Fall 2017

Emergency Diesel-Electric Generator Set Maintenance and Test Periodicity

Stephen John Fehr
Old Dominion University

EMERGENCY DIESEL-ELECTRIC GENERATOR SET

MAINTENANCE AND TEST PERIODICITY

by

Stephen John Fehr
M.S. December 2014, Old Dominion University
B.S. December 1997, Pennsylvania State University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

ENGINEERING MANAGEMENT AND SYSTEMS ENGINEERING

OLD DOMINION UNIVERSITY
November 2017

Approved by:

T. Steven Cotter (Director)

Holly Handley (Member)

Jennifer Michaeli (Member)

Cesar Pinto (Member)

Resit Unal (Member)

ABSTRACT

EMERGENCY DIESEL-ELECTRIC GENERATOR SET
MAINTENANCE AND TEST PERIODICITY

Stephen John Fehr
Old Dominion University, 2017
Director: Dr. T. Steven Cotter

Manufacturer and industry recommendations vary considerably for maintenance and tests of emergency diesel-electric generator sets in emergency standby duty. There is little consistency among generator sets of similar technology, and manufacturers and their representatives often provide contradictory guidance. As a result, periodicity of emergency diesel-electric generator set maintenance and tests varies considerably in practice. Utilizing the framework proposed and tested by Fehr (2014), this research developed a parametric regression survival model of the reliability of modern diesel-electric generator sets in emergency standby duty as a function of maintenance, age, and cumulative run hours. A survival regression technique leveraging Cox's (1972) methods was developed to combine multiple exponential and Weibull (1951) distributions into a single model to represent emergency diesel-electric generator sets and other complex machinery exhibiting multiple independent failure distributions. A generalized model and reliability tables derived from that model are presented along with maintenance and test recommendations to assist managers in determining the optimal maintenance program for a diesel-electric generator set.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

I. INTRODUCTION

Emergency power systems based on packaged emergency diesel-electric generator sets, referred to herein as generator sets, are installed at facilities where a loss of utility power would result in an unacceptable impact to operational capability or present a risk to life or safety. These generator sets are typically configured to start automatically upon electrical utility failure and assume essential facility loads until utility power was restored. Although emergency generator sets in areas with exceptionally poor quality electrical utility power may run 200 hours or more per year, emergency generator sets in areas with very stable utility power may not run operationally at all in a given year. This represents a significant departure from continuous duty applications for which the diesel engines in these generator sets were typically designed. Such structural variability in operational demand also creates challenges in determining optimal maintenance and test periodicity of critical equipment with high reliability requirements.

Maintenance and test recommendations National in Fire Protection Association (NFPA) publications NFPA 70B (2016) and NFPA 110 (2016) represent the standard for power component and emergency diesel-electric generator set maintenance. These document series were referenced by Department of Defense guidance in Joint Departments of the Army, the Navy, and the Air Force Technical Manual TM 5-683 (1995) *Facilities Engineering Electrical Interior Facilities* and by the Institute of Electrical and Electronic Engineers (IEEE) commercial recommendations (IEEE 3007.2-2010). These documents also included non-specific statements to follow manufacturer recommendations. Standards for recommended preventative maintenance have changed little in past revisions of NFPA 70B and NFPA 110 and are largely based on studies conducted on diesel-electric generator sets in use between 1971 and 1998 (Hale & Arno, 2009; IEEE 493-2007). However, in practice, recommendations for periodicity of maintenance and testing of emergency diesel generator systems differ significantly between organizations, publications, and manufacturers of similar technology systems. Recommendations have even varied between individual manufacturer representatives and publications for a specific model generator set. While manufacturers are often the most knowledgeable about the design of their own equipment, manufacturers' abilities to conduct robust long-term failure modes and effects analyses on fielded units have been limited, and manufacturer maintenance recommendations have often been highly speculative (Moubray, 1997). The result has been inconsistent maintenance and test practices on similar systems. This inconsistency provided an opportunity to quantitatively determine the empirical impact of historic maintenance practices on emergency diesel-electric generator system reliability.

PURPOSE

The purpose of this research was to develop a general model for determining the reliability and optimal test and maintenance periodicities for emergency diesel-electric generator sets supporting critical operations facilities and other facilities requiring highly reliable emergency power. Per NFPA 70 (2017), critical operations power systems facilities encompass Department of Homeland Security and Department of Defense command, control and communication centers, as well as hospitals, police stations, and fire stations. These facilities required the highest levels of readiness, with expectations of one hundred percent mission availability driving power availability requirements in excess of 99.9999% (JIE Operations Sponsor Group, 2014). Such high availability requirements push the limit of what is possible with current technology, even with redundancy and near-elimination of single points of failure.

The emergency power systems that support these facilities have to maintain the highest levels of reliability and availability to meet mission availability requirements while still minimizing unnecessary costs.

This research and the new modeling methods developed for it were intended to provide managers the qualitative data needed to confidently optimize the staffing level, and generator set maintenance plan for each facility. This research was intended to also allow managers to more accurately calculate power system reliability as a function of not only design, but also maintenance. This would give managers flexibility to consider installation design with long-term maintenance plans to achieve reliability goals.

This research was supported by the United States Navy in close cooperation with Old Dominion University. This research was intended to guide future policy for test and maintenance periodicity for United States Navy emergency diesel-electric generator systems and to permit the update of maintenance practices in NFPA 110 (2016) and engineering data in TM 5-968-5 (2006), NFPA 70B (2016), and IEEE 493-2007 (2007). The views expressed herein do not necessarily represent the views of the United States Navy or Old Dominion University.

RESEARCH QUESTIONS AND HYPOTHESES

This research sought to produce new knowledge to answer a series of questions defining the relationship between emergency diesel-electric generator set maintenance, tests, and other properties that may impact reliability. These relationships were modeled mathematically in regression Equation 1, with a descriptive list of database and regression variables in Table 1. As the high reliability of emergency diesel-electric generator sets resulted in a low-occurrence rate of failure events, even with large quantities of data (TM 5-968-5, 2006), high confidence intervals would result in a high risk of Type II errors rejecting valid predictors. Therefore, a significance level of $\alpha \leq 0.10$ was chosen for a confidence interval of 90% for hypothesis testing. The associated risk of Type I errors was considered when analyzing and interpreting results.

$$\begin{aligned}
&\log h(\{t, n_s, T_{rt}\}|x_i) \\
&\qquad = \log(h_0(\{t, n_s, T_{rt}\})) + \sum_{i=1}^{22} \beta_i x_i + \sum_{i=1}^{21} \sum_{j=i+1}^{22} \beta_{ij} x_i x_j \\
&\qquad + \sum_{i=1}^{22} \beta_{bi} x_i x_{bi} + \beta_{make} M_{ake} + \beta_{model} M_{odel} + \beta_{kW} k_W \\
&\qquad + \beta_{age} \log\left(T_{age}\right) + \beta_{rtfv} log\left(T_{rtfv}\right) + \beta_L x_L + \beta_{src} x_{src} + \varepsilon
\end{aligned} \tag{1}$$

Primary research question: What is the relationship between emergency diesel-electric generator set reliability and maintenance, test periodicity, training, make, model, size, age, run time, and load?

Null Hypothesis H_{o0}: Maintenance periodicity, test periodicity, training, make, model, size, age, run time, and load have no impact on generator set reliability.

H_{o0}: $\beta_1 = \beta_2 = \ldots \beta_{22} = \beta_{0,1} = \beta_{0,2} = \ldots \beta_{21,22} = \beta_{make} = \beta_{model} = \beta_{kW} = \beta_{age} = \beta_{rtfv} = \beta_L = 0$, $\alpha \leq 0.10$

Alternate Hypothesis H_{a0}: At least one predictor has an impact on generator set reliability.

H_{a0}: At least one $\beta \neq 0$, $\alpha \leq 0.10$

The case for rejection of H_o in favor of H_a indicates that survival regression models can be applied toward the development of optimal test and maintenance policies for critical equipment operated under high reliability requirements.

Research sub-question 1: What is the relationship between emergency diesel-electric generator set maintenance periodicity and reliability?

Null Hypothesis H_{o1}: Maintenance periodicity has no impact on emergency diesel-electric generator set reliability.

H_{o1}: $\beta_1=\beta_2=\beta_3=\ldots=\beta_{18}=0$, $\alpha \leq 0.10$

Hypothesis H_{a1}: Maintenance periodicity has a significant impact on emergency diesel-electric generator set reliability.

H_{a1}: $\beta_i \neq 0$, $\alpha \leq 0.10$, for any value, i = 1 to 18

Research sub-question 2: What is the relationship between emergency diesel-electric generator set test periodicity and reliability?

Null Hypothesis H_{o2}: Test periodicity has no impact on generator set reliability.

H_{o2}: $\beta_{18}=\beta_{19}=\beta_{20}=\beta_{21}=0$, $\alpha \leq 0.10$

Hypothesis H_{a2}: Test periodicity has a significant impact on emergency diesel-electric generator set reliability.

H_{a2}: $\beta_i \neq 0$, $\alpha \leq 0.10$, for any value, i = 19 to 22

Research sub-question 3: What is the relationship between emergency diesel-electric generator set size and reliability?

Null Hypothesis H_{o3}: Size has no impact on emergency diesel-electric generator set reliability.

H_{o3}: $\beta_{kW} = 0$, $\alpha \leq 0.10$

Hypothesis H_{a3}: Size has a significant impact on emergency diesel-electric generator set reliability.

H_{a3}: $\beta_{kW} \neq 0$, $\alpha \leq 0.10$

Research sub-question 4: What is the relationship between emergency diesel-electric generator set age and reliability?

Null Hypothesis H_{o4}: Age has no impact on emergency diesel-electric generator set reliability.

H_{o4}: $\beta_{age} = 0$, $\alpha \leq 0.10$

Hypothesis H_{a5}: Age has a significant impact on emergency diesel-electric generator set reliability.

H_{a4}: $\beta_{age} \neq 0$, $\alpha \leq 0.10$

Research sub-question 5: What is the relationship between emergency diesel-electric generator set cumulative chronometer run-time and reliability?

Null Hypothesis H_{o5}: Cumulative chronometer run-time has no impact on emergency diesel-electric generator set reliability.

H_{o5}: $\beta_{rtfv} = 0$, $\alpha \leq 0.10$

Hypothesis H_{a5}: Cumulative chronometer run-time has a significant impact on emergency diesel-electric generator set reliability.

H_{a5}: $\beta_{rtfv} \neq 0$, $\alpha \leq 0.10$

Research sub-question 6: What is the relationship between emergency diesel-electric generator set load and reliability?

Null Hypothesis H_{o6}: Load has no impact on emergency diesel-electric generator set reliability.

H_{o6}: $\beta_L = 0$, $\alpha \leq 0.10$

Hypothesis H_{a6}: Load has a significant impact on emergency diesel-electric generator set reliability.

H_{a6}: $\beta_L \neq 0$, $\alpha \leq 0.10$

Research sub-question 7: What is the relationship between the training of service personnel and emergency diesel-electric generator set reliability?

Null Hypothesis H_{o7}: Training of servicing personnel has no impact on emergency diesel-electric generator set reliability.

H_{o7}: $\beta_{b1}=\beta_{b2}=\ldots=\beta_{b22}=0$, $\alpha \leq 0.10$

Hypothesis H_{a7}: Training of service personnel has a significant impact on emergency diesel-electric generator set reliability.

H_{a7}: $\beta_{bi} \neq 0$, $\alpha \leq 0.10$, for any value, i = 1 to 22

Research sub-question 8: What is the relationship between the make and model of emergency diesel-electric generator set and emergency diesel-electric generator set reliability?

Null Hypothesis H_{o8}: Make and model have no impact on emergency diesel-electric generator set reliability.

H_{o8}: $\beta_{make}=\beta_{model}=0$, $\alpha \leq 0.10$

Hypothesis H_{a8}: Make and/or model have a significant impact on emergency diesel-electric generator set reliability.

H_{a8}: $\beta_{make} \neq 0$ or $\beta_{model} \neq 0$, $\alpha \leq 0.10$

Table 1. Table of Predictor, Data and Variable Descriptions

Symbol	**Description**
I_D	Unique record identification number for each generator set rational subgroup
F_{ID}	Failure ID, unique within each generator set rational subgroup ($I_D.F_{ID}$)
D_{ate}	Date the record was recorded in the survey format
N_{ame}	The assigned name or designation of a particular generator set
M_{ake}	Generator set manufacturer
M_{odel}	Generator set model
k_W	Generator full load rating, in electrical kilowatts (ekW)
k_{VA}	Generator full load rating, in kilovolt-amps
n_s	Number of generator starts in the reporting period
I_{nstall_Date}	Installation date (calendar), the date the generator set was installed
T_s	Start date (calendar), date of the start of the reporting period
T_e	End date (calendar)
T_{rts}	Run-time start (hours), generator set chronometer (run-hours) at the start of the reporting period
T_{rte}	Run-time end (hours), generator set chronometer (run-hours) at the end of the reporting period
T_{rt}	Total run-time (hours), total run-hours in the reporting period
F_s	Total failures to start in the reporting period
F_r	Total failures while running in the reporting period
F_{st}	Number of failures to start during testing
F_{rt}	Number of failures while running during testing
F_{so}	Number of operational failures to start
F_{ro}	Number of operational failures while running
T_{fv}	Failure date (calendar), the date the failure event was observed
T_{age}	Age (yrs), the generator set age at the failure event
T_{rtfv}	Run hours at failure (hrs), the chronometer (run-hours) at the failure event
T_{ttrv}	Time to repair (hrs) for this failure event
F_{sv}	Failure to start (Boolean), for this failure event
F_{rv}	Failure while running (Boolean), for this failure event
F_{tv}	Failure during testing (Boolean), for this failure event
F_{ov}	Failure during operation (Boolean), for this failure event
F_v	Any failure (Boolean), for this failure event; $F_v=F_{tv}+F_{ov}$

Table 1. Continued

Symbol	Description
x_L	Typical load, as percent of generator full load kW rating
x_1	Maintenance periodicity (hrs.), contractor service visit; details not known
x_2	Maintenance periodicity (yrs.), check alarms
x_3	Maintenance periodicity (yrs.), check switch & breaker positions
x_4	Maintenance periodicity (yrs.), visual inspection for leaking fluids
x_5	Maintenance periodicity (yrs.), visual inspection of hoses, cables, etc.
x_6	Maintenance periodicity (yrs.), check fuel level
x_7	Maintenance periodicity (yrs.), check oil level
x_8	Maintenance periodicity (yrs.), check coolant level
x_9	Maintenance periodicity (yrs.), check air filter
x_{10}	Maintenance periodicity (yrs.), battery voltage & physical condition
x_{11}	Maintenance periodicity (yrs.), check fan belt(s)
x_{12}	Maintenance periodicity (yrs.), battery resistance or impedance test
x_{13}	Maintenance periodicity (yrs.), clean unit exterior (including radiator & louvers)
x_{14}	Maintenance periodicity (yrs.), fuel cleaning (or fluid analysis)
x_{15}	Maintenance periodicity (yrs.), oil change (or fluid analysis)
x_{16}	Maintenance periodicity (yrs.), check electrical tightness
x_{17}	Maintenance periodicity (yrs.), engine intensive maintenance
x_{18}	Maintenance periodicity (yrs.), generator (electrical) intensive maintenance
x_{19}	Test periodicity (yrs.), generator set no-load test
x_{20}	Test periodicity (yrs.), generator set load test on load bank
x_{21}	Test periodicity (yrs.), generator set load test on operational load
x_{22}	Test periodicity (yrs.), generator set dead-bus test on operational load
x_{b1}	Servicing personnel training (factor), contractor service visit; details not known
x_{b2}	Check alarms
x_{b3}	Servicing personnel training (factor), check switch & breaker positions
x_{b4}	Servicing personnel training (factor), visual inspection for leaking fluids
x_{b5}	Servicing personnel training (factor), visual inspection of hoses, cables, etc.
x_{b6}	Servicing personnel training (factor), check fuel level
x_{b7}	Servicing personnel training (factor), check oil level
x_{b8}	Servicing personnel training (factor), check coolant level
x_{b9}	Servicing personnel training (factor), check air filter
x_{b10}	Servicing personnel training (factor), battery voltage & physical condition
x_{b11}	Servicing personnel training (factor), check fan belt(s)
x_{b12}	Servicing personnel training (factor), battery resistance or impedance test
x_{b13}	Servicing personnel training (factor), clean unit exterior (including radiator & louvers)
x_{b14}	Servicing personnel training (factor), fuel cleaning (or fluid analysis)
x_{b15}	Servicing personnel training (factor), oil change (or fluid analysis)
x_{b16}	Servicing personnel training (factor), check electrical tightness
x_{b17}	Servicing personnel training (factor), engine intensive maintenance
x_{b18}	Servicing personnel training (factor), generator (electrical) intensive maintenance
x_{b19}	Servicing personnel training (factor), generator set no-load test
x_{b20}	Servicing personnel training (factor), generator set load test on load bank

Table 1. Continued

Symbol	Description
x_{b21}	Servicing personnel training (factor), generator set load test on operational load
x_{b22}	Servicing personnel training (factor), generator set dead-bus test

SYSTEM COMPONENTS

The types of emergency diesel-electric generator sets investigated in this research were packaged diesel-electric generator sets with the following characteristics:

- turbocharged fuel-injected diesel piston engine prime-mover;
- operating speed of 1500 or 1800 revolutions per minute;
- direct coupled to an alternating current brushless three-phase electrical generator with 120Y/208, 230Y/400 or 277Y/480 volt output at 50 or 60 Hertz;
- a low-voltage electric starting system with lead-acid batteries operating at between 12V-48V;
- an air-to-water/glycol radiator-based cooling system; and
- a diesel fuel oil system.

The focus of this research was on high-efficiency low-emission units of these characteristics between 60kW and 2.5MW electrical capacity that have been installed in the past twenty years at critical operations power facilities and that run fewer than two hundred hours per year. These generator sets generally include optional components to increase reliability such as jacket water heaters, strip heaters and dual electric starters. A photograph of a pair of typical generator sets included in this research is shown in Figure 1. The process flow for emergency power system reliability is shown in Figure 2. Some generator sets that differed in some way but were still appropriate to include, such as diesel engines featuring pneumatic start, or engines with diesel blocks adapted to natural gas, were included when data was available.

Figure 1. Typical packaged emergency diesel-electric generator sets.

Although electronic control systems and automatic transfer switches play an important role in overall emergency power system performance, NFPA 70B (2016) maintenance periodicity recommendations exceed one year for most preventative maintenance actions, and detailed maintenance records are rarely kept for this equipment. The combination of long-interval maintenance and lack of records would make application of the Fehr (2014) framework difficult for this equipment. However, the primary serviceable components comprising these systems, batteries and breakers, are used in other applications for which reliability-centered maintenance failure modes and effects analysis can be performed. Many control system components have no applicable preventative maintenance beyond cleaning. The characteristics of preventative maintenance and primary failure modes of controls and automatic transfer switches are beyond the scope of this research and are not considered herein.

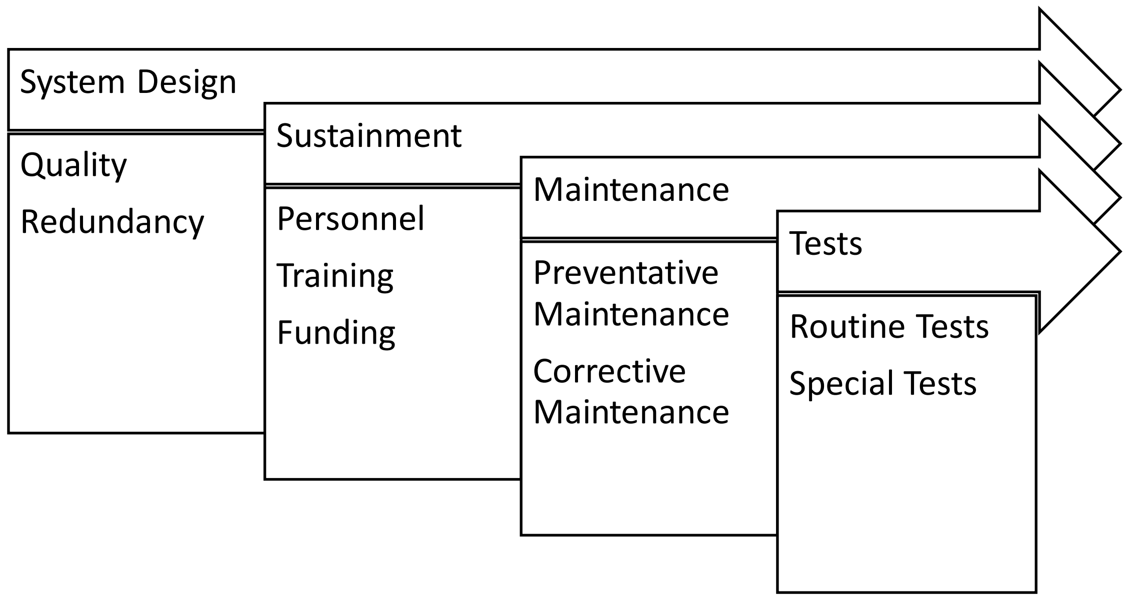

Figure 2. Emergency generator system reliability process flow.

MAINTENANCE

Maintenance represents a combination of preventative maintenance and corrective maintenance. Preventative maintenance is performed at regular intervals and is intended to reduce the failure rate. Corrective maintenance is not performed at regular intervals and involves repairs that are discovered and corrected before resulting in an operational failure. A typical preventative maintenance plan includes very simple items at frequent intervals, such as visual inspections to ensure vents and louvers are not blocked, with more intensive items at longer intervals, such as replacing piston liners and main crankcase bearings. In some context, maintenance includes routine testing as well. This study focuses on routine preventative maintenance actions recommended by NFPA 110 (2016) with intervals of a year or less, as shown in Appendix D.

TESTS

Routine tests of emergency diesel-electric generator sets are categorized in this research as one of four general tests, which will be referred to as no-load tests, load bank tests, operational load tests, and dead-bus tests (Fehr, 2014).

NO-LOAD TESTS

No-load tests of emergency diesel-electric generator sets involve starting the generator set, allowing it to run at idle for a short period, typically between 15 and 60 minutes, and then turning it off. This tests the starting system, engine, and some aspects of the control system, but this does not place the engine under load and does not test the transfer switch. No-load tests represents low-risk to the operator because a test failure has little direct impact on ongoing operations of the facility. This test is often run at weekly or biweekly intervals, but running a diesel engine at low loads and low operational temperatures can cause unburned diesel fuel to build up in the exhaust stack, high moisture content in the lubricating oil, and other unwanted conditions (Loehlein, 2007; Tufte, 2014). While some maintenance manuals recommend no-load tests as part of routine maintenance (Caterpillar, 1997; Caterpillar, 2010a; Caterpillar, 2010b; Caterpillar, 2010c; Caterpillar, 2010d), and use of weekly exercisers is commonly used to automatically run no-load tests at many facilities, other maintenance manuals and many technicians recommend against it, believing it does more harm than good (Loehlein, 2007).

LOAD-BANK TESTS

Load-bank tests of emergency diesel-electric generator sets involve starting the generator set and using a load bank to simulate station loads. This provides a more thorough operational test than a no-load test with a similarly low level of risk but does not exercise or test the transfer switch. Use of a load bank is often the most practical way to test a generator set to full rated operational load. While most load banks are purely resistive, reactive load banks can simulate the power factor of many inductive or capacitive loads. NFPA 110 (2016) recommends performing a stepped load-bank test to 100% of rated capacity at system commissioning and following intensive maintenance, but NFPA 110 only recommends routine load-bank testing if site operational loads are low. For sites with low operational loads, a load bank permits testing of the site power equipment at higher loads than would normally be possible.

OPERATIONAL LOAD TESTS

Operational load tests of emergency diesel-electric generator sets involve starting the generator set and transferring the facility load to the generator system. This test is frequently accomplished by momentarily paralleling the generator sets with utility power to avoid a break in facility power or by synchronizing the generator phase angle to match utility power and then performing an open-transition transfer with an interruption of power lasting no more than 100 milliseconds. A monthly load test including the exercising of automatic transfer switchgear is legally required by NFPA 110 (2016) for generator sets in some applications including life-safety and for Department of Defense generator sets by Joint Departments of the Army, the Navy, and the Air Force Technical Manual TM 5-683 (1995).

DEAD-BUS TESTS

A dead-bus test involves simulating a utility failure and is the most comprehensive and operationally realistic generator test. This test, by its nature, requires a momentary break in facility power and increases the risk of an uninterruptible power supply (UPS) failure causing an uninterruptible critical power outage. It also results in nuisance outages to equipment not supplied with uninterruptible critical power. The generator system experiences the full in-rush and magnetization currents of station loads during a dead bus test, so this test can reveal problems not apparent during paralleled transfer or open-transition synchronized operational load tests.

GENERATOR SET RATINGS

ISO 8528-1 (2005) defines generator set duty ratings by four categories. Emergency standby rated generator sets are capable of delivering up to 200 hours of operation per year at an average of up to 70% of the generator set rating over any 24 hour period. Limited-time running rated generator sets are capable of delivering up to 500 hours of operation per year at 100% of the generator set rating. Prime rated generator sets are capable of unlimited annual running time at an average of up to 70% of the generator set rating over any 24 hour period. Continuous rated generator sets are capable of unlimited annual running time at 100% of the generator set rating. Although prime and continuous rated generators are not restricted in annual run time by the manufacturers, they cannot be run continuously in practice due to maintenance requirements that require shut-down to perform. The type of generator sets included in this research are not typically used in prime or continuous power applications, as they are not typically economical in those applications, but prime or continuous rated 1500-1800rpm generator sets are often selected for emergency standby use if there is risk that extended utility power outages might occur that could require generator sets to operate for more than 200 hours in one year.

It is common practice for manufacturers to dual-rate a diesel-electric generator set model at one capacity rating for standby duty and at a 10% lower kW rating for prime duty. For example, a generator set might be rated 1.2MW for prime duty and 1.32MW for standby duty and may even have both ratings listed on the nameplate. Generator sets are sized at some sites by their prime rating to allow an emergency plant to operate for extended periods of utility outage without violating manufacturer ratings but, in all other respects, perform as an emergency standby generator set. The Fehr (2014) framework considered prime rated generators running in emergency standby duty as if they were emergency standby generators and does not differentiate between these two ratings. This research included standby, prime and continuous rated generator sets but was delimited exclusive to those generator sets that operate normally in emergency standby duty and have not exceeded 200 hours of operation in any one year since installation.

Another common practice among generator manufacturers is to de-rate one model and sell it as a lower-rated model. For example, an 800kW generator set may also be sold as a 650kW generator set for marketing and price stratification purposes with only minor differences in programming and construction between the 650kW and 800kW models. The Fehr (2014) framework did not differentiate based on the potential capacity of various frame sizes and treated each generator set by its reported nameplate rating. The Fehr (2014) framework was structured to detect statistically significant differences in performance between different generator makes and models, although it cannot discern between manufacturing tolerances and design or material changes made during a production run of a particular model series.

II. LITERATURE REVIEW

Despite the ubiquity of emergency diesel-electric generator sets in commercial and industrial facilities, there has been very little published research on the impact of maintenance and tests on the reliability of diesel-electric generator sets manufactured in the last twenty years. Hale and Arno (2009) indicated maintenance quality level was influenced equipment availability in previous studies, but viewed it as a source of potential bias and but those studies did not attempt to quantify equipment reliability or availability as a function of maintenance quality. In the generator reliability studies Hale and Arno (2009) performed in the 1990s, they carefully chose diversified data sets to reduce the potential of bias from maintenance quality.

Although there is some published research on older diesel-electric generator sets in service during the 1970s, 1980s and 1990s, stringent emissions and environmental restrictions in the United States and Europe have driven design changes in the diesel engines powering emergency diesel-electric generator sets produced since the early 1990s when the Euro and Tier emissions standards came into effect in Europe and the United States. Design changes in these modern diesel engines include increased fuel injection pressures, retarded injection timing, exhaust gas recirculation, higher peak combustion pressures, and articulated pistons with steel crowns and high top rings (Margaroni, 1999; Walbolt, 2010), as well as sophisticated emissions monitoring systems and digital controls. Advances in metallurgical techniques, emissions reduction techniques, and component designs continue to improve performance (Walbolt, 2010). Changes since the 1990s are known to impact the life of lubrication oils (Margaroni, 1999), but impact of this and other changes with respect to reliability as a function of maintenance and testing of units in emergency standby duty is not yet well known.

NFPA 70B (2016) and NFPA 110 (2016) recommendations represent the standard for power component and emergency diesel-electric generator set maintenance and are referenced by Department of Defense guidance (TM 5-683, 1995) and IEEE publications (IEEE 3007.2-2010), along with recommendations to follow manufacturer recommendations. The commercial standards for recommended preventative maintenance have changed little in past revisions of NFPA 70B and NFPA 110 and are largely based on studies conducted on diesel-electric generator sets in use between 1971 and 1998 (Hale & Arno, 2009; IEEE 493-2007).

Manufacturers of large emergency diesel-electric generator sets include Caterpillar, Cummins Power Generation, Generac Power Systems, Detroit Diesel/MTU Friedrichshafen, Volvo-Penta, SDMO, and Kohler. Some of these manufacturers directly reference NFPA 110 for recommended maintenance and tests, but others have model-specific maintenance and test recommendations. These engines have a lot of similarity of design and often include components manufactured by the same suppliers as their competitors (Walbolt, 2010), and it’s possible to find major components such as entire engines in generator sets of different manufacturers.

General recommendations published by Caterpillar (SEBU6042-04, 1997) closely match most of the maintenance recommendations of NFPA 110 (2016) including weekly inspection, weekly fluid checks, and additional maintenance at one-year and three-year intervals. While NFPA 110 (2016) requires monthly generator load tests, Caterpillar only recommends weekly no-load tests, with no mention of monthly tests that are legally required on units supporting life-safety equipment. Other specific maintenance recommendations differ between similar models of the same family of generator sets (Caterpillar, 1997; Caterpillar, 2010a; Caterpillar, 2010b; Caterpillar, 2010c; Caterpillar, 2010d). Manufacturer-certified technicians often contradict manufacturers’ published recommendations with respect to tests. Some technicians feel that no-load tests damage the engine and should be avoided, while others strongly advocate weekly no-load tests, and still others recommend only quarterly maintenance. While Caterpillar publications recommend weekly and monthly maintenance and tests on all emergency diesel-electric generator sets, Caterpillar honors the manufacturer's warranty on generators that receive only quarterly service provided by qualified technicians.

Table 2. Cummins Power Generator Recommended Maintenance (Loehlein, 2007)

Maintenance Items	Service time				
	Daily	Weekly	Monthly	6 Months	Yearly
Inspection	X				
Check coolant heater	X				
Check coolant level	X				
Check oil level	X				
Check fuel level	X				
Check charge-air piping	X				
Check/clean air cleaner		X			
Check battery charger		X			
Drain fuel filter		X			
Drain water from fuel tank		X			
Check coolant concentration			X		
Check drive belt tension			X		
Drain exhaust condensate			X		
Check starting batteries			X		
Change oil and filter				X	
Change coolant filter				X	
Clean crankcase breather				X	
Change air cleaner element				X	
Check radiator hoses				X	
Change fuel filters				X	
Clean cooling system					X

The published recommendations of Cummins Power Generation (Loehlein, 2007), shown in Table 2, are more stringent than NFPA 110 (2016). Cummins Power Generation recommends performing daily checks for a number of items that Caterpillar and NFPA 110 (2016) recommend performing weekly. Cummins Power Generation explicitly recommends holding periods of no-load operation to a minimum and recommends a 30-minute generator load test once a month, similar to the monthly load test required by NFPA 110 (2016) and Joint Departments of the Army, the Navy, and the Air Force Technical Manual TM 5-683 (1995).

