5-7-2016

Investigation of Human Subjectivity during Failure Mode Effects Analysis (FMEA)

Marc D. Banghart

Investigation of human subjectivity during Failure Mode Effects Analysis (FMEA)

By

Marc D. Banghart

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Industrial and Systems Engineering
in the Department of Industrial and Systems Engineering

Mississippi State, Mississippi

May 2016

Investigation of human subjectivity during Failure Mode Effects Analysis (FMEA)

By

Marc D. Banghart

Approved:

_____

Kari Babski-Reeves
(Major Professor)

_____

Linkan Bian
(Committee Member)

_____

Lesley Strawderman
(Committee Member)

_____

Stanley F. Bullington
(Graduate Coordinator)

_____

Jason M. Keith
Dean
Bagley College of Engineering

Name: Marc D. Banghart

Date of Degree: May 6, 2016

Institution: Mississippi State University

Major Field: Industrial and Systems Engineering

Major Professor: Dr. Kari Babski-Reeves

Title of Study: Investigation of human subjectivity during Failure Mode Effects Analysis (FMEA)

Pages in Study 51

Candidate for Degree of Master of Science

Several concerns with Failure Modes and Effects Analysis (FMEA), including acknowledgement that the process contains human subjectivity, can be found in literature; however very little research has been conducted to identify where and to what extent this variation is found. This thesis investigated sources of variation related to human decision making within FMEA. Participants were required to determine the effects of given failure modes by selection of a severity level given varied input information. The study found that participants were not able to sift through the provided information and identify the appropriate cues relating data relevance to the failure mode under analysis. Thus, it appeared that more information will reduce conservatism – however the quality of the information and experience level does not have an effect. The study concluded that FMEAs contain significant subjectivity and data quality assessment must form part of the FMEA framework.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 State of Reliability

In order to attain high quality and reliability in large organizations and complex systems, it is important to ensure customer needs are translated through identifiable activities into a useful and effective product. This collection of activities goes by several names, such as Systems Engineering or more commonly the Quality Function in a Quality Engineering environment [1]. These activities are crucial to ensure reliable, cost-effective and sustainable products both in the commercial and government sector.

The U.S. Department of Defense (DoD) has recognized alarming trends in the acquisition of new products or systems. First, the percentage of new acquisitions failing to meet reliability requirements is increasing. Second, the percentage of fielded systems that have decreasing durability and reliability performance is also increasing [2]. These trends were clearly expressed in a memorandum to the DoD community from the Director of Systems Engineering in the Office of the Secretary of Defense in 2010. The memo stated that over a 25 year period a staggering 25% of defense systems were not found suitable in operational testing. The memo further stated that "there is no question the systems emerging from our design and development efforts are often not reliable" [3]. Resultant poor performance has been illustrated with several examples, including the Early-Infantry Brigade Combat Team (E-IBCT) unmanned aerial system which only

achieved 1/10th of its mean time between system aborts requirement [3], and more recently the F-35 aircraft [4]. As a result, the DoD has instituted several initiatives aimed at reversing the trend. These initiatives include:

- Strengthening of oversight and accountability through the Weapon Systems Acquisition Reform Act of 2008
- Reduction of risk through revitalized Systems and Reliability Engineering processes
- Standardization of reliability best practices in new programs though a reliability scorecard
- Shift to Physics of Failure (PoF) based reliability prediction [2].

Additionally, a report from the Defense Science Board (DSB) Task Force on Developmental Test and Evaluation concluded that:

"The single most important step necessary to correct high suitability failure rates is to ensure programs are formulated to execute a viable systems engineering strategy from the beginning, including a robust RAM program, as an integral part of design and development. No amount of testing will compensate for deficiencies in RAM program foundation" [2].

1.2 Achieving Reliability, Availability and Maintainability (RAM)

As stated in the DSB Task Force report [2], development of a robust RAM program must be accomplished early in design in order to reduce life cycle cost while achieving performance objectives.

In order for a product or system to achieve RAM goals the following activities must be performed (adapted from [5]) as part of a RAM program:

- Identification of reliability, availability and maintainability goals or requirements
- Iteratively forecast, measure and verify that these requirements are met
- Ensuring quality problems are not induced by manufacturing and assembly procedures

One major aspect of a RAM program is measuring progress towards requirements, which is intrinsically coupled with risk assessment. Risk assessment must identify both technical and programmatic risks to successfully meeting or exceeding these requirements. These must then be further decomposed into high, medium and low risks – and addressed early in design. Identification of these technical risks is not a trivial task, and requires both pro-active thinking and rigorous analysis tools or processes.

Risk assessment must also be applied throughout the product life-cycle. For any project, five phases can be identified in a generic fashion [5]:

- Product concept phase
- Design phase
- Produce development phase
- Manufacturing and Support phase

Application of risk assessment during the product concept phase, when little or no data is available, is problematic. At this stage in the design, testing has not been performed nor does a physical design exist; however, this is a crucial time to incorporate RAM. In this phase, FMEA is frequently utilized, particularly since data is not available in order to develop failure modes based on the proposed design. The resulting analysis can then be utilized to evaluate weak areas of the design and/or compare other design alternatives. The variables within FMEA are largely quantified by individuals and/or teams.

Risk management consists of several elements, to include assessment, management and risk communication [5]. As illustrated in Figure 1.1, risk assessment consists of assessment, management and communication of risks. FMEA cradles all parts of this risk management framework. For example, failure modes are identified, along with probability and severity classifications, all of which forms part of risk assessment. However, mitigation strategies for failure modes form part of risk management, while the process itself facilitates risk communication.

FMEA can be a very useful tool in risk management – but may also be heavily plagued by human subjectivity. Thus, due to the proactive nature of FMEA it can be utilized to establish a risk management policy [6].

Figure 1.1 Risk Assessment Framework

1.2.1 Risk Assessment Tools in RAM

There are several prevailing tools available which can be used to proactively assess risk in terms of RAM [1]. These include (but are not limited to) [5]:

- Failure Mode and Effects Analysis (FMEA)
- Design Guidelines and Best Practices
    - Design for Reliability (DfR)
    - Design for Manufacturing (DfM)

    - Design for Tests (DfT)
- Modeling and Simulation
- Highly Accelerated Life Tests (HALT) and Accelerated Life Tests (ALT)

FMEA is one of the most widely utilized tools due to:

- Well defined process
- Ability to be tailored to systems or process analyses
- Ability to capture expert opinion
- Ability to be performed with little or no data
- Provision of the capability to analyze and compare different systems or concepts

FMEA has shown success across many industries [6-10]. One benefit of FMEA as an analysis tool is that it is structured and usually includes a diverse group of people from different background and experiences, with the ultimate goal of identifying potential threats to a complex system. This potentially allows different viewpoints to be incorporated leading to a more robust analysis. The analysis is accomplished by analyzing failure modes, or technical risks that are reasonable likely to occur, along with their associated consequences (or severity). Further, in methodologies such as RCM it is further extended to develop mitigation strategies [11].

FMEA also has problems. These problems have received extensive attention in literature, with several changes proposed in order to address possible identified shortfalls within the methodology. Humans make subjective decisions [12], and their inter-rater reliability is not known in FMEA. Thus, the reliability or the "extent to which … any measuring procedure yields the same results on repeated trials" [13] is largely unknown for FMEA. This subjectivity is driven by internal biases and has been shown to occur

during probability estimation (for example in fault trees) [12]. Additionally, the subjectivity occurs in most (if not all people) to include experts and novice users [12]. The extent and effects of this subjectivity is largely unknown. FMEA requires several variables to be quantified by either individuals, teams or experts and can be seen as a complex task. The task complexity is rooted in many related variables (FM 1 influencing the probability of FM 2 for example), undefined variables (data not available) and that several possible solutions exist (for example mitigation strategies) [14]. This bias has been recognized and a plethora of techniques adapted from evidence theory, grey theory and fuzzy logic [7-8, 15-17] have emerged. Recent research has also indicated that our biases and behavior is closely tied to our genetic make-up and the underlying cognitive processes [18]. Decision making in a clinical setting was explored by Smith et al [19]. In their research they highlighted that understanding of the problem at hand and contextual factors are both important in determining outcomes. Thus, they recommend that both the individual's attributes as well as context must be taken into account.

