressed by analytical processes apart from RCM)." (SAE JA1011, 5.3.5)

Deterioration occurs when the capability of an asset is above desired performance to begin with, but then drops below the desired performance after the asset is put into service. It covers all forms of "wear and tear," such as fatigue, corrosion, abrasion, erosion, evaporation, degradation (especially of insulation, lubricants, etc.), and so on. These failure modes should of course be included in a list of failure modes wherever they are thought to be reasonably likely, at the level of detail that is most appropriate as discussed in 8.3.

In some cases, the design of an asset or the configuration of a system might render it incapable of fulfilling the full range of functional requirements in the context in which it is expected to operate. If such deficiencies are known to affect existing equipment, or if in the case of new equipment it is thought that existing design and construction management processes are unlikely to detect and rectify such deficiencies, these failure modes should be listed so that appropriate failure management policies can be identified later in the analysis.

Many functional failures are caused when the stress applied to an asset increases beyond its ability to withstand the stress. In practice, these stress increases are often applied by human beings. The literature on the subject classifies such human errors in a wide variety of ways. However in the world of physical assets these errors usually fall into one of the following categories:

a. Incorrect operation. This usually takes two forms. The first is sustained, often deliberate, overloading (for instance, if a machine is operated at performance levels that approach or exceed its initial capability, such as an automobile engine that is persistently operated at excessively high RPM, causing it to fail prematurely). The second is sudden, usually unintentional, overloading (for instance, if an asset is simply operated incorrectly, such as a vehicle that is put into reverse while it is moving forward, damaging the gearbox).

b. Incorrect assembly (for instance, if a mechanic leaves a tool behind in a gearbox or an electrician wires a switch incorrectly)

c. External damage (for instance, if the casing of a pump is smashed by a forklift truck)

If such increases in applied stress are thought to be reasonably likely in the context under consideration (and if they are not already being dealt with by a separate analytical process), they should also be incorporated in the list of failure modes, so that appropriate failure management policies can be identified.

9. FAILURE EFFECTS

An RCM process that conforms to SAE JA1011 then asks the question, "What happens when each failure occurs (failure effects)?" This section discusses the following two key concepts concerning failure effects that are listed in 5.4 of SAE JA1011:

a. Basic assumptions

b. Information needed

9.1 Basic Assumptions

"Failure effects shall describe what would happen if no specific task is done to anticipate, prevent or detect the failure" (SAE JA1011, 5.4.1)

A failure effect statement describes what would happen if a failure mode were to occur. Note that RCM makes a clear distinction between a failure effect (what happens) and a failure consequence (how, and how much, the failure mode matters).

As explained in Section 10 of this Guide, failure effects statements are used to assess the consequences of each failure mode. They also provide the basic information needed to decide what failure management policies must be implemented to avoid, eliminate or minimize these consequences to the satisfaction of the owners/users of the asset.

The main failure management policy options include proactive maintenance tasks (on-condition, scheduled restoration, and scheduled discard), together with associated frequencies. If we wish to identify these tasks correctly, it is essential to assume that no proactive maintenance is being carried out, when identifying the failure modes and associated effects. In other words, in order to start from a true zero base, it is essential to assume that the failure mode does in fact cause the associated functional failure. Failure modes need to be described, and failure effect statements need to be written, accordingly.

9.2 Information Needed

"Failure effects shall include all the information needed to support the evaluation of the consequences of the failure, such as:

a. What evidence (if any) that the failure has occurred (in the case of hidden functions, what would happen if a multiple failure occurred)

b. What it does (if anything) to kill or injure someone, or to have an adverse effect on the environment

c. What it does (if anything) to have an adverse effect on production or operations

d. What physical damage (if any) is caused by the failure

e. What (if anything) must be done to restore the function of the system after the failure"

(SAE JA1011, 5.4.2)

9.2.1 Evidence That Failure Has Occurred

A failure effect statement should describe whether there is any evidence that the failure mode under consideration has occurred on its own. If so, it should describe what form this evidence takes. For instance, it should mention whether the behavior of the equipment changes noticeably as a result of the failure mode (warning lights, alarms, change in speed or noise levels, etc.). It should also describe whether the failure mode is accompanied (or preceded) by obvious physical effects such as loud noises, fire, smoke, escaping steam, unusual smells, or pools of liquid on the floor.

When dealing with protection, failure effect descriptions should state briefly what would happen if the protected function were to fail while the protection was in a failed state.

9.2.2 Threat to Safety or Environment

If there is a possibility that someone could get injured or killed as a direct result of the failure mode, or an environmental standard or regulation could be breached, the failure effect should describe how this could happen. A selected list of examples includes:

a. Increased risk of fire or explosion

b. The escape of hazardous chemicals

c. Electrocution

d. Vehicle accidents or derailments

e. Ingress of dirt into food or pharmaceutical products

f. Exposure to sharp edges or moving machinery

When listing these effects, care should be taken not to say that the failure mode “has safety consequences” or “affects the environment.” Simply state what happens, and leave the evaluation of the consequences to the next stage of the RCM process.

9.2.3 Effect Upon Production or Operations

Failure effect descriptions should indicate how production or operations is affected (if at all), and for how long. The following issues should be considered.

a. Downtime: How much time the asset would be out of service owing to this failure mode, from the moment it fails until the moment it is fully operational again. In order to ensure that the failure management program is reasonably conservative (but not too conservative), it should be assumed that the failure mode occurs in a “typical worst case” situation, for instance, late at night in a factory, or if mobile equipment is in a more remote location than usual.

b. Speed of operation: Whether the equipment has to slow down as a result of the failure mode and if so by how much

c. Quality: Whether the failure mode affects the quality with which the function is performed, such as the accuracy or precision of guidance or control systems, product quality parameters, and even customer service issues (on-time performance, etc.). The failure effect statement should also indicate whether the failure mode increases reject or scrap rates, causes a mission abort, or incurs significant contractual financial penalties.

d. Other systems: Whether other equipment or processes have to stop, slow down, or are otherwise affected by the failure mode.

e. Overall operating costs: Whether the failure mode causes any other increases in operating costs, such as increased energy consumption or excessive attrition of process materials

9.2.4 Secondary Damage

If the failure mode under consideration causes significant damage to other components or systems, the effects of this secondary damage should also be recorded.

9.2.5 Corrective Action Required

The failure effects description should include a brief description of the action that is required to correct the failure mode after it has occurred.

10. FAILURE CONSEQUENCE CATEGORIES

10.1 Consequence Categories

“The consequences of every failure mode shall be formally categorized...” (SAE JA1011, 5.5.1)

After each reasonably likely failure mode and its effects have been identified at an appropriate level of detail, the next step in the RCM process is to assess the consequences of each failure mode. The primary source of information used to assess failure consequences is the description of the failure effects.

Some failure modes affect output, product quality or customer service. Others threaten safety or the environment. Some increase operating costs, for instance by increasing energy consumption, while a few have an impact in four, five or even all six of these areas. Still others may appear to have no effect at all if they occur on their own, but may expose the organization to the risk of much more serious failure modes.

If any of these failure modes are not anticipated or prevented, the time and effort that need to be spent correcting them also affects the organization, because repairing them consumes resources that might be better used elsewhere.

The nature and severity of these effects govern the way in which each failure mode is viewed by the organization. The precise impact in each case—in other words, the extent to which each failure mode matters—depends on the operating context of the asset, the performance standards that apply to each function, and the physical effects of each failure mode.

This combination of context, standards and effects means that every failure mode has a specific set of consequences associated with it. If the consequences are very serious, then considerable efforts will be made to prevent the failure mode, or at least to anticipate it in time to reduce or eliminate the consequences. On the other hand, if the failure mode only has minor consequences, it is possible that no proactive action will be taken and the failure mode will simply be corrected each time it occurs.

This means that the consequences of failure modes are more important than their technical characteristics. It also suggests that the whole idea of failure management is not so much about anticipating or preventing failure modes per se as it is about avoiding or reducing their consequences.

The remainder of this section considers the criteria used to evaluate the consequences of failure modes, and hence to decide whether any form of failure management is worth doing. These consequences are divided into four categories in two stages. The first stage separates hidden failures from evident failures.

10.1.1 Hidden and Evident Failures

“The consequence categorization process shall separate hidden failure modes from evident failure modes.” (SAE JA1011, 5.5.1.1)

Some failure modes occur in such a way that nobody knows that the item is in a failed state unless, or until, some other failure (or abnormal event) also occurs. These are known as hidden failures. A hidden failure is a failure mode whose effects do not become apparent to the operating crew under normal circumstances if the failure mode occurs on its own. Conversely, an evident failure is a failure mode whose effects become apparent to the operating crew under normal circumstances if the failure mode occurs on its own.

The RCM approach to the evaluation of failure consequences begins by separating hidden failures from evident failures. Hidden failures can account for up to half the failure modes that could affect modern, complex equipment, so they need to be handled with special care. The following paragraphs explain the relationship between hidden failures and protection, and introduce the concept of a “multiple failure.”

Hidden Failures and Protection: 6.2.2.8 of this Guide mentioned that the function of any protection is to ensure that the consequences of the failure of the protected function are much less serious than they would be if there were no protection. So any protective function is in fact part of a system with at least two components:

a. The protective function

b. The protected function

The existence of such systems creates two sets of failure possibilities, depending on whether the failure of the protection is evident or not. The implications of each set are considered in the following paragraphs, starting with devices whose failure is evident.