HISTORICAL EMERGENCY DIESEL-ELECTRIC GENERATOR RESEARCH

Fehr and Cotter (2014) proposed the methodology used herein to determine the relationship between generator set maintenance and testing and reliability. Fehr (2014) expanded and tested this methodology, but data acquisition was limited only to a small number of well-maintained generator sets to test Fehr's methods, and did not include enough operational failure data to achieve statistically significant results. Nevertheless, Fehr's initial small-scale data analysis validated the methodology and the data provided important information on the mean reliability of well-maintained generator sets, even if the 2014 study was insufficient to determine relationships between maintenance and test predictors and reliability.

The United States Army Corp of Engineers' Power Reliability Enhancement Program (PREP) investigated the reliability and availability of emergency generators from studies compiled from multiple sources in the early 1970s. PREP discovered these earlier studies contained confusing information, and the database often contradicted itself (Hale & Arno, 2009). Those studies are obsolete now, but they were the foundation that commercial and governmental policies for emergency diesel-electric generator set maintenance was built upon.

The most recent large-scale study on generator set reliability was conducted on behalf of PREP in the mid-1990s, and the results were compiled in TM 5-698-5 (2006). This was a broad study looking to update previous records by looking at the contemporary technology equipment installed since 1971. The PREP study forming the basis for TM 5-698-5 (2006) recognized that there are differing levels of maintenance for different generators but did not differentiate between the reliability of each maintenance plan. Instead, the authors chose a cross-section of generators of differing maintenance to reduce bias so that they could present a single set of reliability and availability numbers for each category of equipment for the purpose of system reliability and availability calculations. This PREP study assumed exponential failure distributions and calculated an annual reliability factor of 0.8838 for packaged diesel engine generators of 250kW-1.5MW rated capacity in standby duty, based on 672.1 unit-years of operation with 83 failures. This PREP study calculated an annual reliability factor of 0.5310 for unpackaged diesel engine generators of 750kW-7MW rated capacity in standby duty, based on 235.4 unit-years and 149 failures.

IEEE 493-2007, often referred to as the *IEEE Gold Book*, is the commercial standard for design of emergency and critical power plants. It contains methods for calculating overall power system reliability and contains reliability and availability values for making those calculations. IEEE 493-2007 references a 1980 generator survey that states the failure rate of emergency and standby generators is 0.00536 failures per run-hour and 0.0135 failures per start attempt, with an aggregated failure rate of 0.1691 failures per year and an average downtime of 478.0 hours per failure. This differs from the results of a later study presented in the same IEEE document and TM 5-698-5 (2006), which found 0.1235 failures per year and a mean time to repair of 18.28 hours. While the reduced failure rate from the 1980 study to the late-90s study could be explained as increased reliability from technological advances, the order of magnitude disparity in downtime and mean time to repair is difficult to ignore and could be a result of using small pools of data or including out-of-control data in the average. The disparity and scarcity of data from the two primary studies puts the reliability of this data in question. If the reliability of data for such high profile equipment as generators is in question, the reliability of data for other equipment with lower incidence of record keeping is also in question.

Fehr (2014) used TM 5-698-5 (2006) methods to calculate an annual reliability factor of 0.921 and 0.074 failures per year based on 126.71 unit-years of well-maintained generator sets in standby service operation and 9 test and operational failures. This is a much lower failure rate than 0.1235 (TM 5-698-5, 2006) or 0.1691 (IEEE, 2007) recorded by previous studies. Fehr (2014) estimated the inherent availability $A_i = 0.999712$ for the well-maintained generator sets in that data set, which is an order of magnitude lower unavailability than the $A_i = 0.9974$ listed in the PREP database (TM 5-698-5, 2006). It is not clear from prior research whether the discrepancies between these results is due to higher reliability of the latest models of generators, due to different maintenance practices, or due to some combination of these or other conditions.

DIESEL ENGINE MAINTENANCE RESEARCH IN OTHER APPLICATIONS

There have been several studies researching preventative maintenance and replacement cycles for diesel engines in transportation and construction fleets, but emergency generator sets run at a much different duty cycle with fewer run hours than most other diesel engines and exhibit different wear profiles. Though the findings of these studies are not directly applicable to emergency diesel-electric generator sets, the structure of the studies, models used, and other aspects of this research are useful.

Márquez and Herguedas (2004) investigated the failure rate of diesel engines powering earthmoving equipment in mining operations in Spain, concentrating on cylinder liner failures in 1.8MW, 16-cylinder diesel 1900 rpm engines similar to those used for emergency generators. Through their research, they discovered 50% of the failures were occurring in 24% of the cylinders and worked with the manufacturer to determine the assignable cause was excessive vibrations in the crankshaft at high engine inclinations. This allowed the manufacturer to address the problem in future designs and for the mine maintenance departments to increase preventative maintenance on the problem cylinders. Márquez and Herguedas (2004) used maintenance records to conduct this analysis. The records included fifteen trucks with twenty-three failures. They recognized data censoring and devised a model that was insensitive to the data censoring. For the analysis, they simplified the data results and performed a bi-parameter Weibull plot of engine run-hours with linear and quadratic trend regression. They were then able to select a tri-parameter Weibull to analyze the failure of specific cylinders.

Leung and Lai (2003) investigated the preventative maintenance and replacement of diesel engines powering city buses in Hong Kong. They reviewed a subset of 2,282 repair records from buses powered by 171.5kW 1900 rpm turbocharged diesel engines and 134.2kW 1850 rpm naturally aspirated diesel engines. They used the maximum-likelihood density estimation (MLDE) and nearest-neighbor density estimation (NNDE) procedures with the sequential method to determine optimal preventative maintenance and replacement intervals. By these means, they calculated the lowest combined total cost of preventative maintenance, corrective maintenance, and opportunity costs lost during maintenance and repair. Leung and Lai determined that the sequential method was better than the non-homogeneous Poisson process (NHPP) model for analyzing this engine data, as the NHPP model assumed repairs returned the system to original condition while the sequential model accounted for slow degradation of engine components due to use.

The duty cycle of city busses and emergency diesel-electric generator sets differ considerably. The city buses in Leung and Lai's (2003) study experience more run-hours in one or two days than a typical emergency diesel-electric generator experiences in a year, so the assumptions made by Leung and Lai regarding slow degradation may not entirely apply to emergency diesel-electric generator sets. The sequential method may not provide any advantages over NHPP for analyzing emergency diesel-electric generator set reliability.

Table 3. Equipment Failure Rate Multipliers vs. Maintenance Quality (IEEE 493-2007)

Maintenance Quality	**Transformers**	**Circuit Breakers**	**Motors**
Excellent	0.95	0.91	0.89
Fair	1.05	1.06	1.07
Poor	1.51	1.28	1.97
All	1.00	1.00	1.00
Perfect Maintenance	0.89	0.79	0.84

A study published in 1974 regarding the impact of maintenance on the reliability of electrical equipment in industrial plants found that maintenance quality and periodicity had a significant impact on failure reduction. Failure rate multipliers were calculated from this data showing that excellent maintenance could increase reliability of those power system components 40-120% more than similar components receiving poor maintenance (IEEE, 1974). These values are shown in Table 3.

The airline industry and FAA found that preventative maintenance was only effective for items with certain failure patterns and had no benefit for other areas, forming the basis of reliability-centered maintenance and failure modes and effects analysis (Moubray, 1997; IEEE 493-2007).

RARE EVENTS SURVIVAL ANALYSIS

While the field of risk analysis includes a great deal of research on rare events, rare events are considered differently in statistical analysis and risk analysis, as risk analysis considers events as a combination of occurrence rate and consequence (Osterloh & Jaenish, 2016) while statistical analysis only considers occurrence rate. Another difficulty in researching statistical techniques for rare events is that there is no single universal definition. Some risk analysis researchers calculate numerical probabilities while others use subjective definitions and rely on polls of experts to develop quantitative results. While subjective polling techniques may be applicable to extremely complex systems that are difficult to mathematically model, or events so rare there is little hard quantitative data available, they're less precise in areas where numerical studies are possible, and can be off by many orders of magnitude (Osterloh & Jaenish, 2016).

Rowe (2006) defines a rare event as one that has $np < 0.01$ chance of occurring per year, and defines all others as ordinary events. While total power system failures may be considered rare events in facilities with power systems designed with redundancy and high availability, the failure rate of an individual generator set is reported by previous studies as and $\lambda = 0.074$ (Fehr, 2014) and $\lambda = .1235$ (TM 5-695-5, 2006), neither of which are considered rare events per Rowe. However, statistical techniques describing ordinary events create difficulties for events that while perhaps not considered rare, remain uncommon. Some of the challenges present in rare events analysis are present in the field of generator set maintenance, such as belief overcoming hard data.

Xiao and Xie (2016) found maximum likelihood estimator techniques resulted in ill-posed estimates when no failures were observed and resulted in very high variances when the time, *t*, is small. However, Xiao and Xie also found that given enough time, the maximum likelihood estimate can provide good results. While generator set failures may be observed as rare by maintainers whose experience is limited to a small set of units, generator set failures are considered ordinary rate events in statistics. Therefore, increasing the amount of data in the study may be a reasonable means to achieve high quality results. This finding may help address the concern raised by Fehr (2014) that poorly-maintained generator sets often have poor quality logs and are more difficult to acquire data for than well-maintained generator sets. Even if data is harder to acquire for poorly-maintained generator sets, less data is needed due to the higher failure rates of those units as compared to well-maintained generator sets for which data is more readily available, but failure rates are lower.

ANALYSIS METHODS IN RELATED FIELDS

To determine the optimum maintenance cycle of diesel bus engines, Leung and Kai (2003) used a maximum-likelihood density estimation (MLDE) procedure, assuming a Weibull estimation. They used η as the scale parameter and m as the shape parameter, with a cumulative Weibull distribution equation of the form:

$$F\left(t\middle|\vec{\theta}\right) = F(t|\eta, m) = 1 - e^{-1\left(\frac{t}{\eta}\right)^{m}}, t \geq 0$$

Leung and Kai (2003) estimated the value of $\vec{\theta} = (\eta, m)$ using the nearest neighbor distribution estimation (NNDE).

Márquez and Herguedas (2004) also used a Weibull distribution to determine failure rates for earthmoving equipment but took a more straightforward approach to calculate Weibull scale and shape parameters by using spreadsheet software. They graphed distribution function data as an x/y plot of $x = \ln(t)$ and $y = \ln(\ln(1 \,/\, (1 - F(t)\,))\,)$, where $F(t)$ is the life distribution function based on manufacturer laboratory testing, and they used the spreadsheet software's built-in functions to calculate quadratic and linear trend lines. If the two trend lines were similar, they estimated the bi-parameter Weibull distribution using the parameters from the linear regression. Where the two were not similar, they modified the time origin to reach a better fit and chose a tri-parameter Weibull instead.

Hale and Arno (2009) used an exponential failure distribution to model emergency diesel-electric set reliability as $R(t)=e^{-\lambda t}$, with the failure rate per year (λ) calculated as the total number of failures divided by the calendar time the records were collected. They calculated availability as a function of mean time between failures (MTBF), mean down-time (MDT), and mean time to repair (MTTF), where operational availability (A_o) is A_o = MTBM/(MTBM+MDT) and inherent availability (A_i) is A_i = MTBF/(MTBF+MTTR). TM 5-698-5 (2006) and IEEE 493-2007 use the same methods and nomenclature. For time-truncated data sets where no failures were recorded, TM 5-698-5 (2006) utilized a χ^2 60% single side confidence interval to calculate λ and *MTBF*.

Zhou et al. (2014) attempted to develop a survival model for a highly censored sample of 800 utility transformers to gain knowledge regarding transformer lifecycles. Zhou et al. found 44 failures, only one of which was determined to be age related. With a failure rate of approximately 0.2% per year, transformer failures in this study were rare events, and Zhou et al. required special techniques to overcome censoring. Zhou et al. attempted to analyze this sample using a Weibull distribution, but met with considerable challenge due to the high level of censoring and dominance of random failures over age related failures. The failure distribution for this sample more closely resembles an exponential distribution than a Weibull distribution, as is common in the mid-life cycle of a Weibull distributed group, and Zhou et al. concluded the sample size was too heavily censored and did not include enough units reaching the end of the Weibull wear-out cycle.

Relevance vector machine (RVM) (Tipping, 2001) and support vector machine (SVM) machine learning techniques have been used in several studies related to engine maintenance (Jia & Zhao, 2006; Wang et al, 2013). RVM and SVM utilize a combination of mathematical regression and Bayesian statistics to map data points into two different categories. While these techniques have applications in condition-based maintenance and predictive maintenance, they are difficult to apply to survival data with high degrees of left truncation and right censorship and do not appear to have advantages over Cox (1972) regression for this type of survival analysis.

Many research teams including Amorim and Cai (2015); Hosmer, Lemeshow, and May (2008); Kalbfleisch and Prentice (2002); and Kelly and Lim (2000) have discussed methods developed for recurring events in the medical field. The Andersen and Gill (1982) method is a counting method formulated in terms of increments in the number of events along the timeline, but it is restrictive for application to emergency generator set reliability as it requires a right-continuous process without left truncation and is intolerant to treatment changes over the life of the process. The Anderson and Gill (1982) method is used as a basis for many related methods that share many of the same application restrictions (Amorim & Cai, 2015; Kelly & Lim, 2000). The only emergency generator sets that could be analyzed by these methods would need to have complete records available back to installation and no changes in maintenance or test periodicity over potentially decades of operation. However, placing such a restriction on data acquisition to only generator sets with such records and history could introduce bias. The most appropriate methods for analyzing generator set records must be insensitive to left truncation and right censorship. Other models discussed by research teams extend from Andersen and Gill (1982)and share the same limitations or require special cases not applicable to emergency generator set reliability. Hosmer, Lemeshow, and May (2008) criticize the Andersen and Gill (1982) method for assuming complete independence between events to be unrealistically simple for events such as cancer reoccurrence, but this is not necessarily the case for a repairable machine, and models dependent on stratification of recurrent failure data are difficult to apply to generator sets which fail in a multitude of ways.

Kelly and Lim (2002) discuss conditional approaches to analyzing recurrent event data and observed that while conditional approaches assume the current event is unaffected by earlier events, that assumption can in some cases be relaxed by means of additional covariates to represent prior events or other dependencies. This approach appears similar in concept to the introduction of generator set age and run hour covariates independently proposed by Fehr (2014) to analyze time dependent covariates for generator set age and cumulative run hours within a Cox linear regression of arbitrary start time. One conditional approach is the Prentice, Williams, and Peterson (1981) gap-time method, which is similar to Anderson and Gill (1982) but resets the clock after each event. Gap-time includes stratification in the model, with each failure representing a different stratum. Unfortunately, gap-time is still sensitive to left truncation and requires full knowledge of prior failures to assign strata to each data set, and is thus not suitable to apply to emergency generator set reliability modeling.

POST-HOC POWER ANALYSIS AND VALIDITY

Numerous methods have been developed for pre-hoc power analysis for statistical models, but little research has been accomplished regarding post-hoc power analysis of parametric survival models based upon unconstrained and heavily censored data sets. Existing literature was reviewed for methods of calculating post-hoc Type II error and realized power. Schoenfeld (1983), Hsieh and Lavori (2007), Cohen and Cohen (1983), Cohen (1987), Lachin (1984), and Liu (2014) provided relevant methods for pre-hoc regression power analysis, but not post-hoc, and not specific to survival analysis. Lachin (1984) spoke directly to survival analysis, but Lachin's methods are only applicable to groups with a single binary independent variable and an exponential failure distribution. Of these methods, Hsieh and Lavori (2007) were the most relevant to this research and were used herein for pre-hoc power analysis and the estimation of data required. Although none of the power analysis methods reviewed herein were discussed by their authors for post-hoc analysis, methods for potential applicability were reviewed.

Lachin (1984) discussed different methods for F-test based calculations of power for uncensored and censored survival data, but both methods require the calculation of exponential failure rates λ_c and λ_t for the control and treatment groups. While Lachin's discussed methods are applicable to designed experiments or data sets where subjects are randomly assigned to either subpopulation, the Fehr (2014) model includes 26 predictors which are unlikely to all be fully independent. Without independence of those predictors, single values of λ cannot be accurately calculated for subpopulations without a large risk of sampling bias in the results. For example, the subpopulation for a low-cost generator set model may include a disproportionate sample of low-cost maintenance, and attempts to calculate a subpopulation λ_{model} would be biased by the difference in maintenance. As Lachin's method is highly sensitive to such sampling bias but does not provide a means of accounting for it, it cannot be used to calculate power for emergency diesel-electric generator set research.

Cohen and Cohen (1983) provided a method to calculate n* for multiple regression correlation based upon the regression variance R^2 and tables of values for different degrees of freedom and several Type I and Type II error levels. Cohen and Cohen (1983) provided equations both for population R^2 and for sample R^2 and an equation that, given known n^* and R^2, can be solved for power. Cohens' method, in Cohens' Equation 3.7.2, was designed for pre-hoc determination of the quantity of data required and uses the population effects size f^2 (Cohens' Equation 3.7.1), although Cohen and Cohen cautioned against using this for a sample and stated the F equation (Cohen's Equation 3.6.1) should be used for sample values.

$$f^2 = \frac{R^2}{1 - R^2}$$ Cohens' (1983) Equation 3.7.1

$$n^* = \frac{L}{f^2} + k + 1$$ Cohens' (1983) Equation 3.7.2

$$F = \left(\frac{R^2}{1 - R^2}\right)\left(\frac{n - k - 1}{k}\right)$$ Cohens' (1983) Equation 3.6.1

Post-hoc, $n^* = n$, and are R^2 and k are known. Substituting these known values into Cohen's Equation 3.6.1 and 3.7.2 and solving for L instead of n^* gives Equation 2. As noted by

Cohen (1987), this permits use of Cohen's L tables in reverse to look up the observed power β associated with the experimental results. A direct calculation of this is also supported by the *f2.test* function in the *R* library *pwr* (Champeley et al., 2016).

$$L = \left(\frac{R^2}{1 - R^2}\right)\frac{(n - k - 1)^2}{k} \tag{2}$$

Cohen and Cohen (1983) provided a similar method to calculate the n* required for each independent regression coefficient for desired power prior to the experiment, but not a means to calculate the observed power after data is acquired. Hoenig and Heisey (2001) provided a method to estimate observed power for regression coefficients directly from the *p*-values automatically calculated and provided in many software packages. This method is only applicable to significant predictors and cannot be used to estimate the observed power for predictors found not significant in the regression results; attempts to use the *p*-value to calculate β from post-hoc predictors excluded from the final regression model do not give valid results. None of the literature reviewed provided a method for determining the post-hoc observed power for these excluded predictors and the failure to reject the null hypothesis, only pre-hoc estimations of power as a function of anticipated R^2; Steidl, Hayes, and Schauber (1987) contends such a measurement is meaningless. Murphy, Myors, and Wolach (2014) discussed use of post-hoc findings to create an estimate of the potential power of prior research, but did not present any methods for calculating realized power. Thus, these power estimates can be refined using experimental data to increase the accuracy of pre-hoc estimates, but offer little benefit for model validation.

Hoenig and Heisey (2001) argue that observed power is not useful in post-hoc analysis due to what they call the power approach paradox, that higher observed power does not imply stronger evidence for a null hypothesis that is not rejected. In support of this assertion, they provide several examples where post-hoc observed power analysis presents logical flaws and often nonsensical results. They further contend that once the confidence interval is calculated, power analysis provides no further useful information, and recommend post-hoc power analysis should not be done. Steidl, Hayes, and Schauber (1987) states this similarly, contending that retrospective power analysis provides results with no relation to true power. These arguments are themselves not without criticism, but there remains little published literature on the subject. This is unfortunate given the importance of quantifying observed power and Type II error probabilities for excluded predictors, but is consistent with the lack of post-hoc power analysis in prior multiple regression/correlation analysis research. Liu (2014) discusses the relationship between confidence intervals, power, and precision, but does not provide a post-hoc method to determine observed β. Steidl, Hayes, and Schauber (1987) and Hoenig and Heisey (2001) recommend the use of confidence intervals as useful for determining what range the effects size may be and potentially the low probability than a specified effects size exists. Therefore, while it may not be possible to directly calculate the probability of falsely failing to reject the null hypothesis, by calculation of the confidence interval it is possible to make a reasonable calculation of the probability that the effects size is so small as to not be of practical importance (Steidl, Hayes, & Schauber, 1987; Lenth, 2007).

Another question raised by this research relates to validation of the model for applicability to various subpopulations. Specifically, make, model, and size were included in the model as predictors, but were not anticipated to have significant results. Cox (1972) non-parametric regression methods can calculate α for these predictors but cannot return any direct information about β or the risk of falsely failing to reject the null hypothesis that there is no relationship between generator set make, model, and size and reliability. Steidl, Hayes, and Schauber (1997), Hoenig and Heisey (2001), and Murphy, Myors, and Wolach (2016) advocated utilizing the model confidence intervals to qualitatively judge model adequacy, but this does not provide much useful information on the validity of smaller subgroups. Cohen (1987) provided a method for analyzing data sets that is applicable to Cox regressions; although this method is only discussed for pre-hoc power analysis, solving with known values from experimental data can be used to estimate experimental power and permits to assessment of model validity with more confidence.

Fehr (2014) made a general *a priori* assumption that generator set make, model, and size within the delimitations of this research represent a common population with a common response to maintenance. Present maintenance guidance (NFPA 70B, 2016) does not make distinctions of differing maintenance on make, model, or size and there are reasonable arguments for this assumption supported by many technicians (Walbolt, 2010). Despite such industry confidence, there is no solid research to support the assertion of common response. Fortunately, this assumption can be easily tested within the framework of this research by inclusion of make and model factors and a size covariate in the model. If the *p*-values for these predictors are found to be non-significant, then the *a priori* assumption of a common population with common response to maintenance holds. However, if any of these subpopulation predictors are found to be significant, model validity for these subpopulations must be determined.

If the common population assumption fails, there is little research into subpopulation validity in survival distributions with high levels of left truncation and right censorship that can be directly applied to address it. Cohen (1983) presents a related method intended to estimate the proportion of the variance contribution of the subpopulation *Y·B* in equation 4.5.2. *B* represents the subpopulation being analyzed and *A* represents the population containing all others. *Y·B* is the experimental sample data set for subpopulation *B*, *Y·A* is the experimental sample set for the remaining population, and *Y·AB* represents the entirety of the experimental data. Utilizing the *F* equation for samples instead of the population effects size f^2, and solving for L results in Equation 3 and permits use of Cohen's *L* tables in reverse to look up the observed power *β* associated with this data set.

$$F = \frac{R^2_{Y\cdot AB} - R^2_{Y\cdot A}}{1 - R^2_{Y\cdot AB}} \frac{n - k_A - k_B + 1}{k_B} \qquad \text{Cohen (1983) Equation 4.4.2}$$

$$n^* = \frac{L}{f^2} + k_A + k_B + 1 \qquad \text{Cohen (1983) Equation 4.5.2}$$

$$L = \left(\frac{R^2_{Y\cdot AB} - R^2_{Y\cdot A}}{1 - R^2_{Y\cdot AB}}\right) \frac{(n - k_A - k_B - 1)^2}{k_B} \qquad (3)$$

However, Cohen's (1983) *F*-test statistic for data sets has a weakness when data sets are not similar size. For example, if *Y·AB* represented student GRE scores and *Y·A* consisted of 20,000 graduate students while *Y·B* consisted of 20 preschoolers, Cohen (1983) *F*-test Equation

4.4.2 would inaccurately indicate high power that the two subpopulations were the same. As the data sets anticipated from this study are likely to include some makes, models, or sizes in much larger or smaller numbers than others, use of this equation could be problematic.

Cohen's (1983) more general Equation 3.6.1 does not share the same weakness for proportionately small sample subpopulations, and Equation 2 derived from this can be directly applied to each sample subpopulation. However, as the sample subpopulations may be too small or too homogenous in test and maintenance periodicity to independently derive significant models, methods comparing Cox (1972) regression models cannot be used with any confidence of high quality results. As a model already exists from the general population regression, the regression sum of squares for the final model can be calculated for each sample subpopulation, and an *F* test performed to compare each sample subpopulation to the general population. If the model is valid for the sample subpopulation and the relationship with the subpopulation can be represented as a treatment in the model, the weighted difference between the sum of squares response for the subpopulation and general population should be statistically insignificant. If this test fails, additional interaction terms can be explored that may better represent the subpopulation, or the subpopulation removed from the study if appropriate.

SECONDARY DATA

Enormous quantities of data exist in the form of historical logs and records from operations, maintenance, and repair of emergency diesel-electric generator sets, similar to other existing data across a myriad of applications and industries. The primary advantage of such data is that it already exists and is therefore often easier to collect than devising and conducting new experiments. This data also includes real-world performance that's difficult to replicate in a lab.

However, those advantages are also disadvantages, as historic logs and records were rarely recorded under controlled conditions for the sake of pure research, and this secondary data is often of questionable scientific value for positivistic research due to questionable internal validity and reliability (Souza-Posa, 2015). Crowder and Lancaster (2008) suggest that such secondary data is only useful for exploratory research for collection of primary data, but Crowder and Lancaster were focused more on business data and reports and did not consider the specific application of secondary data in the form of logs and records.

Historic logs and records for operations, maintenance, and repair represent a special case subset of secondary data and can be a powerful and economical source of data for analysis if properly used. Even if they were not originally created for pure scientific research, such logs are generally intended as an objective series of measurements for engineering studies, and often have significant practical value to aid managers and maintenance personnel in decision-making (Márquez & Herguedas, 2004). Equipment logs and records have been used successfully for research by several teams (Gat & Eisenbeis, 2000; Mathur, 2002; Márquez & Herguedas, 2004; Devaney et al, 2005; Fehr, 2014). Devaney et al. (2005) conclude that “maintenance logs contain a potential wealth of information that can be used to improve the maintenance of complicated machinery, reduce downtimes, and prevent failures.” Herein, the use of historic logs and records as research data is discussed as well as a review of issues found by researchers and a review of techniques used to mitigate those issues. A consolidated methodology is presented to minimize the risk inherent in using handling historic logs and records for scientific research, and to maximize the research utility from these potentially rich sources of data.

Márquez and Herguedas (2004) examined maintenance records as a tool for root cause analysis and found unqualified record-keepers, incompetent handling of maintenance data, and lack of computerized records represented common problems. They found that data from equipment had to be screened properly to ensure the data represented the same failure modes in homogenous equipment under similar conditions. They stressed the importance of computerized maintenance management and standardized record keeping as very important to structuring records into a manner to aid decision-making for the maintenance personnel involved, and for successful exploitation of such data for broader studies. They found the expert judgment of maintenance engineers to be a significant benefit.

Crowder and Lancaster (2008) asked "What do I need to know?" and "How accurate and detailed does the data need to be?" and discussed the general use of secondary data in detail. The techniques Crowder and Lancaster presented were not specific to historic logs and records, but suggested internal secondary data was more reliable than external secondary data due to the greater level of access researchers are typically granted for internal data. Crowder and Lancaster (2008) further stated that secondary data was rarely useful in answering research questions, but this may be because they were primarily considering financial, sales, marketing and personnel data and did not list maintenance and operational logs and records or other engineering data among examples of secondary data considered in their discussion. Some of the techniques presented are nevertheless applicable to engineering research and describe creation of a methodology able to take advantage of this secondary data including identifying the problem, developing an approach, and formulating a research design.

Gat and Eisenbeis (2000) found the service life and low failure rate of some samples to be a challenge for analysis, and excluded several samples from the data set due to short service length or the small number of logged failures.

Fehr (2014) found that even in a homogenous group with similar equipment and common guidance and policy mandating specific logs be kept, such records were rarely complete and accessible in practice and mostly existed in handwritten form. Fehr (2014) further found that that poorly-maintained equipment often also had poor record keeping, potentially introducing selection bias and skewing studies towards those of well-maintained equipment by the simple fact that those well-maintained units had the most complete records and were more easily accessible. Maintenance performed by third parties introduced additional challenges, as each organization kept records in a different way, and records on some units were split between the operational organization and the maintenance organization. A lot of manual effort was required by the researchers to overcome the challenges presented by the inconsistent record keeping, and the analysis had to be explicitly designed to minimize bias from the data.

Devaney et al. (2005) data-mined equipment maintenance logs of complex machinery, which largely consisted of terse free-text input. They found that even when such logs are recorded for the explicit use of long-term tracking of equipment condition, performance, and reliability, such records included large variation in the input and were often difficult to consistently interpret via automation. They found that the vocabulary was inconsistent and often included jargon, and the terminology used didn’t always correspond to systems of interest. They also found logs to be full of spelling errors, grammar errors, typographical errors, and extremely terse non-standard abbreviations. They also found that most machines have statistically few actions logged per year, making such sets too sparse for data-intensive approaches. Despite the challenges, Devaney et al. (2005) developed a software algorithm using text analytics, clustering techniques and a case-based reasoning to analyze the maintenance data, demonstrating that automated methods are possible to apply to historic logs and records.

Souza-Poza (2015) asserts that the reliability and validity of secondary data must be verified, and the secondary data must be demonstrable as mind-independent and sufficiently positivistic to accomplish the research goals.

Márquez and Herguedas (2004), Fehr (2014), and Gat and Eisenbeis (2000) found data censoring by preventative maintenance to be a significant issue for failure modes and effects analysis (Moubray, 1997); few managers permit equipment to proceed to the point of failure, and maintenance actions modified the failure rate distribution. All three research teams used mathematical techniques tolerant of censoring to analyze the censored data, but such techniques often require assumptions be made and may not be applicable in all cases or for all failure distributions.

Márquez and Herguedas (2004), Fehr (2014), and Gat and Eisenbeis (2000) all also experienced missing data, and small sample sizes meant that only a portion of possible failure combinations were represented in their data sets. Márquez and Herguedas (2004) and Gat and Eisenbeis (2000) chose to address this deficiency by only analyzing data for which sufficient data was available; Márquez and Herguedas (2004) only analyzed cylinders which exhibited at least three failures within the fleet, and Gat and Eisenbeis (2000) excluded data sets for which limited data was available. Márquez and Herguedas (2004) developed a data map, but neither Márquez and Herguedas nor Gat and Eisenbeis attempted any techniques to reconstruct missing data. Fehr (2014) utilized inductive reasoning to reconstruct portions of incomplete data records from partial generator data records, such as using knowledge of generator test frequency and duration to estimate the cumulative generator run hours at different points in time; such techniques may increase the quantity of analyzable data, but introduce error and risk of corruption.

III. METHODOLOGY

Failure modes and effects analysis is a powerful tool for analyzing equipment failure, but this analysis relies upon observing and measuring failure rates of each system component to determine failure patterns (Moubray, 1997). This cannot be accurately performed on operational equipment receiving high levels of preventative and predictive maintenance, as this maintenance introduces bias and censorship into the data. Known high failure rate items such as batteries, filters, and lubricating oil are rarely allowed to degrade to the point of system failure in well-maintained units. Research by Alion Science and Technology (2006) supports this and indicates other failure modes may dominate. This right censorship of failure data by preventative maintenance limits the effectiveness of reliability-centered maintenance studies proposed by Moubray (1997) for this type of application. The effectiveness of the implemented maintenance and test plans at each facility can instead be measured by survival analysis of the overall emergency diesel-electric generator system.

Where failure rates exhibit a traditional Weibull wear-out pattern, there is an expected optimal maintenance periodicity. For such components, there is a point where longer intervals between maintenance would result in a significant increase in failures but where shorter intervals show little or no reduction in failures. The point at which longer intervals represent an increase in failure rate but shorter intervals do not is the point of optimum maintenance periodicity. For failure modes that result in detectable degradation just prior to failure, this optimal predictive maintenance periodicity is known as the Nett P-F interval (Moubray, 1997).