However, as noted by Bozdag et al. very little research has been focused on understanding uncertainty and variations ratings amongst experts [11]. Bozdag et al. further develop a failure mode assessment and prioritization model based on fuzzy logic that both incorporates individual and consensus judgment into a risk rating [11]. As innovative as this model is, there is almost no literature available that illustrates where in FMEA human variability is most prevalent. The healthcare community has performed a few small studies in order to probe this question, which are detailed in Chapter 2 [20,21]. Understanding where variation is most prevalent will aid in developing models that can be validated and ultimately improve the FMEA process.

1.3 Selection of FMEA

As described in section 1.2 several tools are available during the risk analysis and reliability engineering process. These tools were evaluated for inclusion in this study, however FMEA was ultimately selected as the most appropriate tool. Table 1.1, provides a comparison of various tools available based on the opinion of the researchers. FMEA was ultimately utilized within this study due to its widespread application and ease of use.

Table 1.1 Comparison of FMEA to other Common Engineering Tools

Tool	Interpretation Easy	Cross-industry application	Mathematical Skill Requirements	Prevalence in industry	Input Requirements
FMEA	High	High	Low	High	Low
Reliability Block Modeling	Low	Low	High	Low	High
Fault Tree Analysis	High	High	High	High	High
Probabilistic Risk Assessment	Low	High	High	Low	High

1.4 Summary of Main Contributions

This thesis will delve into human subjectivity during FMEA, both of experienced and non-experienced users. The main goal is to fill the literature gap with empirical evidence indicating the role of FMEA experience levels of analysts and the role of available –information on variable selection. The thesis will focus on variables associated with risk quantification. This will provide empirical data that can be used to refine proposed models within the literature.

CHAPTER II

FAILURE MODE EFFECTS ANALYSIS (FMEA)

2.1 Background

The efforts to strengthen RAM activities early in design emphasize various Design for Reliability (DfR) tools such as FMEA. DfR embraces the customer needs and is a process though which customer satisfaction is maximized. The process utilizes numerous integrated tools in order to support a product (and/or design) from cradle to grave, while ensuring the highest reliability at the lowest life cycle cost. Failure Mode Effects Analysis (FMEA) is used extensively throughout industry in order to improve system reliability and aid in risk assessment, and is a well-recognized DfR tool. It was utilized as early as the 1960s by the U.S. National Aeronautics and Space Administration (NASA) on programs such as the Apollo, Viking, Voyager and Skylab explorations. The process was also adopted early on by the Society for Automotive Engineers (SAE) in 1967. The use of FMEA spread rapidly to other industries during the 1970s and subsequent years, and is now utilized in a variety of industries including military, semiconductors, and the food service industry. More recently FMEA has been adopted within the healthcare industry in order in order to assess the high-risk process of care [22]. FMEA is useful in understanding the failure modes of systems or products, qualifying the effects of failure, and aiding in the development of mitigation strategies. It

is a useful tool in improving quality, reliability, and the maintainability of designs, and is a critical analysis component in risk management.

2.2 Process and Analysis Procedure

FMEA is a methodology to determine potential failure modes of an end item. The end item can be a system, process, software or components. The FMEA process starts with the identification of the various functions of the end item, which are closely tied to the mission and operating context. These functions naturally flow out of the inherent design of the process or concept. Once the functions and functional failures have been identified, the analysis develops failure modes. Figure 2.1 illustrates a typical FMEA process [11,23].

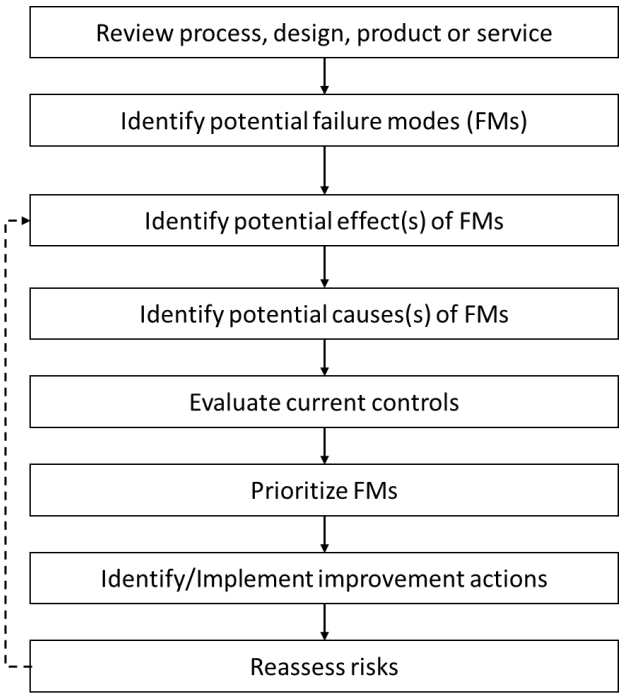

Figure 2.1 Steps of the FMEA Process [11]

Failure modes are the result of a failure mechanism and are a specific manner by which function is lost [24]. Each failure mode has many variables to include local, next higher and end effects if the failure mode occurs. The effects are utilized to determine the severity of the failure mode. Several scales are proposed; however they all rank the consequences of the failure mode from negligible to severe. The consequences are usually defined in terms of cost and safety in compliance with customer safety programs. For example, if a failure mode results in high cost and/or large safety impact it may be ranked with a high severity. Additionally, each failure mode is given a probability of failure (which can be based on the Mean Time Between Failure), detection method, classified as safety or non-safety.

Once a failure mode has been developed, it is scored, or ranked relative to all other failure modes. There are several approaches, such as criticality assessment and Risk Priority Numbering (RPN). The RPN method multiplies the probability of occurrence, detection probability and severity in order to determine a relative ranking. An alternate method (and preferred by the researcher) is criticality assessment. This method utilizes a Hazard Risk Index (HRI) that is color coded red, green, and yellow corresponding to the associated risk level as determined by the customer. Each failure mode is then plotted on this HRI according to their associated probability of occurrence and assigned severity.

There are two key outputs from an FMEA. The first is the associated risk ranking, which aids in identification of high priority failure modes. The second is the associated mitigation strategy. Methodologies such as Reliability Centered Maintenance (RCM) have utilized FMEAs along with criticality assessment in order to develop cost-

effective maintenance programs. RCM compares the associated risk and cost of failure (or doing nothing) with the cost of preventive maintenance or redesign (failure mode mitigation). This comparison is then utilized in order to determine cost effective mitigation strategies. In cases where safety is a concern, failure modes should always include some form of mitigation.

2.2.1 Biases and Team Dynamics Affect Quality

FMEAs are developed utilizing different data sets, as well as anecdotal information and expert knowledge; however, even in cases where a large amount of data is available, the process contains significant amounts of subjectivity. For example, one analyst could air on the side of caution and rank failure modes with a severe consequence, while another might not. This often results the determination of different risk levels and mitigation strategies based on the analysts' own cognitive biases, background knowledge and technical expertise. These differences can potentially lead to a large amount of variation in the process and thus reduce overall quality.

One method to improve quality is utilizing a thorough review process, as well as a team-based approach. Conceptually, addition of team members with varied expertise would improve the overall quality of the product due to the addition of varied levels of expertise and viewpoints. The formation of teams, however, can present a new set of problems. These problems could include communication challenges, group think, ineffective leadership and lack of participation. The analysis is, therefore, not only impacted by individual subjectivity and bias but also team dynamics. The error is in part dependent on the experience of team members in failure analysis and system familiarity, as well as known cognitive biases. The situation is frequently compounded when known

data regarding failure mode occurrence and/or effects is limited. In these situations additional subjectivity is introduced since subject matter expertise and prediction techniques must be utilized. One proposed method to reduce individual bias is to incorporate the right mix of team members, including the customer, engineers, operators, maintainers, and management; however, while these measures are recommended, they do not necessarily reduce human error [25, 26]. Team formation also poses additional concerns, such as team size and inter-personal dynamics. Integration of FMEA within a large quality control system; which includes in-process audits can reduce this risk (assuming that the larger corporate culture does not introduce its own set of bias). One method to improve quality is utilizing a thorough review process, as well as a team-based approach.