10.1.1.1 Evident Failures of Protective Functions

In this context, an "evident" failure of a protective function is where the effects of the failure mode on their own will become apparent to the operating crew under normal circumstances. The existence of such failure modes creates three possible scenarios in any period, as follows.

The first possibility is that neither the protective function nor the protected function fails. In this case everything proceeds normally.

The second possibility is that the protected function fails before the protection. In this case the protection carries out its intended function and, depending on the nature of the protection, the consequences of failure of the protected function are reduced or eliminated.

The third possibility is that the protective function fails before the protected function. Because this failure is "evident," the loss of the protection would become apparent. In this situation, the chance of the protected function failing while the protective function is in a failed state can be almost eliminated, either by shutting down the protected function or by providing alternative protection until the failed protective function is restored, as illustrated in Figure 5. This in turn means that the consequences of an evident failure of a protective function usually fall into the "operational" or "non-operational" categories, as discussed as in 10.1.2.

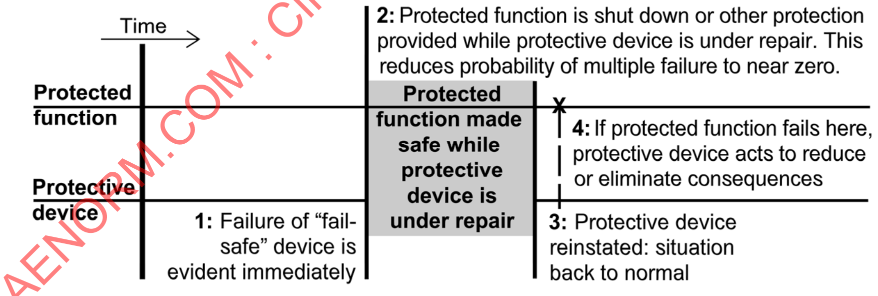
Time
Protected function
Protective device
2: Protected function is shut down or other protection provided while protective device is under repair. This reduces probability of multiple failure to near zero.
Protected function made safe while protective device is under repair
X
4: If protected function fails here, protective device acts to reduce or eliminate consequences
1: Failure of "fail-safe" device is evident immediately
3: Protective device reinstated: situation back to normal

FIGURE 5 - EVIDENT FAILURE OF A PROTECTIVE FUNCTION

10.1.1.2 Protective Functions Whose Failure Is Not Evident

Hidden failures can be identified by asking the following question:

Will any of the effects of this failure mode become evident to the operating crew under normal circumstances if the failure mode occurs on its own?

If the answer to this question is no, the failure mode is hidden. If the answer is yes, it is evident. Note that in this context, "on its own" means that nothing else has failed. Note also that we assume at this point in the analysis that no attempts are being made to check whether the associated function is still working. This is because such checks are a form of scheduled maintenance, and the whole purpose of the analysis is to find out whether such maintenance is necessary.

If such a failure mode occurs, the fact that the protection is unable to fulfil its intended function will not become apparent under normal circumstances. The existence of such failure modes creates four possible scenarios in any period, two of which also applied to evident failures of protective functions. The first is where neither function fails, in which case everything proceeds normally as before.

The second possibility is that the protected function fails at a time when the protection is still functional. In this case the protection also carries out its intended function, so the consequences of the failure of the protected function are again reduced or eliminated altogether.

The third possibility is that the protection fails while the protected function is still working. In this case, the loss of protection has no direct consequences. In fact no-one even knows that the protection is in a failed state.

The fourth possibility during any one cycle is that the protection fails, then the protected function fails while the protection is in a failed state. This situation is known as a multiple failure. (This is a real possibility simply because the failure of the protection is not evident, so no-one would be aware of the need to take corrective—or alternative—action to avoid the multiple failure.)

The sequence of events leading to a multiple failure is summarized in Figure 6.

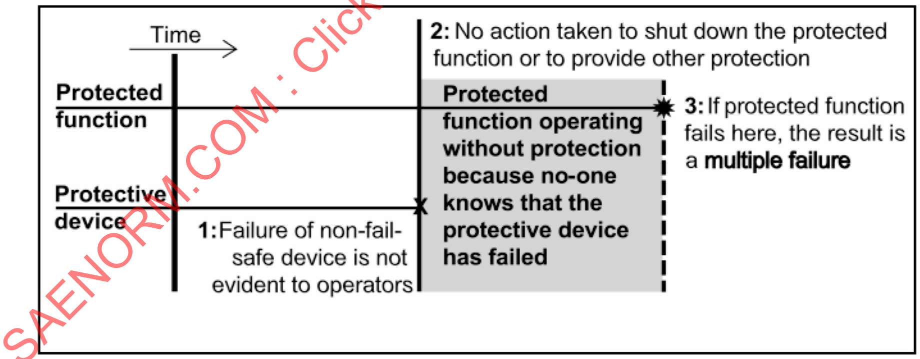
Time
Protected function
Protective device
1: Failure of non-fail-safe device is not evident to operators
2: No action taken to shut down the protected function or to provide other protection
Protected function operating without protection because no-one knows that the protective device has failed
3: If protected function fails here, the result is a multiple failure

FIGURE 6 - HIDDEN FAILURE OF A PROTECTIVE FUNCTION

10.1.2 Safety, Environmental, Operational, and Non-Operational Consequences

The consequence categorization process shall clearly distinguish events (failure modes and multiple failures) that have safety and/or environmental consequences from those that only have economic consequences (operational and non-operational consequences)." (SAE JA1011, 5.5.1.2)

NOTE: Throughout this section, "failure" refers to a failure mode or a multiple failure.

10.1.2.1 Safety consequences

A failure has safety consequences if there is an intolerable probability that it could kill or injure a human being. The distinction between a “tolerable” and an “intolerable” probability is discussed in more detail in 12.1.3 of this Guide.

10.1.2.2 Environmental Consequences

At another level, “safety” refers to the safety or well-being of society in general. Such failures tend to be classed as “environmental” issues. Society's expectations take the form of municipal, regional and national environmental standards. Some organizations also have their own even more stringent corporate standards. As a result, a failure has environmental consequences if there is an intolerable probability that it could breach any known environmental standard or regulation.

10.1.2.3 Operational Consequences

The primary function of most equipment in commerce and industry is usually connected with the need to earn revenue or to support revenue-earning activities. Failures that affect the primary functions of these assets affect the revenue-earning capability of the organization. The magnitude of these effects depends on how heavily the equipment is utilized and the availability of alternatives. However, in nearly all cases the costs of these effects are greater—often much greater—than the cost of repairing the failures, and these costs need to be taken into account when assessing the cost-effectiveness of any failure management policy. In general, failures affect operations in four ways:

a. They affect total output or throughput.

b. They affect product quality.

c. They affect customer service (and may incur financial penalties).

d. They increase operating costs in addition to the direct cost of repair.

In non-profit enterprises like military undertakings, many failures also affect the ability of the organization to fulfil its primary function, sometimes with devastating results. While it may be difficult to cost out the results of losing a battle or even a war, failures that affect operational capability still have economic implications. If they occur too often, it may be necessary to deploy (say) 60 battle tanks instead of 50, or six aircraft carriers instead of five. Redundancy on this scale can be very expensive.

For this reason, if an evident failure does not pose a threat to safety or the environment, the RCM process focuses next on the operational consequences of failure.

Because these consequences tend to be economic in nature, they are usually evaluated in economic terms. However, in more extreme cases (such as losing a war), the “cost” may have to be evaluated on a qualitative basis. In practice, overall economic effect of any failure that has operational consequences depends on two factors:

a. How much the failure costs each time it occurs, in terms of its effect on operational capability plus the cost of repairing the failure (and any secondary damage)

b. How often it happens

10.1.2.4 Non-Operational Consequences

The consequences of an evident failure that has no direct adverse effect on safety, the environment or operational capability are classified as non-operational. The only consequences associated with these failures are the direct costs of repairing the failure itself and any secondary damage, so these consequences are also economic.

10.1.3 RCM and Safety Legislation/Regulations

A question often arises concerning the relationship between RCM and tasks specified by regulatory authorities (environmental legislation is dealt with directly).

Most regulations governing safety merely demand that users are able to demonstrate that they are doing whatever is prudent to ensure that their assets are safe. This has led to rapidly increasing emphasis on the concept of an audit trail, which basically requires users of assets to be able to produce documentary evidence that there is a rational, defensible basis for their maintenance programs. In the vast majority of cases, RCM wholly satisfies this type of requirement.

However, some regulations demand that specific tasks should be done on specific types of equipment at specific intervals. It quite often happens that the RCM process suggests a different task and/or a different interval, and in most of these cases, the RCM-derived task is a superior failure management policy. However, in such cases, it is wise to continue doing the task specified by the regulations and to discuss the suggested change with the appropriate regulatory authority.

10.2 Assessing Failure Consequences

“The assessment of failure consequences shall be carried out as if no specific task is currently being done to anticipate, prevent, or detect the failure.” (SAE JA1011, 5.5.2)

For reasons explained in 9.1 of this Guide, it is essential to assume that no proactive maintenance is being carried out when identifying failure consequences.

