NASA/TM–20250004074
NESC-RP-23-01904

A Methodology to Evaluate the Feasibility of Descoping Nondestructive Evaluation (NDE) Fracture Control Requirements in NASA-STD-5019A

Patrick E. Leser, Peter A. Parker, William H. Prosser/NESC (retired)
Langley Research Center, Hampton, Virginia

Bence B. Bartha
Kennedy Space Center, Florida

K. Elliott Cramer/NESC
Langley Research Center, Hampton, Virginia

April 2025

NASA STI Program Report Series

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- Help desk contact information:

https://www.sti.nasa.gov/sti-contact-form/
and select the "General" help request type.

NASA/TM–20250004074
NESC-RP-23-01904

A Methodology to Evaluate the Feasibility of Descoping Nondestructive Evaluation (NDE) Fracture Control Requirements in NASA-STD-5019A

Patrick E. Leser, Peter A. Parker, William H. Prosser/NESC (retired)
Langley Research Center, Hampton, Virginia

Bence B. Bartha
Kennedy Space Center, Florida

K. Elliott Cramer/NESC
Langley Research Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

April 2025

Acknowledgments

The NASA Engineering and Safety Center (NESC) team would like to express appreciation to SpaceX for providing access to NDE inspection information as well as technical and programming support in exporting this inspection information into manageable formats. The team appreciates the detailed NASA peer reviews by Steven Gentz, Jon Holladay, Heather Koehler, Bryan McEnerney, Peter Spaeth, James Warner, Sara Wilson, and Aerospace Corporation peer reviews from Vinay Goyal, John Klug, Leeland Shimizu, and Evgueni Todorov. The team appreciates external organization practice insights from Michael Gorelik (Federal Aviation Administration), Charles Babish (United States Air Force), and James Sobotka (Southwest Research Institute).

Available from:

NASA STI Program / Mail Stop 050
NASA Langley Research Center
Hampton, VA 23681-2199

NASA Engineering and Safety Center
Technical Assessment Report

A Methodology to Evaluate the Feasibility of Descoping Nondestructive Evaluation (NDE) Fracture Control Requirements in NASA-STD-5019A

TI-23-01904

NESC Lead – K. Elliott Cramer

March 20, 2025

Report Approval and Revision History

Approved: *Original signature on file.*	
NESC Director	

Version	Description of Revision	Office of Primary Responsibility	Effective Date
1.0	Initial Release	K. Elliott Cramer, NESC Chief Engineer, LaRC	03/20/2025

Table of Contents

List of Figures

List of Tables

Technical Assessment Report

1.0 Notification and Authorization

SpaceX suggested that rationale may be found for achieving equivalent risk posture without using the traditional approach to damage tolerance in NASA-STD-5019A, "*Fracture Control Requirements for Spaceflight Hardware*" [ref. 1] by reviewing materials manufacturing and inspection data (e.g., nondestructive evaluation (NDE) inspection results, raw material receipt inspections, manufacturing scrap rate information, statistical process control data, etc.) for parts made of a single wrought material. They excluded additively manufactured parts, castings, forgings, and welds. The objective of this assessment was to develop a probabilistic analysis method for NASA programs and projects to estimate risk associated with descoping NASA-STD-5019A NDE requirements of wrought materials, demonstrate the method using SpaceX-provided data, perform sensitivity studies as to the future acceptability of descope requests, and identify minimum supporting data required for approval. The primary stakeholders are the NASA Office of the Chief Engineer (OCE), the Commercial Crew Program, Human Landing System, and other NASA programs and projects receiving requests from SpaceX or other commercial providers to eliminate or descope NDE inspections in lieu of NASA-STD-5019A requirements.

2.0 Signatures

Submitted by: NESC Lead

All original signatures on file.

Mr. K. Elliott Cramer

Significant Contributors:

Dr. Patrick E. Leser

Dr. Peter A. Parker

Dr. William H. Prosser

Dr. Bence B. Bartha

Signatories declare the findings, observations, and NESC recommendations compiled in the report are factually based from data extracted from program/project documents, contractor reports, and open literature, and/or generated from independently conducted tests, analyses, and inspections.

3.0 Team Members

Name	Discipline	Organization
Core Team		
K. Elliott Cramer	NESC Chief Engineer	NESC/LaRC
Bence Bartha	NDE/Fracture Control	KSC
Joachim Beek	Fracture Control	JSC
Jonathan Burkholder	Fracture Control	MSFC
Steve Gentz	Materials/Structures, NESC Chief Engineer (retired)	NESC/MSFC
Patrick Leser	Co-Lead, Probabilistic Structural Analysis	LaRC
Dave Lubas	NDE	KSC
Atef Muzaffar	Fracture Control	JSC
Peter Parker	Co-Lead, Statistics	LaRC
William Prosser	NASA NDE Technical Fellow (retired)	NESC/LaRC
Gregory Swanson	Fracture Control	MSFC
Will Tilson	Materials	MSFC
Consultants		
Heather Hickman	Structures	NESC/GRC
Alison Park	Materials	NESC/LaRC
Business Management		
April Albert	Program Analyst	MTSO/LaRC
Assessment Support		
Anissa Proctor	Project Coordinator	Barrios/LaRC
Linda Burgess	Planning and Control Analyst	AMA/LaRC
Dee Bullock	Technical Editor	AMA/LaRC

3.1 Acknowledgements

The NASA Engineering and Safety Center (NESC) team would like to express appreciation to SpaceX for providing access to NDE inspection information as well as technical and programming support in exporting this inspection information into manageable formats. The team appreciates the detailed NASA peer reviews by Steven Gentz, Jon Holladay, Heather Koehler, Bryan McEnerney, Peter Spaeth, James Warner, Sara Wilson, and Aerospace Corporation peer reviews from Vinay Goyal, John Klug, Leeland Shimizu, and Evgueni Todorov. The team appreciates external organization practice insights from Michael Gorelik (Federal Aviation Administration), Charles Babish (United States Air Force), and James Sobotka (Southwest Research Institute).