2.2.2 Lack of FMEA Validation Research

While some of the available literature does evaluate the FMEA process and explore possible sources of error, research-based validation of FMEA value and effectiveness is severely lacking, as are conclusive recommendations for improvement of the process, in terms of the human factor. Some work has been done within the healthcare community, which previously relied largely upon retroactive risk management, with the goal of quantifying the reliability and validity of FMEA as a technique of risk analysis. Specifically, results from recent studies demonstrated little reliability and validity within the healthcare setting of FMEA. The researchers did not disregard the potential value or conclude that the FMEA process offered no benefit whatsoever, but did clearly reveal flaws which necessitate process refinement [21, 26-29]. These studies highlighted the discrepancies between the severity ratings selected by different groups as well as the lack

of correlation of risks identified. Shebl et al. found that the two different participant groups not only identified different risks, but also rated overlapping risks differently [21]. Differences were also found between conditions which used consensuses to determine an overall severity rating and those that simply averaged individual ratings by Ashley and Armitage [30].

Potts et al. investigated the validity of structured risk analysis methods. They were specifically interested comparing the resultant outcomes of two conditions: analysis using different techniques, and the replicated analysis by different groups using the same technique. They investigated the Structured What If Technique (SWIFT) and HFMEA (FMEA tailored to the healthcare industry) in a workshop setting [31]. Teams of five participants per group worked together to make decisions using each of the two techniques. Additionally, all the participants were new to risk and task analysis with no experience in either technique. The participants were also provided with a hierarchical task analysis diagram previously developed. First, the risks identified in each technique were compared for overlap (both in terms of the actual risk and the associated severity rating) as well as overlap with current risk management processes. The participants identified 61 total risks, with three deemed critical when utilizing the SWIFT approach. The HFMEA resulted in a total of 72 risks, with 12 deemed high risk. The researchers compared the identified risks and concluded that 33 (54.1 %) risks identified by SWIFT were not identified by HFMEA. In turn, HFMEA had 42 (58.3 %) risks that were not identified by SWIFT. Additionally, the researchers concluded that there was little overlap of high risks items between the two analysis techniques [31].

2.2.3 The Risk Priority Number (RPN)

The Risk Priority Number (RPN) that is utilized within FMEA is a method aimed at ranking and prioritizing failure modes – in order to develop mitigation strategies and reduce the overall consequences of the failure mode occurring. The higher the RPN, the higher the possible risk. The goal of the analysis is threefold. First, it provides a way to measure which failure modes pose significant risk and should be removed or mitigated from the system. Second, it provides insight into which failure modes can be corrected or mitigated. Finally, it provides a basis to determine which failure modes can be ignored (and allowed to occur) with no adverse safety or cost implications [32].

Depending on the approach followed, there are slight deviations in how the RPN is calculated. If the Reliability Centered Maintenance (RCM) methodology is followed, criticality is calculated by only including the severity rating and probability of occurrence. Detectability calculations, though not utilized in the RCM methodology, are included in other methods. Bowles [32] points out in his mathematical analysis of the RPN that there are significant flaws within that approach. In the case where the RPN is calculated by multiplying Severity (S) with probability of occurrence (O) and detectability (D), and assuming a range of 1 to 10 for each factor a resulting RPN range from 1 to 1000 is obtained. As explained by Bowles, three factors within the RPN are ordinal scales where items are ranked in series; however, the interval size between measurements are not specified [32]. Siegel (1956) further states that the intervals are determined subjectively and not identical to each other [33]. Thus, conducting multiplication utilizing these factors violates basic mathematical principles [32, 33].

Bowles further highlights four additional concerns with utilizing the RPN in its current form. These include:

- Holes in the scale. The RPN scale is not continuous and various numbers between 1 and 1000 cannot be formed by the product of S, O and D. This is specifically evident in higher numbers (600+). Only 120 unique numbers can be formed with 88 % of the range empty.
- Duplication of RPNs. RPNs can be formed with many combinations of S, O and D thus making the inaccurate assumption that each factor is equally important.
- Sensitivity to small changes. The RPN can be affected significantly by a small change in one factor, especially if the other factors are large numbers.
- Utilizing a single dimension RP to quantify and rank a design encourages management to set arbitrary thresholds- which may not be realistic.

2.2.4 Other FMEA Concerns

The healthcare industry has also questioned both the reliability and validity of FMEA. Reliability has been explored in several studies, with little research performed on FMEA validity [21, 30, 34]. Validity is important since, without it analysis techniques such as FMEA are prone to skepticism regarding their value, and it is difficult to provide feedback in order to improve the tool. Furthermore, as stated by Kirwan (1996) techniques that depend on significant amounts of judgement may not adequately and

accurately quantify risk [35]. In spite of the concerns with the FMEA methodology, the approach has been shown to improve quality and safety in the healthcare field [36].

Validity is defined as a measure to assess whether an instrument measures what it was designed to measure [37, 38]. Thus, in the context of FMEA, validity aims to measure whether or not risks are proactively and thoroughly identified, parameter estimates line up with observed data and identify if the tool is indeed appropriate. Shebl et al. (2012) investigated FMEA validity utilizing several measures [20]. These included face validity, which refers to how relevant experts view the tool. Context validity was defined as a measure of how well the FMEA results mirror information available in the application domain. Criterion validity refers to the level of correlation of the FMEA outputs to other measurement systems of the same variables. And finally, construct validity was used to "determine the extent to which a particular measure relates to other measures consistent with theoretically derived hypotheses concerning the concepts that are being measured" [20].

Their study, conducted during 2012, utilized two groups from a hospital setting that were tasked to complete a FMEA on prescribing, administering and monitoring of two prescription drugs [20]. Both groups were familiar with the process, and provided the same input information. The researchers observed how these drugs were administered in a clinical setting, and mapped the subsequent process to a flow chart. This was then compared to the mapped processes as determined by the FMEA teams in order to address face validity. Although there were differences between the FMEA flowcharts and the observation flow chart, the major steps within the process correlated well. Thus, the team concluded that face validity appeared adequate for FMEA.

Criterion validity was deemed low by Shebl et al. due to several reasons. First, 59 % of failures observed in clinical data were identified by the FMEA team. Additionally, the probability of failure and the actual observed frequency in the clinical data showed little correlation. In general, the FMEA team appeared to estimate the probabilities higher than actual observed data indicated. The team also scored failure modes higher in terms of severity than the data indicated [20]. The research team concluded that FMEA should not be solely utilized in order to understand and quantify risk in a healthcare setting.

CHAPTER III

RESEARCH DESIGN AND METHODOLOGY

3.1 Introduction

This thesis investigated sources of variation in terms of human decision making within the FMEA process, and how both related and unrelated data are categorized and used. It was hypothesized that known cognitive biases, such as anchoring and the ambiguity effect, play a significant role within the FMEA process. The following hypotheses were investigated:

- Hypothesis 1: Inter-rater reliability within the FMEA process will decrease based on the amount of input information (failure and mishap data) provided within the analysis.
- Hypothesis 2: Type of information influences inter-rater reliability in varying degrees and experienced users are less affected by the availability of superfluous information.
- Hypothesis 3: More experienced users will be better able to vet through irrelevant information and be less distracted by it.

Participants were classified as experienced or inexperienced, and tasked to complete three trials of FMEA analysis. In each trial, participants from both groups (inexperienced and experienced) were required to complete the analysis of three failure modes, for a total of nine failure modes per participant. Participants were asked to

determine the qualitative effects and severity level of three failure modes in each trial, given a data set, and if a mitigation strategy was needed for each failure mode. Different input data sets were provided in each trial in order to bias participants towards a more conservative analysis. Details of the study are provided in the following sections.

3.2 Study Definitions

The study utilized standard definitions, familiar to all participants. These definitions were based on industry and government standards such as the NAVAIR 00-25-403 manual [39].

3.2.1 Function

"A function is the intended purpose of an item as described by a required standard of performance [39]."

3.2.2 Functional Failure

"A functional failure is defined as the inability of an item to perform a specific function within the specified limits. A functional failure may not necessarily be a complete loss of the function [39]."