11. FAILURE MANAGEMENT POLICY SELECTION

11.1 The Relationship Between Age and Failure

“The failure management selection process shall take account of the fact that the conditional probability of some failure modes will increase with age (or exposure to stress), that the conditional probability of others will not change with age, and the conditional probability of yet others will decrease with age.” (SAE JA1011, 5.6.1)

One of the most important factors that affects the selection of any failure management policy is the relationship between age (or exposure to stress) and failure. There are six sets of ways in which the conditional probability of failure varies as an item gets older, as shown in Figure 7.

Patterns A and B both display a point at which there is a rapid increase in the conditional probability of failure (sometimes called a “wear-out zone”). Pattern C shows a steady increase in the probability of failure, but no distinct wear-out zone. Pattern D shows low conditional probability of failure when the item is new or just out of the shop, then a rapid increase to a constant or very slowly increasing level, while pattern E shows a constant conditional probability of failure at all ages (random failure). Pattern F starts with high infant mortality, dropping to a constant or very slowly decreasing conditional probability of failure.

In general, age-related failure patterns apply to items that are very simple, or to complex items that suffer from a dominant failure mode. In practice, they are commonly associated with direct wear (most often where equipment comes into direct contact with the product), fatigue, corrosion, oxidation and evaporation.

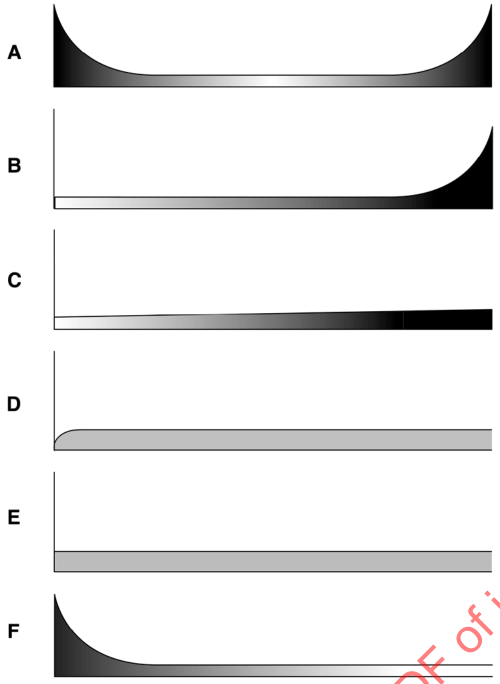

FIGURE 7 - SIX PATTERNS OF FAILURE

11.2 Technically Feasible and Worth Doing

“All scheduled tasks shall be technically feasible and worth doing (applicable and effective) ... “ (SAE JA1011, 5.6.2)

Any scheduled task is only worth doing if it reduces (avoids, eliminates or minimizes) the consequences of the failure mode to an extent that justifies the direct and indirect costs of doing the task. (Note that in this context built-in monitoring devices constitute a “scheduled task,” albeit one that is being done automatically—either continuously or at predetermined intervals—by the monitoring device. Therefore they should be subjected to the same selection criteria as any other form of scheduled task. Note too that such devices themselves require design, installation and maintenance, which should also be considered when assessing their cost-effectiveness.)

If an appropriate scheduled task cannot be found, and if the consequences of the failure mode are not acceptable to the owner or user of the asset, then some other way must be found to deal with the failure consequences.

Of course, it also has to be technically possible for any failure management policy to influence the failure consequences. Whether or not any such policy is technically feasible (or applicable) depends on the technical characteristics of the policy and of the failure mode under consideration. The criteria governing technical feasibility are discussed in more detail in Sections 12 through 14 of this Guide.

11.3 Cost Effectiveness

“If two or more proposed failure management policies are technically feasible and worth doing (applicable and effective), the policy that is most cost-effective shall be selected.” (SAE JA1011, 5.6.3)

Given the number of failure management policy options (especially predictive maintenance or condition-monitoring techniques) that are currently available, it is often tempting to select a policy purely on the basis of technical sophistication rather than on the basis of cost-effectiveness. When more than one failure management policy option is technically appropriate, correctly-applied RCM always strives to select the policy that deals satisfactorily with the consequences of the failure mode in the most economical fashion, rather than always selecting the policy with the greatest technical sophistication.

11.4 Failure Management Policy Selection

"The selection of failure management policies shall be carried out as if no specific task is currently being done to anticipate, prevent or detect the failure." (SAE JA1011, 5.6.4)

Again, for reasons explained in 9.1 of this Guide, it is essential to assume that no proactive maintenance is being carried out when selecting failure management policies.

12. FAILURE CONSEQUENCE MANAGEMENT

12.1 Evident Failure Modes with Safety or Environmental Consequences

"In the case of an evident failure mode that has safety or environmental consequences, the task shall reduce the probability of the failure mode to a level that is tolerable to the owner or user of the asset." (SAE JA1011, 5.7.1.1)

Much as most people would like to live in an environment where there is no possibility at all of death or injury, there is in fact an element of risk in everything we do. In other words, "zero" is unattainable. So what is attainable?

To answer this question, the question of risk must be considered in more detail.

Risk assessment consists of three elements. The first asks what could happen if the event under consideration did occur. The second asks how likely it is for the event to occur at all. The combination of these two elements provides a measure of the degree of risk. The third—and often the most contentious element—asks whether this risk is tolerable.

For example, consider a failure mode that could result in death or injury to 10 people (what could happen). The probability that this failure mode could occur is 1 in 1000 in any one year (how likely it is to occur). On the basis of these figures, the risk associated with this failure mode is:

10 x (1 in 1000) = 1 casualty per 100 years

Now consider a second failure mode that could cause 1000 casualties, but the probability that this failure mode could occur is 1 in 100 000 in any one year. The risk associated with this failure mode is:

1000 x (1 in 100 000) = 1 casualty per 100 years

In these examples, the risk is the same although the figures upon which it is based are quite different. Note also that these examples do not indicate whether the risk is tolerable—they merely quantify it. Whether or not the risk is tolerable is a separate question which is dealt with later.

(The terms "probability" (1 in 10 chance of a failure mode in any one period) and "failure rate" (once in 10 periods on average, corresponding to a Mean Time Between Failures (MTBF) of 10 periods) are often used as if they are interchangeable when applied to random failures. Strictly speaking, this is not true. However, if the MTBF is greater than about four periods, the difference is so small that it can nearly always be ignored.)

The following paragraphs consider each of the three elements of risk in more detail.

12.1.1 What Could Happen if the Failure Mode Occurred?

Exactly what happens if and when each failure mode occurs should be recorded as part of the list of failure effects. In other words, the failure effect statement should record whether any one occurrence of the failure mode (say) has a one in ten chance of killing one person or is likely to kill up to 10 people or is likely to cause an operator to lose a limb. In order to be reasonably conservative, note that failure effect statements should reflect the "typical worst case scenario" (but not the extreme worst case, because this would be excessively conservative). If in doubt, the people performing the analysis should ask themselves, if the worst comes to the worst, what viewpoint will be defensible before whatever authority is ultimately likely to hold them and their superiors accountable.

12.1.2 How Likely is the Failure Mode to Occur?

Section 8.1 of this Guide mentioned that only failure modes that are reasonably likely to occur in the context in question should be recorded in the FMEA. As a result, if the FMEA has been prepared on a realistic basis, the mere fact that the failure mode has been listed suggests that there is a finite probability that it could occur. Ideally this probability should be quantified, either as part of the failure effect statement or in a separate database, so that the risk can also be quantified. (Note that, in practice, accurate historical failure data are often unavailable, especially in the case of new equipment that incorporates substantial amounts of new technology. In these cases, the assessment must be based on intelligent estimates by people who clearly understand the equipment and the context in which it is being used.)

12.1.3 Is the Risk Tolerable?

As mentioned previously, risk is measured by multiplying probability by severity. It is usually expressed on an annualized basis (although it can be expressed in terms of events per a given number of cycles or operating hours or whatever else makes sense in the context in question). Deciding what is tolerable is another matter entirely.

Beliefs about what is a tolerable level of risk of death or injury vary widely from individual to individual and from group to group. Many factors influence these beliefs. The two most dominant are the degree of control that any individual thinks he or she has over the situation and the benefit that people believe they will derive from exposing themselves to the risk. This in turn influences the extent to which they might choose to expose themselves to the risk. This view then has to be translated into a degree of risk that might be tolerated by the whole population (all the workers on a site, all the citizens of a town or even the entire population of a country).

In other words, if I tolerate a probability of 1 in 100 000 (10^{-5}) of being killed at work in any one year and I have 1000 co-workers who all share the same view, then we all accept that on average one person on our site will be killed at work every 100 years—and that person may be me, and it may happen this year.

Bear in mind that any quantification of risk in this fashion can only ever be a rough approximation. In other words, a tolerable probability of 10^{-5} is never more than an approximation. With this in mind, the next step is to translate the probability that one individual and his or her co-workers are prepared to tolerate that any one of them might be killed by any event at work into a tolerable probability for each single event (failure mode or multiple failure) that could kill someone.