Many failure-finding tasks can reduce the operational failure rate at ever decreasing periodicity intervals with continuous monitoring providing the shortest possible interval and near-instant detection of hidden failures. But for intrusive actions or actions that must remove the generator set from operational service to perform, excessive maintenance results in an increase in maintenance down time and a decrease in inherent availability. Excessive maintenance can also increase the risk from human error, increase component wear, or increase the risk of installation of a defective component causing a failure that otherwise would not have occurred. Thus, excessive maintenance is not necessarily erring on the side of caution and may result in an increase of operational risk. For failure-finding tasks where this is the case, the minima of the failure rate is the optimal point.

The research herein utilizes and refines the methodology developed and tested by Fehr (2014) to model emergency diesel-electric generator set reliability as a function of maintenance and test periodicity with the intention of determining optimal maintenance and test periodicities for maximum operational diesel-electric generator set reliability. This methodology was based upon examining existing emergency-diesel electric generator set logs for generator sets supporting critical operations power systems, high-reliability applications, and life safety applications. These logs were required to be maintained per government regulations (TM 5-683, 1995; NFPA 110, 2016; EPA, 2016) and include records of generator set maintenance, tests, repairs, run-hours, starts and failures. However, the means of maintaining these records varied, as no specific formats were mandated. Data gathered from these logs was compiled into the standardized form in Appendix A (Fehr, 2014), which allowed this data to be combined into a single database for analysis.

A generalized survival function was created from the generator set history data with focus on operational failures as the key survivability event. A distinction was made between failures to start and failures while running. A distinction was also made between failures during tests and failures during actual emergency operation. To avoid biasing against sites with high levels of testing, corrective maintenance and failures during scheduled routine tests were not included as operational failures in the survival analysis.

Reliability-centered maintenance practices would consider all critical component failures to be system failures regardless of whether the failures occurred during tests or operation, but such a philosophy is not appropriate for determination of optimal maintenance levels as it can introduce bias against sites with more frequent maintenance. In an extreme example, if a component in an emergency generator set exhibiting hidden failure characteristics is found to have a mean time to failure of 30 days, sites that perform weekly tests are likely to log a higher total number of annual failures and annual maintenance downtime than sites performing no testing. However, the sites performing weekly testing are much more likely to have units that function during actual emergency operation. From a holistic facility perspective, failures of emergency diesel-electric generator sets during emergency operation result in facility operational downtime, while failures during testing do not. As this study was interested in determining optimal practice for operational reliability and availability, it considered failures discovered and corrected during testing to be maintenance actions and not failure events. Such a test failure was only considered an operational failure if the resultant deficiency could not be corrected in time to avert failure during subsequent emergency operation. Occurrences of operational failures stemming from test failures were treated as operational failures within this study.

The ability of existing maintenance practices to address failures of individual subsystem components was implicit within the observed survival function of the system. This approach was a conservative approach with limited ability to determine the true optimal periodicity of some subsystem components, especially where conflation of maintenance or test actions is occurring. This approach was designed to err on the side of excessive rather than insufficient maintenance.

RESEARCH SCOPE

This research was focused on answering the research question and subquestions for the subset of emergency generator sets most common at the government and commercial facilities participating in this study. These facilities predominantly utilize emergency backup electrical power from water-cooled diesel fuel-injected turbocharged piston engines driving permanent-magnet excited electrical alternators. While such diesel-electric generator sets can be small enough to roll around a jobsite or large enough to support entire industrial complexes, this research was focused on fixed generator sets of sizes and duty cycles commonly found in small and medium data centers, telecommunications sites, hospitals, commercial facilities, and critical operations facilities. To avoid introducing variances from unique issues that may occur in very large, very small, or uncommon systems, data acquisition for this study was limited to the subset of units described in Chapter I, SYSTEM COMPONENTS. Generator sets with similar characteristics were also included when data was readily available but units such as diesel-turbine generator sets and portable generator sets used in prime duty for temporary power were excluded even when data was available.

Emissions requirements for non-road diesel engines went into effect in 1996 in the United States and in 1999 in Europe, with phased implementations of increasingly stringent regulations in subsequent years. Compliance prior to these dates was not required by the United States' Clean Air Act of 1990 (Environmental Protection Agency, 2013) or similar European regulations, but many manufacturers began fielding units well in advance of the required dates, often due to commonalities in product lines with road-going diesel engines and other markets which faced earlier compliance requirements. While diesel-electric generator sets are complex systems-of-systems, the basic diesel engine technology used is similar across all major manufacturers and models, as are reliability centered maintenance philosophies. This results in many common routine maintenance actions between disparate makes and models, and often similar recommendations for periodicity of those maintenance actions.

The scope of this study was primarily limited to modern, high-efficiency, low emission generator sets. Generator sets manufactured since the passage of the 1990 Clean Air Act Amendments were the most common subset in use at the facilities participating in this study, but little research has been published on the reliability or maintenance of this generation of generator sets. These generator sets may respond to maintenance and tests differently than older generator sets from which industry standard maintenance and test policies were developed (Margaroni, 1999). While there is benefit of extending this research to older generator sets that remain in service, such generator sets are rapidly approaching the end of their design lives, and the return on investment for data acquisition is greater for newer sets which were anticipated to remain in service well into future decades. Answering the research question regarding impact of the age of generator sets on reliability required some older units to be included in the study, but the emphasis of data acquisition was on units manufactured since 1990.

The scope of this research also extended to determining the impact of human performance on generator set reliability as related to the training level of servicing personnel. Specific training of servicing personnel is not typically recorded in maintenance logs, so training was categorized as factors of one of three general types: staff collateral duty, staff subject-matter expert (SME) and service visit SME. Collateral duty personnel are comprised of site personnel with a minimal level of training who are often expected to perform routine low-difficulty maintenance and test actions. Both staff and service visit SMEs were assumed to have deep understandings of the emergency diesel-electric generator sets at the facility. This study examined the interaction between personnel training and the maintenance and test actions performed. Knowledge gained from this interaction was anticipated to indicate if total productive maintenance (TPM) utilizing operators or facilities staff for simple high-frequency maintenance tasks could be as effective for these generator sets as dedicated maintenance personnel and to allow managers to quantitatively determine optimal training and manning levels for each facility.

One aspect that was not within the scope of this study was the cost of the maintenance performed. While knowledge of generator set maintenance and test costs would increase the usefulness of the findings of this research to cost-sensitive commercial applications, generator records maintained by sites do not typically include the associated cost of maintenance or repair actions and would pose significantly increased challenges to data acquisition. This research was designed to exclusively investigate reliability. This is appropriate for critical operations power systems where cost is a secondary concern and for facilities where the cost of maintenance is small compared to the financial impact of a loss of power. Economic calculations beyond the scope of this research would be required to determine optimal maintenance and test procedures for non-critical applications where cost is a primary consideration. The reliability model developed by this research will assist managers of emergency diesel-electric generator systems to calculate the risk inherent in differing levels of maintenance, tests, and training for their facility, and to make risk decisions based upon their own estimated costs.

A secondary objective of this research was to estimate the operational availability (A_o) and inherent availability (A_i) of optimally maintained and tested emergency diesel-electric generator sets. While a very important characteristic, determination of the complete emergency power system A_o and A_i includes many variables that fall outside the scope of this research. While reliability may be a generalizable property, emergency power system availability is not, and there are many challenges to be overcome to accurately determine time to repair. For instance, if an emergency generator set receives no maintenance and then suffers an operational failure due to a dead battery, is time to repair measured by the number of hours it takes to replace the battery after the failure was discovery, or should time to repair include the unknown number of months the failure sat hidden? Response time for staff and availability of spare parts varies significantly, further complicating generalization.

Maintenance down-time was not included in the survey form, but, except for fluid changes, valve lashing, and other intensive maintenance, none of the maintenance or test actions included in the survey require taking the generator out of service or result in operational maintenance down-time and have negligible impact on A_o. Inspections and fluid level checks do not impede the ability of generator sets to automatically start, and, even if units are switched off for worker safety while examining belts and filters, the servicing technicians are available to immediately restore the units to operational condition. Only sites with the most comprehensive logs contain sufficient information to accurately calculate repair and maintenance downtimes, but using these numbers exclusively may have introduced bias.

Response and repair time are anticipated to be shorter for failures discovered during testing due to technicians being on-site and able to immediately initiate repair actions. The actual point a failure occurred was rarely recorded, only when the failure was discovered, and the precise time of repair was not always recorded, so any calculations of availability developed from the data in this study would include an amount of uncertainty.

Availability calculations based exclusively on operational reliability would be inaccurate, as they would not consider unavailability during corrective maintenance. The question of availability calculation becomes more complicated when considering hidden failures, as it is impossible in many cases to know when a failure occurred, only when that failure was discovered. Utilization of Fehr (2014) methods to investigate test failures for the purposes of availability calculation run the risk of bias against sites with frequent testing. Calculating availability is outside the scope of this research.

RELIABILITY PREDICTORS

Twenty-one maintenance and test predictors proposed by Fehr (2014) were investigated in this study, each representing a maintenance or test action with the potential to contribute to overall system output. These actions, shown in Table 4, were chosen based on periodicity of one year or less per industry recommendations in NFPA 110 (2016) and are of the form:

$x_i = j$, periodicity of each action, where:

i = action, per Table 4

j = action periodicity, per Table 5

The periodicity of actions x_i were measured in years and modeled as covariates. Modeling these actions as factors would have the advantage of independence from any required knowledge of the relationship between periodicity and reliability, and would allow for complex relationships such as bathtub-shaped hazard functions which prior research indicates may be appropriate for some actions. However, modeling these actions as factors with limited data exceeded the model capacity for significant results and required the actions instead be treated as covariates to reduce the model to a manageable number of degrees of freedom. Table 5 shows each action periodicity j measured in years between maintenance periods. Continuous monitoring by watch personnel twenty-four hours a day is approximated as 0.25 hours to reflect response time. Continuous monitoring only during the normal forty-hour work week is approximated as a mean response time of 15 hours. This mean assumes that while response time may be within 0.25 hours for forty hours per week, average response time will be eight hours during sixty-four hours of weeknights per week, and an average of thirty-two hours response time during the sixty-four hours of weekends. The covariate value for actions performed at “More than three years” is approximated as five years, and the covariate value for actions marked “Not Routine” is approximated as seven years. Hours were converted to days for modeling by dividing by twenty-four hours per day, and days to years by dividing by three hundred and sixty-five days per year.

Table 4. Table of Test and Maintenance Actions

i=	**Test and Maintenance Actions**
1	Maintenance performed, details not known
2	Check alarms
3	Check switch & breaker positions
4	Visual inspection for leaking fluids
5	Visual inspection of hoses, cables, etc.
6	Check fuel level
7	Check oil level
8	Check coolant level
9	Check air filter
10	Battery voltage & physical condition
11	Check fan belt(s)
12	Battery resistance or impedance test
13	Clean unit exterior (including radiator & louvers)
14	Fuel cleaning (or fluid analysis)
15	Oil change (or fluid analysis)
16	Check electrical tightness
17	Engine intensive maintenance
18	Generator (electrical) intensive maintenance
19	Generator set no-load test
20	Generator set load test on load bank
21	Generator set load test on operational load
22	Generator set dead-bus test on operational load

Table 5. Table of Test and Maintenance Action Periodicity

j=	Action Periodicity
0.25 hr	continuous watch
15 hr	40hr/week watch
1 day	Daily
7 days	Weekly
14 days	Biweekly
1/12 yr	Monthly
¼ yr	Quarterly
½ yr	6-month
1 yr	Annual
2 yr	2 years
3 yr	3 years
5 yr	More than 3 years
7 yr	Not Routine

DATA ACQUISITION

The data acquisition framework from Fehr (2014) was used to acquire data from operational records of emergency generator systems with diverse maintenance policies. Managers and maintainers are asked to provide data on the Fehr (2014) *Microsoft Excel* form included in Appendix A. As maintenance policies and periodicity for a given unit can change over time, managers and maintainers were asked to fill out a new form every time maintenance periodicity for a generator set changed so that each form contained a period of consistent maintenance. Thus, each form represented a rational subgroup of a single specific generator set with consistent maintenance and test practices. As an increase in maintenance may be a managerial response to failure, regardless of whether maintenance practices played a role in the failure, each instance of change in maintenance policy was individually investigated to prevent the inadvertent inclusion of bias in the study. The completed and validated forms were compiled into a single database for analysis.

The United States Department of Defense and other authorities require facilities to keep accurate and complete generator set maintenance and operation records from commissioning to decommissioning, and the Environmental Protection Agency has the authority to levy fines against organizations that do not keep accurate run logs, but logs were often found to be incomplete, damaged, or lost. Incomplete data could in some cases be accurately estimated from interviews with maintenance personnel and other data. Where data such as this was estimated, it was annotated in the comment field. For example, in a few cases it was necessary to estimate unknown chronometer run hours at the start of observation period by combining knowledge of test periodicity and duration with knowledge about historic power failures. Where only the month or year of the generator set installation date was available, it was estimated for regression purposes as occurring in the middle of the year or month, and this estimation was annotated in the comments field. Where no installation date information was known, the installation date was often able to be accurately estimated using information about the building's construction date or extrapolated from chronometer run hours through the known event period. Where a survey form was incomplete, and data required for regression analysis could not be accurately estimated, the observation was excluded from the database.

The initiation date of logbooks was often arbitrary, such as the first of the year, or whenever the previous logbook was filled up. Earlier logbooks were not always available, so logs for many observations were left truncated. While failure data is important, reporting periods that do not end in failure are equally as important and were also recorded to avoid introduction of bias. The resultant data sets from this data acquisition contained significant amounts of left truncation and right censoring.

Many critical operations power systems support sensitive mission functions, and specific information about facility emergency power systems is often sensitive as well. Care was taken to sanitize the data of all identifying information such as site name, location, or function. Where site managers provided identifiable descriptive names for generator sets on the survey forms, the identifying information was replaced with alternate generic generator name. Likewise, descriptive information was removed from the comments field when unit data was imported into the database. The database did not include information about any site's system configuration or redundancy; such information is not relevant to the intent of this of research. Instead, a generic identification number was assigned to each observation, with the identification key and original survey form maintained in a separate database on a secure government server.

A spreadsheet was developed in *Microsoft Excel 2016* to organize the data in a format that could be easily exported into *R* for analysis. To eliminate transcription errors, a *Visual Basic for Applications* script was written to import data from the *Excel* forms into the spreadsheet. This script automatically converted check-box matrices of descriptive predictor periodicities into numerical periodicities and generated individual observation entries for each logged failure. The data was thoroughly reviewed for completeness and errors, exported in comma-delimited *.csv* format, and imported into *R* for analysis.

DATA REQUIRED

This study sought to determine the relationship of many predictors, and two important questions that were asked prior to conducting research was how much data was required to achieve statistically significant results, and if that data was reasonable to acquire.

Some knowledge about the survival distributions of the generator sets was required to estimate the amount of data required. Fehr (2014) applied the PREP (TM 5-695-5, 2006) formula of λ = [Failures]/[Time] and Reliability $R(t) = e^{-\lambda t}$, to a sample set of well-maintained generator sets, and determined the annual reliability of the sample set of well-maintained generator sets was $R(1) = 0.931$ and $\lambda = 0.074$ failures per year. This was a higher reliability than the published annual reliability of $R(1) = 0.8838$ and $\lambda = .1235$ failure per year as measured from a diverse data set of well, average and poorly-maintained units during the PREP study. The mean and standard deviation of an exponentially distributed function can be calculated from this data as $\mu = \sigma = 1 / \lambda$, with the results $\mu_{FEHR} = 13.514$ and $\mu_{PREP} = 8.097$. The hazard rate ratio of the well-maintained and average-maintained generator sets Δ can be calculated as

$\Delta = (1 / \mu_{FEHR}) / (1 / \mu_{PREP}) = 0.599$. Data on poorly-maintained units was not available and was estimated by extrapolation from the PREP and Fehr data. An extrapolated value of $\mu_{POOR} = 2.684$ in combination with $\mu_{FEHR} = 13.514$ resulted in an estimated average $\mu_{PREP} = 8.097$. From this extrapolation, the hazard ratio Δ was estimated as $\Delta = (1 / \mu_{FEHR}) / (1 / \mu_{POOR}) = 0.199$.

Schoenfeld (1983) derived the equation $(Z_\beta + Z_{1-\alpha})^2 / (P_A P_B \log^2\Delta)$ where $Z_\beta + Z_{1-\alpha}$ are the desired $1 - \alpha$ and β percentiles of the hazard distribution, P_A and P_B are the proportion of patients randomized to treatments A and B, and Δ is the hazard rate ratio. This method used the log hazard ratio to determine the number of deaths (failures) required for statistically significant results and did not require knowledge about the distribution beyond the proportionality assumption. Schoenfeld (1983) was silent on the impact of censoring and truncation on this calculation, but the generator data in this study included a great deal of truncation and censoring. The only hazard rate data available pre hoc assumed an exponential distribution and was used for pre-hoc data estimation. With $\alpha=0.1$, $\beta=0.2$, $P_A = P_B = 0.5$, and $\Delta = 0.599$, Schoenfeld's equation indicated 94 failures were required to achieve statistically significant results. Assuming the PREP hazard rate of λ = .1235 failure per year was representative of the average generator set failure rate, documenting 94 operational failures would require a sample size of 761 standby-years. With an extrapolated Δ = 0.199 based off the extrapolated μ_{POOR}, however, only 9 failure events would be required from a sample size of 72 standby-years. Schoenfeld's method requires the assumption of binary covariates though, which is not entirely applicable to generator set data with a spectrum of covariate values.

Hsieh and Lavori (2007) proposed a variation on the Schoenfeld (1983) equation utilizing the variance instead of the sample proportions, $(Z_\beta + Z_{1-\alpha})^2 / (\sigma^2 \log^2\Delta)$. Hsieh and Lavori's (2007) model supports non-binary covariates and was found by Hsieh and Lavori to be insensitive to data censoring, although Hsieh and Lavori were silent on the impact of truncation. Like Schoenfeld's method, this method does not require assumptions of the distribution beyond the proportional hazard, but this method still requires knowledge of the hazard rates. If the acquired data was entirely binary with equal parts 0 and 1, the normalized covariate variance is σ^2 = 0.25, the same as the value of $P_A P_B$ used in Schoenfeld's equation where $P_A = P_B$ = 0.50. If, however, it is assumed that the distribution of normalized covariates is equally split between high-quality maintenance (0), average maintenance (0.5), and poor quality maintenance (1), then σ^2 = 0.1667. Including this variance into Hseih and Lavori's equation indicates 141 failures are required for Δ = 0.599 and 18 for Δ = 0.199. Assuming that the PREP hazard rate of λ = 0.1235 failure per year is representative of the average of this study, documenting 141 operational failures would require a sample size of 1142 standby-years. Documenting 18 operational failures would require a sample size of 146 standby-years.

Hseih and Lavori's method is more appropriate than Schoenfeld's method for estimating the data required for this study. The exact amount of data required depends on the covariates of the acquired data, but the amount of data required to achieve statistically significant results was estimated pre hoc to be achievable with as little as 72 standby years of data if the right samples were found. Acquisition of 1000-2000 standby years of data was anticipated for this study, which was anticipated to achieve statistically significant results for at least one research question.

COX/WEIBULL REGRESSION TECHNIQUE

The Cox (1972) proportional hazards regression model was developed by Cox to represent survival as a function of time. One complication for the application to emergency diesel-electric generator sets is that there are multiple ways that time can be measured, including the calendar time in service and the run hours of the unit. While Hale and Arno (2009) assumed an exponential survival distribution based on calendar time in service when creating their model of generator set reliability, there is no existing research to show which way to measure time is the most appropriate. Elapsed calendar time since manufacture T_{age} is one of the ways in which time can be measured. However, this presents difficulties within the methods available for this analysis, and appropriately addressing T_{age} within the model was one of the earliest and most difficult problems dealt with by Fehr (2014) in developing the methods used for this research.

One of challenges in utilizing T_{age} as the Cox survival function time variable is the repairable nature of the system and potential for each unit to suffer failure multiple times over its operational life and the restoration of condition by repair. The original Cox function (Cox, 1972) does not have a mechanism for multiple deaths (failures); it treats mortality as an event that happens only once in the life of a subject.

Successive failures beyond the first cannot be entered into a Cox regression utilizing T_{age} as the time variable without erroneous results. Assuming an average failure rate of $\lambda = .1235$ failures per year as found by PREP (TM 5-695-5, 2006), the average generator set from that study may have accumulated three failures over a twenty-five-year period. Every operational failure has operational impact regardless of the age of the generator set that failed, and restricting this study to only the first operational failure could bias results, so it is appropriate that this methodology can analyze multiple failures of each unit throughout its operational life. The methodology must be capable of handling significant amounts of left truncation and right censorship to analyze data which often has arbitrary start and end dates.

Multiple methods have been developed to adapt the Cox function for recurring or multiple events (Amorim & Cai, 2015; Hosmer, Lemeshow, & May, 2008; Kalbfleisch & Prentice, 2002; Kelly & Lim, 2000; Andersen & Gill, 1982), but these methods can only be applied to a right-continuous process that represents everything that happens up to time *t* (Andersen & Gill, 1982) and cannot be applied to left truncated data such as required for this research.

Fehr (2014) proposed a different method using Cox's non-parametric distribution. Fehr (2014) proposed this method because the data was anticipated to have a high degree of left truncation and a high degree of right censorship and because the baseline hazard model for generator set reliability was not known and Cox's non-parametric model does not require any assumption of baseline hazard model be made (Cox, 1972). However, some assumptions are required to use Cox's model to analyze generator set reliability. One assumption required to apply Cox's model to a multiple mortality application is the assumption of an exponential distribution with a base hazard rate that is constant and the same before and after the failure.

Under this assumption, the time variable may be reset following each failure or change in maintenance periodicity; the number of starts in the reporting period n_s is reset to zero, the time in service in the reporting period t is reset to zero, and the run hours in the reporting period T_{rt} is reset to zero. For the first failure of each unit, $t = T_{age}$, but for subsequent failures, attempting to use T_{age} would result in invalid regression results as Cox's model would interpret it not as a single individual suffering multiple failures, but as multiple individuals with increasingly long survival life. Therefore, T_{age} cannot be directly used as a time variable in this method, but T_{age} can be included as a covariate predictor to retain incorporation of generator set age in the model. Similarly, the total cumulative run hours T_{rtfv} can also be included as a covariate predictor. Both generator set age T_{age} and total cumulative run hours T_{rtfv} are time dependent properties, but the end of the observation period is a fixed point in the record, so age and run hours at the end of the observation period can be treated as static time independent covariates in the Cox (1972) regression. These properties change over the operational life of the generator set, but selecting record entry and exit points for the regression permits inclusion of observations taken at any point in time.

Due to the impact of corrective repairs and data truncation in real world data sets, the structure of the analysis requires the assumption that emergency diesel-electric generator systems exhibit exponential base hazard rate characteristics so that random or arbitrary starting points for each data set can yield valid results. The Cox non-parametric distribution does not require any assumption of baseline hazard model be made for the regression calculations, but only a failure model with a constant failure rate function can tolerate data sets with random or arbitrary start and stop times and still generate valid results. Inclusion of generator set age T_{age} and total cumulative run hours T_{rtfv} as covariates in the regression corrects for this assumption. This introduces additional error and degrees of freedom into the model, but permits Cox regression analysis despite long-term trending over multiple failure events.

Inclusion of the generator set age T_{age} and cumulative run hours T_{rtfv} as covariate predictors had two advantages. The first advantage is that it allowed inclusion of these predictors in the model despite the assumption of an exponential distribution. The second advantage is that it freed the model from the requirement to treat these variables as linear. As covariates, these predictors could be included into the model as logarithmic, linear, exponential, or other complex relationships, even within an exponential parametric regression that would normally have a constant failure rate λ. Representing the generator set age covariate $\beta_{age}T_{age}$ as the transformed function $constant + \beta_{age}\log(T_{age})$ has the same mathematical response within the model as a Weibull base hazard rate function. The same is true for T_{rtfv}. Other transforms may represent other distributions, although Weibull is the distribution most often associated with engine and mechanical wear-out. This technique allows the resultant Cox nonparametric proportional hazards model to simultaneously exhibit characteristics of an exponential distribution and Weibull distribution with respect to long-term generator set age and cumulative run-hours, and in this case produced a much higher quality model than an exponential base hazard rate alone. Thus, even though a parametric model using this technique will have a constant base hazard rate λ_0, the hazard rate $h(t|x_i)$ calculated for each unit for any point in time can include complicated interactions and non-linear relationships.

Equation 3 shows the Cox regression expression representing the log of generator set calendar age as the *n*th predictor. Equation 4 shows the generalized Cox proportional hazards model.

$$\beta_n x_n := \beta_{age} \log(T_{age}) \tag{3}$$

$$\log(h(t|x_i)) = \log(h_0(t)) + \sum_{i=1}^{n} \beta_i x_i + \varepsilon \tag{4}$$

Equation 5 shows the Weibull baseline hazard function where T_{age} is age, m is the shape parameter and η is the scale parameter. Equation 6 shows this expression transformed algebraically via logarithm into an expression similar to the form used in Cox's proportional hazards model in Equation 4.

$$h_0(T_{age}) := \frac{m}{\eta}\left(\frac{T_{age}}{\eta}\right)^{m-1} \tag{5}$$

$$log\left(h_0(T_{age})\right) = (m-1) log(T_{age}) + \log(m) - m\ log(\eta\ ) \tag{6}$$

If Equation 6 is transformed into to the same notation as a typical Cox covariate by substituting $\beta_T := m - 1$, $x_T := \log(T_{age})$, and $\log(h_0) := \log(m) + m\log(\eta)$, as in Equation 7, it is clear that the expression relating Weibull's hazard function based on age is fundamentally no different than any other regression covariate. In fact, including $\log(T)$ as a predictor is mathematically equal to Weibull's hazard function and retains the time-dependent properties of this hazard function without violating the assumption of an exponential baseline hazard model.

$$log\left(h_0(T_{age})\right) = \beta_T x_T + \log(h_0) \tag{7}$$

Substituting this transformed Weibull expression of the time covariate in as the baseline hazard function in the Cox proportional hazards model of Equation 4 reveals the form of a standard exponential parametric hazard distribution shown in Equation 8, or, as used in this research, Equation 9.

$$\log h(t|x_i) = \log(h_0) + \sum_{i=1}^{n} \beta_i x_i + \beta_T x_T + \varepsilon \tag{8}$$

$$\log h(t|x_i) = \log(h_0) + \sum_{i=1}^{n} \beta_i x_i + \beta_{age} log\left(T_{age}\right) + \varepsilon \tag{9}$$

In this way, *t* may be utilized as the regression time variable for each observation period with arbitrary start time set to 0 while simultaneously retaining any Weibull hazard distribution contribution related to the true age of the generator set. As the Cox log likelihood estimation algorithm only utilizes the final event or truncation time, substituting a fixed value of T_{age} for each observation in place of the normally unconstrained Weibull time variable is valid and does not bias results. Any number of time-dependent variables can be included in the Cox regression in this manner, permitting quantitative calculation of complex systems with multiple independent time dependencies.

Including the transformed Weibull expression as a covariate predictor in a parametric proportional hazards model with a nominal exponential base hazard rate results in a distribution displaying aspects of multiple exponential and Weibull distributions. This parametric model can be represented as a model with both exponential and Weibull base hazard rates. The Weibull shape parameter *m* can be calculated from the β_T values returned by nonparametric Cox regression as $m = \beta_T + 1$. For simple models with a single time dependent predictor of the form log(T), the scale parameter can be calculated from the parametric regression results. If multiple time-dependent predictors of the form log(T) are included in the regression, the Weibull scale parameters become conflated into the h_0 scalar and are not as easily separated. This still results in a useful model, as the individual contributions of each constituent component remain included in the resultant hazard function. Hazard curves may be calculated from this model by Cox (1972) methods.

One weakness of this method is that the limit as T approaches zero increases logarithmically over a very short time spans and becomes infinite at log(0). This could lead to over-estimations in survival functions for samples where T=0. Because of this, time values must be constrained or transformed so that $T > 0$.

Rodriguez (2010) showed that a regression model consisting of a single covariate of the form log(t) is a special case where the proportional hazards model multiplied by an accelerated life expression yields the same result as an accelerated life model multiplied by a proportional hazard expression. This relationship is does not hold when multiple time dependent covariates are added into the models, and is not otherwise similar to the methods herein, but did recognize a connection between a covariate of the form log(t) and the Weibull distribution.

DELIMINATIONS AND ASSUMPTIONS

This study assumed prime and standby rated generators in emergency service share a common response to test and maintenance periodicity and can be directly compared. This assumption was based upon the common practice of de-rating a standby generator by ten percent and labeling it as a prime-rated generator. This assumption may have increased the error of calculations of relationships with generator set size and loading.

This study assumed data gathered will be accurate and not tampered with, filtered, or otherwise corrupted by maintenance personnel trying to hide mistakes or lapses in proscribed maintenance or tests. Failures attributable to human error may be entered in official logs as an equipment failure or an unknown failure by unscrupulous personnel and would be indistinguishable from actual equipment failures by this methodology. As this study was looking at aggregated failures of all causes and was largely blind to specific cause attribution, impact from misattribution of failures was anticipated to have minimal negative impact on study results. However, the methods used were not capable of determining whether maintenance and tests were performed, only that records indicate they were performed. This aspect could potentially bias results if widespread falsification of records occurred.

This study assumed that historic maintenance records were accurately interpreted and that the survey form was consistently understood and reported. All survey forms were reviewed for completeness and consistency with past records prior to incorporation into the study database.

This study assumed the quality of maintenance was consistent between sites of similar personnel training and maintenance and test periodicity. As training and personnel qualifications and competency varied, the quality of maintenance performed may also have varied. This methodology had no means with which to gauge maintenance quality beyond the general training category of servicing personnel.

It was assumed that diesel-electric generator sets adapted for use as spark-ignition natural gas generator sets and otherwise sharing identical parts shared identical characteristics and responses could be included in the sample population as diesel-electric generator sets.

IV. ANALYSIS

DATA

One thousand two hundred and eighty one (1,281) standby-years of generator set data was acquired for this research from 239 generator sets, capturing 58 operational failures from 40,161 run-hours of operation. This data was provided from multiple sources including the Department of Defense and commercial providers of maintenance and repair services. The sample population size exceeded the anticipated minimum 1142 standby-year size, but the 58 operational failure events within that data was much smaller than the 141 operational failures anticipated to be required to achieve for significant results. This appears to be due to the sample population of Department of Defense and commercially maintained units in this study to be much more reliable than anticipated based on previous findings by PREP (TM 5-695-5, 2006) or the small scale study in Fehr (2014).

The sample population included generator sets between 10kW and 2000kW with a 324kW mean generator set rating. A bubble chart depicting the distribution of data from different size units is shown in Figure 3; the small bubbles represent right-censored observations and the large bubbles represent observations ending in operational failure.

The oldest generator sets in this study were 43 years old, and five were older than 35 years, but the mean age of generator set in this study was 11.2 years. The mean age of generator sets experiencing operational failures was 12.2 years. Figure 4 shows a summary of generator set records and failures by age.