3.2.3 Failure Mode

"A failure mode is a specific physical condition that can result in a functional failure. The failure mode statement should include a description of the failure mechanism (e.g., fatigue) in addition to the specific condition whenever possible [39]."

3.2.4 Failure Effects

"Failure effect is described as the result of a functional failure on surrounding items, the functional capability of the end item, and hazards to personnel and the environment. In other words, it is the impact that a functional failure has on the item under analysis, the surrounding environment (to include equipment and personnel), and the functional capability of the end item [39]. "

3.2.5 Severity

"Severity classifications are assigned to failure modes based on the impacts of their failure effects at the end item level. Classifying failure modes in this manner provides a primary source for determining the priority under which each should be addressed, and may also be used by the program to establish the acceptable probability level for failure modes based on categories of effects [39]."

3.2.6 Mitigation/Task Selection

Mitigation strategies are defined as "the best alternative for either preventing the functional failure altogether, mitigating its consequences to an acceptable level if it does occur, or allowing it to occur and accepting the consequences [39]."

3.3 Study Design

A between subjects design was used to test for the effects of trial (3 levels) and experience (2 levels). A between subjects design was selected so that participants (full time employees) would not be overburdened with the study demands (each trial is expected to take a minimum of 15 minutes). Also, as the objective of this study was not

to compare within individuals, but rather to quantify the effects of experience levels and information quality/quantify on FMEA.

3.3.1 Independent Variables

The two main independent variables studied were trial and experience level. Trial consisted of three levels: (1) no information, (2) irrelevant information, and (3) relevant information. During the first trial participants were not be provided any input information beyond a system description, and failure mode descriptions. In the second trial participants were provided a system description along with irrelevant failure and mishap data of the system being analyzed; however the data was not tied to any failure modes being analyzed (irrelevant data). For example, if the analyst was completing a failure mode of the aileron actuator; irrelevant information consisted of failures of other components as well as mishaps related to the flight control system. During the third trial; a system description along with relevant failure and mishap information was provided.

Experience level pertained to the number of years participants had experience in completing FMEAs. Experienced users were defined as having ten or more years' experience performing FMEAs. In-experienced users were those with a minimum of 1 year of experience up to 9 years. Demographics were collected for use during analysis as appropriate using a custom questionnaire. Participants were selected at random from the volunteer pool and participated in each subsequent trial (three total).

3.3.2 Dependent Variables

Two variables commonly utilized within the FMEA process were recorded: failure mode severity and mitigation (Table 3.1). These variables were selected since they are used in determining the risk of a failure mode occurring, often represented by a Risk Priority Number. In methodologies such as Reliability Centered Maintenance (RCM), FMEAs are extended to also determine appropriate mitigation strategies for failure modes based on a criticality assessment.

Table 3.1 Variables

Variable	Data Type	Possible Values
Severity	Categorical	1-4
Mitigation	Attribute	Yes (1) /No (0)

Analysts were asked to select a severity classification based on Table 3.2 which was adapted from MIL-STD-882E [40].

Table 3.2 Severity Definitions

Catastrophic (I)	Could result in death, permanent total disability, total system loss, or irreversible severe environmental damage that violates law or regulation.
Critical (II)	Could result in permanent partial disability, injuries or occupational illness that may result in hospitalization of at least three personnel, or reversible environmental damage causing a violation of law or regulation.
Marginal (III)	Could result in injury or occupational illness resulting in one or more lost work days(s), or mitigatable environmental damage without violation of law or regulation where restoration activities can be accomplished.
Negligible (IV)	Could result in injury or illness not resulting in a lost work day, or minimal environmental damage not violating law or regulation.

Available mitigation strategies provided to the participants are shown in Table 3 for reference. Participants were not be required to select a specific strategy, however were asked to utilize Table 3.3 when determining if they deem mitigation is appropriate.

Table 3.3 Available Mitigation Strategies

Redesign/Other Action	Redesign includes any additional analysis that should be performed or changing the design itself.
Inspection	Any inspection that can be utilized to assess the condition of the component in question. The assumption is made that any inspection has a 100 % probability of finding defects if present.
Scheduled Removal/Discard	Removal and replacement of the component at a fixed interval.

3.3.3 Sample FMEA

Participants were provided a worksheet for completing the FMEA. Typically larger FMEAs are completed in software or utilize industry schemas prescribed by standards, such as the SAE J1739. These layouts vary depending on industry. These layouts were not utilized due to the limited scope of this study and to avoid overburdening participants with fields not applicable in this research. Table 3.4 provides a sample of three completed failure modes as utilized within this study. Supporting rationale was also documented by participants utilizing a blank sheet of paper. Section 3.5.1 provides all failure modes by trial utilized within this study.

Table 3.4 Sample Completed FMEA Worksheet

System	Aircraft Hydraulic System			
Function	To provide 3000 psi of hydraulic pressure to aircraft systems			
Functional Failure	Fails to provide 3000 psi of hydraulic pressure to aircraft systems			
Failure Mode	Severity Classification			
	I Catastrophic	II Critical	III Marginal	IV Negligible
Hydraulic actuator seals worn	X			
Cockpit control stick cracked beyond limits		X		
Cable from cockpit control stick to actuator frayed		X		

3.4 Participant Selection, Grouping and Sources of Bias

The study solicited volunteers that are active practitioners of reliability engineering, specifically developing FMEAs within the Aerospace industry. Volunteers were not necessarily degreed engineers, however, all had at least one year of practical experience developing FMEAs on real systems. Training alone was not substituted for practical experience. Participants were not screened beyond experience level.

A total of 15 participants completed the study protocols. The volunteer pool was divided into two groups based on experience performing FMEAs. As stated earlier, experienced users were defined as having ten or more years' experience performing FMEAs, while in-experienced users had at least 1 year but no more than 9 years of experience.

Potential sources of bias in this study included:

- Test effect due to learning after each subsequent trial
- Fatigue due to trial repetition ("carry-over effect")
- Knowledge of the respective system being analyzed
- Environment and industry bias

These sources of bias were controlled by implementing several controls. In order to minimize potential bias from improved performance due to multiple trials or participant fatigue, a cool down period was utilized. The cool down period consisted of one week (7 days) between successive trials for each participant. Additionally, participants were asked to not disclose their participation or details of the study to any other participant. Participants were not told what systems they would analyzed or what subsequent trials consisted of. Each trial also involved a different system (landing gear versus flight controls) and no participant analyzed the same system twice.

Trials were randomly assigned to participants. Specifically, a random uniform number was generated for each trial by participant. Thus, a table was constructed with three columns (trial 1, trial 2, and trial 3). The participant completed the associated trial with the smallest number first, followed by the middle number and finally the largest number. The constructed table for all participants was visually evaluated to ensure no apparent pattern existed, and that assignment appeared truly random.

It is well known that the environment can significantly impact human information processing. In order to control environmental bias, each participant was provided a secluded space and the same mathematical tools (calculator and Excel) and tasked to complete the analysis in one setting. Participants were not explicitly monitored or

interacted with, though they were required to complete all tasks in one sitting. Each participant was given unlimited time and complete all activities individually.

The study acknowledges that industry bias cannot be controlled at this stage, and that subsequent research must include participants from other industries.

3.5 Study Method

Participants were presented with a case study in each trial and tasked to develop aspects of a FMEA given a specific amount of information. Participants were presented all failure modes for a given trial simultaneously. Participants were required to determine the effects of given failure modes along with several related variables. The participants were tasked to utilize all the resources provided and determine for each failure mode:

- the local through end item effects of the failure mode (qualitative)
- severity classification (category I, II, III or IV),
- Rationale for decisions

The participants were instructed to evaluate each failure mode assuming no other failures exist, zero maintenance was performed and that end effects were reasonable to occur. For example, if the failure mode was a worn landing gear actuator seal, participants assumed that a leak would occur and remain unnoticed until complete actuator failure. Although this is not necessarily the case, the methodology of assuming zero based maintenance is prevalent in engineering standards, such as the NAVAIR 403 manual and JAE 1011/1012. All participants were familiar with these standards and performed their duties as normal.

Participants were required to document their analysis on a worksheet, and were instructed that all fields were mandatory (i.e., all fields required a response). If any fields

were left blank, their data was not utilized during data analysis. Once all participants have completed their tasks, data was transferred to Microsoft Excel and analyzed.