For example, continuing the logic of the previous example, the probability that any one of my 1000 co-workers will be killed in any one year is 1 in 100 (assuming that everyone on the site faces roughly the same hazards). If the activities carried out on the site embody (say) 10 000 events that could kill someone, then the average probability that each event could kill one person must be reduced to 10^{-6} in any one year. This means that the probability of an event that is likely to kill ten people must be reduced to 10^{-7}, while the probability of an event that has a 1 in 10 chance of killing one person must be reduced to 10^{-5}. (The techniques used to move up and down hierarchies of probability in this fashion are known as probabilistic or quantitative risk assessments.)

Although the issues discussed above usually dominate decisions about the tolerability of risk, they are not the only issues. Additional factors that help decide what is tolerable include individual values, industry values, how much is known about the real effects and consequences of each failure mode, the value placed on human life by different cultural groups, religious values, and the age and marital status of the individual.

12.1.4 Who Should Evaluate Risks?

The very diversity of the factors discussed previously mean that at this point in time, it is often impossible for any one person—or even one organization—to decide what is “tolerable” on behalf of all the people exposed to a particular risk. Furthermore, at present few organizations use a formal methodology for determining what constitutes a tolerable risk. In the absence of such a methodology, what is tolerable can be determined by a group representing:

a. People who are likely to have a clear understanding of the failure mechanism, the failure effects (especially the nature of any hazards), the likelihood of the failure mode occurring and what possible measures can be taken to anticipate or prevent it.

b. People who have a legitimate view on the tolerability or otherwise of the risks. This might include representatives of:

   1. The likely victims (operators or maintainers in the case of direct safety hazards, and the community in general in the case of environmental hazards),

   2. Those who have to deal with the consequences if someone is injured or killed or if an environmental standard is breached (such as management).

However if an organization has already established levels of risk that are considered to be tolerable by all the parties involved, then these levels can be used when assessing whether any failure management policy is worth applying to failure modes with safety or environmental consequences.

12.1.5 Safety and Failure Management

If there is an intolerable risk that a failure mode could affect safety or the environment, the RCM process stipulates that we must try to reduce the probability of the failure mode, or its consequences, or both, to the extent that the overall risk drops to a tolerable level. This suggests that, for failure modes that have safety or environmental consequences, a failure management policy is only worth applying if it reduces the risk of the failure mode to a tolerably low level.

Note that when dealing with evident failure modes that have safety or environmental consequences, RCM does not consider the cost of the failure mode. If the risk is intolerable, then it must be reduced to a tolerable level, either by introducing a suitable proactive task (or tasks), or by changing the design or operation of the asset in such a way that the risk is reduced to a tolerable level.

12.2 Hidden Failure Modes with Safety or Environmental Consequences

“In the case of a hidden failure mode where the associated multiple failure has safety or environmental consequences, the task shall reduce the probability of the hidden failure mode to an extent which reduces the probability of the associated multiple failure to a level that is tolerable to the owner or user of the asset.” (SAE JA1011, 5.7.1.2)

As explained previously, a multiple failure only occurs if the protected function fails while the protection is in a failed state. This means that the probability of a multiple failure in any period is given by the probability that the protected function will fail while the protection is in a failed state during the same period. This can be calculated as follows in Equation 1:

$$\begin{array}{c}\text{Probability of a}\\ \text{multiple failure}\end{array} = \begin{array}{c}\text{Probability of failure of}\\ \text{the protected function}\end{array} \times \begin{array}{c}\text{Average unavailability}\\ \text{of the protection}\end{array} \qquad \text{(Eq. 1)}$$

For multiple failures that have safety or environmental consequences, the tolerable probability should be determined as described in 12.1.3 and 12.1.4. The probability of failure (or failure rate) of the protected function is usually a given. So if these two variables are known, the allowed unavailability of the protective function can be expressed as follows in Equation 2:

$$\text{Allowed unavailability of the protection} = \frac{\text{Tolerable probability of a multiple failure}}{\text{Probability of failure of the protected function}} \tag{Eq. 2}$$

So a crucial element of the performance required from any protection that can suffer from a hidden failure mode is the maximum unavailability that can be allowed if the probability of the associated multiple failure is not to exceed the tolerable level. This unavailability is determined in the following three stages:

a. If not already determined as described in 12.1.3 and 12.1.4, establish what probability the organization is prepared to tolerate for the multiple failure.

b. Then determine the probability that the protected function will fail in the period under consideration (this is also sometimes known as the “demand rate”).

c. Finally, determine the unavailability (also known as “fractional dead time”) of the protection that results in the tolerable probability of the multiple failure.

Note that it is usually possible to vary both the probability of unanticipated failure of the protected function and (especially) the unavailability of the protective function by adopting suitable failure management policies. As a result, it is also possible to reduce the probability of the multiple failure to almost any desired level within reason by adopting such policies. (Zero is of course an unattainable ideal.)

12.3 Evident Failure Modes with Economic Consequences

“In the case of an evident failure mode that does not have safety or environmental consequences, the direct and indirect costs of doing the task shall be less than the direct and indirect costs of the failure mode when measured over comparable periods of time.” (SAE JA1011, 5.7.1.3)

Section 10.1.2.3 and 10.1.2.4 described the key elements of the economic consequences of a failure mode. These sections also show that economic consequences comprise operational consequences and non-operational consequences, and that they are assessed on the assumption that no scheduled task is being performed.

If failure consequences are economic, the total cost to the organization over a period of time is affected not only by the magnitude of the consequences that might occur, but also by how often the consequences are likely to occur. Similarly, the total cost to the organization of carrying out any scheduled task is also affected by the total cost of doing the task and by how often it is done. In this context, the total cost of doing the task should take into account the cost of doing the task itself, plus the fact that on occasion it may be necessary to do additional work arising from the task. For instance, it may be necessary to check a bearing for noise once a week and replace a noisy bearing once every four or five years on average.

Consequently, in order to assess the economic viability of any task, it is necessary to compare the total cost of the failure mode over a given period with the total cost of the failure management policy over an equal period. (In most cases, these costs are compared by reducing them to an annualized basis.)

If the cost of doing the task over this period is less than the total cost of the failure mode, then the task is worth doing. If not, then the task is not appropriate and some other failure management policy needs to be considered.

Note that if there is a reasonable degree of certainty that the conditional probability of the failure mode occurring will increase with age, then the period used for comparison should be long enough to encompass both the early life and the period of increased failure probability when assessing whether the scheduled task is worth doing.

Note also that if the remaining useful life of the asset is significantly shorter than the mean time between occurrences of the failure mode (especially in the case of age-related failures), then it may be appropriate to take this into account when assessing the economic viability of a scheduled task.

12.4 Hidden Failure Modes with Economic Consequences

"In the case of a hidden failure mode where the associated multiple failure does not have safety or environmental consequences, the direct and indirect costs of doing the task shall be less than the direct and indirect costs of the multiple failure mode plus the cost of repairing the hidden failure mode when measured over comparable periods of time." (SAE JA1011, 5.7.1.4)

Multiple failures that only have economic (operational or non-operational) consequences cost money. Failure management also costs money. As a result, it is usually possible to identify a failure management policy that reduces the total cost of managing the hidden failure to a minimum. In such cases, the first step is to determine what failure management policy leads to the minimum total cost on an annualized basis, then ascertain whether this (albeit minimized) financial risk can be tolerated by the owners/users of the asset.

13. FAILURE MANAGEMENT POLICIES—SCHEDULED TASKS

13.1 On-Condition Tasks

"Any on-condition task (or predictive or condition-based or condition monitoring task) that is selected shall satisfy the following additional criteria:

a. There shall exist a clearly defined potential failure.

b. There shall exist an identifiable P-F interval (or failure development period).

c. The task interval shall be less than the shortest likely P-F interval.

d. It shall be physically possible to do the task at intervals less than the P-F interval.

e. The shortest time between the discovery of the potential failure and the occurrence of the functional failure (the P-F interval minus the task interval) shall be long enough for predetermined action to be taken to avoid, eliminate or minimize the consequences of the failure mode." (SAE JA1011, 5.7.2)

13.1.1 Potential Failures and the P-F Curve

Most failure modes do not occur absolutely instantaneously. In such cases, it is often possible to detect that the items concerned are in the final stages of deterioration before they reach the failed state. This evidence of imminent failure is known as a "potential failure," which is defined as "an identifiable condition that indicates that a functional failure is either about to occur or is in the process of occurring." If this condition can be detected, it may be possible to take action to prevent the item from failing completely and/or avoid the consequences of the failure mode.

Figure 8 illustrates what happens in the final stages of the failure process. It is called the P-F curve, because it shows how a failure starts, deteriorates to the point at which it can be detected ("P") and then, if it is not detected and corrected, continues to deteriorate—usually at an accelerating rate—until it reaches the point of functional failure ("F").

If a potential failure is detected between point P and point F in Figure 8, this is the point at which it may be possible to take action to prevent the functional failure and/or to avoid its consequences. (Whether or not it is possible to take meaningful action depends on how quickly the functional failure occurs, as discussed later.) Tasks designed to detect potential failures are known as on-condition tasks.