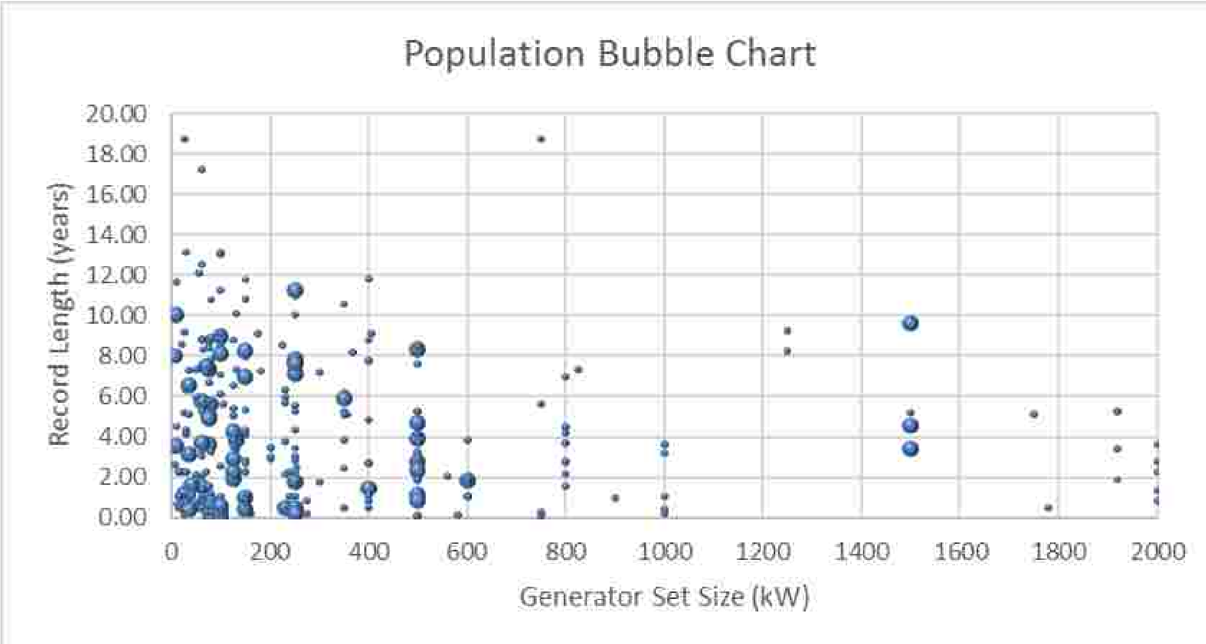

Figure 3. Data distribution, generator size (kW) vs. record length

Most generator sets in this sample population received regular maintenance and could be considered well or very well maintained. Few records were obtained of generators receiving little or no maintenance, as reliable records proved difficult to obtain. To avoid the introduction of selection bias, generator sets whose records were only available because a failed generator was repaired were excluded. Only records for subpopulations believed to be free of selection bias were included. A portion of the generator sets included in the sample population of this study were units where maintenance contracts lapsed due to financial constraints and were later renewed, permitting complete and accurate knowledge of the maintenance history of these units even though no logs were recorded during the period of no maintenance. The only generator sets in this study receiving no maintenance or testing at all were four 600kW generator sets in a new building that were installed without being configured for automatic exercise and with no maintenance provided until several years after installation, and one 60kW generator set wherein the exerciser was not functioning during a period of lapsed maintenance. In the other cases with lapsed maintenance, automatic engine exercisers continued providing weekly no-load tests despite no other maintenance being performed.

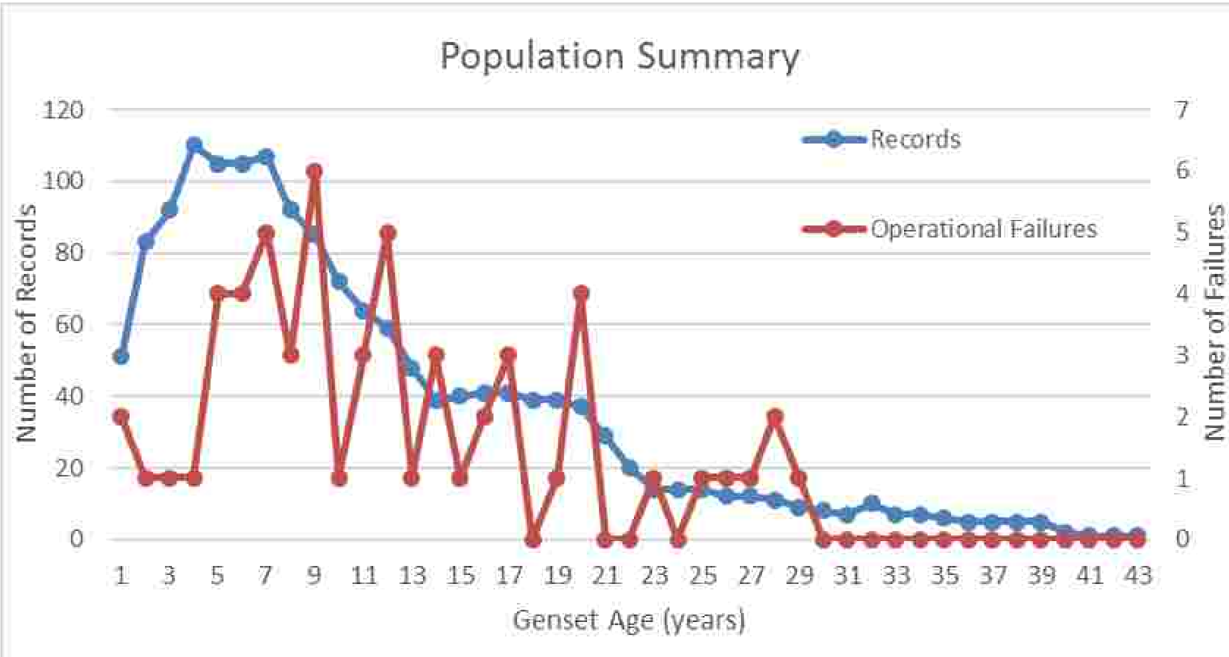

Figure 4. Summary of generator set records and operational failures by age

Maintenance records often listed component failures without clarity on whether the failure occurred during a test or during operation. Descriptions like, “generator failed to start during outage,” were clear, but other descriptions were not always easily interpretable. It was generally assumed that records stating, “unit is alarming” referred to test failures while “generator did not start” referred to operational failures unless occurring during a scheduled maintenance visit. Dispatch personnel of one company indicated electronic maintenance records stating, “customer reported” could also mean, “technician reported” as their work tracking system did not make a distinction when repair orders were called in. Test failure information was not collected for most units, but frequent test failures were noted anecdotally by the research team while reviewing maintenance logs; a thorough review of one subpopulation deemed typical found test failures occurred at a rate of 14:3 compared to operational failures.

INITIAL ANALYSIS

Analysis began with the Cox log likelihood regression model shown in Equation 1, where x_i represented the covariate predictors in Table 4, x_{bi} represented the training factors of the servicing personnel, and β_i and β_{bi} represented the regression coefficients. Make, model, age, run hours, capacity, load, and the predictors in Table 1 were included in the initial regression model. Training was not initially included, as it was intended as an interaction component to be combined with a performed maintenance or test action. The event response was the operational failure Boolean indicator F_{ov}. The periodicity of maintenance and test action predictors in Table 4 was measured in years or fraction of a year. These predictors were treated as covariates with logarithmic and second order predictors tested for significant predictors to determine the mathematical relationship to failure, especially for complicated relationships where failure rates may not only increase with too little maintenance but also with too frequent maintenance.

The regression was performed independently using time t and run-hours T_{rt} as the time variable. There was insufficient data available on the number of starts n_s to include n_s as a time variable. The number of starts was determined to largely be a function of test periodicity within the sample population and was excluded from the regression model as these predictors were already included. The Fehr (2014) framework included a normalized weighted time function, $t_w(t,n_s,T_{rt}) = \omega_t t + \omega_n n_s + \omega_T T_{rt}$, to develop a single general regression model of generator set reliability as a function of time in standby service, the number of starts, and run hours. Statistically significant models were developed using standby time t, but no statistically significant models using time as run-hours T_{rt} were found, leaving standby time t as the only significant time variable. It was concluded from this result that calendar time in emergency standby duty is a more appropriate metric for time than cumulative run hours. Weights were set as $\omega_t = 1$, $\omega_n = 0$, and $\omega_T = 0$, and so $t_w(t,n_s,T_{rt}) = t$.

To remove left truncation of prior failures and the associated bias, time was normalized as $t = 0$ for the start of each observation period and counted in years to the end of each observation period. This was necessary for truncated records as the regression would otherwise assume the generator set had operated to that point since installation without failure, and it would create bias in the regression. Run hours were likewise set as $T_{rt} = 0$ at the start of the observation and counted in run hours accumulated to the end of the observation. Time t and run hours T_{rt} were reset to 0 for the next observation following each failure event. This method was independently developed by Fehr (2014) but is similar to the method proposed by Thomas and Reyes (2014) for Cox regression models with time-dependent covariates. This method removes evidence of left truncation from the observation record, but any potential time-dependent relationship was preserved by the inclusion of the age covariate $\log(T_{age})$ and run hours covariate $\log(T_{rtfv})$ in the model as regression covariates. This further permitted analysis of other orders of these predictors and permitted analysis of complex relationships independent of any assumptions. As shown in Chapter III, inclusion of $\log(T_{age})$ and $\log(T_{rtfv})$ as covariates is mathematically identical to inclusion as Weibull base hazard rates, and the contributions of each to the model can be calculated directly from a standard Cox regression. Thus, this initial model included the simultaneous calculation of one constituent exponential function, two constituent

Weibull functions, and the other maintenance, test, and generator set property predictors. An additional factor, x_{src}, was included to test for bias from the data source. The initial model tested is shown in Equation 10.

$$\log h(\{t\}|x_i) = \log(h_0(\{t\})) + \sum_{i=1}^{22} \beta_i x_i + \beta_{make} M_{ake} + \beta_{model} M_{odel} + \beta_{kW} k_W + \beta_{age} \log(T_{age}) + \beta_{rtfv} \log(T_{rtfv}) + \beta_L x_L + \beta_{src} x_{src} + \varepsilon \quad (10)$$

The statistical software package, *R*, with the *survival* and *eha* packages (Broström, 2011) was selected for this analysis as it supports Cox and parametric regression models and includes functions insensitive to left truncation and right censorship. Data was analyzed using the functions *surv*, *coxph*, *coxreg*, *cox.zph*, *resid*, and *survfit*, as well as *plot*, *summary*, and other standard *R* tools. The *coxph* function was used for bidirectional stepwise regression as it was more tolerant of poorly fitting data than *coxreg*, and *coxreg* was used to generate log likelihood values to compare the best fitting models. The *cox.zph* function was used to test the proportionality assumption and the *resid* function to analyze the residuals. Bidirectional stepwise regression was used beginning with the log likelihood regression model shown in Equation 10. Backwards stepwise regression was used iteratively to eliminate the least significant predictors from the model. Forward stepwise regression was then used to add predictors back into the model one at a time to ensure none were erroneously removed. Bidirectional stepwise regression was further used to investigate other combinations of predictors, with predictors chosen based on interim stepwise models. Predictors with “NaN” Z and p values were removed from the model via backward stepwise regression until the regression results converged. Once the regression results converged, the least significant predictors were removed one at a time through backward stepwise regression until all remaining predictors were significant. The log likelihood method was used in conjunction with p-tests to determine the most appropriate model. The initial regression model was also analyzed with the *stepAIC* function of the *MASS* package (Ripley, 2017) with identical results. As very few predictors were found to be significant, the use of more sophisticated operations research methods for model building like simulated annealing and Markov-chain Monte Carlo analysis was not utilized.

After the model in Equation 2 was reduced to only contain significant predictors, the second order interactions of Equation 1 were added by forward stepwise regression including interactions between test and maintenance predictors and interactions with the training predictors. Alternatives to $\log(T_{age})$ and $\log(T_{rtfv})$ were tested via substitution and forward stepwise regression to eliminate assumptions of mathematical relationship, including $\log(\log(T))$, T, T^2, and e^T.

This model was then fitted to a parametric model via piecewise constant hazard analysis to determine the baseline hazard function value.

COX REGRESSION ANALYSIS RESULTS

Cox regression analysis using the *coxph* and *coxreg* functions in *R* found the two Weibull predictors for age $\log(T_{age})$ and run hours $\log(T_{rtfv})$ to be statistically significant in nearly all interim and final models with $p < 0.01$ in the most significant models. Five maintenance predictors, x_{12}, x_{13}, x_{16}, x_{17}, and x_{18} were found to be significant with $p < 0.1$ when combined with $\log(T_{age})$ and $\log(T_{rtfv})$, as were the generator set parameters for size k_W and load x_L and the control factor for source x_{src}. However, none of these predictors were found to be significant when applied in any combination except one-at-a-time in a model with $\log(T_{age})$ and $\log(T_{rtfv})$. Nor were any training predictors found to be significant in combination with these predictors. Further investigation revealed a large degree of conflation between x_{12}, x_{13}, x_{16}, x_{17}, and x_{18}, as well as some potential conflation with k_W, x_L, and x_{src}. This was due to a large portion of the data coming from either Department of Defense sources which had some variance in periodicity but were still largely homogenous or third-party contractors which also had some variance in periodicity of some maintenance items, but were even more homogenous. The predictors that showed the highest significance in the regression were the predictors that had the largest differences in periodicity between these two sources.

Two new variables were created to adjust for this conflation to represent different archetypes of the conflated predictors, with x_{23} representing the shortest periodicity of touch labor of any sort and x_{24} representing the shortest periodicity of any type of generator run test. Two new accompanying training predictors x_{b23} and x_{b24} were also created. Training predictors x_{23} and x_{24} were found not to be significant; however the accompanying training predictors x_{b23} and x_{b24} were both found to be significant in interim models for staff subject-matter expert maintenance and tests. No statistically significant difference was found between site visit subject-matter expert maintenance and staff collateral duty maintenance for this data. This raised the question: is the mere presence of a staff subject-matter expert as or more important than the periodicity of maintenance that subject-matter expert performs, or is something else occurring?

An in-depth review of the data revealed that the training factors were highly conflated with the most significant predictors, and that the significance of the maintenance predictors appeared to be highest in the predictors that had the least homogeneity across the subpopulations.

The maintenance predictor for electrical tightness x_{16} had the highest predictor significance and overall model quality in interim models and was selected for the final model. No failures related to electrical tightness maintenance were observed in any unit of the sample population in tests or operation, but this predictor was nevertheless found significant in the Cox model with $p = 0.0469$. This indicator is not only highly conflated with the training predictor x_{b23}, but also reflects the largest difference in periodicity between units maintained by staff subject-matter experts and units maintained by contractors. Every site with a staff subject-matter expert performed this maintenance at least annually, while very few sites without staff subject-matter experts performed it at all. This predictor may not have significance from indicating the periodicity of electrical tightness, but significance from reflecting the level of overall generator maintenance intensity, allowing it to effectively act as an analogue representing a combination of maintenance, testing, and training.

Generator set size k_W was statistically significant in an interim model with just $\log(T_{age})$, $\log(T_{rtfv})$, but with both k_W and x_{16} in the model, the statistical significance of generator set size reduced to outside of the the α error threshold, indicating a degree of conflation and risk of Type 1 error. This may be due to stratification and bias in the data, as the mean contractor-maintained units in this study tended to be smaller than the mean units maintained by staff subject-matter experts. Correlating against the training indicator for x_{16}, x_{b16}, the average size of staff subject-matter expert generator sets was 770kW in the data acquired while the average size of service visit subject-matter expert was 233kW. No statistically significant relationship between generator set reliability and generator size was found in the Cox regression analysis aside from one incomplete interim model. There is insufficient evidence to suggest generator size is a statistically significant predictor of reliability.

Generator set load was statistically significant in a model with just log(T_{age}), log(T_{rtfv}), and x_L, but with both x_L and x_{16} in to the model, the statistical significance of generator set load reduced to outside of the α error threshold, indicating a degree of conflation. This may due to stratification of the data and a degree of conflation with x_{16} as most contractor-maintained units in this study utilized automatic weekly no-load tests where x_L = 0kW and reduced the mean load of the subpopulation, while few sites with staff subject-matter experts ran no-load tests and most only ran monthly load tests. Correlating against the training indicator for x_{16}, x_{b16}, the average loading of staff subject-matter expert generator sets was 12% in the data acquired, while the average loading of contractor maintained size was 2%. These numbers are unlikely to be accurate, as accurate load data was difficult to obtain for many generator sets, especially for those units whose history was compiled by contractor maintenance records that did not include information about facility load. These units all utilized regular no-load tests, so typical loads were approximated as 0% when this maintenance data was imported into the database. However, the actual average load should be higher since these generators support their facilities during utility outages and the facilities are unlikely to be at 0% load. There were reports noted anecdotally during data acquisition as indications of wet stacking of some lightly-loaded units, but no clear statistically significant relationship between generator set reliability and generator set load was found in the Cox regression analysis of this data.

Sensitivity analysis was performed on the data with six different test data sets eliminating certain subgroups or injecting erroneous failures to determine if the strong statistical significance of log(T_{age}) and log (T_{rtfv}) from the regression may be resulting from failed repair attempts or the relatively small number of older generators. Regression tests were performed after removing all generator sets older than twenty-five years, younger than five years, and reporting periods of less than six months, and by changing the status of data from some of the oldest units from censored to failure, but the regression relationships varied little during this sensitivity analysis. This increases confidence that the relationships determined by the Cox regression exist in the sample population and are not purely statistical chance.

The resultant Cox model selected from the analysis is show in Equation 11 and includes $\log(T_{age})$, $\log(T_{rtfv})$, and x_{16} with coefficients calculated from the regression.

$$\log h(t|x_i) = \log(h_0(t)) - 1.476\log\left(T_{age}\right) + 0.700\log\left(T_{rtfv}\right) + 0.240x_{16} + \varepsilon \qquad (11)$$

The proportionality assumption was tested on Equation 11 utilizing the *cox.zph* proportionality test function and *resid* residuals function which validated the proportionality assumption. The regression functions, test results, residuals, and survival fits associated with the analysis are shown in Appendix B.

Predictor significance is presented in Table 6 as a listing of the Wald *p*-values of each predictor when added as a fourth predictor to the model of Equation 11 and as a replacement for x_{16}. Different order time predictors replaced the Weibull component were also tested, either as an additional predictor or as a replacement for the relevant time predictor. For the special cases of $\log(T_{age})$ and $\log(T_{rtfv})$, the other time predictor was removed from the model. Predictors with significance within $\alpha \leq 0.10$ are notated with *. Predictors selected for the final model are notated with **.

The survival fit of the Cox regression using mean covariate values is shown in Figure 5. However, while the *coxreg* and *coxph* functions deal properly with $\log(T_{age})$ and $\log(T_{rtfv})$ in the computations, the *survfit* function plot treats both log(T_{age}) and log(T_{rtfv}) as static covariates based on the sample population mean. The survfit function assumes all generator sets are eternally 12.4 years old with 395 run hours, and does not accurately represent the time dependency of these covariates. This issue will be addressed in greater depth during the parametric regression analysis.

Table 6. Table of Predictor Significance When Added to or Replacing a Potentially Conflated Predictor in the Model of Equation 11.

Predictor	Addition *p*-value	Replacement *p*-value	Predictor	Addition *p*-value	Replacement *p*-value
x_1	0.751	0.971	x_{19}	0.504	0.306
x_2	0.914	0.897	x_{20}	0.742	0.157
x_3	0.931	0.833	x_{21}	0.896	0.106
x_4	0.934	0.83	x_{22}	0.993	0.101
x_5	0.931	0.833	x_{23}	0.867	0.658
x_6	0.871	0.661	x_{24}	0.873	0.847
x_7	0.871	0.661	x_{b23}	0.233	0.078*
x_8	0.871	0.661	x_{b24}	0.243	0.078*
x_9	0.872	0.666	k_W	0.506	0.094*
x_{10}	0.872	0.666	x_L	0.173	0.062*
x_{11}	0.866	0.657	x_{src}	0.434	0.039*
x_{12}	0.565	0.048*	log(log(T_{age}))	0.678	0.0008*
x_{13}	0.731	0.086*	log(T_{age})	0.000018**	0.0004*
x_{14}	0.606	0.107	T_{age}	0.313	0.0096*
x_{15}	0.752	0.905	log(log(T_{rtfv}))	0.236	0.026*
x_{16}	0.047**	N/A	log(T_{rtfv})	0.007**	0.023*
x_{17}	0.572	0.059*	T_{rtfv}	0.379	0.025*
x_{18}	0.572	0.059*	log(x_{16})	1.000	0.567

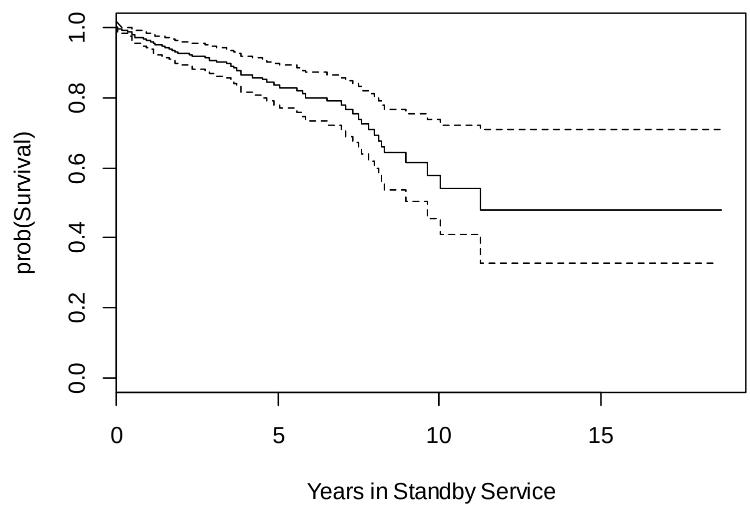

Figure 5. Survival fit of regression by time in standby service, *h(t).*

PARAMETRIC REGRESSION ANALYSIS, FULL MODEL

This study required the assumption of an exponential base hazard rate to permit Cox regression analysis of multiple-failure systems. While Weibull base hazard rates are supported within this method, h_0 must be a constant, $h_0 = \lambda$. This delimitation permitted taking the Cox regression results in Equation 11 to create the parametric hazard function in Equation 12.

$$h(t) = \lambda e^{0.240x_{16} - 1.476\log(T_{age}) + .700\log(T_{rtfv}) + \varepsilon} \quad (12)$$

This hazard rate is time dependent due to the Weibull relationships of $\log(T_{age})$ and $\log(T_{rtfv})$. As time t as used in the Cox regression is not in this equation, the arbitrary measure of time and arbitrary establishments of $t = 0$ as used in the Cox regression have little meaning. A more meaningful unit of measurement for the parametric model is absolute time from the original installation of the generator, $T_{age} = 0$. Therefore, in the parametric mathematical model $T_{age} = t$, it can be represented with t. Run hours T_{rtfv} is time dependent as well, and it will never decrease over time, but T_{rtfv} is not necessarily linear. The mean annual run-time in the sample population was 31.2 hours per year, but this differs randomly from year to year and from different test periodicities. Different run hours are explored later in the discussion, but T_{rtfv} was approximated for this analysis as related to the mean annual run-time for the sample population multiplied by the generator set age, $T_{rtfv} = 31.2T_{age}$. This approximation permits further development of this parametric equation into the more standard time dependent form in Equation 13. In this form, t represents the time passed since generator set installation, as measured in years.

$$h(t) = \lambda e^{0.240x_{16} - 1.476\log(t) + .700\log(31.2t) + \varepsilon} \tag{13}$$

The presence of two time-dependent Weibull predictors $\log(T_{age})$ and $\log(T_{rtfv})$ in the model pose challenges for parametric regression as standard tools like the *phreg* function in *R* are not structured to deal with repairable systems. The parametric regression must use Equation 12 to develop regression coefficients, but simultaneously use Equation 13 to fit the parametric model to the data. None of the techniques proposed by Thomas and Reyes (2014) or Fehr (2014) yielded accurate results due to the inability of the function to accurately model T_{age} and T_{rtfv} as time dependent variables. The *phreg* function extensions required for parametric regression this have not yet been developed. The regression was tested in *R* to see what the results would be, and while the parametric regression results from *phreg* returned very similar predictor coefficients as the Cox non-parametric regression, *phreg* was unable to accurately fit $h_0(t)$ to the data due to function limitations, resulting in nonsensical hazard rates and an extremely poor fit to the actual data.

It was instead chosen to apply a piece-wise constant hazard model in *Microsoft Excel* which effectively permitted treating the complex relationship as a summation series of standard exponential distribution expressions. This method leveraged the predictor coefficients returned by the Cox model and allowed the iterative fitting of $h_0(t)$ values to the data until a good fit was achieved. As this methodology had an underlying assumption of an exponential distribution, it was assumed that $h_0(t) = \lambda$, with a single value of λ for all populations. The only variable left unsolved in $h(t)$ is λ. A piece-wise constant hazard parametric model was constructed using Equation 13 and the Cox (1972) survival equations. The following equations were used to build a piece-wise constant hazard parametric model as s spreadsheet in *Microsoft Excel*.

$$f(t) = h(t)S(t-1)$$

$$F(t) = \sum_{t=0}^{T} f(t)$$

$$R(t) = e^{-h(t)}$$

$$S(t) = e^{-F(t)}$$

To fit the model to the data, the sample population data was stratified into two subpopulations based upon existing stratification of x_{16}, one subpopulation with $x_{16} = 7$ years and one subpopulation $x_{16} \leq 1$ year. The value of 7 was selected to represent units receiving no maintenance by Fehr (2014). The mean value of x_{16} for the latter subpopulation was 0.45 years. Sample mean time between failures MTBF = $\Sigma\, t \,/\, \Sigma\, F_{ov}$ = 20.7961 years as calculated from the raw data for the subpopulation with x_{16} = 7 years, and 60.2178 for the subpopulation where x_{16} < 7 years. The equation $R_{\mathrm{data}} = \mathrm{e}^{-1/\mathrm{MTBF}}$ yields an estimation of 98.4% and 95.3% reliability for each respective subpopulation. These values were compared to the average reliability $R(t)$ of the time-dependent model, which was weighted by a histogram function of the age of the generators in the sample pool. Values for λ were iteratively selected to minimize the variance between the model prediction for the two subpopulations and the sample MTBF. A value of λ = 0.00696 was found to yield the best model fit using these reliability estimators. The final parametric model, where t represents the age of the generator set in years, is shown in Equation 14. The hazard function values for the two subpopulations are plotted in Figure 6 for comparison. The survival functions are shown in Figure 7 and Figure 8.

$$h(t) = 0.00696e^{0.240x_{16} - 1.476\log(t) + .700\log(31.2t)} \tag{14}$$

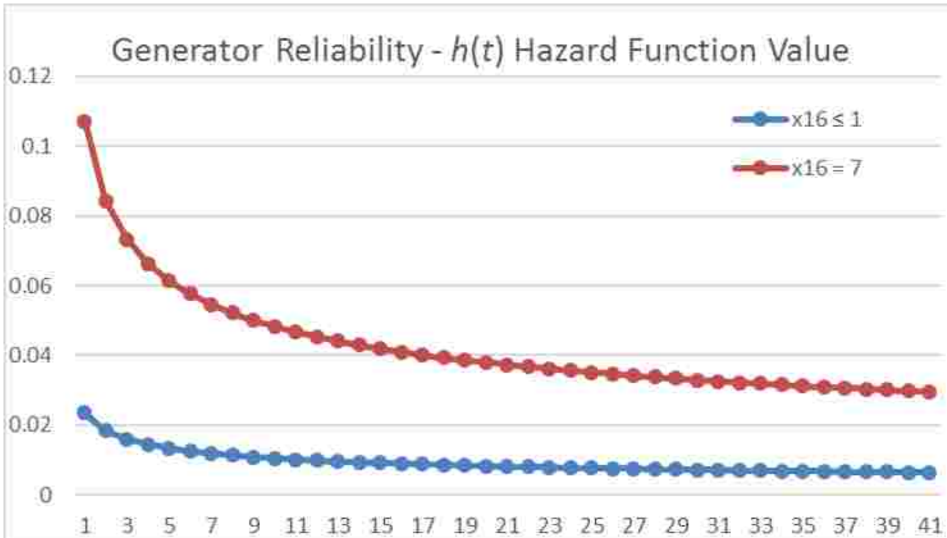

Figure 6. Generator reliability parametric model stratified hazard function values

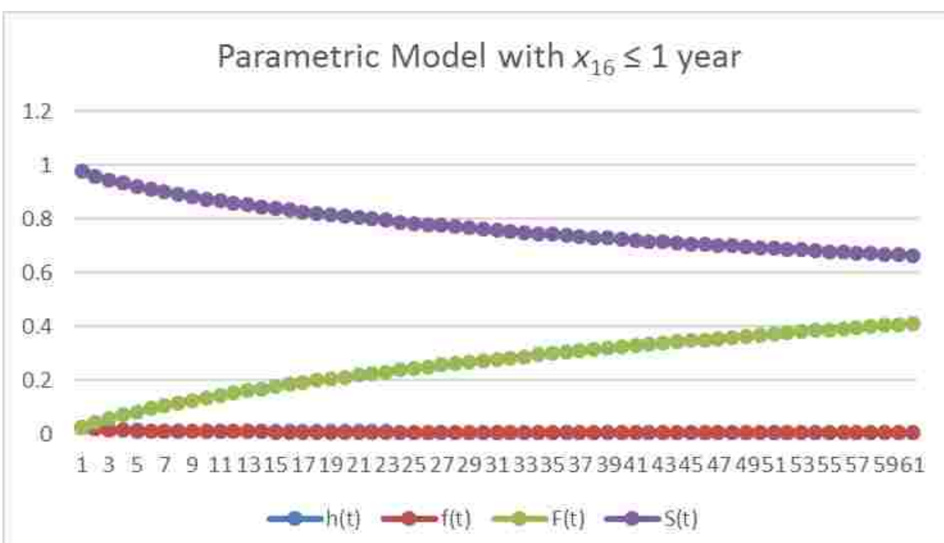

Figure 7. Generator reliability parametric model for the $x_{16} \leq 1$ subpopulation

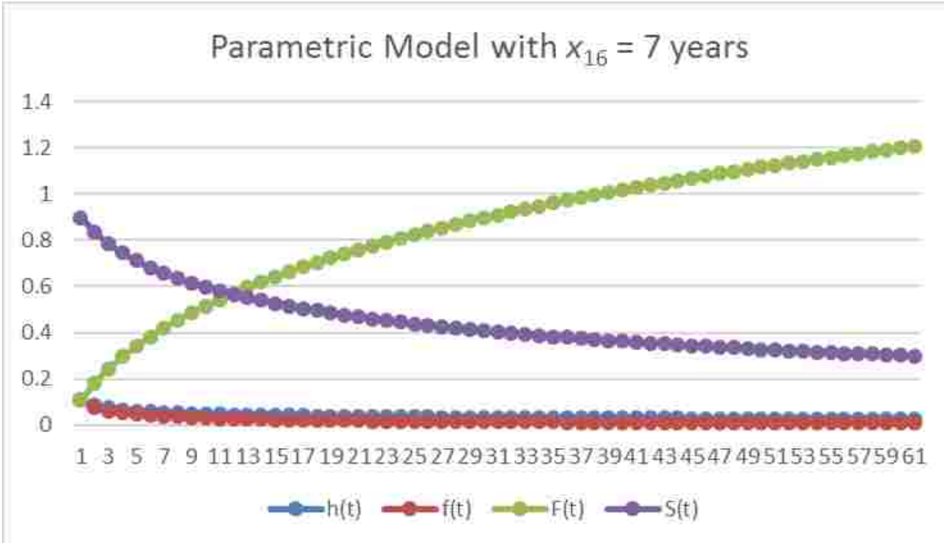

Figure 8. Generator reliability parametric model for the x_{16} = 7 subpopulation

This model yielded a weighted average annual reliability for x_{16} = 7 years of 94.7% compared to the sample subpopulation net annual reliability of 95.3%. This model yielded a weighted average annual reliability for x_{16} = 0.45 of 98.9% compared to the $x_{16} \leq 1$ sample subpopulation net annual reliability of 98.4%.