The following steps were performed in each trial:

1) Participant were briefed on expectations and tasking
2) Participant were provided a secluded space, blank worksheet, calculator and writing utensils
3) Participant were provided handout with applicable information for respective trial:
    a. Trial 1 (Flight Control System): System description and failure modes
    b. Trial 2 (Landing Gear System): System Description, failure modes, Irrelevant failure and mishap data
    c. Trial 3 (Hydraulic System): System description, failure modes, relevant failure and mishap data
4) Participants were provided unlimited time to complete failure modes
5) Handout was collected, steps repeated for next participant

Three different systems (as detailed in section 3.5.1 below) were utilized within this study. Each system utilized three failure modes. Although each participant completed trials in a random order, the systems and associated failure modes remained unchanged across all trials. Thus, trial one always consisted of analyzing a flight control system with three specific failure modes.

3.5.1 Trial 1 Inputs: Aircraft Flight Control System

3.5.1.1 System Description

The analysis utilized a generic fighter jet horizontal stabilator system. The pilot utilizes a control stick in the cockpit (with hydraulic assist) to transmit inputs to the horizontal stabilator actuator mechanically. The actuator receives inputs through mechanical linkage and adjusts the horizontal stabilator appropriately utilizing hydraulic power. The hydraulic actuator includes a cylinder which contains a piston. The hydraulic actuator has several seals to prevent fluid leakage from the piston rod. The stabilator is critical to maintain control of the aircraft during all phases of flight.

3.5.1.2 Failure Modes

- FM 1: Hydraulic actuator seals worn
- FM 2: Cockpit control stick cracked beyond limits
- FM 3: Cable from cockpit control stick to actuator frayed

3.5.2 Trial 2 Inputs: Aircraft Landing Gear System

3.5.2.1 System Description

The analysis utilized a generic fighter jet tri-cycle landing gear system. The landing gear consists of three shock struts that absorb loads during landing. The shock struts are mounted to the airframe utilizing mechanical pins and are operated hydraulically. The landing gear are lowered utilizing hydraulic pressure, however can also be lowered in the event of hydraulic failure utilizing an emergency system. Each strut contains a hydraulic actuator that raises and lowers the gear. The gear are mechanically locked utilizing springs that over-center the respective gear drag brace.

According to the flight manual a pilot should not attempt landing if one main landing gear is not down and locked.

3.5.2.2 Failure Modes

- FM 1: Main landing gear actuator seals worn
- FM 2: Strut hydraulic servicing valve fails open
- FM 3: Nose landing gear strut assembly cracked

3.5.2.3 Failure and Mishap Data (Irrelevant)

The aircraft has experienced 10 mishaps related to landing gear during the past 5 years. Fifty percent of these mishaps have led to serious injury. Thirty percent of the mishaps have unknown root causes. The remaining mishaps have been attributed the wheel and tire assemblies.

Failure data was randomly generated. For each data point an index was provided, which component failed, failure symptom as well as flight hours at failure. A total of 76 data points were provided.

3.5.3 Trial 3 Inputs: Aircraft Hydraulic System

3.5.3.1 System Description

The analysis utilized a generic fighter jet hydraulic system. The hydraulic system powers all aircraft systems to include flight controls. The system operates at 3000 psi and consists of two redundant pumps, hydraulic reservoir, heat exchanger and thermal bypass valve. The bypass valve opens when fluid temperature reaches a pre-defined temperature in order to allow fluid cooling. The aircraft requires at least one hydraulic pump to operator primary flight controls.

3.5.3.2 Failure Modes

- FM 1: Utility hydraulic pump shaft shears
- FM 2: Hydraulic reservoir leaks
- FM 3: Hydraulic thermal-bypass valve fails closed

3.5.3.3 Failure and Mishap Data (Irrelevant)

The aircraft has experienced 9 mishaps related to landing gear during the past 6 years. Fifty percent of these mishaps have led to serious injury. Thirty five percent of the mishaps have unknown root causes. The remaining mishaps have been attributed to hydraulic pumps.

Failure data was randomly generated. For each data point an index was provided, which component failed, failure symptom as well as flight hours at failure. A total of 88 data points were provided.

3.6 Data Coding and Analysis

Next, the data was analyzed by comparing proportions between each trial by group. For example, the proportion of participants that selected a mitigation strategy for each failure mode during each trial was calculated by group (A and B) and compared. Data was compared with all failure modes combined (thus overall proportion of all participants by group for a trial).

The Fisher's exact test was utilized, since the failure mode severity data was categorical in nature with a discrete outcome. The data analysis focused largely on participant selection of either severity class I or IV, with less emphasis on class II and III. Severity class I is considered the most severe consequences (death, loss of system), thus

any failure mode within this category will significantly impact the system design. Severity class IV is on the other end of this spectrum, with most failure modes rated as such requiring no mitigation strategy. Class II and III are generally utilized to relate different financial impacts of the failure mode.

The percent agreement (or proportion of respondents that selected a certain failure mode) was utilized to gain insight into inter-rater reliability (IIR). Several statistics are available to analyze IIR, such as the joint-probability of agreement, Cohen and Fleiss kappa metrics and inter-rater correlation (to name a few). The percent agreement metric was ultimately chosen since the study included multiple participants and categorical data.

\

CHAPTER IV

RESULTS AND ANALYSIS

4.1 Response Rates

The analysis was performed by four experienced users and eleven non-experienced users resulting in 117 analyzed failure modes. The overall response rate, which is defined as the ratio of surveys completed and total provided was also calculated. In this study, each trial was viewed as one survey resulting in a total of 45 surveys given. Surveys were only counted as completed if no questions were left blank by participants. The overall response rate for the study was 87 %. The response rate for the experienced user group was the lowest (75 %) across all three trials. This compared to 91 % for the inexperienced user group. The response rates were also calculated by trial (experienced and inexperienced users combined) resulting in:

- 80 % for trial one,
- 93 % for trial two, and
- 87 % for trial three.

Overall, the high response rates coupled with utilization of simple random sampling (without replacement) indicated the study did not suffer from significant bias.

4.2 Participant Demographics

Demographic data was also collected for all participants for analysis purposes (such as identification of experienced and inexperienced users) and to ensure the sample was representative of the population.

The participants could be divided fairly evenly into two work experience groups (60 % ≥ 15 years, 40 % < 15 years) where work included military experience, technical or non-technical work.

The majority of participants were considered inexperienced (73 % or 11 total) in terms of performing FMEA. However, all participants had at least one year experience with FMEA – thus all participants were familiar with the analysis process and technique.

The sample was mainly drawn from retired military aviation maintainers, thus explaining the difference between work experience and FMEA experience (since participants likely never performed FMEA while on Active Duty), which is illustrated in Figure 4.1. Although work experience was not utilized to identify experience levels, it is an important demographic since it may indicate higher levels of domain knowledge (aviation).

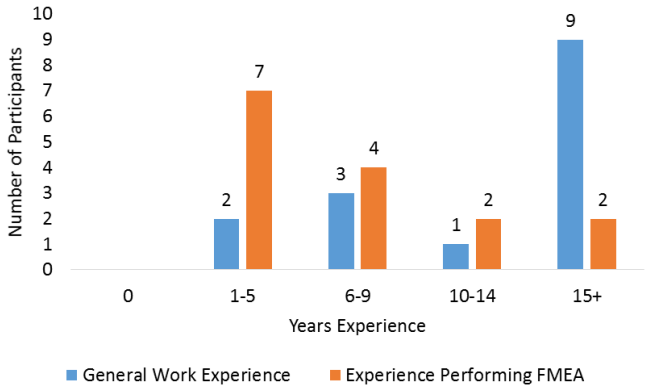

Figure 4.1 Participant Experience (work versus FMEA) by CategoryDemographics

Demographic data such as gender, veteran status, education level and age were also collected as illustrated in Figures 4.2 and 4.3. All participants were male, with the exception of one female. The majority of participants had a college education (93 %) and were veterans of the Armed Services (73 %).