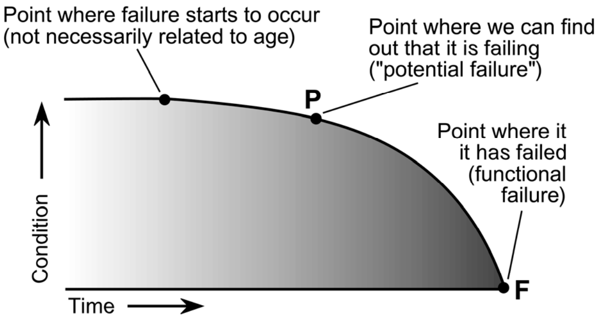

FIGURE 8 - THE P-F CURVE

On-condition tasks are so called because the items that are inspected are left in service on the condition that they continue to meet specified performance standards—in other words, on the condition that the failure mode under consideration is unlikely to occur before the next check. This is also known as predictive maintenance (because we are trying to predict whether—and possibly when—the item is going to fail on the basis of its present behavior) or condition-based maintenance (because the need for corrective or consequence-avoiding action is based on an assessment of the condition of the item.)

13.1.2 The P-F Interval

In addition to the potential failure itself, it is also necessary to consider the amount of time (or the number of stress cycles) that elapse between the point at which a potential failure occurs—in other words, the point at which it becomes identifiable—and the point where it deteriorates into a functional failure. As shown in Figure 9, this interval is known as the P-F interval.

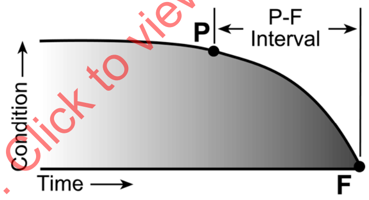

FIGURE 9 - THE P-F INTERVAL

The P-F interval governs how often on-condition tasks must be done. In order to detect the potential failure before it becomes a functional failure, the interval between checks must be less than the P-F interval. It is also essential that the potential failure condition is sufficiently clear for it to be certain that a person who is trained to perform the check will detect the potential failure if and when it occurs (or at least, that the likelihood that the potential failure will not be detected will be sufficiently low to reduce the probability of an unanticipated failure mode to a level that is tolerable to the owner or user of the asset).

The P-F interval is also known as the warning period, the lead time to functional failure or the failure development period. It can be measured in any units that provide an indication of exposure to stress (running time, units of output, stop-start cycles, etc). For different failure modes, it varies from fractions of a second to several decades.

Note that if an on-condition task is done at intervals that are longer than the P-F interval, there is a chance that the potential failure will be missed altogether. On the other hand, if the task is done at too small a fraction of the P-F interval, resources will be wasted on the checking process.

In practice task intervals should always be selected that are shorter than the shortest likely P-F interval. In most cases, it is sufficient to select a task interval equal to half of this P-F interval. However it is sometimes appropriate to select task intervals that are some other fraction of the P-F interval. This may be governed by the net P-F interval required (as discussed as follows), or it may be because the user of the asset has relevant historical data that dictate that a different fraction is appropriate.

13.1.3 The Net P-F Interval

The net P-F interval is the minimum interval likely to elapse between the discovery of a potential failure and the occurrence of the functional failure. This is illustrated in Figure 10, which shows a failure process with a P-F interval of nine months. The figure shows that if the item is checked monthly, the net P-F interval is eight months. On the other hand, if it is checked at six monthly intervals, the net P-F interval would be three months. So in the first case the minimum amount of time available to do something about the potential failure is five months longer than in the second, but the on-condition task has to be done six times more often.

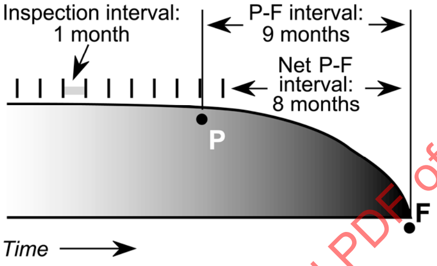

FIGURE 10 - NET P-F INTERVAL

The net P-F interval governs the length of the period available to take whatever action is needed to reduce or eliminate the consequences of the failure mode. For an on-condition task to be technically feasible, the net P-F interval must be longer than the period required to avoid or reduce the consequences of failure mode. If the net P-F interval is too short for sensible action to be taken, then the on-condition task is not technically feasible. In practice, the period required varies widely. In some cases, it may be a matter of hours (say until the end of an operating cycle or the end of a shift) or even minutes (to shut down a machine or evacuate a building). In other cases, it can be weeks or even months (say until a major shutdown). In general, longer P-F intervals are desirable for two reasons:

a. It is possible to do whatever is necessary to avoid the consequences of the failure mode (including planning the corrective action) in a more considered and hence more controlled fashion.

b. Fewer on-condition inspections are required.

For this reason a great deal of energy is being devoted to finding potential failure conditions and on-condition techniques that give the longest possible P-F intervals. However, it is possible to make use of very short P-F intervals in some cases.

13.1.4 The Relationship Between P-F Interval and Age

13.1.4.1 P-F Intervals and Random Failures

When applying these principles for the first time, people often have difficulty in distinguishing between the “life” of a component and the P-F interval. This leads them to base on-condition task frequencies on the real or imagined “life” of the item. If it exists at all, this life is usually many times greater than the P-F interval, so the task achieves little or nothing. In reality, we measure the life of a component forward from the moment it enters service. The P-F interval is measured back from the functional failure, so the two concepts are often completely unrelated. The distinction is important because failure modes that are not related to age (in other words, random failures) are as likely to be preceded by a warning as those that are not.

For example, Figure 11 depicts a component that conforms to a random failure pattern (pattern E). One of the components failed after five years, a second after six months and a third after two years. In each case, the functional failure was preceded by a potential failure with a P-F interval of four months.

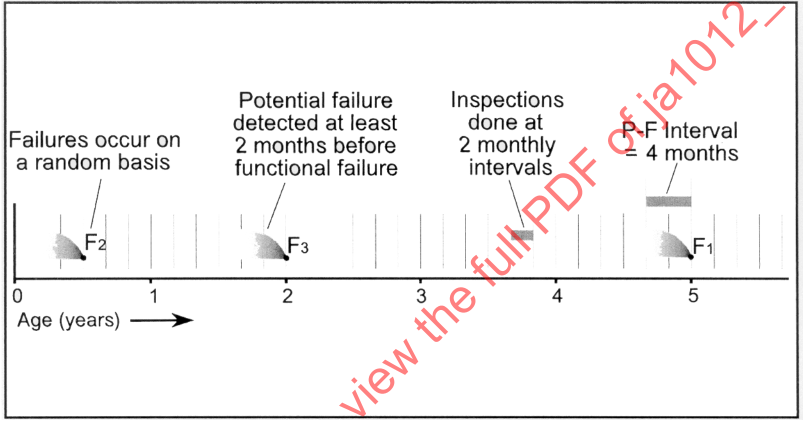
Failures occur on a random basis
Potential failure detected at least 2 months before functional failure
Inspections done at 2 monthly intervals
P-F Interval = 4 months
F2
F3
F1
0
1
2
3
4
5
Age (years)

FIGURE 11 - RANDOM FAILURES AND THE P-F INTERVAL

Figure 11 shows that in order to detect the potential failure, we need to do an inspection task every two months. Because the failure modes occur on a random basis, we don't know when the next one is going to happen, so the cycle of inspections must begin as soon as the item is put into service. In other words, the timing of the inspections has nothing to do with the age or life of the component.

However this does not mean that on-condition tasks apply only to items that fail on a random basis. They can also be applied to items that suffer age-related failure modes, as discussed below.

13.1.4.2 P-F Intervals and Age-Related Failure Modes

If an item deteriorates in a more or less linear fashion over its entire life, it stands to reason that the final stages of deterioration will also be more or less linear. This is likely to be true of age-related failure modes.

For example, consider tire wear. The surface of a tire is likely to wear in a more or less linear fashion until the tread depth reaches the legal minimum. If this minimum is (say) 2 mm, it is possible to specify a depth of tread greater than 2 mm that provides adequate warning that functional failure is imminent. This is of course the potential failure level.

If the potential failure is set at (say) 3 mm, then the P-F interval is the distance the tire could be expected to travel while its tread depth wears down from 3 to 2 mm, as illustrated in Figure 12.

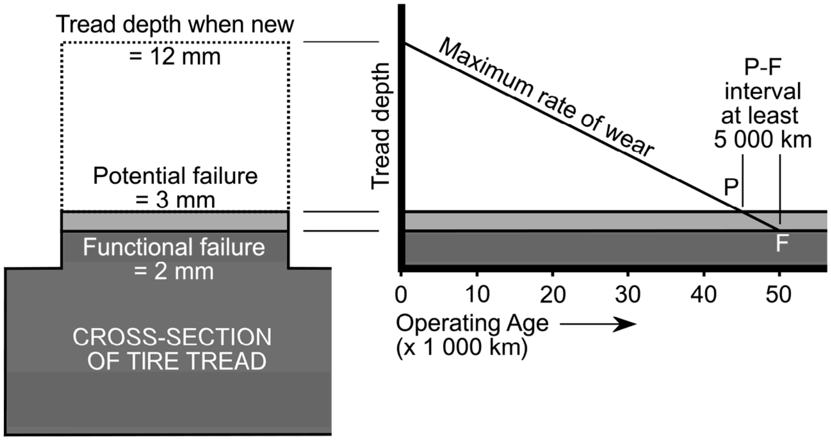

FIGURE 12 - A LIINEAR P-F CURVE

Figure 12 also suggests that if the tire enters service with a tread depth of (say) 12 mm, it should be possible to predict the P-F interval based on the total distance usually covered before the tire has to be retreaded. For instance, if the tires last at least 50 000 km before they have to be retreaded, it is reasonable to conclude that the tread wears at a maximum rate of 1 mm for every 5000 km traveled. This amounts to a P-F interval of 5000 km. The associated on-condition task would call for the driver to: “Check tread depth every 2500 km and report tires whose tread depth is less than 3 mm.”