The fitted models were visually compared to the Nelson-Aalen estimators (Müller, 2004) for goodness-of-fit and were found to be consistent with the data. The Nelson-Aalen estimators were calculated in *Microsoft Excel* using a Boolean summation algorithm with record lengths rounded up to the next whole year. The data was stratified by x_{16} = 7 and $x_{16} \leq 1$ and the Nelson-Aalen curves are shown in Figure 9 and Figure 10. The Nelson-Aalen curves did not reflect as many early deaths as the infant mortality curve of log(T_{age}) plots would indicate, but instead reflected relatively linear rates with sudden knees and low failure rates for older units. Using *Microsoft Excel* to fit a curve to match the shape of the Nelson-Aalen plot for x_{16} = 7 yielded a linear plot as the highest R^2 = 0.96. The exponential trendline fit in Microsoft Excel yielded an

$R^2 = 0.83$ and logarithmic trendline fit yielded an $R^2 = 0.81$. Despite having similar R^2 values, these two trendlines had opposite shaped concave and convex curves. Trendlines fitted by *Microsoft Excel* to the $x_{16} \leq 1$ sample subpopulation yielded lower R^2 values, with a maximum $R^2 = 0.71$ for logarithmic and 0.47 for linear. *Microsoft Excel* would not fit an exponential curve for this data set. In both cases, the logarithmic curve was observed to have a shape closely resembling the model cumulative distribution function; in the case of $x_{16} \leq 1$ the scale was also similar, and the plot nearly identically matched the cumulative distribution function.

One interesting aspect of this model is that the two most significant predictors, $\log(T_{age})$ and $\log(T_{rtfv})$, were found to have opposite signs. This model predicted generator sets will become more reliable with age, but less reliable with more run-hours. The combination of these two predictors resulted in relatively static hazard functions once past the infant mortality stage, and helps explain the observed low failure rate of older generator sets while remaining consistent with logical expectations that generator sets approaching manufacturer design limits of maximum run hours will be less reliable.

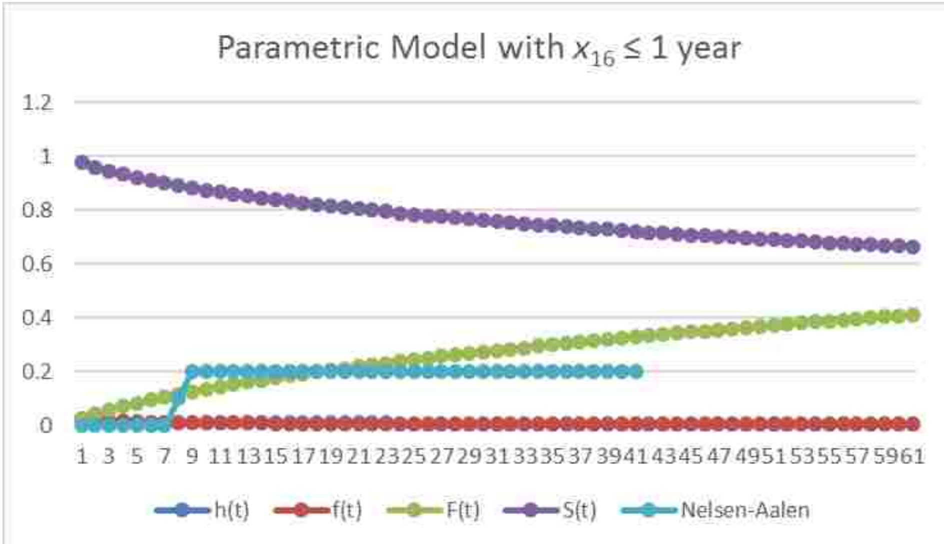

Figure 9. Generator reliability parametric model for the $x_{16} \leq 1$ subpopulation with Nelson-Aalen estimator.

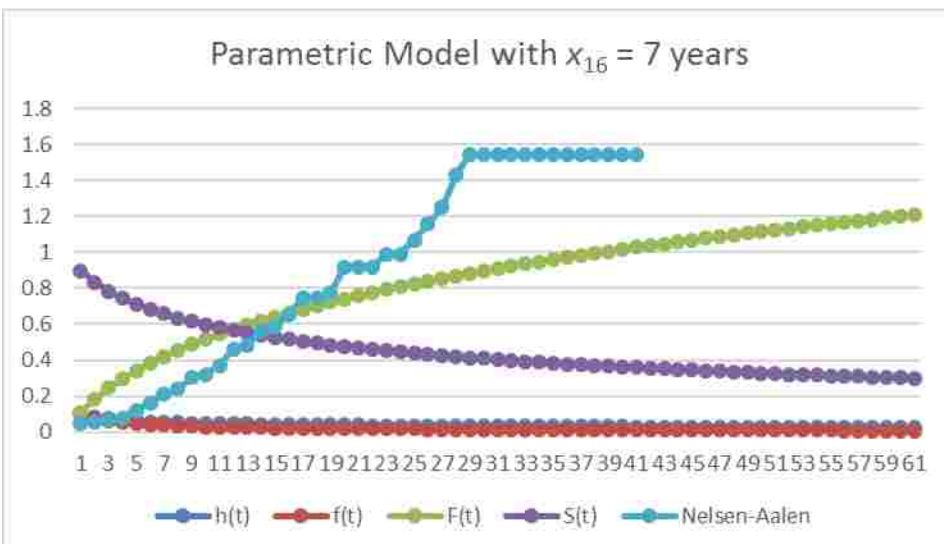

Figure 10. Generator reliability parametric model for the $x_{16} = 7$ subpopulation with the Nelson-Aalen estimator.

PARAMETRIC REGRESSION ANALYSIS, SIMPLIFIED MODEL

The Cox (1972) parametric proportional hazard model combining one exponential and two Weibull base hazard rate relationships resulted in a high quality model fits this data well, but there is potential benefit to also investigating a simpler model that could be more easily applied to practical engineering problems and might better benefit non-academics more interested in practical use of this research than the pure science and mathematics driving it. Therefore, a simplified time-independent model was developed without log(T_{age}) or log(T_{rtfv}) and compared to the full model. One immediate benefit of the simplified model is that removal of the time dependent predictors permitted standard functions in *R* to be used. The Cox regression in *coxreg* in *R* of the simplified model function with x_{16} as the only predictor was significant at $p = 0.084$ for covariate x_{16} and $p = 0.0325$ for the model with a model maximum log likelihood of -279.75. This was much lower quality than the model in Equation 12, but still significant with $\alpha \leq 0.10$ and of a form that can be analyzed in *R* using standard functions. This simplified model from *coxreg* is shown in Equations 15 and 16. The parametric model from *phreg* is shown in Equation 17 and Figure 11. The goodness of fit of this model is shown in Figure 12. The Cox regression using run hours as time yielded an identical model; the full regression information, proportionality test results, and residuals of both models are included in in Appendix B.

$$\ln h(t|x_i) = \ln(h_0(t)) + 0.199x_{16} \tag{15}$$

$$\ln h(t|x_i) = h_0(t)e^{0.199x_{16}} \tag{16}$$

$$h(t) = 0.0116e^{0.199x_{16}} \tag{17}$$

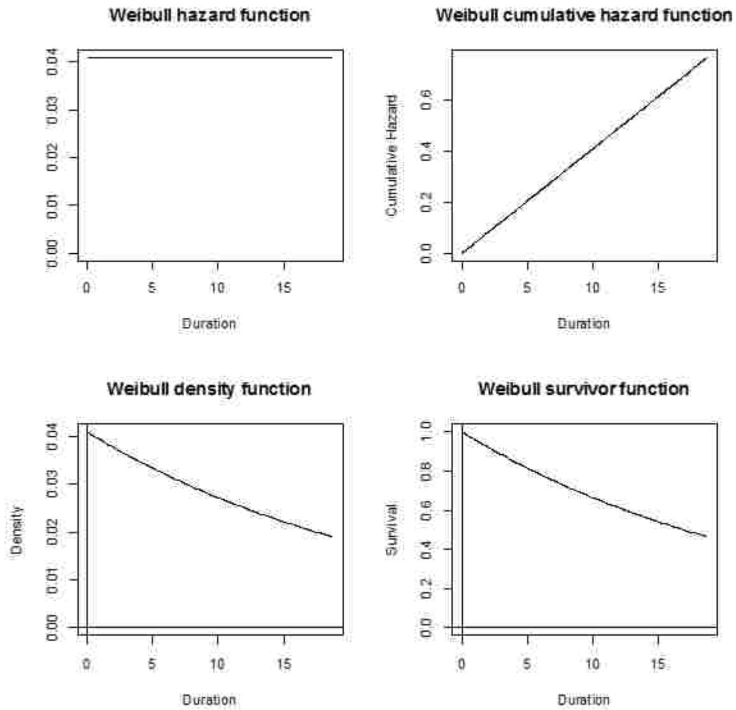

Figure 11. Simplified model parametric functions.

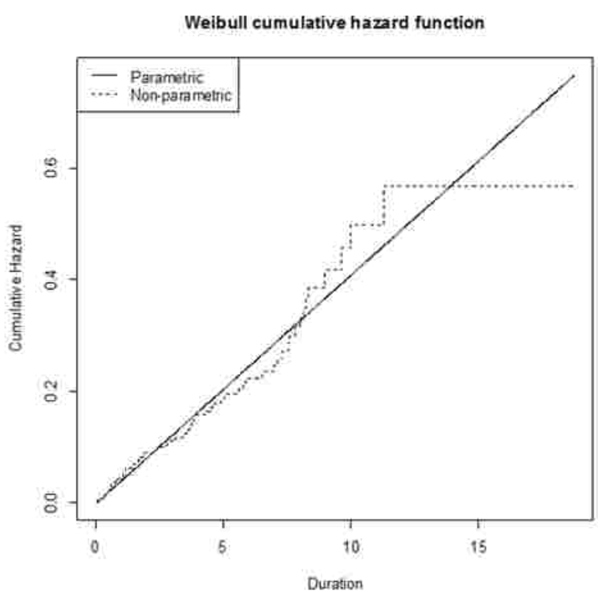

Figure 12. Graphical goodness-of-fit test of the simplified model, Exponential vs Cox

The simplified model yielded an exponential failure distribution with a constant hazard rate of λ = 0.01161. When applied to the weighted reliability piecewise-constant hazard model developed for the full model, the results had less variance from sample MTBF than the full model. The simplified model predicted an annual reliability of 95.4% for $x_{16} = 7$ years compared to the sample subpopulation annual reliability of 95.3%. This model yielded a weighted average annual reliability of 98.8% for $x_{16} \leq 1$ year compared to the sample sub population net annual reliability of 98.4%. The full and simplified model returned very similar values for the $x_{16} \leq 1$ subpopulation, 98.9% and 98.8% respectively, but much larger differences between values for the $x_{16} = 7$ subpopulation, 94.7% and 95.4% respectively.

The visual comparison of the hazard functions of the full and simplified models in Figure 13 reveals the time dependency of the comparison. The full model and simplified model may have yielded similar results for averages of large blocks of generator sets, but the hazard function and reliability predictions differed by considerable amounts, especially for younger and older generator sets where the differences between the model predictions are more pronounced.

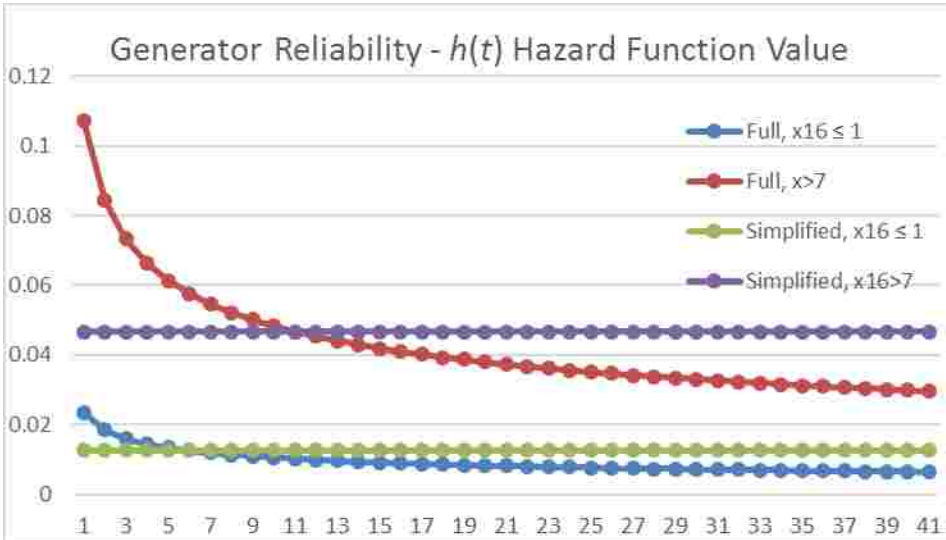

Figure 13. Generator reliability parametric model stratified hazard function values, Full Model and Simplified Model

V. RESULTS AND DISCUSSION

The data acquired for this study revealed a sample population with fewer operational failures and much higher reliability than was anticipated from prior research. The sample population was found to have maintenance practices falling loosely into two stratified subpopulations, each with a large degree of homogeneity. The maintenance of Department of Defense units differed with some respects, but were still largely maintained following Department of Defense standards with monthly load tests and frequent maintenance intervals. The contractor-maintained generator sets were also largely homogenous with nearly every unit receiving an automated weekly no-load test and a contractor visit with a periodicity of one to six months. Nearly all units received what could be categorized as good, very good, or excellent maintenance, and failure rates were lower than anticipated. Statistically significant relationships with reliability could not be determined for all predictors.

Two predictors stood out as the most highly significant, the log of generator age $\log(T_{age})$ and the log of generator cumulative run hours $\log(T_{rftv})$. These predictors both exhibit Weibull base hazard rate characteristics in the parametric proportional hazards model and suggest generator sets experience elevated infant mortality levels but become more reliable with age, although less reliable at higher run hours.

Interim model test results and data analysis suggested a high degree of conflation existed between many of the maintenance and test predictors within the sample population. The interim model test results also suggested conflation between several maintenance and test predictors and generator set size and load. The most significant predictors in the interim models were battery resistance or impedance test x_{12}, check electrical tightness x_{16}, data source x_{src}, and generator load

x_L. Investigation of the data found x_{12} and x_{16} were largely conflated with each other and also with x_{src}. When paired together in a single model, x_{12} and x_{16} yielded nearly equal and opposite coefficients. Two new predictors were created to be generic analogues of the most frequent touch maintenance periodicity of any type x_{23} and most frequent test periodicity of any type x_{24}, but neither of these predictors were found to be significant. However, the associated training predictors x_{b23} and x_{b24} were found to be significant, but only when analyzed independently of the maintenance and test periodicity predictors. This was unexpected as these predictors were intended to exclusively be interaction modifiers to the maintenance and test predictors, and not standalone predictors, but the statistical significance as a standalone predictor in the regression suggested the presence of a staff subject-matter expert performing tasks at any periodicity increased reliability compared to collateral duty staff or subject-matter expert site visits. No statistically significant difference was noted between collateral duty staff and subject-matter expert site visits, but all sites in the sample population with collateral duty staff also relied heavily on subject-matter expert site visits, so this lack of difference may reflect conflation between these two predictors and not equivalence.

Deeper analysis revealed a very high level of conflation between x_{b23}, x_{b24}, x_{src} and electrical tightness x_{16}, with all reflecting high significance in interim models. Maintenance predictor x_{16} is related to re-torqueing lug bolts on the main electrical conductors as these connections can loosen over time. While no failures related to electrical tightness were reported, the only sites that reported performing electrical tightness checks were those with had full-time subject-matter experts on staff, and the periodicity of x_{16} appeared to correlate generally with the intensity of maintenance from those staff subject-matter experts. For this reason, and the high significance of x_{16} in the models and high quality of the model including x_{16}, predictor x_{16} was selected as the analogue predictor to represent the general quality of maintenance and testing in lieu of x_{b23}, x_{b24}, x_{src}, or any other maintenance or test predictor.

The interim regression models discussed in this section are included in Appendix B.

HYPOTHESIS TESTS RESULTS

Generator set reliability was found to be representable as a function of generator set age, run hours, and other maintenance and test predictors. The null hypothesis H_{o0} was therefore rejected for the primary research question in favor of the alternate hypothesis H_{a0} that maintenance periodicity, test periodicity, training, make, model, size, age, run time, or load have an impact on emergency diesel-electric generator set reliability.

Predictors $\log(T_{age})$ and $\log(T_{rtfv})$ were found to be highly significant. It can be stated with a high level of confidence that there is a relationship between generator set age and generator set reliability and that there is a relationship between generator set run hours and generator set reliability; the null hypotheses H_{o4} and H_{o5} were rejected. Predictors x_{12}, x_{13}, x_{16}, x_{17}, x_{18}, x_{b23}, x_{b24}, x_{kW}, x_{xl} and x_{src} each returned significant results in interim models, but there is a degree of conflation evident between these predictors. Due to this conflation, it cannot be stated with complete confidence from the results of the statistical analysis alone what the individual relationship of these predictors was with generator set reliability, but there was sufficient evidence to reject the null hypothesis for maintenance periodicity H_{o1} and to include x_{16} in the final model as the collective analogue for maintenance. Sufficient evidence was found to reject the null hypothesis for test periodicity H_{o2}, but a test predictor specific to test periodicity could not be explicitly included in the model due to conflation.

The regression results of interim models suggested the possibility of a relationship between generator size, load, make, and model, but such relationships may also be due to conflation. As no evidence was found to clearly demonstrate such relationships, the result of the null hypotheses test of H_{o3}, H_{o6}, H_{o7}, and H_{o8} is failure to reject.

The hypothesis tests results are summarized in Table 7.

Table 7. Hypothesis Test Results Summary

	Null Hypothesis	**Result**	***p*-Test**
H_{o0}	Maintenance periodicity, test periodicity, training, make, model, size, age, run time and load have no impact on emergency diesel-electric generator set reliability.	Reject	$p = 0.00002$
H_{o1}	Maintenance periodicity has no impact on emergency diesel-electric emergency diesel-electric generator set reliability.	Reject	$p = 0.047$
H_{o2}	Test periodicity has no impact on emergency diesel-electric generator set reliability.	Reject	See Text
H_{o3}	Size has no impact on emergency diesel-electric emergency generator set reliability.	Fail to reject	$p > 0.10$
H_{o4}	Age has no impact on emergency diesel-electric emergency generator set reliability.	Reject	$p = 0.00002$
H_{o5}	Cumulative chronometer run-hours have no impact on emergency diesel-electric generator set reliability.	Reject	$p = 0.0067$
H_{o6}	Load has no impact on emergency diesel-electric generator set reliability.	Fail to reject	$p > 0.10$
H_{o7}	Training of servicing personnel has no impact on emergency diesel-electric generator set reliability.	Fail to reject	$p > 0.10$
H_{o8}	Make and Model have no impact on emergency diesel-electric generator set reliability.	Fail to reject	$p > 0.10$

HYPOTHESIS H_{o1} TEST RESULTS

The null hypothesis was rejected for maintenance periodicity H_{o1}. The maintenance predictor x_{16} was selected with coefficient $\beta_{16} = 0.240$ for the final model at a significance of $p = 0.047$, which meets the $\alpha \leq 0.10$ criteria for significance. Maintenance predictors x_{12}, x_{13}, x_{17}, and x_{18} and were found to be significant at $\alpha \leq 0.10$ in interim models, but there was a degree of conflation evident between these predictors and predictors x_{b23}, x_{b24}, $k_{W,}$ x_l and x_{src} that made determining the specific relationship of each predictor difficult. The significance of predictors $k_{W,}$ x_L and x_{src} are discussed in detail in subsequent paragraphs, but are believed to be significant in the interim regression models only due to a degree of conflation with significant maintenance predictors. None of the maintenance predictors were significant in combination with each other. Predictor x_{16} had the highest significance in the interim models and yielded the most significant mode and was thus selected to be a general analogue reflecting the conflated maintenance predictors x_{12}, x_{13}, x_{16}, x_{17}, and x_{18} and training predictors x_{b23} and x_{b24}. Predictor x_{16} was viewed as generally reflecting the intensity and quality of maintenance efforts, and not specifically related to electrical tightness.

HYPOTHESES H_{o2} TEST RESULTS

The null hypothesis was rejected for test periodicity H_{o2} due to observations made during data acquisition. The most significant test periodicity predictor for dead-bus operational tests β_{22} yielded $p = 0.101$ in one interim model, suggesting some significance but falling just beyond the criteria for statistical significance. Load test on operational load β_{21} yielded $p = 0.106$ in one interim model, suggesting some significance for this test periodicity as well, but also falling just beyond the criteria for significance. Load testing on a load bank β_{20} yielded $p = 0.157$ in one interim model and no-load testing β_{19} yielded $p = 0.306$ in another interim model. Conflation with other model predictors and subpopulation homogeneity is suspected to have contributed to the regression interim model results as the indications of negative test correlation of no-load tests and load bank tests runs counter to other observations made during this research.

The regression results may be a Type II error from insufficient data to determine a statistically significant relationship between test periodicity and reliability for the frequently tested and very frequently tested generator sets in the sample population, and the absence of available data from infrequently tested generator sets. Test failures were not requested or collected during data acquisition, but much larger numbers of test failures than operational failures were observed in historic records during data acquisition indicating that a positive correlation between test frequency and reliability would likely have been found had frequent tests not been conducted on these units, a correlation not visible in the regression results. A review of a portion of maintenance records acquired for this research covering a subpopulation of 225 standby-years of operation for units receiving weekly no-load tests and contractor maintenance visits every six months revealed fourteen test failures and three operational failures, a ratio of 4.7:1 of test to operational failures, and an annual reliability of 92% with test failures included. This is consistent with reliability levels reported by prior research that included test failures and operational failures in the reliability calculations. The small-scale data set of Fehr (2014) found eight test failures and two test failures in 126 standby-years of records, a ratio of 4:1 of test to operational failures, and 94% reliability for well-maintained units with test failures included. PREP (TM 5-605-5) reported 88% annual reliability for a mix of well, average, and poorly-maintained units, but PREP did not make a distinction between test and operational failures.

HYPOTHESES H_{o3} TEST RESULTS

The predictors for generator set size and data source were included in the model only to test the assumption that this model can be generalized to generator sets of different size and ensure that the source of the data was not a source of bias.

Generator set size and data source were both found to be statistically significant in interim regression models with p = 0.094 and p = 0.039 respectively, but neither was significant when combined in a model with predictor x_{16}. These results are believed to be due to conflation with x_{16} and several other predictors due to stratification in the sample population. However, the results cannot be separated out and tested independently to verify these assumptions due to the conflation.

Despite the inconclusive statistical results, inductive reasoning suggests that the relationship with generator size and source in the interim models was a Type 1 error from conflation with the subpopulation and not a direct cause. Therefore, this work fails to reject the null hypothesis that generator set size impacts reliability, but additional research is required to quantitatively support the assumption that this model is generalizable to different size generator sets.

HYPOTHESES H_{o4} TEST RESULTS

Regression analysis suggests a statistically significant relationship between the log of emergency diesel-electric generator set age $\log(T_{age})$ and reliability of $\beta_{age} = -1.476$, with a significance of $p = 0.000018$, and thus the null hypothesis is rejected. This relationship exhibits a Weibull base hazard rate. The shape parameter m_{age} of the Weibull base hazard rate can be calculated as $m_{age} = \beta_{age} + 1 = -0.476$. The negative coefficient suggests a relationship of infant mortality and higher reliability of older generator sets. This predictor was the most significant predictor in the model by two orders of magnitude and persisted with high significance in nearly every interim model suggesting a low level of conflation with other predictors and a high indication of independence.

HYPOTHESES H_{o5} TEST RESULTS

Regression analysis suggests a statistically significant relationship between the log of emergency diesel-electric generator set cumulative run hours $\log(T_{rtfv})$ and reliability of $\beta_{rtfv} = 0.700$, with a significance of $p = 0.0067$, and thus the null hypothesis is rejected. This relationship exhibits a Weibull base hazard rate. The shape parameter m_{rtfv} of the Weibull distribution can be calculated as $m_{rtfv} = \beta_{rtfv} + 1 = 1.700$. The positive coefficient greater than one suggests a wear-out distribution. Although the coefficients for age and run hours were offsetting for much of the sample population, statistically significant relationships between reliability and age and reliability and run hours persisted in interim models even when only one or the other predictor was present, indicating the offset coefficients are a result of different effects and not conflation.

HYPOTHESES H_{o6} TEST RESULTS

Generator set load presented a challenge to hypothesis testing, as x_L was significant in interim models when included as the only time-independent predictor, but lost significance when paired with the other most significant predictors, x_{16}, x_{12}, and x_{src}. The coefficients suggest conflation of generator set load x_L with x_{12}, x_{16}, and x_{src}. The conflation is explained by data stratification within the sample population. Most sites in the sample population that performed regular tightness tests x_{16} and battery resistance or impedance tests x_{12} performed regular load tests at typical station load, but most sites in the sample population that did not regularly perform these maintenance steps performed no-load testing. Where load information was unavailable but generator sets were reported to receive weekly no-load tests, typical load x_L was approximated as 0 kW. Figure 14 shows a bubble chart with the data distribution comparing x_{16} and x_L and reflects the negative correlation of x_{16} and x_L.; the size of the bubbles in this chart reflect the histogram data. Figure 14 depicts 299 records, 251 of which are at $x_{16} = 7$ and $x_L = 0$ kW. All but two of the operational failures occurred in units receiving regular no-load testing at 0 kW load and no tightness testing, but x_L values were biased by the assumption of 0 kW typical load. These units presumably supported load during utility power failures and should more accurately be reflected as operating at an average of something greater than 0 kW, but accurate load information was not reported for most of the units and thus is biased in the methodology. While x_L is one of the most statistically significant results, the existence of bias and lower quality data suggests these results may be a Type 1 error. For this reason, x_L was excluded from the final model and the null hypothesis test deemed to be a failure to reject.

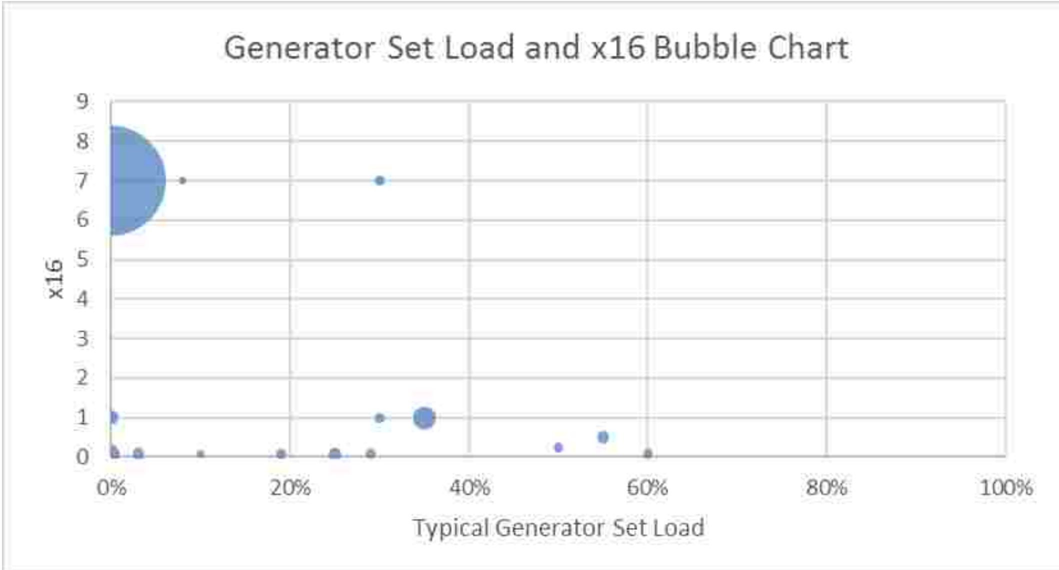

Figure 14. Generator set load and x_{16} bubble chart. The size of the bubble reflects the number of records at each axis point.

HYPOTHESES H_{o7} TEST RESULTS

The null hypothesis for training H_{o7} was rejected as the training predictors x_{b23} and x_{b24} were both significant in interim models for staff subject-matter experts with a significance of $p = 0.078$, which meets the $\alpha \leq 0.10$ criteria for significance. There was no significance in the regression models distinguishing between staff collateral duty and service visit subject-matter expert (contractor) test and maintenance. These results were highly conflated with maintenance predictors x_{12}, x_{13}, x_{16}, x_{17}, and x_{18} and predictors $k_{W,}$ x_l and x_{src}, making individual analysis difficult.

The training predictors x_{bi} were intended by Fehr (2014) for inclusion in the model only as interaction terms to determine of the training level of technicians performing various test and maintenance actions had significance. No training modifier x_{bi} was found to have any significance in any interim model as an interaction with a test or maintenance predictor x_i;

however, the aggregated training predictor for maintenance x_{b23} and the aggregated training predictor for tests x_{b24} were both unexpectedly found to have significance in interim models as standalone predictors.

Predictors x_{b23} and x_{b24} were factors with three categories: staff subject-matter expert, staff collateral duty, and subject-matter expert service visit. The latter, subject-matter expert service visits, were all third-party contractors in the sample population. There was no statistically significant distinction found between staff collateral duty and service visit subject-matter expert, but there was a statistically significant difference between these two factors and staff subject-matter expert. A new investigatory predictor x_{b25} was created as a factor with two categories: staff subject-matter expert, and not staff subject-matter expert. This predictor was found to have higher significance and higher model quality than either x_{b23} or x_{b24}. Predictor x_{b25} was also found to be nearly fully conflated with electrical tightness test predictor x_{16}, as only sites with staff subject-matter experts performed this maintenance action, and the periodicity at which sites performed this action appeared to correlate not simply with this action, but the intensity of maintenance and testing periodicity in general. This hidden correlation with maintenance rigor may explain why predictor x_{16} was found to have such a high significance despite no reported failures in the sample population related to electrical tightness. As the presence of maintenance personnel is meaningless without action, maintenance predictor x_{16} must still represent maintenance actions, but it also appears to be a better analogue for training than any of the training predictors included in the study.

HYPOTHESES H_{o8} TEST RESULTS

The analysis of the sample population suggests no statistically significant relationship between emergency diesel-electric generator set make (M_{AKE}) or model (M_{ODEL}) factors, with $p > 0.10$ in all interim regression models, resulting in a failure to reject the null hypothesis. The complexity of the emergency diesel-electric generator set market made this criterion difficult to evaluate as different manufacturers outsource engines and generators from different companies utilizing different engine manufacturers for different model units. Different makes often source the same engine or generator, and mergers and acquisitions and licensing agreements have blended different makes under different marques. While many technicians offered unsolicited opinions on various marques during this research, no statistically significant difference was found between any make or model. However, this lack of statistically significant difference between makes or models may represent Type II error and does not prove there is no difference in reliability between different makes or models, only that this work failed to reject the null hypothesis. When all makes and models were condensed into nine marques, Caterpillar, Cummins, Onan, MTU, Empire, Generac, SDMO, Olympian, and Kohler, little significance was found.

The regression, included in Appendix B, compared each marque to Caterpillar, but only two marques were found to have statistically significant differences, Olympian and Onan, with $p = 0.075$ and $p = 0.085$ respectively, both with positive coefficients indicating reduced reliability compared to Caterpillar. Olympian is a Caterpillar brand of 30kW-200kW generator sets manufactured at different times by Caterpillar and under license by Generac (Generac Corporation v. Caterpillar, 1999), so there may be some conflation here due to either generator size or size stratification in the sample population. Onan was bought by Cummins in 1986 and

Onan products are produced by Cummins Power Generation (Cummins, 2017). Cummins manufactures small diesel-electric generator sets under the Onan name, as well as large generators coupled to Cummins-branded engines. Several larger generator sets in the sample population were recorded as Onan, but may be referring to Onan generators coupled to Cummins engines, so it remains difficult to determine if Onan performance differs significantly from the others.

Combining Olympian with Caterpillar and Onan with Cummins and re-running the regression resulted in MTU becoming statistically significant with a negative coefficient and $p = 0.036$, a marked change from the significance of $p = 0.427$ in the regression when Olympian and Onan were considered separate marques. Like Caterpillar and Cummins, MTU owns multiple makes including Detroit Diesel and Katolight, which were included under the MTU marque for this analysis. The large change between the two regressions casts doubt on whether MTU is more reliable than the other marques or the second result is Type I error.

Given the differences between the two regressions and potential for Type I error through random chance, the significant regression results for Olympian, Onan, and MTU are insufficient evidence to reject the null hypothesis or to void the assumption that this research is generalizable across many different makes and models of modern low-emissions emergency diesel-electric generator sets.