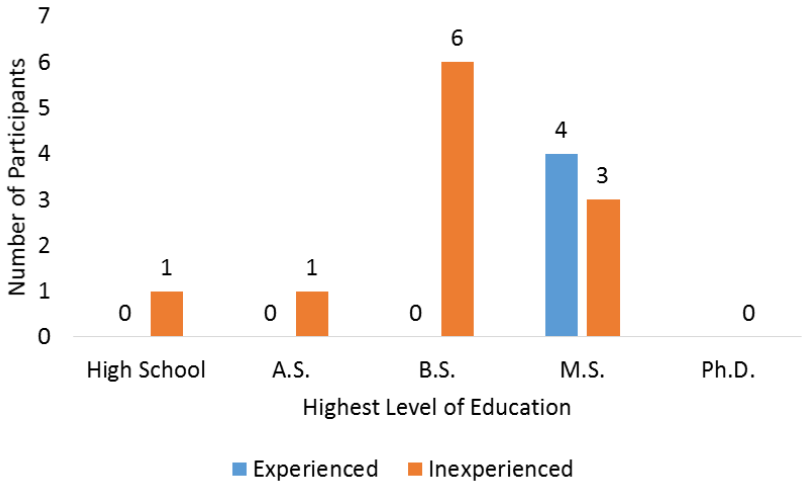

Figure 4.2 Participant Education Level

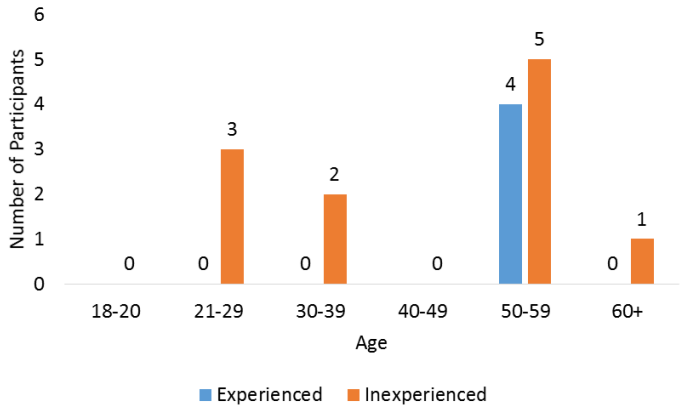

Figure 4.3 Participant Age Distribution

4.3 Results and Discussion

This study sought to establish additional empirical evidence that can be utilized in order to improve the FMEA process. Specifically, the research posed several questions. First, how important is user experience within the FMEA process? Secondly, will the amount of available information impact on the choices made by an analyst during the FMEA process? Thirdly, it was hypothesized experienced will be less distracted by superfluous information. These are important questions and speak to both the reliability and the validity of FMEA. These hypotheses will be discussed concurrently.

The first two hypotheses investigated in this study were related to both the amount of input information (included/not include) and the type (mishap data, failure data) available to the participants.

It was hypothesized that inter-rater reliability within the FMEA process will decrease based on the amount of input information (failure and mishap data) provided within the analysis. Thus, if more information was provided there would be less

consensus amongst participants in terms of severity rating selection. It was also hypothesized that the type of information influences inter-rater reliability in varying degrees, and that experienced users are less affected by the availability of superfluous information. Inter-rater reliability was investigated by first analyzing severity rating selections across the trials regardless of experience level.

As illustrated in Figure 4.4 and Table 4.1, trial 1 had the most consensus (56 % of failure modes classified as severity class I), with the consensus apparently decreasing in trials 2 and 3. In trial 2 the majority of failure modes were still classified as severity class I (36 %). However, in trial 3 there appeared to be the least amount of consensus, with failure modes being evenly divided across all severity classifications. Thus, not only did it appear that inter-rater reliability does decrease as the amount of information is increased, but that the type of information may influence selection in varying degrees.

Table 4.1 Severity Selection across all trials (counts n=117, all participants)

		No Information Provided (Trial 1)	**Irrelevant Information Provided (Trial 2)**	**Relevant Information Provided (Trial 3)**
Severity Rating	**I**	20	15	10
	II	9	6	9
	III	5	12	11
	IV	2	9	9

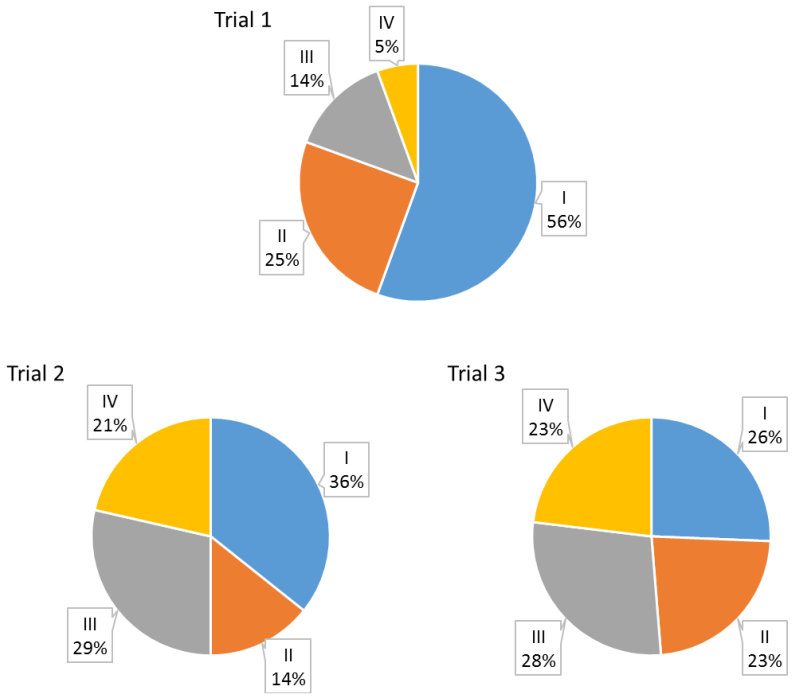

Figure 4.4 Severity Selection across all trials (percentages, all participants)

Each case study was compared to the other case studies in order to determine if there was any statistically significant association between the case study and how many participants selected a severity class of I or severity class of II-IV, regardless of experience level. The input data is provided in Tables 4.2 through 4.4.

Table 4.2 Contingency Table for Trial 1 and 2 to test Association between Trial and Severity Selection

	Number of Participants that Selected Severity Class I	**Number of Participants that Selected Severity Class II - IV**
Trial 1	20	16
Trial 2	15	27

Table 4.3 Contingency Table for Trial 1 and 3 to test Association between Trial and Severity Selection

	Number of Participants that Selected Severity Class I	**Number of Participants that Selected Severity Class II - IV**
Trial 1	20	16
Trial 3	10	29

Table 4.4 Contingency Table for Trial 2 and 3 to test Association between Trial and Severity Selection

	Number of Participants that Selected Severity Class I	**Number of Participants that Selected Severity Class II - IV**
Trial 2	15	27
Trial 3	10	29

The null hypothesis was that there is there is no significant difference between the proportion of participants that selected severity class I or II through IV in case study 1 when compared trial 2 and 3 respectively. Utilizing a two-tailed Fisher's Exact yielded the following results:

- Trial 1 versus 2 was not statistically significant (p-value of 0.1102)
- Trial 1 versus 3 was statistically significant at a significance level of 0.05(p-value of 0.0102)
- Trial 2 versus 3 was not statistically significant (p-value of 0.3479)

Thus, there appeared to be no statistically significant difference between severity class selection when no information (trial 1) versus irrelevant information (trial 2) is provided. However, there did appear to a statistically significant difference between the

severity class selection when no information (trial 1) versus relevant information (trial 3) is provided. There was no statistical significant association when trial 2 and 3 were compared.

Inter-rater reliability was further analyzed by grouping the participant selections in different configurations in order to ascertain whether participant demographics impacted the results previously presented. The data was grouped in the following ways:

- Participants with ≥ 10 years of work experience
- Participants with ≥ 10 years of FMEA experience
- Participants with < 10 years FMEA and work experience

Grouping the data in this manner yielded similar results, with an apparent repeating trend that additional information (trials 2 and 3) and the variability of severity class selections are associated. Specifically, regardless of how the data was grouped, 50 % or more participants selected the same severity class in trial 1 (no information provided). This was contrasted to trials 2 and 3, were it was not typical for more than 50 % of participants selected the same severity class. The only exception was trial 2 for participants with ≥ 10 years of FMEA experience.