Not only will this task ensure that wear is detected before it exceeds the legal limit, but it also allows plenty of time—2500 km in this case—for the vehicle operators to plan to remove the tire before it reaches the limit.

In general, linear deterioration between “P” and “F” is only likely to be encountered where the failure mechanisms are intrinsically age-related.

13.1.5 P-F Interval Consistency

The P-F curves illustrated so far in this section of the Guide indicate that the P-F interval for any given failure mode is constant. In fact, this is not the case—some actually vary over a wide range of values, as shown in Figure 13. In these cases a task interval must be selected that is less than the shortest of the likely P-F intervals. This ensures a reasonable degree of certainty of detecting the potential failure before it becomes a functional failure. If the net P-F interval associated with this minimum interval is long enough for action to be taken to deal with the consequences of the failure mode, the on-condition task is technically feasible.

On the other hand, if the P-F interval is very inconsistent, then it is not possible to establish a meaningful task interval, and the task should be abandoned in favor of some other way of dealing with the failure mode.

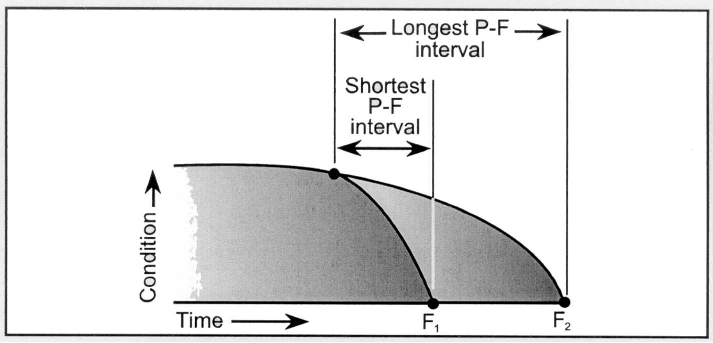

FIGURE 13 - INCONSISTENT P-F INTERVALS

13.1.6 Categories of On-Condition Technique

The four major categories of on-condition techniques are as follows:

a. Techniques based on variations in product quality. In many cases, the emergence of a defect in an article produced by a machine is directly related to a failure mode in the machine itself. Many of these defects emerge gradually, and so provide timely evidence of potential failures.

b. Primary effects monitoring techniques. Primary effects (speed, flow rate, pressure, temperature, power, current, etc.) are another source of information about equipment condition. The effects can be monitored by a person reading a gauge, by a computer as part of a process control system, or by a chart recorder.

c. Techniques based on the human senses (look, listen, feel, and smell).

d. Condition monitoring techniques. These are techniques to detect potential failures that involve the use of specialized equipment (which is sometimes built into the asset that is being monitored). These techniques are known as condition monitoring to distinguish them from other types of on-condition maintenance.

Many failure modes are preceded by more than one—often several—different potential failures, so more than one category of on-condition task might be appropriate. Each of these will have a different P-F interval, and each will require different types and levels of skill. This means that no single category of tasks will always be more cost-effective. So, to avoid unnecessary bias in task selection, it is essential to:

a. Consider all the detectable phenomena that are reasonably likely to precede each failure mode, together with the full range of on-condition tasks that could be used to detect those warnings.

b. Apply the RCM task selection criteria rigorously to determine which (if any) of the tasks is likely to be the most cost-effective way of anticipating the failure mode under consideration.

Note that any built-in devices designed to ascertain whether a failure mode is in the process of occurring or about to occur should satisfy the same criteria for technical feasibility and worth doing as any on-condition maintenance. Note also that when such devices are added to a system they constitute an additional function or functions, with additional failure modes, and should be analyzed accordingly.

13.2 Scheduled Restoration and Scheduled Discard Tasks

“Any scheduled discard task that is selected shall satisfy the following additional criteria:

a. There shall be a clearly defined (preferably a demonstrable) age at which there is an increase in the conditional probability of the failure mode under consideration.

b. A sufficiently large proportion of the occurrences of this failure mode shall occur after this age to reduce the probability of premature failure to a level that is tolerable to the owner or user of the asset.” (SAE JA1011, 5.7.3)

“Any scheduled restoration task that is selected shall satisfy the following additional criteria:

a. There shall be a clearly defined (preferably a demonstrable) age at which there is an increase in the conditional probability of the failure mode under consideration.

b. A sufficiently large proportion of the occurrences of this failure mode shall occur after this age to reduce the probability of premature failure to a level that is tolerable to the owner or user of the asset.

c. The task shall restore the resistance to failure (condition) of the component to a level that is tolerable to the owner or user of the asset.” (SAE JA1011, 5.7.4)

Scheduled restoration and scheduled discard tasks have a number of features in common, so this part of this Guide considers their common features first, then reviews the differences.

Scheduled restoration entails taking periodic action to restore the capability of an item at or before a specified interval (age limit), regardless of its condition at the time, to a level that provides a tolerable probability of survival to the end of another specified interval (which need not be the same as the initial interval). This action usually entails either remanufacturing a single component or overhauling an entire assembly

Scheduled discard means discarding an item or component at or before a specified age limit, regardless of its condition at the time. This is done on the understanding that replacing the old component with a new one will restore the original resistance to failure.

If the failure mode under consideration conforms to Pattern A or B, it is possible to identify the age at which wear-out begins. The scheduled restoration or scheduled discard task must be done at intervals less than this age. In other words, the frequency of a scheduled restoration or scheduled discard task is governed by the age at which the item or component shows a rapid increase in the conditional probability of failure.

In the case of Pattern C, more complex analytical techniques are required. These techniques are beyond the scope of this Guide. Note that two types of life-limits apply to scheduled restoration and scheduled discard tasks. These are safe-life limits and economic-life limits.

13.2.1 Safe-Life Limits

Safe-life limits only apply to failure modes that have safety or environmental consequences, so the associated tasks must reduce the probability of a failure mode occurring before the life limit to a tolerable level. (A method of deciding what is tolerable was discussed in 12.1.3 of this Guide. In practice, probabilities as low as 10^{-6} and sometimes even 10^{-9} are often used in this context.) This requirement means that safe-life limits cannot apply to any failure mode that has a significant probability of occurring when the item enters service.

Ideally, safe-life limits should be determined before a new item is put into service. They should be established by testing a statistically adequate sample of items in a simulated operating environment to determine what life is actually achieved. Some industries apply a conservative fraction of this life (typically a third or a quarter) as the safe-life limit, as illustrated in Figure 14.

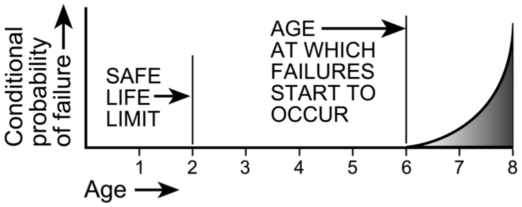

FIGURE 14 - SAFE LIFE LIMITS

13.2.2 Economic-Life Limits

Operating experience sometimes suggests that the scheduled restoration or scheduled discard of an item is desirable on economic grounds. This is known as an economic-life limit. It is based on the actual age-reliability relationship of the item, rather than a fraction of the age at which there is an increase in the conditional probability of failure. A sufficiently large majority of the items must survive to the economic-life limit for the task to be justifiable on economic grounds.

13.3 Failure-finding Tasks

"Any failure-finding task that is selected shall satisfy the following additional criteria (failure-finding does not apply to evident failure modes):

a. The basis upon which the task interval is selected shall take into account the need to reduce the probability of the multiple failure of the associated protected system to a level that is tolerable to the owner or user of the asset.

b. The task shall confirm that all components covered by the failure mode description are functional.

c. The failure-finding task and associated interval selection process should take into account any probability that the task itself might leave the hidden function in a failed state.

d. It shall be physically possible to do the task at the specified intervals." (SAE JA1011, 5.7.5)

13.3.1 Multiple Failures and Failure-finding

As mentioned in 10.1.1.2, a multiple failure occurs if a protected function fails while the protection is in a failed state. This phenomenon was illustrated in Figure 5. Equation 1, repeated below as Equation 3, shows how the probability of a multiple failure can be calculated.

$$\begin{array}{c}\text{Probability of a}\\\text{multiple failure}\end{array} = \begin{array}{c}\text{Probability of failure of}\\\text{the protected function}\end{array} \times \begin{array}{c}\text{Average unavailability}\\\text{of the protection}\end{array} \qquad \text{(Eq. 3)}$$

This led to the conclusion that the probability of a multiple failure can be reduced by reducing the unavailability of the protection—in other words, by increasing its availability.

The best way to do this is to prevent the protective function from getting into a failed state by applying some sort of proactive maintenance. However, few proactive tasks satisfy the criteria for technical feasibility when applied to hidden failures. Nonetheless, although proactive maintenance is often inappropriate, it is still essential to do something to reduce the probability of the multiple failure to the required level. This can be done by checking periodically whether the hidden failure has occurred. Such checks are known as failure-finding tasks.