If this finding is a Type II error and there is a difference between the reliability of different makes, it's likely modest, as large differences in reliability should have been detected by this methodology. There is also potential for bias between makes related to market share of different makes in different size ranges and applications, and the potential for increased managerial attentiveness and prioritization on large premium units as compared to inexpensive smaller units. Additional research is required to validate the assumption of generalizability and determine if there is a relationship between generator make and reliability.

DISCUSSION

By regression analysis, the reliability of emergency diesel-electric generator sets was found to be representable by a parametric proportional hazards model with exponential and Weibull base hazard rates. This model includes an exponential base hazard rate λ, an infant-mortality Weibull base hazard rate based on generator set age $\log(T_{age})$, a wear-out Weibull base hazard rate based on generator set cumulative chronometer run hours $\log(T_{rtfv})$, and a covariate predictor x_{16} that regression and analysis suggests to be an analogue represents the general intensity of maintenance and testing.

Five of the twenty-two test and maintenance predictors were found to have significance in various interim regression models included in Appendix B, but these test and maintenance predictors were found to be highly conflated with each other and with generator set size and load. These seven predictors attained offsetting coefficients and lost significance when applied in any combination in models, indicating a level of conflation. Much of this conflation was found to be due to common stratifications and subpopulation homogeneity in the sample population. Further investigation into the sample population revealed the highest significance maintenance and test predictor, electrical tightness testing x_{16}, appear to be related to the general intensity and rigor of maintenance. The most reliable generator sets in the sample population were those maintained by staffs of full-time generator technicians that performed intensive maintenance like x_{16} at very frequent intervals. Predictor x_{16} appears to have become significant in the model simply because it was one of the few predictors with a large variance within the sample population and thus was more sensitive to very small differences in generator set reliability between the subpopulations. The regression results and analysis suggest that predictor x_{16} is highly representative of the general intensity of maintenance within the sample population.

The regression analysis of emergency diesel-electric generator set data revealed statistically significant relationships with age and run hours suggesting generators become more reliable with age, but less reliable at higher run hours. The relationship between reliability and generator set age has the strongest significance relationship in the model with p = 0.000018. The shape parameter m_{age} of the Weibull base hazard rate can be calculated as $m_{age} = \beta_{age} + 1 = -0.476$. The negative coefficient suggests a relationship of infant mortality and higher reliability of older generator sets. That generators become more reliable with age is at first counter-intuitive, but this funding seems to reflect the statistical impact of infant mortality and shows that well-maintained generator sets do not exhibit wear-out characteristics based on age alone. This research suggests the operational life of older generator sets with good maintenance and repair support may be extended with little risk. However, reliability is not the only parameter that matters, and operational availability of units is a function of mean time to repair. Retirement of units of any age may be prudent when spare parts and maintenance support are no longer available or units incur excessive time to repair.

The model also reflects generator sets with greater run-hours as being less reliable, which follows logic, as machinery of this sort exhibits wear characteristics and nearly identical generator sets in prime-power generation applications are well known to wear out from use. The mean annual run-time of generator sets in this study was 31.2 hours per year, which results in an interesting parity; the infant mortality of $\log(T_{age})$ was found to dominate the model the first five years in service, but is largely offset by $\log(T_{rtfv})$ for much of the remaining life. The relationship of cumulative run hours to reliability was found to be highly statistically significant with $p = 0.0067$.

The shape parameter m_{rtfv} of the $\log(T_{rtfv})$ Weibull base hazard rate can be calculated as $m_{rtfv} = \beta_{rtfv} + 1 = 1.700$. The positive coefficient suggests a wear-out relationship which generators of greater run-hours exhibiting reduced reliability. The mode of this pattern is important for failure prediction and determination in the statistically optimal retirement date for units with high hours, but the mode unfortunately cannot be determined directly from the regression results, as the λ value from the regression represents the sum of the contributions from the base exponential distribution, the $\log(T_{age})$ Weibull parameters and $\log(T_{trfv})$ Weibull parameters. Without the ability to extract the specific λ contribution of each piece, the predictive power of this portion of the model is limited. One manufacturer representative who did not wish to be cited said his company expects their longest-lived model of emergency generator set to be completely overhauled every 11,000 hours and replaced after 22,000 hours of use. The typical generator set in this study would take 32 years to accumulate 1000 hours and even the most used generator set in this study only accumulated 4680 hours. Only generator sets in areas with exceptionally poor-quality utility power or operated in duty cycles other than emergency duty could experience this many hours before replacement. It would take 110 years for an emergency generator set to reach 22,000 hours at the ISO 8528-1 (2005) maximum rated 200 hours per year.

No statistically significant relationships were found in the data for make, model, size, or load, but insufficient power was realized to conclude the absence of these relationships due to the limited amount of data. Likewise, confirmation could not be confidently made that there were no interactions between these predictors and other maintenance and test predictors. Some of the interim models suggest there may be some dependencies; however, any particularly strong dependencies would be expected to be clear despite the limited data. Further, the absence of such results in the model and the large numbers of predictors involved suggest the significant results in some interim models were more likely from Type I error than actual relationships. Some terms were included in the regression to avoid reliance upon testable *a priori* assumptions, but similarities in technology and construction of generator sets in this size range permits the *a priori* assumption to be made that the reaction to maintenance will be similar for all makes and models of generator sets of the size range in this study. This assumption permits treating the sample population as a single population and permits drawing additional conclusions from the data without the need for make, model, or size covariates.

When the parametric model results were compared to the sample MTBF from the data, the model results appeared to slightly over-estimate the effect on reliability of maintenance, with the model over-predicting the reliability of units receiving high intensity maintenance by 0.6% but under-predicting the reliability of units receiving lower intensity maintenance by 0.6%. The difference between the sample MTBF and model suggests there may be influence from other sources not included in this model. This was unsurprising considering the amount of conflation between x_{16} and other predictors within the sample population and the imprecision of x_{16} in representing all test and maintenance for these generator sets. The significance of this predictor in the statistical models and effect size strongly suggests a relationship between emergency diesel-electric generator set reliability and the maintenance and test intensity. While there may be some error and uncertainty in the model results, there was enough evidence conclude that maintenance and testing have an impact on generator set reliability. This knowledge can be used to aid managers in determining the appropriate intensity of maintenance and testing for their generator sets and for systems engineers calculating emergency power system reliability.

This research initially sought to develop a general model to quantitatively determine the optimal level of test and maintenance periodicity for emergency diesel-electric generator sets. More generally speaking, however, the model was a means to an end, and the true intention was to develop recommendations for managers to improve the performance of emergency power systems supporting critical operations facilities and other facilities requiring highly reliable emergency power. The optimal periodicity of NFPA 110 (2016) maintenance and test predictors cannot be precisely determined from this data, but the significant relationship in this data strongly suggests that more intensive maintenance and testing yields higher reliability than less intensive maintenance and testing. The sample subpopulation with full-time subject-matter experts on staff and the highest intensity maintenance had the highest operational reliability and an observed failure rate of 1.647% per year. The sample population receiving outsourced maintenance also had high operational reliability, much higher than previous studies (Hale & Arno, 2009; IEEE 493-2007), but the observed failure rate for the lower intensity subpopulation was about 2.85 times higher than the higher intensity subpopulation, at 4.695% per year. The new knowledge from this work will help managers make staffing and maintenance and testing decisions. There is a roughly order of magnitude cost difference going from outsourced maintenance to a full-time staff, so the decision is not necessarily an easy one.

One concern mentioned multiple times by technicians during data acquisition and by sources such as Loehlein (2007) is that frequent no-load tests damage engines and cause wet-stacking, a condition where cylinder and exhaust temperatures are insufficient to achieve complete diesel combustion, resulting in deposits in the cylinder and a build-up of unburnt diesel fuel in the exhaust system. The only record was reviewed during data acquisition with a failure attributed to wet-stacking was a 10kW unit at a police station that did not receive regular load testing, only weekly no-load testing, and may have had another attributable cause such as very low load during emergency operation. This was one of the smallest generator sets in the sample population and no other generator sets in this study were reported to have any failures or serious problems related to wet stacking or low-load operation, although a few units not regularly tested under load were found during maintenance to show signs that wet stacking was occurring. NFPA 110 (2016) recommends, and in some cases requires, monthly load tests with ATS transfer and annual load-bank tests for units that operate below 30% load and do not achieve adequate exhaust temperature during these tests. Tufte (2014) recommended limiting loads below 30% to no more than 8 hours before loading the generator set at minimum 50% load. Tufte (2014) found that while older generator sets and early low-emissions units were highly susceptible to wet-stacking, the newest generation of generator sets can run at much lower levels for longer periods before running into wet stacking or similar low-load problems and can run up to 8 hours at below 10% load and 24 hours below 30% load before running above 50% load is necessary. None of the units in the sample population receiving monthly operational load tests were reported to have any issues related to wet stacking.

NFPA 110 (2016) does not require regular no-load tests, and Cummins recommends no-load tests be held to a minimum (Loehlein, 2007). However, technicians at the Cummins factory could not recall any of their newest engines suffering from wet-stacking when asked, and no-load tests are required by Caterpillar to perform certain preventative maintenance checks (Caterpillar, 1997; Caterpillar, 2010a; Caterpillar, 2010b; Caterpillar, 2010c; Caterpillar, 2010d). There are some negative effects on diesel engines from running at low load, but these negative effects are largely neutralized by running periodically at higher loads (Tufte, 2014). Additionally, large numbers of hidden failures were discovered during regular no-load tests that likely would have resulted in operational failures had those no-load tests not been conducted. Thirty-minute weekly no-load tests will accumulate twenty hours per year in addition to the monthly load tests required by NFPA 110 for certain applications. Similar increases in run hours were found by the regression to result in a statistical increase in failure rate. However, the relationship found between run-hours and reliability is logarithmic and the hazard rate increases only slightly with an additional twenty hours of run time per year. This is a very small impact compared to the contribution of maintenance and test intensity on overall generator set reliability. This analysis suggests the benefits of no-load testing are greater than the incurred wear or negative effects, and a test plan with regular weekly no-load tests and monthly load tests with ATS transfer will result in an overall increase in generator set reliability.

Another important consideration is that tests are intended to detect hidden failures so they can be quickly corrected, but tests must be monitored to be useful. Numerous incidents were recorded in logs of automatic no-load tests failing but going unnoticed for weeks or months because nobody noticed or reported the test failure. In one instance, a generator set in a remote portion of a university campus suffered a controller failure and ran for 23 days at idle until it ran dry of oil and catastrophically seized, requiring replacement. If the scheduled weekly testing of this unit had been monitored, it would likely still be in service. In other instances, alarms reported to technicians could not be replicated or troubleshot, and the result was that multiple failures occurred before the prudent corrective maintenance action could be completed. If technicians had been on-site during the first test that exhibited problems, corrective maintenance actions may have been more quickly taken and problems corrected. Other issues such as the potential for oil and coolant leaks increase the risk of unmonitored testing causing environmental problems. Unmonitored automatic tests can also fail to detect frequency oscillations (hunting), mild overheating, squealing belts, bearing noise, or other indications of pending failures that would not necessarily result in an alarm. The benefit of no-load tests is reduced and risk is increased if exercisers automatically run no-load tests without active monitoring. All tests should be actively monitored to achieve maximum benefit.

This study did not directly investigate operational availability, but it was recognized that no-load test failure alarms going un-noticed for months would yield much lower availability than no-load test failures that received prompt response. The data acquired for this study primarily consisted of logs recorded by subject-matter experts and did not include information on local monitoring of no-load tests, making it difficult to determine how quickly automatic no-load test failures were detected and responded to with appropriate corrective action. Some of the higher failure rates of the units receiving automatic no-load tests and less frequent visits by servicing personnel may have been biased by lack of monitoring. Sites were recorded as performing no-load tests if the exerciser was configured, but if nobody was monitoring those tests, much of the benefits of these tests were lost. As these sites were included in the regression as if they were performing regular no-load tests, the bias would result in a model that underestimates the reliability of generator sets receiving regular monitored no-load tests.

GENERALIZABILITY

The purpose of this research was to develop a general model for determining the reliability and optimal test and maintenance periodicities for emergency diesel-electric generator sets supporting critical operations facilities and other facilities requiring highly reliable emergency power, but with the intent of creating a general model applicable to all emergency diesel generator sets between 60 kW and 2.5 MW electrical capacity and of the characteristics described in the system subcomponents section of the introduction. As data was readily available for a number of units in the 10 kW to 60 kW size range and the decision to delimit to 60 kW was arbitrary, the regression analysis was expanded to cover units 10 kW to 2.5 MW. No statistically significant evidence was found during the regression analysis of the sample population to indicate a lack of generalizability of this model across this size range.

The pre-hoc power analysis predicted that a sample size of 1142 standby-years including 141 operational failures would achieve adequate power at $\alpha = 0.1$ and $\beta = 0.2$. The sample population included 1281 standby-years of generator set data from 239 generator sets, capturing 58 operational failures in 40,161 run-hours of operation. As the sample population proved to be more reliable than anticipated during pre-hoc power analysis, insufficient operational failures were observed to achieve adequate power at $\alpha = 0.1$ and $\beta = 0.2$. No statistically significant relationship was found between reliability and generator set make, model, or size. However, due to data failing to achieve the predicted power for this study, the absence of evidence of a relationship between reliability and generator set make, model, or size is insufficient statistical evidence to conclude the absence of such relationships without an unacceptable risk of Type 2 error.

Other methods were attempted to test generalizability of the model within the sample population including sensitivity analysis. Sensitivity analysis removing all the small units from the sample population revealed little change to regression results. The sensitivity analysis tested various intentionally biased models where portions of the population were removed, new failures were introduced, and failures were removed. None of the sensitivity analysis tests resulted in large changes to any significant model parameter, which suggested relationships in the final model are robust and the model is insensitive to noise and random chance in the data.

Present maintenance guidance (NFPA 110, 2016) does not make distinctions based on make, model, or size and there are reasonable arguments for this assumption supported by many technicians (Walbolt, 2010). Fehr (2014) made a general *a priori* assumption that generator sets of make, model, and size within the delimitations of this research represent a common population with a common response to maintenance. Given this rationale and the lack of evidence to the contrary, it is appropriate to continue the *a priori* assumption of generalizability of make, model, and size from the Fehr (2014) framework and extend the generalizability of this model to all emergency diesel generator sets between 10 kW and 2.5 MW electrical capacity with of the characteristics described in Chapter I. This can also be extended to apply to spark ignition natural gas-powered generator sets based on modified diesel engines as these engines are identical in nearly every way to diesel engines, and no evidence was found in this study that natural gas and diesel engines responded differently to NFPA 110 maintenance.

LIMITATIONS

The Cox (1972)-based regression failed to calculate the specific contribution of each of the twenty-two maintenance and test predictors. Although 1281 standby-years of generator set data from 239 generator sets was acquired for this data, only 58 operational failures were captured, fewer than the 141 failures anticipated to yield statistically significant results at $\alpha = 0.10$ and $\beta = 0.20$ by Hsieh and Lavori (2007) methods. This was likely exacerbated by a lack of diversity in the sample population which consisted largely of Department of Defense and contractor-maintained commercial units being maintained in accordance with Department of Defense policies and standard commercial practices. The sample population included very few units that varied significantly from NFPA 110 (2016) recommendations and included very few units that received no regular tests or poor maintenance, data essential to teasing significant results from generator sets that have been found to be far more reliable in practice than previous research suggested. The lack of available data on poorly-maintained units further reduced the realized normalized covariate variance σ^2 used in the Hsieh and Lavori (2007) power calculations to estimate the amount of data required for this research.

The Fehr (2014) framework relied upon several common characteristics of emergency diesel-electric generator sets to work including a large data pool, diversity of maintenance practices, and good records. While there was some diversity of maintenance practices, the sample population included a large degree of stratification and conflation, making it difficult to discern statistically significant results with the available data. To better differentiate the impact of test and maintenance periodicity, more data must also be drawn from units of average and poor-quality maintenance.

Insufficient data prevented the support of generalizability of this model by statistical power analysis. Make, model, and size were not found to have statistical significance, but the risk of Type II error exceeds the threshold of $\beta = 0.20$ due to the higher than expected reliability of the sample population and relative homogeneity of maintenance. While the reliability model developed herein is believed to be generalizable to all models of diesel-electric generator sets within the size range of 10 kW to 2.5 MW in emergency service, this belief requires an *a priori* assumption.

Another limitation of this research is that it was restricted to NFPA 110 (2016) recommended maintenance with periodicity of one year or less. Maintenance such as engine overhauls, battery replacement, thermostat replacement, and block heater replacement were not investigated, but numerous failures related to these items were discovered during data acquisition. The choice to focus on maintenance with recommended periodicity of one year or less was made to reduce the scope of this study to a reasonable length observation period and is consistent with NFPA 110 (2016) recommended practice, but as such, this study is unable to measure the effectiveness of maintenance actions with longer periodicity. Future research should be extended to include such additional maintenance items.

AVAILABILITY CALCULATIONS

A secondary objective of this framework was to use the data to calculate the availability of generator sets with varying level of maintenance for inclusion into the US Army Corps of Engineers Power Reliability Enhancement Program (PREP) database and future editions of TM 5-698-5 (2006), NFPA70B (2016), and IEEE 493 (2007). This would allow engineers to better design emergency generator systems to meet availability requirements and allow managers to make well informed risk decisions when planning maintenance. This research has produced the reliability model in Equation 14, but was unable to produce calculations of inherit availability A_i or operational availability $A_{o,}$ as repair time data was not collected, nor was data collected on unavailability periods due to scheduled preventative maintenance or corrective maintenance following test failures. While the operational reliability statistics determined by this research can be used to estimate operational reliability based on generator set age, run hours, and maintenance for the purposes of engineering calculations, similar estimations of availability will require additional research.

Previous research estimated inherent availability for emergency diesel-electric generator sets as A_i = 0.999712 for well-maintained generator sets (Fehr, 2014) and A_i = 0.9974 for a general population of generator sets including well, average, and poorly-maintained units (TM 5-

698-5, 2006). This order of magnitude disparity between prior research findings on availability and these new research findings on reliability suggest the availability of well-maintained units and poorly-maintained units may differ by a large degree and use of PREP availability numbers is likely highly conservative for well-maintained generator sets.

Mean time to repair (MTTR) will differ between sites as well. Previous research estimated MTTR as 18.3 hours (TM 6-698-5, 2006). This MTTR was pooled from an average of well, average, and poorly-maintained generator sets. PREP did not publish information on staffing at these facilities, but it can be inferred that a site with full time subject-matter experts on staff and well-stocked inventories of spare parts on-site will have a shorter mean time to repair than sites that experience contracting, travel, and shipping delays. Sites that fail to monitor no-load generator tests might have effective MTTR measured in weeks as failures may not be noticed until the next scheduled maintenance.

Another complication is the difficulty in determining the time between the occurrence of hidden failures and subsequent repair as only the time of discovery of the failure is typically known, not the point where the failure occurred. For example, it may not be known when a starter battery died, for instance, only that the unit did not start the next time it was attempted.

OTHER APPLICATIONS

The survival regression technique discussed herein using a Cox (1972) proportional hazards model to simultaneously combine multiple exponential and Weibull relationships as predictors is believed to have a wide number of applications for describing complex machinery and other populations exhibiting survival distributions based upon multiple independent base hazard rates. These statistical modeling methods can be calculated using standard *coxreg* and *coxph* statistical modeling packages in the *R* libraries *eha* and *survival* (Broström, 2011) or other statistical packages that use Cox regression algorithms. This survival regression technique could be directly applied to empirical research into transformer statistical lifetime modeling similar to the simulated modeling of Zhou, Wang, and Li (2014), or to any number of complex machinery where the lack of such methods to account for time dependent properties necessitated assumptions of exponential relationships (Moubray, 1997; Hale & Arno, 2009). This type of model could potentially also be used to better represent populations presently represented by pure Weibull failure modes but that are also subject to unrelated random failures like lightning strikes and accidents.

The Fehr (2014) framework for this research has other potential uses as well. While this framework was developed to provide a means of determining optimal maintenance of emergency diesel-electric generator systems, the Fehr (2014) methodology would apply equally well to create general models for other high-reliability systems that, due to a combination of low failure rates and the censorship actions of preventative maintenance, are difficult to analyze with conventional failure modes and effects analysis techniques. This framework could be adapted to analyze system subcomponents as well as whole systems. Potential applications of these techniques include uninterruptible power supply systems; heating, ventilation, and air conditioning systems; maritime shipping industry systems; military applications; aviation industry systems, and others.

This framework is most effective where the data pool is large compared to the failure rate, where maintenance practices vary, where training of maintenance crews vary, where good records are maintained in a consistent fashion, and where the units have been in service long enough to develop a history. For example, this framework may work well to compare failure rates of similar models of military aircraft in service at different organizations that receive different maintenance practices or utilize different training. This includes widely produced and internationally sold aircraft such as the General Dynamics F-16 Fighting Falcon, the Sukhoi Su-27, Sikorsky UH-60 Black Hawk, and unmanned aerial systems (UAS) such as the General Atomics MQ-1 Predator and Boeing Insitu ScanEagle.

VI. CONCLUSIONS

Diesel-electric generator sets in emergency standby duty receiving average or better levels of maintenance were found to have significantly higher operational reliability than previous research (Fehr, 2014; TM 5-698-5, 2006) indicated. The Fehr (2014) framework was successfully applied to perform Cox (1972) regression and piecewise constant hazard model analysis on 1281 standby-years of generator set data from 239 generator sets with 58 operational failures in 40,161 run-hours to develop a parametric Cox proportional hazards model with exponential and Weibull base hazard rates representing generator set reliability as a function of age, run hours, and maintenance intensity. This model can be used to estimate the reliability of generator sets of various age, run hours, and maintenance intensity.

This research found that generator sets exhibited characteristics of multiple survival distributions including exponential random failure, Weibull wear-out, and Weibull infant mortality. The regression model found generator sets in this study suffered elevated failure rates in the first few years after installation but become more reliable as they aged. This result was unexpected, but was a highly statistically significant finding with $p = 0.000018$. The model also found generator sets became less reliable as cumulative run hours increased, offsetting much of the age-related increase in reliability for units near the near 31.2 hours annual run hours in the sample population. Statistical significance for run hours was also very high at $p = 0.0067$. These highly significant results provide confidence that the Weibull $\log(T_{age})$ and $\log(T_{rtfv})$ relationships are representative of the population and not a result of Type I error. The regression failed to return information that will allow precise estimates of the optimal periodicity of all NFPA 110 (2016) recommended maintenance and test predictors, but the regression found the intensity of maintenance has a very strong effect on generator set reliability with a significance of p = 0.047. This research found that common commercial maintenance plans with weekly no-load generator tests and monthly, bi-monthly, or quarterly maintenance achieve mean annual reliability levels higher than 96% for well-established units with typical run hours. This modeling further shows that sites with full-time staffs of subject-matter experts and highly intense maintenance plans can reduce the rate of operational failures by another 66-80%, achieving greater than 99% annual operational availability for similar generator sets.

The maintenance predictor for check electrical tightness x_{16} was found to function as an analogue for the general intensity of maintenance and testing for the final model, and it was the most significant maintenance or test predictor at p = 0.0067. Several other maintenance predictors showed some statistical significance in some interim models, but the individual contribution of each of these predictors could not be determined due to conflation and stratification in the data, and statistical significance was lost when these predictors were included together in any combination in other interim models. The predictor x_{16} was found to be heavily conflated with predictors for the personnel conducting the most frequent touch maintenance x_{b23} (staff subject-matter expert vs. subject-matter expert site visit), data source x_{src}, and maintenance periodicities for battery resistance or impedance test x_{12}, clean unit exterior x_{13}, engine intensive maintenance x_{17} and generator electrical intensive maintenance x_{18}. Interpretation of this finding, guided by the statistical significance of each predictor and other associated knowledge, is that x_{16} represented not just checks for electrical tightness, but the general intensity of all maintenance. Sites in the sample population with large subject-matter expert staffs performing the shortest periodicity tests had the shortest periodicity of x_{16} while sites with contract maintenance did not perform maintenance x_{16} at all. Intensity cannot be quantified as specific periodicities of each test and maintenance predictors, but the relationship can be estimated by a categorization as average, high, and extremely high maintenance intensity represented respectively by x_{16}=7 years, x_{16}=1 year, and $x_{16}={}^{1}/_{12}$ year. In this context, average maintenance refers to generator sets receiving weekly automatic no-load tests and contractor site visits with a periodicity of one to six months. Highly intensive maintenance refers to sites adhering to NFPA 110 (2016) requirements and generator sets receiving weekly maintenance and monthly load tests with ATS transfer. Extremely high intensity maintenance refers to sites well exceeding NFPA 110 requirements with daily maintenance, weekly load tests, and intensive monthly maintenance. The only sites in this study receiving extremely high intensity maintenance were critical operations power system facilities with large full-time on-site staffs of generator technicians, electricians, and mechanics tasked exclusively with maintenance and operation of a small number of emergency diesel-electric generator sets and uninterruptible power supply systems.

No statistically significant evidence was found that contradicts the *a priori* assumption that generator set make, model, or size within the range in this study have no significant impact to reliability or the related assumption that this research can be generalized to all diesel-electric generator sets 10 kW to 2.5 MW in emergency duty operating fewer than 200 hours per year. It was sought to confirm this assumption quantitatively to increase confidence, but there was insufficient data to do so.

A regression technique combining exponential and Weibull distribution components was successfully used with standard Cox (1972) regression functions *coxreg* and *coxph* in the *Survival* package for *R*. This technique included time dependent covariate predictors for age and cumulative run hours in the regression analysis in the form log(T), which is mathematically equivalent to a Weibull base hazard rate. This technique permitted straightforward Cox regression of multiple exponential and Weibull relationships simultaneously. A weakness was found in the unavailability of parametric regression tools to complete the analysis; the *phreg* function in the *Survival* package in *R* accurately estimated regression coefficients but was unable to properly fit the model due to the inability of the function to treat time dependent variables as both static and time dependent in different portions of the algorithm. Utilizing piecewise constant hazard (PCH) modeling in *Microsoft Excel* to fit the model to empirical subpopulation reliability calculations permitted development of a parametric model, but this modelling could be improved with adaptation of R functions that can better handle time dependent variables. As logarithmic functions become asymptotic as the limit approaches zero, the PCH model was calculated with one-year cuts starting at year one.

The shape parameter of constituent Weibull distribution components within the combined model can be directly calculated from the β coefficients returned by the regression, but the scale parameter returned by the regression cannot be easily separated from the scalar h_0 which contains the product of the baseline exponential distribution λ and all other Weibull scale parameters in the model. Overall model performance is unimpacted, as h_0 still contains all these coefficients, but not knowing what the scale parameter values are prevents direct calculation of the Weibull distribution modes or other predictions that would be useful for better understanding system performance over time.

RECOMMENDATIONS

The reliability of diesel-electric generator sets in emergency standby applications should be calculated by engineers and managers using Equation 19 as a function of generator set age, run hours, and maintenance intensity, where maintenance intensity is 7 for average, 1 for high, and 0.0833 for very high. Reliability tables are provided in Appendix C for average, high, and extremely high intensity maintenance of generator sets at different ages and run hours. The reliability tables in Appendix C reflects the findings of this research that generator reliability is not static over time, and that new generator sets are statistically less reliable than generator sets that have been in service for ten or more years.

$$h(t) = \frac{0.00885e^{(Maint\ Intensity)}(Run\ hours)^{0.7}}{(Age)^{1.476}} \tag{19}$$

The model developed by this research predicts generator sets will become more reliable as they age, but this does not necessarily mean generator sets should remain in service indefinitely, as reduced availability of spare parts and qualified maintenance personnel for older units may increase time to repair and decrease operational availability to unacceptable levels and necessitate generator set replacement based on obsolescence. Managers should consider extending the life of generator sets with low cumulative run hours and plan replacement based on criteria other than just age. Other options such as refurbishment or replacement of obsolete ancillary components may be the most optimal solution for some aging generator sets. Changing emissions requirements or other local requirements may also play a role.

One of the questions originally driving this research was whether no-load tests are beneficial or harmful to generator set reliability from the additional wear from testing and risk of wet stacking. Test failure data was not specifically investigated in this research, but a review of a subpopulation receiving weekly no-load tests revealed 82% of total failures were found during testing, all of which would likely have resulted in an operational failure during the next outage had those tests not occurred. The only evidence of wet-stacking reported in the generator sets of this sample population were in a small number of units that did not receive any regular load tests or transfer tests. Only one unit was reported to have suffered a failure related to wet stacking, but this unit may have had another attributable cause. No generator sets also conducting monthly transfer load tests were reported to have any evidence or symptoms of wet stacking. Other research has found the effects of wet stacking are minimal in the latest generation of generator sets and mitigated on all diesel engines by regularly running under load (Tufte, 2014). Conducting weekly no-load tests in addition to monthly transfer load tests will incur an addition twenty run-hours of no-load tests per year, a rate that has only a small negative impact to long-term generator reliability due to the logarithmic relationship between cumulative chronometer run hours and generator reliability.

All managers interested in improving emergency generator set reliability should conduct weekly monitored no-transfer tests and monthly load tests with automatic transfer switch transfer. The no-transfer tests may be no-load tests or load bank tests for sites configured with load banks. These tests should not be conducted by automatic exerciser, but should be manually initiated and monitored by qualified personnel to reduce the time to repair and increase the likelihood that problems will be discovered, and corrective actions taken. Even monitoring of tests by minimally trained site personnel has advantages over unmonitored automatic tests. Such tests also give personnel an opportunity to gain and maintain proficiency in generator set operation. NFPA 110 (2016) should be updated to require weekly no-transfer tests for legally required units and should clarify that mandatory tests must be monitored. All generator sets receiving regular no-load tests should also receive monthly load tests, even if not required to by NFPA 110. Typical generator set load during normal operation should be maintained above 10%

of rated load for the newest units and above 30% of rated load for older units, as it is important to ensure damage from low-load operation does not occur. Periods of up to eight hours at low loads should be followed by loading the units to at least 50% (Tufte, 2014). Legally required generator sets must be loaded to a minimum 30% per NFPA110 (2016) during the required monthly load test with ATS transfer.

Generator sets utilizing contract subject-matter expert site visits yielded very high reliability, but the highest reliability levels were at facilities with staff subject-matter experts conducting generator set maintenance and tests at much more frequent intervals than required by NFPA 110 (2016). Generator sets receiving such extremely high intensity maintenance were found to have 66-80% fewer operational failures than sites receiving average levels of maintenance. While full time staffs may not be financially viable or justifiable for all facilities, full time staffs or highly intensive contracts should be considered for sites where failure has catastrophic consequences.

The United States Army Corp of Engineers' Power Reliability Enhancement Program (PREP) should update TM 5-698-5 (2006), NFPA should updated NFPA 70B (2016), and IEEE should update IEEE 493-2007 (2007) to reflect the impact of maintenance intensity, age, and run hours on generator set operational reliability. Additional research is needed to better determine emergency generator set availability based on maintenance, testing, age, and run hours. Additional investigation should also be made into the applicability of the Fehr (2014) methodology for the reliability of uninterruptible power supply systems; heating, ventilation, and air conditioning systems; maritime shipping industry systems; military applications; aviation industry systems; and other complex systems which share similar traits of maintenance-censored failure data and diverse maintenance practices for similar or identical equipment.

Survival models using a Cox proportional hazards model with exponential and Weibull base hazard rates were developed with relative ease by software implementing Cox regression techniques and described emergency diesel-electric generator sets better than more traditional models assuming a single survival distribution. This survival distribution technique including Weibull components as logarithmic covariate predictors should be considered for other systems whose overall survival distribution may be best represented as a combination of multiple independent distributions. Better software tools should be developed to enable better development of non-parametric regression models with time dependent covariates.

Lastly, the research herein should be continued until sufficient data is acquired to quantitatively determine specific optimal test and maintenance periodicities. This research should be expanded to contain test failures and times to repair so that availability can be investigated with similar rigor.