The apparent trend between decreasing consensus and the amount of available information can have several reasons. Firstly, it was well known in the cognitive phycology field that too much information can overwhelm the available processing we have when complex tasks are performed. This may result in information cues being missed by some participants resulting in less consensus. It is also important to consider how the information is organized and presented. In this study the information was presented in raw format to the analysis – since this is likely how they will perform the

analysis in a real-world setting. However, participants were not restricted and could have organized the data into other formats if they desired. Interestingly it did not appear that any of the participants reorganized the data. Future work should investigate the role of how information is presented and/or organized within the FMEA process.

Cognitive biases such as anchoring or confirmation biases likely also play a role within the FMEA process. For example, decision makers may bias their severity rating around a certain risk level (for example severity rating of I). An additional bias that may also play a role during risk analysis is the confirmation bias, where information is interpreted in such a way that validates the analyst's priori. These biases were investigated by grouping the proportion of failure modes assigned to each severity level by trial. It is important to remember that each participant completed trials in a random order. For example, the first participant may perform the trials in the order 1-2-3, while another participant may perform them in order 2-3-1 and so forth. As illustrated in Figure 4.5, the proportion of failure modes assigned a severity of I was appeared higher, when compared to trials 2 and 3. However, when trials 2 and 3 are compared a similar proportion of failure modes was assigned to each severity class.

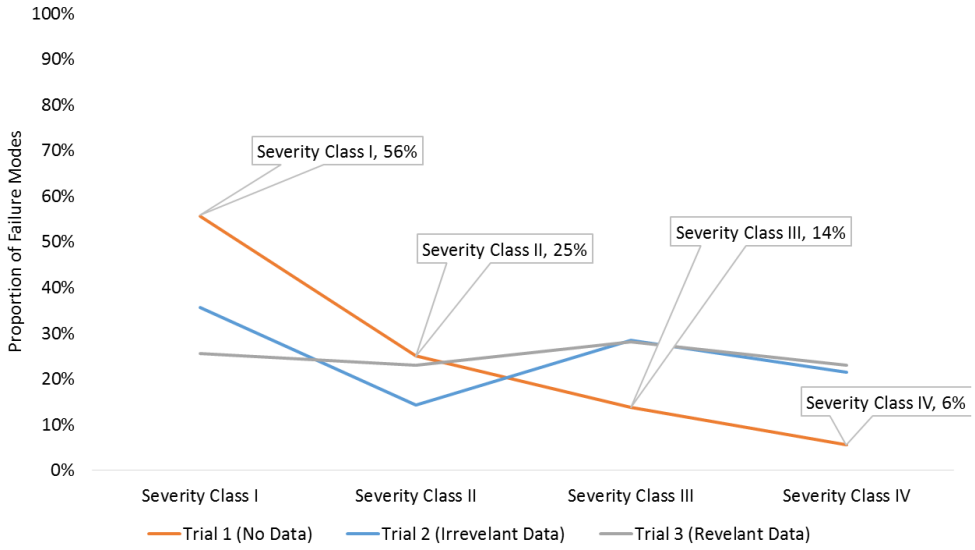

Figure 4.5 Severity Classification by Trial (regardless of Experience)

Although cognitive biases cannot be ruled out as a significant contributor to severity selection, the analysis suggest they did not play a large role in this study due to the low level of consensus and study design. This is due to several reasons to include the randomized order of trial completion and the cooldown period between trials.

However, there appeared to be a clear distinction between trials 1 and 2/3, which was supported by the statistical analysis presented earlier. In trial 1 where no data was provided participants clearly aired on the side of caution and thus attributing a larger portion of failure modes as high risk. However, interestingly this pattern did not hold in trials 2/3 where the proportions were fairly evenly distributed across all severity levels. In these trials each severity category was assigned approximately 25 percent of the failure modes. Similar effects were observed by researchers such as Fox and Clemen [41] who investigated an ignorance prior. The ignorance prior is a cognitive phenomenon where assessor assign probabilities based on the number of categories available. They

concluded and illustrated through five studies that it appears the number of available categories sets the initial probability assessment. For example, in the case of severity classification a failure mode can fit into four categories. Thus, an ignorance prior could be computed as 1/4 or 25 %.

In this study appeared that participants were not able to sift through the provided information and identify the appropriate cues relating to failure relevance to the failure mode under analysis. Thus, it appeared that larger amounts of information will generally reduce conservatism. However, this is concerning since even information not related to the failure mode will influence the decision maker in a similar manner. The researchers further hypothesized that more experienced users would more effectively vet through irrelevant information and be less distracted by it – thus expecting to see higher consistency in experienced users. Table 4.5 provides the contingency table utilized in the statistical analysis.

Table 4.5 Contingency Table for to test Association between Experience Level and Severity Selection

	Number of Participants that Selected Severity Class I	**Number of Participants that Selected Severity Class II - IV**
Experienced Users	12	15
Non-Experienced Users	33	57

The null hypothesis was that there is there is no significant difference between the proportion of participants that selected severity class I or II through IV based on user experience. Based on a two-tailed Fisher's Exact test there wasn't a statistically significant association between the experience level and the number of severity class selection (p-value of 0.5038).

Thus, the null hypothesis was not rejected and it was concluded that experience level did not play a role in severity class selection within this study. This result should be further investigated due to the small number of experienced users that participated in this study (n = 4). This result may also be explained by considering the background of most participants. Almost 70 % of participants came from an aviation background, thus equipped with a large amount of aerospace domain knowledge. The failure modes in this study were all aerospace related. Although, this provided participants with a large amount of domain knowledge related to the study – it may have introduced bias as well. For example, participants may have encountered similar failure modes during their career biasing their selection towards that prior.

4.3.1 Improvement of FMEA

This study did not measure if modifications to the FMEA process would yield improvements to the analysis quality. However, several recommendations, both from literature and author experiences, are provided below [25].

FMEAs should always be performed utilizing a team environment. Although team dynamics may introduce its own set of biases, this study clearly illustrates individual selections will vary significantly. Thus, capturing a group of individual responses and reaching team consensus is important.

Expert opinion may also be subject to bias. This study did not clearly indicate that FMEA and domain experience significantly reduces subjectivity within FMEA. Although expert opinion should not be discarded, it is important to include a wide variety of backgrounds within the FMEA process.

Ensure a strong focus on data quality and not quantity. This research illustrated that the availability of input data influences the severity selections of participants – and that participants may not be able to identify ques distinguishing unrelated versus related data. Thus, the FMEA process should include an assessment of data quality. Additionally, it may be beneficial to segregate data and not provide irrelevant data to analysts. Finally, the analysis must be revisited once data becomes available (if it was not initially available).

CHAPTER V

CONCLUSION, LIMITATIONS AND FUTURE WORK

The results from this study raise important concerns. Is there any utility in performing a FMEA from a risk standpoint if no failure information is available? Based on the results in this study, it appears that participants will be more conservative in severity selection if no information provided. Thus, the analysis may become overly conservative – adding little value in terms of risk mitigation while increasing cost. Secondly, even if information is available but of unknown quality – are results valid? Is experience performing FMEAs (thus process familiarity) actually a good predictor of FMEA quality? Or, instead should subject matter experts on the system being analyzed form a significant portion of the team? These are questions that cannot be conclusively answered by this study alone, however are critical to understand in order to improve the FMEA process.

Although this study utilized a much larger sample size than similar work within the healthcare field the authors acknowledge that a larger group of experienced users must be included. Additionally, the target industry should be expanded beyond aerospace and the results need to be replicated.

However, this study provides further empirical evidence that FMEAs are subject to bias and should be viewed in that light. The authors are not stating that FMEAs have

no value – however based on the work of researchers to include this study the true value from a risk mitigation perspective should be questioned.

This study proposes several potential improvements to the FMEA process. These improvements must be further investigated through research studies. As discussed throughout this thesis, the current FMEA literature lacks validation research. Additionally, no research papers were found at the time of publication that empirically tested modifications to FMEA. Thus, although improvements to FMEA are provided in literature – the majority of these are anecdotal.

REFERENCES

[1] F. M. Gryna, R. C. Chua and J. A. DeFeo, Juran's Quality Planning and Analysis, McGraw-Hill, 2007.