13.3.2 Technical Aspects of Failure-Finding

The objective of failure-finding is to ascertain whether a hidden failure mode or combination of hidden failure modes has rendered a protective function incapable of providing the required protection if it is called upon to do so. (This is why failure-finding tasks are also known as functional checks.) The following paragraphs consider some of the key issues in this area.

13.3.2.1 Check the Protective Function in Its Entirety

A failure-finding task must be sure of detecting all the hidden failure modes at which it is directed. This is especially true of complex devices, such as those made up of sensors, electrical circuits, and actuators. Ideally, this should be done by simulating the conditions the sensor should detect, and checking whether the actuator gives the right response. The failure-finding interval should be established accordingly.

13.3.2.2 Do Not Disturb

Dismantling anything always creates the possibility that it will be put back together incorrectly. If this happens to a protective device that could suffer from hidden failures, the fact that the failure modes are hidden means that no one will know it has been left in a failed state until the next check (or until it is needed). For this reason, always look for ways of checking the functions of protective devices without disconnecting or otherwise disturbing them.

This having been said, some devices simply have to be dismantled or removed altogether to check if they are working properly. In these cases, great care must be taken to do the task in such a way that the devices will still work when they are returned to service.

13.3.2.3 It Must be Physically Possible to Check the Function

In a very small but still significant number of cases, it is impossible to carry out a failure-finding task of any sort. These are:

a. Where it is impossible to gain access to the protective device in order to check its function (this is almost always a result of thoughtless design), and

b. When the function of the device cannot be checked without destroying it (as in the case of fusible devices and rupture discs). In most such cases, other technologies are available (such as circuit breakers instead of fuses). However, in one or two cases the only options are to find some other way of managing the risks associated with untestable protection until something better comes along, or to abandon the processes concerned.

13.3.2.4 Minimize Risk While the Task is Being Done

It should be possible to carry out a failure-finding task without significantly increasing the risk of the multiple failure. If a protective device has to be disabled in order to carry out a failure-finding task, or if such a device is checked and found to be in a failed state, then alternative protection should be provided or the protected function should be shut down until the original protection is restored.

13.3.2.5 The Frequency must be Practical

It must be practical to do the failure-finding task at the required intervals. This is discussed in 13.3.3. However, before we can decide whether a required interval is practical, we need to determine what interval is actually “required.”

13.3.3 Failure-Finding Task Intervals

13.3.3.1 Failure-Finding Intervals, Availability, and Reliability

Not one but two variables—availability and reliability—are used to set failure-finding intervals. It can be shown that if there is a linear correlation between the unavailability, the failure-finding interval and the reliability of the protective function as given by its MTBF, as follows in Equation 4:

$$\text{Unavailability} = 0.5 \times \frac{\text{Failure-finding interval}}{\text{MTBF of the protective function}} \qquad \text{(Eq. 4)}$$

It can also be shown that this linear relationship is valid for all unavailabilities of less than 5%, provided that the protective function conforms to an exponential survival distribution.[1]

13.3.3.2 Excluding Task Time and Repair Time

The "unavailability" of the protective function in Equation 4 does not include any unavailability incurred while the failure-finding task is being performed, nor does it include any unavailability caused by the need to restore the function if it is found to be failed. This is so for two reasons:

a. The unavailability required to carry out the failure-finding task and to effect any repairs is likely to be very small relative to the unrevealed unavailability between tasks, to the extent that it will usually be negligible on purely mathematical grounds.

b. Both the failure-finding task and any repairs that might be needed should be carried out under tightly controlled conditions. These conditions should greatly reduce—if not completely eliminate—the chance of a multiple failure while the intervention is under way. This entails either shutting down the protected system or arranging alternative protection until the system has been fully restored. If this is done properly, the unavailability resulting from the (controlled) intervention can be ignored in any assessments of the probability of a multiple failure.

In the RCM decision process, the latter point is covered by the criteria for assessing whether a failure-finding task is worth doing. If there is a significant increase in the likelihood of a multiple failure while the task is under way, the answer to the question "Does the task reduce the probability of a multiple failure to a tolerable level" will be "no."

13.3.3.3 Calculating FFI Using Availability and Reliability Only

If we use the abbreviation "FFI" to describe the failure-finding interval and "MTIVE" to describe the MTBF of the protective function, Equation 4 can be rearranged to give the Equation 5:

$$\text{FFI} = 2 \times \text{Unavailability} \times \text{Mtive} \qquad \text{(Eq. 5)}$$

This means that in order to determine the failure-finding interval for a single protective function, it is necessary to ascertain its mean time between failures and the desired availability of the function (from which it is possible to compute the unavailability to be used in the formula). For those who are uncomfortable with mathematical formulae, Equation 5 can be used to develop a simple table, as shown in Figure 15:

[1] See Cox and Tait or Andrews and Moss.

Availability we require for the protective function	99.99%	99.95%	99.9%	99.5%	99%	98%	95%
Failure-finding interval (as a % of the MTBF)	0.02%	0.1%	0.2%	1%	2%	4%	10%

FIGURE 15 - FAILURE-FINDING INTERVAL, AVAILABILITY, AND RELIABILITY

13.3.3.4 Rigorous Methods for Calculating FFI

A single formula for determining failure-finding intervals that incorporates all the variables considered so far can be developed by combining Equations 1 and 5, as explained in the following paragraphs. To do so, more terms need to be defined as follows:

a. A probability of a multiple failure of 1 in 1 000 000 in one year implies a mean time between multiple failures of 1 000 000 years. If this is called M_{MF}, the probability of a multiple failure occurring in any one year is $1/M_{MF}$.

b. If the demand rate of the protected function is (say) once in 200 years, this corresponds to a probability of failure for the protected function of 1 in 200 in any one year, or a mean time between failures of the protected function of 200 years. If this is called M_{TED}, the probability of failure of the protected function in any one year will be $1/M_{TED}$. This is also known as the demand rate.

c. M_{TIVE} is the mean time between failures of the protective function and FFI is the failure-finding task interval.

d. U_{TIVE} is the allowed unavailability of the protective function.

If the previous expressions are substituted, Equation becomes:

$$1/M_{MF} = (1/M_{TED}) \times U_{TIVE} \tag{Eq. 6}$$

This can be rearranged as follows in Equation 7:

$$U_{TIVE} = \frac{M_{TED}}{M_{MF}} \tag{Eq. 7}$$

Substituting U_{TIVE} from Equation 7 into Equation 5 gives Equation 8:

$$FFI = \frac{(2 \times M_{TIVE} \times M_{TED})}{M_{MF}} \tag{Eq. 8}$$

This formula allows a failure-finding interval to be determined in one step for a single, independent protective function.

13.3.3.4.1 Multiple Failure Modes of a Single Protective Function

Throughout this section, all the failure possibilities that could cause any protective function to fail have been grouped together as one single failure mode ("stand-by pump fails"). The vast majority of protective functions can be treated in this way, because all the failure modes that could cause a protective function to fail are checked when the function of the device as a whole is checked.

However, it is sometimes appropriate to carry out a detailed FMEA of the protective function in order to identify individual failure modes, each of which on its own might cause the protective device or system to be unable to provide the required protection. This is usually done under not one but two sets of circumstances:

a. When some of the failure modes are known to be susceptible to on-condition maintenance or scheduled restoration or scheduled discard tasks, but others are neither predictable nor preventable. In these cases, the appropriate on-condition or scheduled restoration/discard task should be applied to the failure modes that qualify, and failure-finding applied to the remaining failure modes.

b. When the protective device is new and the only failure data that are available (from data banks, component suppliers or wherever) apply to parts of the device but not to the device as a whole.

In these circumstances, Equation 8 should be modified to accommodate the combination of individual failure modes that are the object of the failure-finding task, by determining a composite mean time between failures of the protective function based on the MTBFs of each failure mode.

13.3.3.4.2 Methods of Calculating Failure-Finding Intervals for Other Types of Protective Functions

The techniques for setting failure-finding intervals described previously are risk-based approaches to single protective functions. The management of multiple protective functions and the management of multiple failures that have only economic consequences are beyond the scope of this Guide.

13.3.3.5 The Practicality of Failure-Finding Task Intervals

The methods described so far for calculating failure-finding intervals sometimes produce very short or very long intervals, with the following implications:

a. A very short failure-finding task interval has two main implications:

   1. Sometimes the interval is simply far too short to be practical. An example would be a failure-finding task that calls for a major item of a process plant to be shut down every few days.

   2. The task could cause habituation (which might happen if a fire alarm is tested too often).

   In these cases, the proposed task must be rejected and some other way found to reduce the probability of the multiple failure to a tolerable level.

b. A very long failure-finding task interval also has two main implications:

   1. For intervals that are substantially longer than the projected remaining useful life of the asset: Such intervals suggest that there is no need to do a scheduled failure-finding task at all (though it is still necessary to ascertain during commissioning that the devices have been installed correctly).