BIBLIOGRAPHY

Alion Science and Technology. (2006). *CMMS selection and standby diesel generator maintenance study.* Retrieved from Alion Science and Technology, 1750 Tysons Boulevard Suite 1300 McLean, VA 22102.

Amorim, L., Cai, J. (2015). Modelling recurrent events: A tutorial for analysis in epidemiology. *International Journal of Epidemiology, 2015.* doi: 10.1093/ije/dyu222

Andersen, P., Gill, R. (1982). Cox's regression model for counting processes: A large sample study. *Ann Stat, 10*. Retrieved from http://www.jstor.org/stable/2240714

Broström, G. (2011). *Event history analysis with R.* Boca Raton, Florida: CRC Press.

Caterpillar. (1997). *Maintenance management schedules/recommended preventative maintenance schedule for standby generator sets* (SEBU6042-04). Retrieved from: http://www.thomsonequipment.com/Diesel%20Plant/CATERPILLAR%20MAINTENANCE%20MANAGEMENT%20SCHEDULES%20STANDBY%20GENERAT.pdf

Caterpillar. (2010a). *Maintenance intervals, operational and maintenance manual excerpt, C15 generator set* (SEBU7909-09). Retrieved from: http://safety.cat.com/cda/files/2450470/7/SEBU7909-09%20M.pdf

Caterpillar. (2010b). *Maintenance intervals, operational and maintenance manual excerpt, C27 and C32 generator sets* (SEBU8088-10). Retrieved from: http://safety.cat.com/cda/files/2450492/7/SEBU8088-10%20M.pdf

Caterpillar. (2010c). *Maintenance intervals, operational and maintenance manual excerpt, C175 generator sets* (SEBU8100-12). Retrieved from: https://safety.cat.com/cda/files/2450523/7/SEBU8100-12%20M.pdf

Caterpillar. (2010d). *Maintenance intervals, operational and maintenance manual* excerpt, *3500 generator sets* (SEBU7899-05). Retrieved from: http://safety.cat.com/cda/files/2450381/7/SEBU7509-11%20M.pdf

Caterpillar. (2012). *Maintenance intervals, operational and maintenance manual excerpt, C18 generator sets* (SEBU7898-13). Retrieved from: https://safety.cat.com/cda/files/2450591/7/SEBU7898-13%20M.pdf

Cohen, J., Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1987). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cox, D. (1972). Regression models and life tables. *Journal of the Royal Statistical Society Series B (with discussion), 34*, 187–220.

Crowther, D., Lancaster, G., (2008). *Research methods in management* (2nd ed.). London, UK: Routledge. Retrieved from https://books.google.com/books?id=0f8rBgAAQBAJ&pg=PA79&dq=lancaster+%22secondary+data%22&hl=en&sa=X&ved=0CCkQ6AEwAmoVChMI4L_hwOyTyQIVUy2ICh2PaAKf#v=onepage&q=lancaster%20%22secondary%20data%22&f=false

Cummins. (2017). *History of Cummins, Inc*. Retrieved from http://training.cummins-distributors.com/calendar/03742/about_us

Department of the Army. (2006). *Survey of reliability and availability information for power distribution, power generation, and heating, ventilating and air conditioning (HVAC) components for commercial, industrial, and utility installations.* (Technical Manual TM 5-698-5). Washington, DC: US Government Printing Office.

Devaney, M., Ram, A., Qiu, H., Lee, J. (2005). Preventing failures by mining maintenance logs with case-based reasoning. *Proceedings from MFPT-59: 59th Meeting of the Society for Machinery Failure Prevention Technology*.

Environmental Protection Agency. (2013). *Overview - the Clean Air Act amendments of 1990.* Retrieved from: http://epa.gov/oar/caa/caaa_overview.html

Environmental Protection Agency. (2016). *Reciprocating internal combustion engines (RICE).* https://www3.epa.gov/region1/rice/

Fehr, S. (2014). *Emergency diesel-electric generator set maintenance and test periodicity* (Masters Thesis). Retrieved from Old Dominion University, Norfolk VA.

Fehr, S., Cotter, T. (2014). Emergency diesel-electric generator set maintenance and test periodicity. Proceedings from ASEM 2014: *International Annual Conference*. Virginia Beach, VA.

Fleet Cyber Command. (2011). *Test and maintenance requirements for facilities mission support equipment* (Instruction COMFLTCYBERCOM/COMTENTHFLTINST 11311.1). Ft. Meade, MD: US Government Printing Office.

Gat, Y., Eisenbeis, P., (2000). Using maintenance records to forecast failures in water networks. *Urban Water 2 (2000)*. Retrieved from http://www.elsevier.com/locate/urbwat

Generac Corporation v. Caterpillar. 7 U.S. 97-1404. (1999). Retrieved from http://caselaw.findlaw.com/us-7th-circuit/1022151.html

Hale, P., & Arno, R. (2009). Operational and maintenance data collection. *IEEE Industry Applications Magazine, 15*(5), 21-4. doi: 10.1109/MIAS.2009.933400

Hoenig, J., Heisey, D. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *The American Statistician (55)*1.

Hosmer, D., Lemeshow, S., May, S. (2008). *Applied survival analysis: Regression modeling of time to event data, 2nd Edition.* New York: Wiley.

Hsieh, F., Lavori, P. (2000). Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. *Controlled Clinical Trials 21*.

Institute of Electronic and Electrical Engineers. (2007). *Design of reliable industrial and commercial power systems* (IEEE 493-2007 Redline Version). Retrieved from: http://ieeexplore.ieee.org/servlet/opac?punumber=6044682

Institute of Electronic and Electrical Engineers. (2010). *IEEE recommended practice for the maintenance of industrial and commercial power systems* (IEEE 3007.2-2010). Doi: 10.1109/IEEESTD.2010.5618906

International Organization for Standardization. (2005). *Reciprocating internal combustion engine driven alternating current generator sets – Part 1: Application, ratings and performance* (ISO 8528-1:2005). New York, NY: ISO.

JIE Operations Sponsor Group. (2014). *Joint information environment operations concept of operations.* Washington, DC: US Government Printing Office.

Joint Departments of the Army, the Navy, and the Air Force. (1995). *Facilities engineering electrical interior facilities.* (Technical Manual TM 5-683/NAVY NAVFAC MO-116/AFJMAN 32-1083). Washington, DC: US Government Printing Office.

Kalbfleisch, J., & Prentice, R. (2002). *The statistical analysis of failure time data, 2nd ed*. New York: Wiley.

Kelly, P., Lim, L. (2000). Survival analysis for recurrent event data: An application to childhood infectious diseases. *Statistics in Medicine, 19.*

Lenth, R., (2007). Post hoc power: Tables and commentary. Retrieved from: https://stat.uiowa.edu/sites/stat.uiowa.edu/files/techrep/tr378.pdf

Leung, F., & Lai, K. (2003). A case study on maintenance of bus engines using the sequential method. *International Journal of Quality & Reliability Management*, *20*(2), 255-267. doi: 10.1108/02656710310456644

Liu, X. (2014). *Statistical power analysis for the social and behavioral sciences*. New York, NY: Rutledge.

Loehlein, T. (2007). *Maintenance is one key to diesel generator set reliability* (PT- 7004). Cummins Power Generation. Retrieved from: http://cumminspower.com/www/literature/technicalpapers/PT-7004-Maintenance-en.pdf

Margaroni, D. (1999). Extended drain intervals for crankcase lubricants. *Industrial Lubrication and Tribology*, *51*(2), 69-76. doi: 10.1108/00368799910261478

Márquez, A., & Herguedas, A. (2004). Learning about failure root causes through maintenance records analysis. *Journal of Quality in Maintenance Engineering*, *10*(4), 254-262. doi: 10.1108/13552510410564873

Mathur, A. (2002). Data mining of aviation data for advancing health management. Proceedings from SPIE 4733: *Component and Systems Diagnostics, Prognostics, and Health Management II*. doi:10.1117/12.475495

Moubray, J. (1997). *Reliability centered maintenance (2nd ed)*. United Kingdom: Butterworth-Heinemann.

Müller, M. (2004). Goodness-of-fit criteria for survival data. *Sonderforschungsbereich, 386*. Retrieved from http://epub.ub.uni-muenchen.de/

Murphy, K., Myors, B., Wolach, A. (2016). *Statistical power analysis. A simple and general model for traditional and modern hypothesis tests.* New York, NY: Routledge.

National Fire Protection Association. (2016). *Recommended practice for electrical equipment maintenance 2016 edition* (*NFPA 70B)*. Quincy, MA: National Fire Protection Association.

National Fire Protection Association. (2016). *Standard for emergency and standby power systems 2016 edition* (*NFPA 110)*. Quincy, MA: National Fire Protection Association.

Osterloh, K., Jaenish, G. (2016). Rare events – a probability approach or an ill-posed problem? *IEEE Insight*, *58*(1). DOI: 10.1784/insi.2016.58.1.46

Prentice, R., Williams, B., Peterson, A. (1981). On the regression analysis of multivariate failure time data. *Biometrika, 68*. Retrieved from http://www.jstor.org/stable/2335582

Jia, Q., Zhao, Q. (2006). A SVM-based method for engine maintenance strategy optimization. *Proceedings of the 2006 IEEE International Conference on Robotics and Automation*. DOI: 10.1109/ROBOT.2006.1641851

Ripley et al. (2017). *Package 'MASS.'* Retrieved from https://cran.r-project.org/web/packages/MASS/MASS.pdf

Rodriquez, G. (2010). *Parametric survival models*. Retrieved from: http://data.princeton.edu/pop509/ParametricSurvival.pdf

Rowe, W. (2006). Rare-event risk analysis. *Encyclopedia of Statistical Sciences.* John Wiley & Sons, Inc. doi: 10.1002/0471667196.ess2194.pub2

Schoenfeld, D. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics, 39*.

Sousa-Poza, A. (2015). ENMA 821, foundations of research [lecture]. (Available from Old Dominion University, 2101 Engineering Systems Building, Norfolk, VA.)

Steidl, R., Hayes, J., Schauber, E. (1997). Statistical power analysis in wildlife research. *Journal of Wildlife Management, 61*(2).

Tipping, M. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research, 1*. DOI: 10.1162/15324430152748236

Thomas, L., Reyes, E. (2014). Tutorial: Survival estimation for Cox regression models with time-varying coefficients using SAS and R. *Journal of Statistical Software, 14*(61).

Tufte, E. (2014). *Impacts of low load operation of modern four-stroke diesel engines in generator configuration.* (Masters Thesis). Norwegian University of Science and Technology, Norway. Retrieved from: https://daim.idi.ntnu.no/masteroppgaver/011/11083/masteroppgave.pdf

Walbolt, J. (2010). The evolution of diesel piston designs. *Engine Builder Magazine, June 2010.* Retrieved from http://www.enginebuildermag.com/2010/06/the-evolution-of-diesel-piston-designs/

Wang J., & Yang D., Duan X., Ji J., Bai P., (2013). Application of relevance vector machine in the engine oil wear particle fault diagnosis. *Proceedings of 2013 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation.*

Weibull, W. (1951). A statistical distribution function of wide applicability. *ASME Journal of Applied Mechanics, September 1951*. Retrieved from http://www.barringer1.com/wa.htm

Xiao, X., & Xie, M. (2016). A shrinkage approach for failure rate estimation of rare events. *Quality and Reliability Engineering International, 2016.* DOI: 10.1002/qre.1732

Zhou, D., Wang, Z., Li, C. (2014). Data requisites for transformer statistical lifetime modelling—part II: Combination of random and aging-related failures. *IEEE Transactions on Power Delivery, 29*(1). DOI: 10.1109/TPWRD.2013.2270116

APPENDIX A

SURVEY FORM

Generator Test and Maintenance Study

Fleet Cyber Command/Commander Tenth Fleet is conducting a study on the impact of test and maintenance intervals on the reliability of modern high-efficiency diesel-electric generator sets. As manufacturers' preventative maintenance recommendations differ significantly, and often contradictory depending on which manufacturer representative or document you reference, actual implementation has varied widely. This presents us an opportunity to review your logs and scientifically determine the benefit of more frequent tests and preventative maintenance. This study will shape future FCC/C10F policy for test and maintenance frequency.

Please complete this survey and email it to Steve Fehr, FCC/C10F power engineer, at stephen.fehr@navy.mil. FCC/C10F commands should consider this a data call; other organizations (public or private) are encouraged to participate, as the more data we have, the better this study will be. Results of the study will be provided to all who participate. Sensitive information will not be shared.

POINT OF CONTACT INFORMATION

Name*	
Email*	
Phone*	
Organization*	
Date	

** This information is for contact purposes only*

UNIT INFORMATION *Please submit a separate form for each genset.*

Site Name*	
Generator Name ("#1" etc)	
Genset Make	
Genset Model	
Genset kW/kVA Rating	
Genset Installation Date**	

** This information will not be shared outside of US Navy. ** This study is looking primarily at "modern" emergency diesel generator sets manufactured after 1990.*

REPORTING PERIOD *Please submit a separate form for each period*.*

Start date of reporting period	
End date of reporting period	
Run hours at start of reporting period	
Run hours at end of reporting period	
Number of starts	
Typical operational loads (%)	

** Please report as many years of service as you have records for. Each reporting period should cover a single genset over a period of consistent maintenance and testing. If maintenance or test procedures or frequency changed, please submit a separate form. For instance, if a genset installed in 1998 changed from weekly to monthly testing in 2004, 1998-2004 is one reporting period, and 2004-2014 is another.*

ROUTINE PREVENTATIVE MAINTENANCE	Routine Maintenance Frequency															Done by		
	24/7 watch	40hr/week watch	Daily	Weekly	Biweekly	Monthly	Quarterly	6-month	Annual	2 year	3 year	More than 3 year	Not Routine	N/A	Unknown	Collateral Duty	Staff SME	Service Visit SME
Contractor service visit(s); details not known																		
Check alarms																		
Check switch & breaker positions																		
Visual inspection for leaking fluids																		
Visual inspection of hoses, cables, etc.																		
Check fuel level																		
Check oil level																		
Check coolant level																		
Check air filter																		
Battery voltage & physical condition																		
Check fan belt(s)																		
Battery resistance or impedance test																		
Clean unit exterior (including radiator & louvers)																		
Fuel cleaning (or fluid analysis)																		
Oil change (or fluid analysis)																		
Check electrical tightness																		
Engine intensive maintenance																		
Generator (electrical) intensive maintenance																		

ROUTINE TESTING	Test Frequency															Done by		
	24/7 watch	40hr/week watch	Daily	Weekly	Biweekly	Monthly	3-month	6-month	Annual	2 year	3 year	More than 3 year	Not routine	N/A	Unknown	collateral duty	in-house SME	service visit SME
Generator set no-load test																		
Generator set load test on load bank																		
Generator set load test on operational load																		
"Dead bus" operational load test																		

Form Version 9DEC13

OTHER COMMENTS *Any information that might help us properly interpret your data, such as information about unit damage, special maintenance, corrosive climate, etc., as well as any other information not covered on this form that may be pertinent to the study. If historic logs were low on detail, or had to be reproduced by memory, please note that here.*

DESCRIPTION OF FAILURE *Briefly describe what failed. Expand this field if needed. You do not need to report corrective maintenance (repairs) discovered early and repaired to prevent a failure from occuring; this section should be used to list events that actually resulted in a failure to start or failure while running. For failures while running, please include how long the genset was running prior to failure. If details of the failure are unknown, enter as much as is known; check the box(es) and leave the rest blank if that's all you know.*	Date of Failure	Run hours at failure	Time to repair (hrs)	Failure to start	Failure while running	Test Failure	Operational Failure

APPENDIX B
REGRESSION OUTPUT

COX REGRESSION TIME ANALYSIS, x_{16} FULL TIME-DEPENDENT MODEL

The following output is from *R* during analysis of the data set by time (t) and operational failure events (F_{ov}). The predictors are as defined in Table 1. This utilized data set "g41r" which represents the final data set reduced by the removal of units that lacked run-hour chronometers. Time at start (T_{2s}), was used to represent time at the start of the observation period and was set as T_{2s} = 0 for all observations. T_{2ey} represents time (in years) accumulated during each observation period as $T_{2ey} = (T_{fv} - T_{sv})/365$. This model was the culmination of the bidirectional stepwise regression process described in Chapter 4 and represents the final non-parametric model. Plots for this model are shown in Figure 15 through Figure 19.

```
> fit <- coxreg(Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) + x16, data =
g41r)

> summary(fit)
Call:
coxreg(formula = Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) +
    x16, data = g41r)

Covariate           Mean       Coef     Rel.Risk   S.E.    Wald p
log(Tagey)          2.499    -1.476     0.229     0.344     0.000
log(Trtfv)          5.888     0.700     2.014     0.258     0.007
x16                 6.273     0.240     1.271     0.121     0.047

Events                    58
Total time at risk        1281.3
Max. log. likelihood      -269.94
LR test statistic         24.2
Degrees of freedom        3
Overall p-value           2.27095e-05

> cox.zph(fit)
               rho  chisq     p
log(Tagey) -0.1099 0.7528 0.386
log(Trtfv)  0.0412 0.0907 0.763
x16        -0.0837 0.4160 0.519
GLOBAL          NA 1.5785 0.664
```

```
> plot(survfit(fit),ylab="prob(Survival)",xlab="Years in Standby Service")
> plot(resid(fit))
> plot(cox.zph(fit)[1])
> plot(cox.zph(fit)[2])
> plot(cox.zph(fit)[3])
```

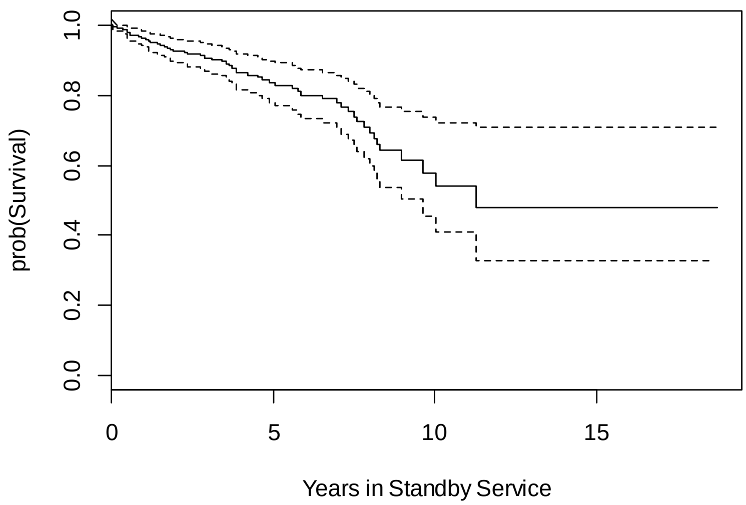

Figure 15. Survival fit of full model regression by time in standby service, *h(t)*.

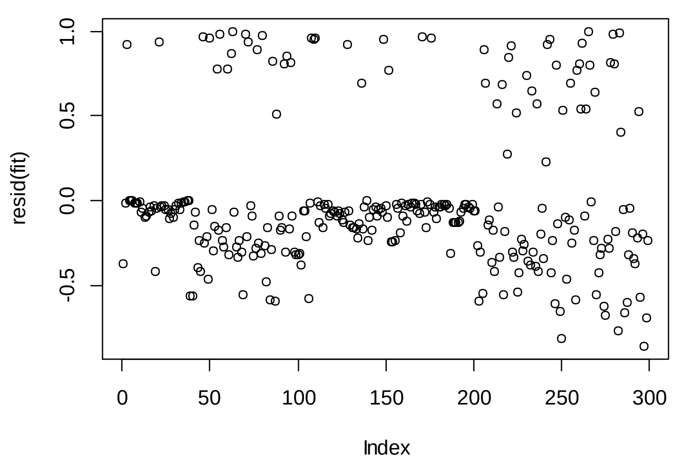

Figure 16. Martingale residuals of full model regression fit by time in standby service, *h(t)*.

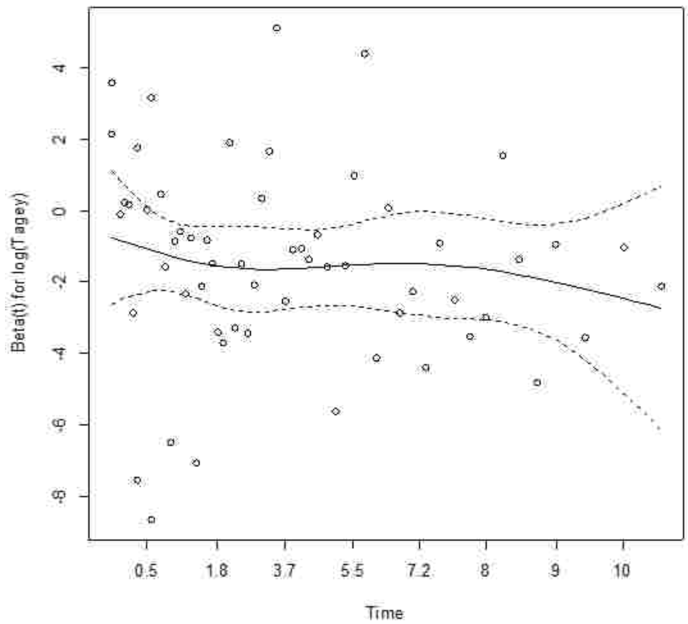

Figure 17. Schoenfeld residuals of *cox.zph* fit test for log(T_{age}), full model by time

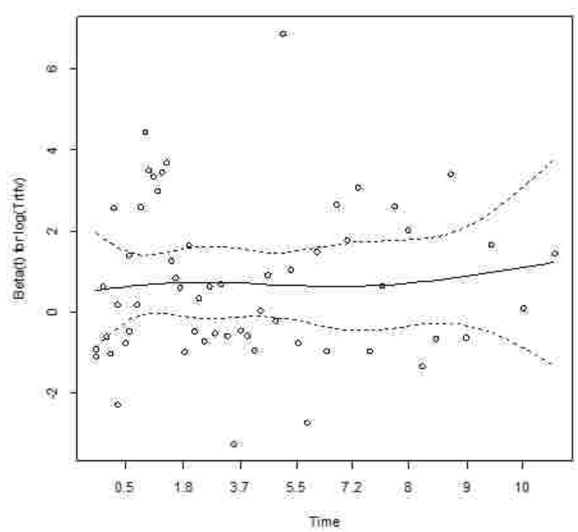

Figure 18. Schoenfeld residuals of *cox.zph* fit test for log(T_{rtfv}), full model by time

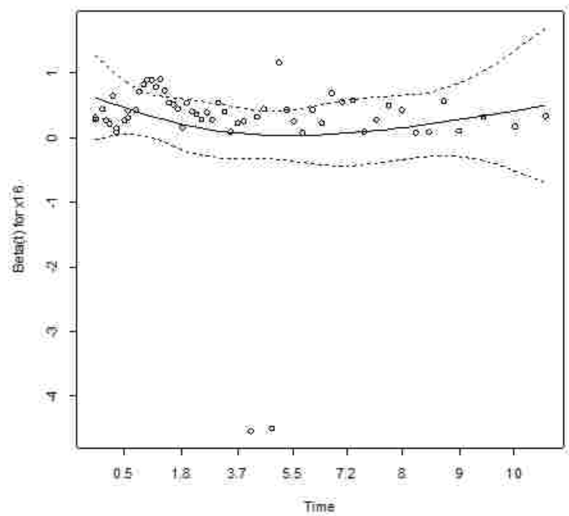

Figure 19. Schoenfeld residuals of *cox.zph* fit test for x_{16}, full model by time

COX REGRESSION RUN HOURS ANALYSIS, x_{16} FULL TIME DEPENDENT MODEL

The following output is from *R* during analysis of the data set by run hours (T_{rt}), and all operational failure events (F_{ov}). This utilized data set "g41r" which represents the final data set reduced by the removal of units that lacked run-hour chronometers. Time at start (T_{2s}), was used to represent run hours at the start of the observation period and was set as $T_{2s}=0$ for all observations. T_{3e} represents the run hours accumulated during each observation period as $T_{3e}=T_{rtfv}-T_{rtsv}$. This model was the culmination of the bidirectional stepwise regression process described in Chapter 4. As no covariates were significant in this model and the log likelihood of this model strength weak compared to the model by time, this model was dropped from the analysis. The survival fit is shown in Figure 20.

```
> fit <- coxreg(Surv(T2s, T3e, Fov) ~ log(Tagey) + log(Trtfv) + x16, data = g
41r)

> summary(fit)
Call:
coxreg(formula = Surv(T2s, T3e, Fov) ~ log(Tagey) + log(Trtfv) +
    x16, data = g41r)

Covariate           Mean       Coef     Rel.Risk   S.E.    Wald p
log(Tagey)          2.458     -0.240     0.787     0.324    0.460
log(Trtfv)          6.098     -0.396     0.673     0.293    0.176
x16                 6.342      0.155     1.168     0.118    0.187

Events                    58
Total time at risk         40151
Max. log. likelihood        -265
LR test statistic         13.5
Degrees of freedom        3
Overall p-value           0.00365329

> cox.zph(fit)
               rho chisq      p
log(Tagey)  0.0808 0.357 0.5499
log(Trtfv) -0.1779 2.255 0.1332
x16        -0.2524 3.800 0.0512
GLOBAL          NA 5.849 0.1192

> plot(survfit(fit),ylab="prob(Survival)",xlab="Run Hours")
> plot(survfit(fit.x16.c),ylab="prob(Survival)",xlab="Years in Standby Servic
e")
> plot(resid(fit))
> plot(cox.zph(fit)[1])
```

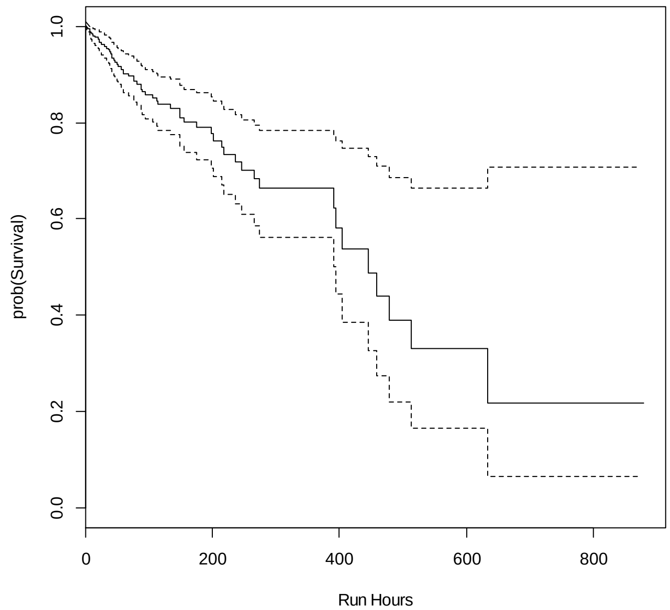

Figure 20. Survival fit of full model regression by run hours, $h(T_{rt})$

COX REGRESSION TIME ANALYSIS, x_{16} SIMPLIFIED MODEL

The following output is from *R* during analysis of the data set by time (t) and all failure events (F_v). The predictors are as defined in Table 1. This utilized data set “g41r” which represents the final data set reduced by the removal of units that lacked run-hour chronometers. Time at start (T_{2s}), was used to represent time at the start of the observation period and was set as T_{2s} = 0 for all observations. T_{2ey} represents time (in years) accumulated during each observation period as $T_{2ey} = (T_{fv} - T_{sv})/365$. This model removed the time dependent Weibull predictors from the full model. Plots for this model are shown in Figure 21 through Figure 23.

```
> fit.x16.c <- coxreg(Surv(T2s, T2ey, Fov) ~ x16, data = g41r)

> summary(fit.x16.c)
Call:
coxreg(formula = Surv(T2s, T2ey, Fov) ~ x16, data = g41r)

Covariate           Mean       Coef     Rel.Risk   S.E.    Wald p
x16                 6.273     0.199     1.220     0.115     0.084

Events                    58
Total time at risk        1281.3
Max. log. likelihood      -279.75
LR test statistic         4.57
Degrees of freedom        1
Overall p-value           0.0325215

> cox.zph(fit.x16.c)
     rho chisq     p
x16 -0.1 0.541 0.462

> plot(survfit(fit.x16.c),ylab="prob(Survival)",xlab="Years in Standby Servic
e")
> plot(resid(fit.x16.c))
> plot(cox.zph(fit.x16.c)[1])
```

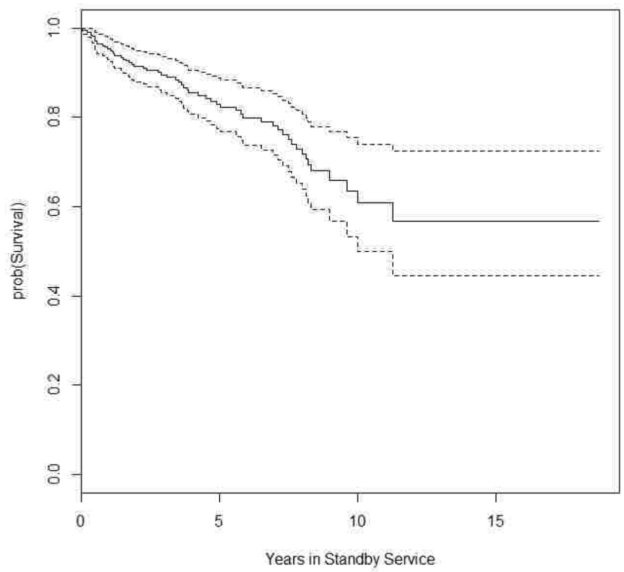

Figure 21. Survival fit of simplified model regression by time in standby service, *h(t)*

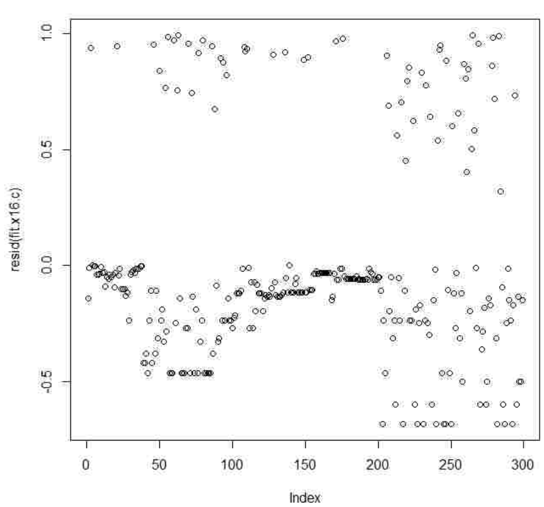

Figure 22. Martingale residuals of simplified model regression fit by time in standby service, *h(t)*

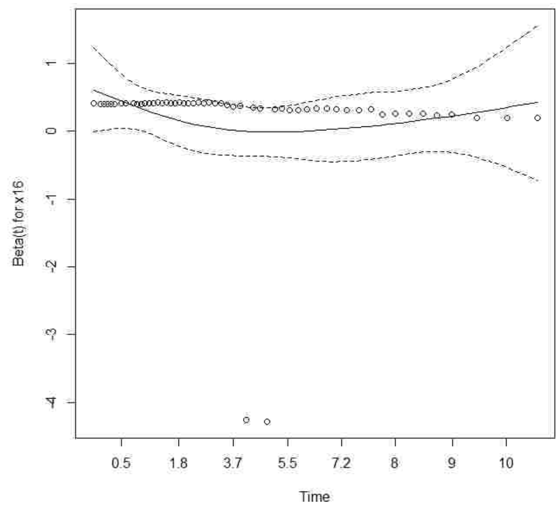

Figure 23. Schoenfeld residuals of *cox.zph* fit test for x_{16}, simplified model by time