[2] J. G. McLeish, "Enhancing MIL-HDBK-217 reliability predictions with physics of failure methods," in *Reliability and Maintainability Symposium (RAMS)*, San Jose, 2010.

[3] M. Gilmore, "State of Reliability: Memorandum for Principal Deputy Under Secretary of Defense (AT&L)," Department of Defense, Washington, D.C., 2010.

[4] Gilmore, Michael, "Observations on the Marine Corps F-358 Demonstration on USS Wasp," Office of the Secretary of Defense, Washington, D.C., 2015.

[5] J. Pulido, "Effective Implementation of an Enterprise Reliability Program with Supplies," in *Reliability and Maintainability Symposium*, 2013.

[6] H. Arabian-Hoseynabadi, H. Oraee and P. Tavner, "Failure Modes and effects analysis (FMEA) for wind turbines," *International Journal of Electrical Power & Energy Systems,* vol. 32, pp. 817-824, 2010.

[7] C. L. Chang, P. H. Liu and C. C. Wei, "Failure mode and effects analysis using grey theory," *Integrated Manufacturing Systems,* vol. 12, pp. 211-216, 2001.

[8] C. L. Chang, C. C. Wei and Y. H. Lee, "Failure mode and effects analysis using fuzzy method and grey theory," *Kybernetes,* vol. 28, pp. 1072-1080, 1999.

[9] J. K. Chen, "Utility priority number evaluation for FMEA," *Journal of Failure Analysis and Prevention,* vol. 7, pp. 321-328, 2007.

[10] C. Welborn, "Applying failure mode and effects analysis to supplier selection," *The IUP Journal of Supply Chain Management,* vol. 7, pp. 7-14, 2010.

[11] E. Bozdag, U. Asan, A. Soyer and S. Serdarasan, "Risk prioritization in Failure Mode and Effects Analysis using interval type-2 fuzzy sets," *Expert Systems with Applications,* vol. 42, pp. 4000-4015, 2015.

[12] C. R. Fox and Y. Rottenstreich, "Partition Priming in Judgment under Uncertainty," *Psychological Science,* vol. 14, no. 3, pp. 195-200, 2003.

[13] E. G. Carmines and R. A. Zeller, Reliability and validity assessment, USA: Sage, 1979.

[14] J. Pohl, "Elements of Human Decision-making," in *18th International Conference on Systems Research, Informatics and Cybernetics*, Baden-Baden, 2006.

[15] J. Yang, H. Z. Huang, L. P. He, S. P. Zhu and D. Wen, "Risk evaluation in failure mode and effects analysis of aircraft turbine rotor blades using Dempster-Shafer evidence theory under uncertainty," *Engineering Failure Analysis,* vol. 18, pp. 2084-2092, 2010.

[16] W. Song, X. Ming, Z. Wu and B. Zhu, "A rough TOPSIS approach for failure mode and effects analysis in uncertain environments," *Quality and Reliability Engineering International,* vol. 30, pp. 473-486, 2014.

[17] K. H. Chang and C. H. Cheng, "A risk assessment methodology using intuitionistic fuzzy set in FMEA," *International Journal of Systems Science,* vol. 41, pp. 1457-1471, 2010.

[18] J. P. Forgas, M. G. Haselton and W. V. Hippel, Evolution and the social mind: Evolutionary Psychology and Social Cognition, New York: Psychology Press, 2007.

[19] M. Smith, J. Higgs and E. Ellis, "Factors infuencing clinical decision making," *Clinical reasoning in the health professions,* pp. 89-99, 2008.

[20] N. A. Shebl, B. D. Franklin and N. Barber, "Failure mode and effects analysis outputs: are they valid?," *BMC Health Services Research,* vol. 12, no. 150, 2012.

[21] N. A. Shebl, B. D. Franklin and N. Barber, "Is failure mode and effect analysis reliable?," *Journal of Patient Safety,* vol. 5, no. 2, pp. 86-94, 2009.

[22] B. D. Franklin, N. A. Shebl and N. Barber, "Failure mode and effects analysis: too little for too much?," *BMJ Quality and Safety,* 2012.

[23] U. Asan and A. Soyer, Intelligent Decision Making in Quality Management, Atlantis Press, 2015.

[24] J. W. Langford, Logistics: Principles and Applications, McGraw Hill, 1995.

[25] C. S. Carlson, Effective FMEAs: Achieving safe, reliable, and economical products and processes using failure mode and effects analysis, Hoboken: Wiley, 2012.

[26] M. Konstandinidou, Z. Nivolianitou, C. Kiranoudis and N. Markatos, "A Fuzzy Modeling Application of CREAM Methodology for Human Reliability Analysis," *Reliability Engineering and System Safety,* vol. 91, no. 6, pp. 706-716, 2006.

[27] D. Phipps, G. H. Meakin, P. C. Beatty, C. Nsoedo and D. Parker, "Human factors in anaesthetic practice: Insights from task analysis," *British Journal of Anaesthesia,* vol. 100, no. 3, pp. 333-343, 2008.

[28] M. Apkon, L. Probst, J. Leonard, L. DeLizio and R. Vitale, "Design of a safer approach to intravenous drug infusions: failure mode effects analysis," *Quality and Safety in Healthcare,* vol. 13, no. 4, pp. 265-271, 2004.

[29] M. Lyons, S. Adams, M. Woloshynowych and C. Vincent, "Human reliability analysis in healthcare: A review of techniques," *International Journal of Risk & Safety in Medicine,* vol. 16, pp. 223-237, 2004.

[30] L. Ashley and G. Armitage, "Failure mode and effects analysis: An empirical comparison," *Journal of Patient Safety,* vol. 6, no. 4, pp. 210-215, 2010.

[31] H. W. Potts, J. E. Anderson, L. Colligan, P. Leach, S. Davis and J. Berman, "Assessing the validity of prospective hazard analysis methods: a comparison of two techniques," *BMC Health Services Research,* vol. 14, no. 41, 2014.

[32] J. Bowles, "An Assessment of RPN Prioritization in a Failure Modes Effects and Criticality Analysis," in *Reliability and Maintainability Symposium*, 2003.

[33] S. Siegel, Nonparametric statistics for the behavioral sciences, Tokyo: McGraw-Hill Kogakusha Ltd., 1956.

[34] N. A. Shebl, B. D. Franklin, N. Barber, S. Burnett and A. Parand, "Failure Mode Effects Analysis (FMEA): The views of UK hospital staff," *BMC Health Services Research Journal,* vol. 17, no. 1, pp. 34-37, 2011.

[35] B. Kirwan, "). The validation of three human reliability quantification techniques-THERP, HEART and JHEDI: part 1-technique descriptions and validation issues," *Applied Ergonomics,* vol. 27, no. 6, pp. 359-373, 1996.

[36] J. B. Battles, N. M. Dixon, R. J. Borotkanics, B. Robin-Fastmen and H. S. Kaplan, "Sense making of patient safety risks and hazards," *Health Services Research Journal,* vol. 41, no. 4, pp. 1555-1575, 2006.

[37] A. Bowling, Research methods in health-investigating health and health services, Buckingham: Open University Press, 2002.

[38] F. Smith, Research methods in pharmacy practice, London: Pharmaceutical Press, 2002.

[39] United States Navy, NAVAIR 00-25-403 Manual: Guidelines for Naval Aviation Reliability-Centered Maintenance Process, United States Navy, 2005.

[40] Department of Defense, "MIL-STD-882E," Department of Defense, Washington, D.C., 2012.

[41] C. Fox and R. Clemen, "Subjective Probability Assessment in Decision Analysis: Partition Dependence and Bias Toward the Ignorance Prior," *Management Science,* vol. 51, no. 9, pp. 1417-1432, 2005.

[42] S. Guarino and E. Roth, "Modeling human reasoning about meta-information," *International Journal of Approximate Reasoning,* vol. 50, no. 3, pp. 437-449, 2009.

[43] K. M. Tay and C. P. Lim, "Fuzzy FMEA with a Guided Rules Reduction System for Prioritization of Failures," *International Journal of Quality and Reliability,* vol. 23, no. 8, pp. 1047-1066, 2006.