   2. For intervals that are longer than the maximum planning horizon of existing maintenance planning systems, but are shorter than the projected remaining useful life of the asset: In these cases, care should be taken not to reduce the associated intervals simply to accommodate the limitations of existing planning systems, if only because failure-finding tasks can sometimes induce the failure modes that they are meant to detect.

c. Note that a failure-finding task interval can emerge that exceeds the average interval between failures of the protected function. As the amount by which the failure-finding interval exceeds the failure interval increases, the value of failure-finding diminishes rapidly, until a point is reached where it has little or no effect on the probability of the multiple failure. If any of the above formulae yield an interval at or beyond this point, some other way must be found to reduce the probability of the multiple failure to the tolerable level.

13.4 Combination of Tasks

If a failure mode or a multiple failure could affect safety or the environment and no scheduled task can be found that on its own reduces the risk of failure to a tolerably low level, it is sometimes possible that a combination of tasks (usually from two different task categories, such as an on-condition task and a scheduled discard task), might reduce the risk of the failure mode to the tolerable level.

When considering such combinations, care must be taken to ensure that each task on its own will satisfy the technical feasibility criteria appropriate to that kind of task, and that each task is carried out at the frequency appropriate for that task. Care must also be taken to ensure that the two tasks combined will in fact reduce the consequences to a tolerable level. However it must be stressed that situations in which combinations of tasks are necessary are very rare, and care should be taken not to employ such combinations indiscriminately.

14. FAILURE MANAGEMENT POLICIES—ONE-TIME CHANGES AND RUN-TO-FAILURE

14.1 One-Time Changes

“The RCM process shall endeavor to extract the desired performance of the system as it is currently configured and operated by applying appropriate scheduled tasks.” (SAE JA1011, 5.8.1.1)

“In cases where such tasks cannot be found, one-time changes to the asset or system may be necessary, subject to the following criteria:

a. In cases where the failure is hidden, and the associated multiple failure has safety or environmental consequences, a one-time change that reduces the probability of the multiple failure to a level tolerable to the owner or user of the asset is compulsory.

b. In cases where the failure mode is evident and has safety or environmental consequences, a one-time change that reduces the probability of the failure mode to a level tolerable to the owner or user of the asset is compulsory.

c. In cases where the failure mode is hidden, and the associated multiple failure does not have safety or environmental consequences, any one-time change must be cost-effective in the opinion of the owner or user of the asset.

d. In cases where the failure mode is evident and does not have safety or environmental consequences, any one-time change must be cost-effective in the opinion of the owner or user of the asset.” (SAE JA1011, 5.8.1.2)

Earlier sections of this Guide stressed that the initial capability (or inherent reliability) of any asset is established by its design and by how it is made, and that maintenance cannot yield reliability beyond that inherent in the design. This led to two conclusions.

Firstly, if the initial capability of an asset is greater than the desired performance, maintenance can help achieve the desired performance. Most equipment is adequately specified, designed, and built, so it is usually possible to develop a satisfactory maintenance program, as described previously. In other words, in most cases, RCM helps us to extract the desired performance from the asset as it is currently configured.

Secondly, if desired performance exceeds initial capability, then no amount of maintenance can deliver the desired performance. In these cases "better" maintenance cannot solve the problem, so it is necessary to look beyond maintenance for the solutions. In most cases, this entails changing the capability of one of three elements of the system:

a. A change to the physical configuration of the asset (what is normally referred to as a "redesign" or "modification"). This is any action that should result in a change to a drawing or a parts list. It includes changing the specification of a component, adding a new item, replacing an entire machine with one of a different make or type, or relocating a machine. (Note that if any such changes are made, the RCM process will need to be applied in full to the new design in order to ensure that it continues to function as intended.)

b. A change to a process or procedure that affects the operation of the asset.

c. A change to the capability of one of the people involved in operating or maintaining the equipment (this usually entails training the person concerned as a method of dealing with a specific failure mode).

The term "one-time change" is used in this Guide to refer to these interventions because they are usually made only once to any specific system, as opposed to scheduled tasks that are performed at regular intervals. The following paragraphs outline the specific objectives of one-time changes for each of the main categories of failure consequences.

14.1.1 Safety or Environmental Consequences

If a failure mode could affect safety or the environment and no scheduled task or combination of tasks can be found that reduce the risk of failure to a tolerable level, something must be changed, simply because we are now dealing with a safety or environmental hazard that cannot be adequately prevented. In these cases, redesign is usually undertaken with one of two objectives:

a. To reduce the probability of an unanticipated failure mode occurring to a level that is tolerable. This is usually done either by replacing the affected component with one that is stronger or more reliable, or by making it possible to anticipate the failure mode.

b. To change the item or the process in such a way that the failure mode no longer has safety or environmental consequences. This is most often done by installing a suitable protective device. Remember that if such a device is added, its maintenance requirements must also be analyzed.

Safety or environmental consequences can also be reduced by eliminating hazardous materials from a process, or even by abandoning a dangerous process altogether. In essence, if the level of risk associated with any failure mode is regarded as intolerable, RCM obliges us either to prevent the failure mode from occurring, or to make the process safe. The alternative is to accept conditions that are known to be unsafe or environmentally unsound. This is no longer acceptable in most industries.

14.1.2 Hidden Failures

In the case of hidden failures, the risk of a multiple failure can be reduced by making any of the following one-time changes:

a. Make the hidden failure evident by adding another device: Certain hidden failures can be made evident by adding another device (such as “Built-In Test Equipment,” or BITE), that draws the attention of the operator to the hidden failure. Special care is needed in this area, because the failures of extra functions installed for this purpose also tend to be hidden. If too many layers of protection are added, it becomes increasingly difficult—if not impossible—to define sensible failure-finding tasks. A more effective approach is to substitute an evident function for the hidden function, as explained in the next paragraph.

b. Substitute a protective function whose failure is evident for the hidden function: In most cases this means substituting a device or system whose failure is genuinely evident for one whose failure is not evident.

c. Substitute a more reliable (but still hidden) device for the existing protective device: A more reliable device (in other words, one that has a higher mean time between failures) will enable the organization to achieve one of three objectives:

   1. To reduce the probability of the multiple failure without changing the failure-finding task intervals. This increases the level of protection.

   2. To increase the interval between tasks without changing the probability of the multiple failure. This reduces resource requirements.

   3. To reduce the probability of the multiple failure and increase task intervals.

d. Duplicate the hidden function: If it is not possible to find a single protective device that has a high enough MTBF to give the desired level of protection, it is still possible to achieve any of the above three objectives by duplicating (or even triplicating) the hidden function. However, bear in mind that the function of all these devices would still need to be subjected to analysis in order to identify an appropriate failure management policy.

e. Make it possible to perform a scheduled task (for instance by improving access to the protective device or system).

f. Reduce the demand rate of the protected function: Depending on the failure modes that lead to the demand for protection, change the physical configuration of the system and/or the capability of the operator or maintainer in such a way that the system is likely to require the protection less often.

14.1.3 Operational and Non-Operational Consequences

For some failure modes with operational or non-operational consequences, the most cost-effective failure-management policy may be to change the system to reduce total costs. To achieve this, the changes could seek to:

a. Reduce the number of times the failure mode occurs, or possibly eliminate it altogether, again by making some element of the system more robust or more reliable.

b. Reduce or eliminate the consequences of the failure mode (for example, by providing a stand-by capability).

c. Make a scheduled task cost-effective (for instance, by making a component more accessible).

Note that in this case, modifications must be cost-justified, whereas they are compulsory if there is no other way to reduce the risk of failures that have safety or environmental consequences to a tolerable level.

14.2 Run To Failure

“Any run-to-failure policy that is selected shall satisfy the appropriate criterion as follows:

a. In cases where the failure is hidden and there is no appropriate scheduled task, the associated multiple failure shall not have safety or environmental consequences.

b. In cases where the failure is evident and there is no appropriate scheduled task, the associated failure mode shall not have safety or environmental consequences.” (SAE JA1011, 5.8.2)

In the case of some failures that are evident and that do not affect safety or the environment, or that are hidden and the multiple failure does not affect safety or the environment, the most cost-effective failure management policy might simply be to allow the failures to occur and then take appropriate steps to repair them. In other words, “run to failure” is only valid if:

a. A suitable scheduled task cannot be found for a hidden failure, and the associated multiple failure does not have safety or environmental consequences, and

b. A cost-effective proactive task cannot be found for failures with operational or non-operational consequences.

15. FAILURE MANAGEMENT POLICY SELECTION

15.1 Two Approaches

The last three questions in the RCM process, discussed in Sections 10 through 14 of this Guide, entail the selection of appropriate failure management policies for each failure mode identified in the FMEA. Two distinct approaches can be used to select failure management policies. The first is a rigorous approach and the second is a decision diagram approach.

The rigorous approach is more thorough and produces a fully cost-optimized failure management policy for dealing with each failure mode in the FMEA. Decision diagrams are popular because they are quicker and cheaper to apply than the rigorous approach. However any decision diagram approach must fully address the safety and environmental consequences of each failure mode. It should also be borne in mind that the use of decision diagrams does introduce an element of sub-optimization to the failure management policy selection process, from the cost point of view.

Note that when applying either approach, most decisions have to be made in the absence of complete data. This can lead to a temptation to start relying excessively on “default logic,” in which decisions are made automatically if comprehensive data are not readily available. However the application of such logic can lead to incorrect decisions, especially in the assessment of consequences. In practice the view should be taken that, if the possible repercussions of too much uncertainty cannot be tolerated, then action should be taken to change the consequences of the failure mode—rather than rely upon “default” decisions.