COX REGRESSION RUN HOURS ANALYSIS, x_{16} SIMPLIFIED MODEL

The following output is from *R* during analysis of the data set by run hours (T_{rt}), and all operational failure events (F_{ov}). This utilized data set "g41r" which represents the final data set reduced by the removal of units that lacked run-hour chronometers. Time at start (T_{2s}), was used to represent run hours at the start of the observation period and was set as $T_{2s}=0$ for all observations. T_{3e} represents the run hours accumulated during each observation period as $T_{3e}=T_{rtfv}-T_{rtsv}$. This model removed the time dependent Weibull predictors from the full model. Plots are shown in Figure 24 through Figure 26.

```
> fit.x16.c <- coxreg(Surv(T2s, T3e, Fov) ~ x16, data = g41r)

> summary(fit.x16.c)
Call:
coxreg(formula = Surv(T2s, T3e, Fov) ~ x16, data = g41r)

Covariate           Mean       Coef     Rel.Risk   S.E.    Wald p
x16                 6.342     0.200     1.221     0.117     0.087

Events                      58
Total time at risk           40151
Max. log. likelihood        -269.54
LR test statistic           4.43
Degrees of freedom          1
Overall p-value             0.0353569

> cox.zph(fit.x16.c)
       rho chisq      p
x16 -0.227  2.92 0.0876

> plot(survfit(fit.x16.c),ylab="prob(Survival)",xlab="Years in Standby Servic
e")
> plot(resid(fit.x16.c))
> plot(cox.zph(fit.x16.c)[1])
```

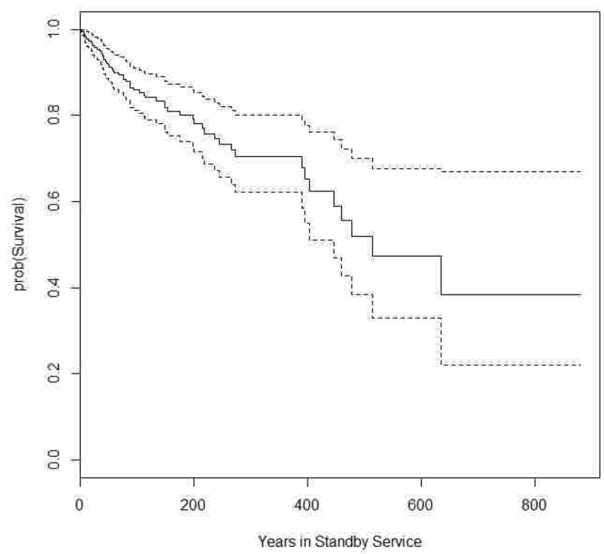

Figure 24. Survival Fit of simplified model regression by run hours, *h(t)*

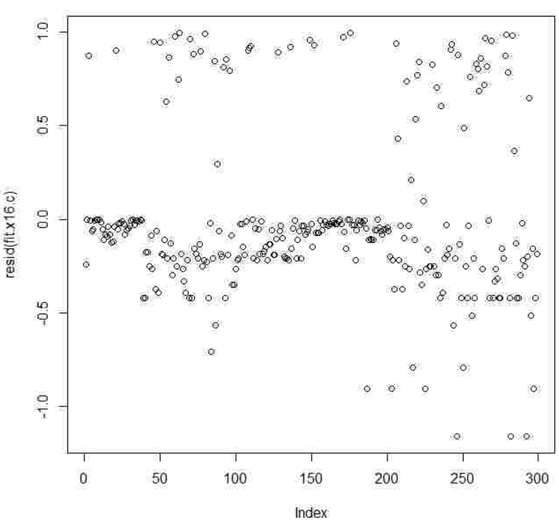

Figure 25. Martingale residuals of simplified model regression fit by run hours, *h(t)*

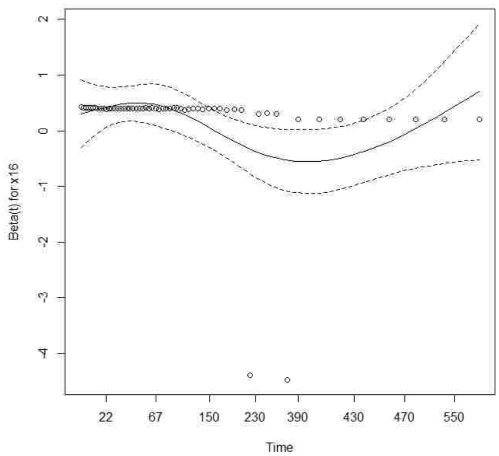

Figure 26. Schoenfeld residuals of *cox.zph* fit test for x_{16}, simplified model by run-hours

COX REGRESSION TIME ANALYSIS, INTERIM TIME-DEPENDENT MODELS

The following output is from *R* during analysis of the data set by time (t) and all operational failure events (F_{ov}). The predictors are as defined in Table 1. This utilized data set "g41r" which represents the final data set reduced by the removal of units that lacked run-hour chronometers. Time at start (T_{2s}), was used to represent time at the start of the observation period and was set as $T_{2s} = 0$ for all observations. T_{2ey} represents time (in years) accumulated during each observation period as $T_{2ey} = (T_{fv} - T_{sv})/365$. The interim models in this section show models with combinations of the predictors with the highest statistical significance in model development, x_{12}, x_{16}, x_L and *src*.

Alone in combination with log(T_{age}) and log(T_{rtfv}), x_{12}, x_{16}, x_L and *src* all yield highly significant results.

```
> coxph(Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) + x12, data = g41r)
Call:
coxph(formula = Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) +
    x12, data = g41r)

             coef exp(coef) se(coef)     z       p
log(Tagey) -1.479     0.228    0.345 -4.29 1.8e-05
log(Trtfv)  0.703     2.020    0.258  2.72  0.0065
x12         0.217     1.242    0.110  1.97  0.0484

Likelihood ratio test=23.9  on 3 df, p=2.58e-05
n= 299, number of events= 58

> coxph(Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) + x16, data = g41r)
Call:
coxph(formula = Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) +
    x16, data = g41r)

             coef exp(coef) se(coef)     z       p
log(Tagey) -1.476     0.229    0.344 -4.29 1.8e-05
log(Trtfv)  0.700     2.014    0.258  2.71  0.0067
x16         0.240     1.271    0.121  1.99  0.0469

Likelihood ratio test=24.2  on 3 df, p=2.27e-05
n= 299, number of events= 58

> coxph(Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) + xl, data = g41r)
Call:
```

```
coxph(formula = Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) +
    xl, data = g41r)

                coef exp(coef) se(coef)     z       p
log(Tagey) -1.53743   0.21493  0.35433 -4.34 1.4e-05
log(Trtfv)  0.74718   2.11103  0.26519  2.82  0.0048
xl         -6.02614   0.00241  3.22697 -1.87  0.0618

Likelihood ratio test=25.8  on 3 df, p=1.05e-05
n= 299, number of events= 58

> coxph(Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) + src, data = g41r)
Call:
coxph(formula = Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) +
    src, data = g41r)

              coef exp(coef) se(coef)     z       p
log(Tagey) -1.494     0.224    0.347 -4.31 1.7e-05
log(Trtfv)  0.708     2.031    0.261  2.72  0.0066
srcB        1.537     4.651    0.744  2.07  0.0389

Likelihood ratio test=24.6  on 3 df, p=1.84e-05
n= 299, number of events= 58
```

When placed in combination, x_{12}, x_{16}, x_L and *src* all lost statistical significance and the model significance worsened. There were no models where two or more of these predictors remain significant together. The opposite coefficients in these paired models suggest conflation is impacting the regression results.

```
> coxph(Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) + x12 + x16 + xl + src
, data = g41r)
Call:
coxph(formula = Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) +
    x12 + x16 + xl + src, data = g41r)

                coef exp(coef)  se(coef)     z     p
log(Tagey) -1.47e+00  2.30e-01  3.57e-01 -4.11 4e-05
log(Trtfv)  6.94e-01  2.00e+00  2.73e-01  2.54 0.011
x12        -4.04e+01  2.74e-18  1.43e+04  0.00 0.998
x16         4.17e+01  1.23e+18  1.27e+04  0.00 0.997
xl         -5.85e+00  2.88e-03  4.25e+00 -1.38 0.169
srcB        2.26e+01  6.57e+09  9.19e+04  0.00 1.000

Likelihood ratio test=29.1  on 6 df, p=5.94e-05
n= 299, number of events= 58
Warning message:
In fitter(X, Y, strats, offset, init, control, weights = weights,  :
  Loglik converged before variable  3,4,6 ; beta may be infinite.
```

```
> coxph(Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) + x12 + x16, data = g4
1r)
Call:
coxph(formula = Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) +
    x12 + x16, data = g41r)

                coef exp(coef)  se(coef)     z       p
log(Tagey) -1.39e+00  2.48e-01  3.47e-01 -4.02 5.9e-05
log(Trtfv)  6.23e-01  1.86e+00  2.63e-01  2.36   0.018
x12        -4.29e+01  2.36e-19  1.19e+04  0.00   0.997
x16         4.84e+01  1.03e+21  1.34e+04  0.00   0.997

Likelihood ratio test=26.3  on 4 df, p=2.76e-05
n= 299, number of events= 58
Warning message:
In fitter(X, Y, strats, offset, init, control, weights = weights,  :
  Loglik converged before variable  3,4 ; beta may be infinite.

> coxph(Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) + x16 + xl, data = g41
r)
Call:
coxph(formula = Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) +
    x16 + xl, data = g41r)

               coef exp(coef) se(coef)     z       p
log(Tagey) -1.53516   0.21542  0.35320 -4.35 1.4e-05
log(Trtfv)  0.75149   2.12016  0.26532  2.83  0.0046
x16         0.07402   1.07682  0.14233  0.52  0.6030
xl         -4.86990   0.00767  3.84834 -1.27  0.2057

Likelihood ratio test=26.1  on 4 df, p=3e-05
n= 299, number of events= 58

> coxph(Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) + x16 + src, data = g4
1r)
Call:
coxph(formula = Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) +
    x16 + src, data = g41r)

               coef exp(coef) se(coef)     z       p
log(Tagey) -1.49375   0.22453  0.34708 -4.30 1.7e-05
log(Trtfv)  0.70838   2.03070  0.26063  2.72  0.0066
x16         0.00979   1.00984  0.46795  0.02  0.9833
srcB        1.47827   4.38535  2.90636  0.51  0.6110

Likelihood ratio test=24.6  on 4 df, p=5.96e-05
n= 299, number of events= 58


> coxph(Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) + x12 + xl, data = g41
r)
Call:
coxph(formula = Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) +
    x12 + xl, data = g41r)
```

```
                  coef exp(coef) se(coef)     z       p
log(Tagey) -1.53675   0.21508  0.35335 -4.35 1.4e-05
log(Trtfv)  0.75194   2.12110  0.26534  2.83  0.0046
x12         0.05786   1.05956  0.13056  0.44  0.6577
xl         -5.04269   0.00646  3.85424 -1.31  0.1908

Likelihood ratio test=26  on 4 df, p=3.13e-05
n= 299, number of events= 58

> coxph(Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) + x12 + src, data = g4
1r)
Call:
coxph(formula = Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) +
    x12 + src, data = g41r)

                coef exp(coef) se(coef)     z       p
log(Tagey) -1.494     0.224    0.347 -4.31 1.7e-05
log(Trtfv)  0.706     2.025    0.261  2.70  0.0069
x12        -0.260     0.771    1.065 -0.24  0.8074
srcB        3.260    26.047    7.228  0.45  0.6520

Likelihood ratio test=24.8  on 4 df, p=5.66e-05
n= 299, number of events= 58

> coxph(Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) + xl + src, data = g41
r)
Call:
coxph(formula = Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) +
    xl + src, data = g41r)

                  coef exp(coef) se(coef)     z       p
log(Tagey) -1.53433   0.21560  0.35334 -4.34 1.4e-05
log(Trtfv)  0.74907   2.11504  0.26569  2.82  0.0048
xl         -4.69266   0.00916  4.06455 -1.15  0.2483
srcB        0.48335   1.62150  0.94431  0.51  0.6088

Likelihood ratio test=26.1  on 4 df, p=3.01e-05
n= 299, number of events= 58
```

Dataset g41r4 was created with a new Marque factor which sorted the myriad of makes and models cluttering the original Make factor into nine specific marques, Caterpillar, Cummins, Onan, MTU, Empire, Generac, SDMO, Olympian, and Kohler. Dataset g41r5 reduced this further by incorporating Onan into the Cummins parent brand and Olympian into the Caterpillar parent brand. SDMO and Kohler were left separate in all models.

```
> coxph(Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) + x16 + Marque, data =
g41r4)
Call:
coxph(formula = Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) +
    x16 + Marque, data = g41r4)

                    coef exp(coef)  se(coef)     z       p
log(Tagey)     -1.65e+00  1.91e-01  3.51e-01 -4.71 2.4e-06
log(Trtfv)      6.10e-01  1.84e+00  2.70e-01  2.26   0.024
x16             1.87e-01  1.21e+00  1.28e-01  1.46   0.144
MarqueCummins   7.70e-01  2.16e+00  6.36e-01  1.21   0.226
MarqueEmpire   -1.36e+01  1.25e-06  2.62e+03 -0.01   0.996
MarqueGenerac   1.18e-01  1.12e+00  6.32e-01  0.19   0.852
MarqueKohler    4.06e-01  1.50e+00  6.51e-01  0.62   0.533
MarqueMTU      -6.39e-01  5.28e-01  8.05e-01 -0.79   0.427
MarqueOlympian  1.16e+00  3.18e+00  6.50e-01  1.78   0.075
MarqueOnan      1.14e+00  3.12e+00  6.59e-01  1.73   0.085
MarqueSDMO      9.48e-02  1.10e+00  9.23e-01  0.10   0.918

Likelihood ratio test=38.9  on 11 df, p=5.43e-05
n= 299, number of events= 58
Warning message:
In fitter(X, Y, strats, offset, init, control, weights = weights,  :
  Loglik converged before variable  5 ; beta may be infinite.


> coxph(Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) + x16 + Marque, data =
g41r5)
Call:
coxph(formula = Surv(T2s, T2ey, Fov) ~ log(Tagey) + log(Trtfv) +
    x16 + Marque, data = g41r5)

                   coef exp(coef)  se(coef)     z       p
log(Tagey)    -1.51e+00  2.21e-01  3.41e-01 -4.43 9.5e-06
log(Trtfv)     5.76e-01  1.78e+00  2.63e-01  2.19   0.029
x16            2.68e-01  1.31e+00  1.22e-01  2.19   0.028
MarqueCummins  2.10e-01  1.23e+00  3.65e-01  0.58   0.564
MarqueEmpire  -1.44e+01  5.74e-07  2.60e+03 -0.01   0.996
MarqueGenerac -5.80e-01  5.60e-01  4.11e-01 -1.41   0.158
MarqueKohler  -3.21e-01  7.25e-01  4.54e-01 -0.71   0.479
MarqueMTU     -1.36e+00  2.57e-01  6.47e-01 -2.10   0.036
MarqueSDMO    -6.04e-01  5.46e-01  7.85e-01 -0.77   0.442

Likelihood ratio test=34.9  on 9 df, p=6.1e-05
n= 299, number of events= 58
Warning message:
In fitter(X, Y, strats, offset, init, control, weights = weights,  :
  Loglik converged before variable  5 ; beta may be infinite.
```

PARAMETRIC MODEL SPREADSHEET REGRESSION ANALYSIS

The parametric model could not be analyzed by the *R* function *phreg* due to the different time dependencies required for predictor coefficient calculations and final model fitting. The parametric model was instead developed by piecewise constant hazard regression in *Microsoft Excel*.

The final parametric model below was developed by separating the sample population into two stratified subpopulations, one with $x_{16} = 7$ and one with $x_{16} \leq 1$. Different values of λ were selected until the difference between the weighted reliability mean of the model matched the sample reliability from the data. The mean value of $x_{16} = 0.45$ for the $x_{16} \leq 1$ subpopulation was used to represent x_{16} for the $x_{16} \leq 1$ subpopulation. The histogram field was used for weighting the mean and represents the number of observations of generator sets of that age included in the sample population data set.

The piecewise constant hazard model spreadsheet equations used were as follows, with *T* representing the age of the generator set from *T* = 1 to 43 years, the oldest set in the sample population. The calculated values are shown in Table 8 and Table 9.

h(t) = \$B7*EXP(\$C7+\$D7+\$E7) = $\lambda e^{-1.476\log(T) + .7\log(31.2T) + 0.1044 \text{ [or } 1.624 \text{ for } x16=7]}$

f(t) = I7*L6 = h(t)S(t-1)

F(t) = K6+J7 = f(t) + F(t-1)

S(t) = EXP(-K7) = $e^{-F(t)}$

R(t) = EXP(-I7) = $e^{-h(t)}$

Table 8. Model Parameter Calculations From the Spreadsheet Regression for $x_{16} \leq 1$

x16<=1	**h(t)**	**f(t)**	**F(t)**	**S(t)**	**R(t)**	
t(years)	**$x16 \leq 1$**	**PDF**	**CDF**	**Surv**	**Rel**	**Histogram**
1	0.023061915	0.023062	0.023062	0.977202	0.977202	6
2	0.018257634	0.017841	0.040903	0.959922	0.981908	7
3	0.015925737	0.015287	0.056191	0.945359	0.9842	8
4	0.014454185	0.013664	0.069855	0.932529	0.98565	11
5	0.013407065	0.012502	0.082358	0.920943	0.986682	7
6	0.012608071	0.011611	0.093969	0.910311	0.987471	10
7	0.011969796	0.010896	0.104865	0.900446	0.988102	10
8	0.011443075	0.010304	0.115169	0.891215	0.988622	10
9	0.010997746	0.009801	0.12497	0.882523	0.989063	10
10	0.010614092	0.009367	0.134338	0.874295	0.989442	6
11	0.010278577	0.008987	0.143324	0.866473	0.989774	4
12	0.009981545	0.008649	0.151973	0.859012	0.990068	4
13	0.009715888	0.008346	0.160319	0.851872	0.990331	2
14	0.009476236	0.008073	0.168391	0.845023	0.990569	2
15	0.009258442	0.007824	0.176215	0.838438	0.990784	5
16	0.009059242	0.007596	0.183811	0.832093	0.990982	9
17	0.008876028	0.007386	0.191196	0.82597	0.991163	7
18	0.008706685	0.007191	0.198388	0.820052	0.991331	8
19	0.008549474	0.007011	0.205399	0.814322	0.991487	8
20	0.008402954	0.006843	0.212242	0.808769	0.991632	9
21	0.008265915	0.006685	0.218927	0.803381	0.991768	9
22	0.008137334	0.006537	0.225464	0.798146	0.991896	0
23	0.008016338	0.006398	0.231862	0.793055	0.992016	0
24	0.00790218	0.006267	0.238129	0.788101	0.992129	1
25	0.00779421	0.006143	0.244272	0.783275	0.992236	1
26	0.007691865	0.006025	0.250297	0.77857	0.992338	1
27	0.007594652	0.005913	0.25621	0.77398	0.992434	1
28	0.007502137	0.005807	0.262016	0.769499	0.992526	0
29	0.007413938	0.005705	0.267721	0.765121	0.992613	0
30	0.007329714	0.005608	0.273329	0.760842	0.992697	0
31	0.007249162	0.005515	0.278845	0.756657	0.992777	0
32	0.007172012	0.005427	0.284271	0.752562	0.992854	3
33	0.007098019	0.005342	0.289613	0.748553	0.992927	3
34	0.007026965	0.00526	0.294873	0.744626	0.992998	3
35	0.006958652	0.005182	0.300055	0.740778	0.993066	3
36	0.006892899	0.005106	0.305161	0.737005	0.993131	3
37	0.006829545	0.005033	0.310194	0.733304	0.993194	3
38	0.006768439	0.004963	0.315158	0.729674	0.993254	3
39	0.006709446	0.004896	0.320053	0.72611	0.993313	3
40	0.006652442	0.00483	0.324884	0.722611	0.99337	0
41	0.006597312	0.004767	0.329651	0.719175	0.993424	0
42	0.006543951	0.004706	0.334357	0.715798	0.993477	0
43	0.006492262	0.004647	0.339004	0.712479	0.993529	0

Table 9. Model Parameter Calculations From the Spreadsheet Regression for x_{16} =7.

x16 =7 t(years)	h(t) *x* 16 = 7	f(t) PDF	F(t) CDF	S(t) Surv	R(t) Rel	Histogram
1	0.1054021	0.105402	0.105402	0.899963	0.899963	45
2	0.083444629	0.075097	0.180499	0.834853	0.919942	76
3	0.072786935	0.060766	0.241266	0.785633	0.929799	84
4	0.06606136	0.0519	0.293166	0.745899	0.936073	99
5	0.061275607	0.045705	0.338871	0.712574	0.940564	98
6	0.057623887	0.041061	0.379932	0.683908	0.944005	95
7	0.054706716	0.037414	0.417347	0.658793	0.946763	97
8	0.052299391	0.034454	0.451801	0.636481	0.949045	82
9	0.050264063	0.031992	0.483793	0.616441	0.950978	75
10	0.048510611	0.029904	0.513697	0.59828	0.952647	66
11	0.046977175	0.028105	0.541803	0.581699	0.954109	60
12	0.045619621	0.026537	0.568339	0.566465	0.955405	55
13	0.044405464	0.025154	0.593494	0.552394	0.956566	46
14	0.043310158	0.023924	0.617418	0.539335	0.957614	37
15	0.042314751	0.022822	0.64024	0.527166	0.958568	35
16	0.04140433	0.021827	0.662067	0.515784	0.959441	32
17	0.04056697	0.020924	0.68299	0.505104	0.960245	34
18	0.039793003	0.0201	0.70309	0.495053	0.960988	31
19	0.039074488	0.019344	0.722434	0.485569	0.961679	31
20	0.038404832	0.018648	0.741082	0.476598	0.962323	28
21	0.037778509	0.018005	0.759087	0.468093	0.962926	20
22	0.037190843	0.017409	0.776496	0.460015	0.963492	20
23	0.036637846	0.016854	0.79335	0.452327	0.964025	14
24	0.036116096	0.016336	0.809686	0.444998	0.964528	13
25	0.03562263	0.015852	0.825538	0.437999	0.965004	13
26	0.035154873	0.015398	0.840936	0.431307	0.965456	11
27	0.034710571	0.014971	0.855907	0.424898	0.965885	11
28	0.034287743	0.014569	0.870476	0.418752	0.966293	11
29	0.033884637	0.014189	0.884665	0.412852	0.966683	9
30	0.0334997	0.01383	0.898496	0.407182	0.967055	8
31	0.033131546	0.013491	0.911986	0.401726	0.967411	7
32	0.032778938	0.013168	0.925154	0.39647	0.967752	7
33	0.032440763	0.012862	0.938016	0.391404	0.96808	4
34	0.032116018	0.01257	0.950586	0.386514	0.968394	4
35	0.031803799	0.012293	0.962879	0.381792	0.968697	3
36	0.031503285	0.012028	0.974907	0.377228	0.968988	2
37	0.031213729	0.011775	0.986681	0.372812	0.969268	2
38	0.030934451	0.011533	0.998214	0.368537	0.969539	2
39	0.030664831	0.011301	1.009515	0.364396	0.969801	2
40	0.030404299	0.011079	1.020594	0.360381	0.970053	2
41	0.030152333	0.010866	1.031461	0.356486	0.970298	1
42	0.029908452	0.010662	1.042123	0.352705	0.970534	1
43	0.029672214	0.010466	1.052588	0.349033	0.970764	1

The weighted reliability of each subpopulation was calculated by the equation below, where n(t) is the histogram value for each year.

$$R_{model} = \frac{\sum_{t=1}^{43} R(t)n(t)}{\sum_{t=1}^{43} n(t)}$$

This was compared to the observed sample reliability, calculated as $R_{data} = e^{-1/MTBF}$, where MTBF = Total Time / Number of Failures for the sample population.

For $x_{16} \leq 1$, MTBF = 143.77 / 2 = 71.88 years

For $x_{16} = 7$, MTBF = 1137.45 / 56 = 20.3116 years

A value of $\lambda = 0.00696$ resulted in the smallest net difference between the model and data reliability figures.

$$h(t) = 0.00696e^{0.240x_{16} - 1.476\log(t) + .700\log(31.2t)}$$

For $x_{16} \leq 1$, $R_{data} = 0.95305$, $R_{model} = 0.94703$

For $x_{16} = 7$, $R_{data} = 0.98353$, $R_{model} = 0.98956$

SIMPLIFIED PARAMETRIC REGRESSION MODEL ANALYSIS

A simplified time-independent parametric regression model fitted to an exponential survival distribution using x_{16} as the only predictor was developed. The following output is from *R* during analysis of the data set using the *phreg* function in *R*. The predictors are as defined in Table 1. This utilized data set "g41r" which represents the final data set reduced by the removal of units that lacked run-hour chronometers. Time at start (T_{2s}), was used to represent time at the start of the observation period and was set as $T_{2s} = 0$ for all observations. T_{2ey} represents time (in years) accumulated during each observation period as $T_{2ey} = (T_{fv} - T_{sv})/365$.

The $p = 0.94$ value calculated for the fixed shape exponential distribution by *phreg* appears to be in error as *phreg* is comparing the log likelihood to a test statistic of 0. The *p*-value calculated for the unrestricted Weibull distribution returns a shape nearly identical to the exponential function but a statistically significant value of $p = 0.022$. As the predictor β_{16} value (0.207 and 0.208 respectively), shape (1 and 0.962 respectively) and log likelihood values (-234.98 and -234.91 respectively) for these two functions are nearly identical, the *p*-value for the exponential model should also be very close to $p = 0.022$.

The plots of these functions are shown in Figure 11 and Figure 12 in the Chapter IV.

```
> fit.x16.c <- coxreg(Surv(T2s, T2ey, Fov) ~ x16, data = g41r)
> fit.x16.c
Call:
coxreg(formula = Surv(T2s, T2ey, Fov) ~ x16, data = g41r)

Covariate           Mean       Coef     Rel.Risk   S.E.    Wald p
x16                 6.273     0.199     1.220     0.115     0.084

Events                    58
Total time at risk        1281.3
Max. log. likelihood      -279.75
LR test statistic         4.57
Degrees of freedom        1
Overall p-value           0.0325215
```

```
> fit.x16.e <- phreg(Surv(T2s, T2ey, Fov) ~ x16, data = g41r, dist="weibull",
shape=1)
> fit.x16.e
Call:
phreg(formula = Surv(T2s, T2ey, Fov) ~ x16, data = g41r, dist = "weibull",
    shape = 1)

Covariate           W.mean      Coef Exp(Coef)  se(Coef)    Wald p
x16                  6.273     0.207     1.230     0.115     0.072

log(scale)                     4.456    86.165     0.792     0.000

 Shape is fixed at  1

Events                    58
Total time at risk        1281.3
Max. log. likelihood      -234.98
LR test statistic         0
Degrees of freedom        1
Overall p-value           0.945997

> fit.x16.w <- phreg(Surv(T2s, T2ey, Fov) ~ x16, data = g41r, dist="weibull")
> fit.x16.w
Call:
phreg(formula = Surv(T2s, T2ey, Fov) ~ x16, data = g41r, dist = "weibull")

Covariate           W.mean      Coef Exp(Coef)  se(Coef)    Wald p
x16                  6.273     0.208     1.232     0.115     0.070

log(scale)                     4.576    97.116     0.891     0.000
log(shape)                    -0.040     0.961     0.109     0.717

Events                    58
Total time at risk        1281.3
Max. log. likelihood      -234.91
LR test statistic         5.17
Degrees of freedom        1
Overall p-value           0.0229996
```

Using this model with the spreadsheet regression method returns the following:

For $x_{16} \leq 1$, $R_{data} = 0.98353$, $R_{model} = 0.98739$

For $x_{16} = 7$, $R_{data} = 0.95305$, $R_{model} = 0.95434$

APPENDIX C

RELIABILITY TABLES

Tables are provided for reliability based upon the model in Equation 18. Table 10 reflects typical values assuming an average of 31.2 hours annual run time, the average for the sample population in this study. Table 11 reflects units an average 100 hours annual run time. Generator sets receiving weekly no-load tests and bimonthly or quarterly maintenance visits are represented by the average intensity column. Generator sets receiving NFPA 110 (2016) recommended maintenance are represented by the high intensity column. Units well exceeding NFPA 110 (2016) recommendations are represented by the extremely high intensity column. Reliability for specific generator set age, run hours, and maintenance intensity can be calculated using Equation 18. Values of 0.08333, 1, and 7 were used respectively for average, high, and extremely high maintenance intensity.

$$h(t) = 0.00696e^{0.240(Maint\ Intensity) - 1.476\log(Age) + .700\log(Run\ Hours)} \qquad (18)$$

Table 10. Emergency Diesel-Electric Generator Set Typical Reliability Table for 31.2 Annual Run Hours

	Maintenance Intensity		
Age	Extremely High	High	Average
0 < t ≤ 5	0.985201	0.981593	0.924670
5 < t ≤ 10	0.989952	0.987493	0.948271
10 < t ≤ 15	0.991500	0.989419	0.956098
15 < t ≤ 20	0.992390	0.990525	0.960616
20 < t ≤ 25	0.992995	0.991278	0.963701
25 < t ≤ 30	0.993445	0.991837	0.965998
30 < t ≤ 35	0.993798	0.992277	0.967805
35 < t ≤ 40	0.994085	0.992635	0.969280

Table 11. Emergency Diesel-Electric Generator Set Typical Reliability Table for 100 Annual Run Hours

	Maintenance Intensity		
Age	Extremely High	High	Average
$0 < t \leq 5$	0.978981	0.973879	0.894474
$5 < t \leq 10$	0.985712	0.982227	0.927113
$10 < t \leq 15$	0.987910	0.984958	0.938033
$15 < t \leq 20$	0.989173	0.986527	0.944356
$20 < t \leq 25$	0.990033	0.987595	0.948679
$25 < t \leq 30$	0.990672	0.988389	0.951904
$30 < t \leq 35$	0.991173	0.989013	0.954441
$35 < t \leq 40$	0.991582	0.989522	0.956514

APPENDIX D

NFPA 110 (2016) RECOMMENDATIONS

ROUTINE PREVENTATIVE MAINTENANCE	NFPA110 (2013) Maintenance Frequency														
	24/7 watch	40hr/week watch	Daily	Weekly	Biweekly	Monthly	Quarterly	6-month	Annual	2 year	3 year	More than 3 year	Not Routine	N/A	Unknown
Check alarms							X								
Check switch & breaker positions						X									
Visual inspection for leaking fluids				X											
Visual inspection of hoses, cables, etc.				X											
Check fuel level				X											
Check oil level				X											
Check coolant level				X											
Check air filter							X								
Battery voltage & physical condition						X									
Check fan belt(s)						X									
Battery resistance or impedance test													X		
Clean unit exterior (including radiator & louvers)									X						
Fuel cleaning (or fluid analysis)									X						
Oil change (or fluid analysis)									X						
Check electrical tightness									X						
Engine intensive maintenance									X						
Generator (electrical) intensive maintenance									X						

ROUTINE TESTING	NFPA110 (2013) Test Frequency														
	24/7 watch	40hr/week watch	Daily	Weekly	Biweekly	Monthly	3-month	6-month	Annual	2 year	3 year	More than 3 year	Not routine	N/A	Unknown
Generator set no-load test													X		
Generator set load test on load bank													X		
Generator set load test on operational load							X								
"Dead bus" operational load test													X		

VITA

Stephen John Fehr

Engineering Management & Systems Engineering

2101 Engineering Systems Building

Norfolk, VA 23529

757-683-4558 (office)

Stephen John Fehr earned a Master of Science in Engineering Management from Old Dominion University in 2014, specializing in emergency generator set reliability modeling; a Bachelor of Science in Engineering Science from Pennsylvania State University in 1997, specializing in semiconductors and microelectrical mechanical systems; and is conducting doctoral research at Old Dominion University into emergency generator set reliability modeling and risk management of high reliability complex systems. Stephen is the Emergency Power Program Manager at United States Navy Information Forces, an adjunct instructor at Old Dominion University teaching engineering economics, and a licensed Professional Engineer in the Commonwealth of Virginia.