4.0 Executive Summary

As a result of inquiries from SpaceX, the NASA Office of the Chief Engineer (OCE) requested the NASA Engineering and Safety Center (NESC) to assess a proposal related to descoping (i.e., eliminating or reducing) nondestructive evaluation (NDE) inspections on fracture-critical spaceflight hardware. Such an approach would be in lieu of the requirements of NASA-STD-5019A "*Fracture Control Requirements for Spaceflight Hardware*" [ref. 1] and require programmatic or Agency acceptance of waivers or eventual changes to this NASA standard. The proposal was motivated by the observation that historical NDE data from a large number of parts found few flaw indications. Therefore, it was asserted that the inspections were not 'value added'[1] in the manufacturing process. Upon initial discussions with SpaceX, during which information was being solicited to develop the scope and plan for the assessment, it was determined that no formal proposal to descope NDE existed, nor had they quantitatively analyzed the NDE inspection data. Instead, SpaceX was offering the NESC access to their NDE database and requesting assistance with formulating an approach that might lead to NASA program and project acceptance of descoping NDE for future programs. Eventually, the NESC settled upon a scope for the assessment in which statistical techniques and tools would be developed that would allow NASA programs and projects to assess the increased risk associated with descoping NDE for a particular wrought part based on historical NDE inspections and associated rejection rates. Further, the NESC would perform a targeted review of the available SpaceX data for example parts to exercise the risk-evaluation tools. However, the NESC would not assess the acceptability of added risk for descoping NDE for any specific part or application as that evaluation is the purview of the affected NASA programs and projects. This assessment scope was agreed upon by the NESC, the OCE, and SpaceX.

It was found that descoping NDE for a fracture-critical part that meets the requirements of NASA-STD-5019A would result in an increased risk of failure that is proportional to the probability of critical initial flaw size (CIFS) defect existence. A quantitative risk-evaluation framework was developed that uses the probability of NDE detectable flaw existence estimated from historical inspection data. The methodology is applicable to descoping a single NDE method applied to wrought metallic materials where measurable/monitored time-invariant process control is established to ensure the estimated probabilities will not appreciably change over time. The methodology is not applicable to additive manufacturing, castings, forgings, and welds due to higher variability in inspection and manufacturing. It was noted that descoping NDE may remove a means of monitoring process control. To improve the statistical estimation of risk, it was also noted that similarity can be applied to aggregate multiple parts which will increase the number of samples evaluated. However, qualitative and quantitative measures should be considered by NASA programs and projects to establish similarity. This risk-evaluation framework includes the effect of process escape defects if their frequency of occurrence is captured in the NDE database. If process escapes are not captured in the database (e.g., parts are scrapped without documentation) and the method being descoped is the only means of process escape detection,[2] then the evaluation of risk may be non-conservative.

[1] For the purposes of this report, value added is the inspection cost (i.e., resources, time, and risk associated with rework) versus benefit (i.e., reliability change).

[2] Visual inspections are assumed to remain in production and may detect large flaws induced by process escapes, but they typically do not have quantified probability of detection. Therefore, benefits associated with visual inspection are not included in this risk-evaluation framework.

As a conservative assumption, the likelihood of part failure is assumed to be equal to the probability of a detectable flaw existing. This likelihood can be used in conjunction with the appropriate programmatic or hazard risk matrix to enable the appropriate responsible Technical Authority[3] to make risk acceptance decisions. For a fracture-critical part that satisfies the requirements of NASA-STD-5019A, the Consequence is assumed to be 5 in the typical 5 Likelihood X 5 Consequence risk matrix.[4] Thus, the evaluation will determine where the risk resides on the N x 5 Consequence column based on the assessment of the category of Likelihood N. As a starting point, the baseline primary structural failure likelihood for a part that meets NASA-STD-5019A is assumed to be at a 1 (i.e., lowest risk) even though this is not quantitatively demonstrated by test or analysis. Thus, using the NESC risk matrix[5] for crewed missions [ref. 2] (see Appendix A), this equates to a likelihood of failure of less than 0.000001 (or 0.0001%). The case study and analysis of the graphs of estimated likelihood of failure versus sample size herein suggest that it might be reasonable to obtain a likelihood of failure of level 2 (i.e., between 0.0001 and 0.1%) with sufficient NDE inspections (i.e., greater than 5000) and if NDE detected flaws are historically rare (<0.02%). However, it is unlikely that sufficient historical NDE inspections will be available to reach the baseline Likelihood 1 category, as this would require at least 3 million inspections without detecting any flaws. Thus, descoping NDE is expected to result in an elevated risk acceptance posture even if a sufficiently large NDE database is available and there are no flaw detections in the inspection history.

As a result of this assessment, the NESC recommends the Agency/programs/projects implement the proposed risk-evaluation methodology if an NDE descope is requested for a wrought structural part that meets NASA-STD-5019A. In doing so, it is recommended that rationale for data aggregation across multiple parts to establish similarity and quantitative evidence for time-invariant process control be required. It is recommended that evidence and/or rationale be required to establish that the risk-evaluation database is representative of future parts, under consistent process control, and continues to meet NASA-STD-5019A based on the NDE inspection technique used to generate the database. Also, it is recommended that evidence be required that process escapes are captured in the risk-evaluation database. Lastly, it is recommended that the appropriate NASA Fracture Control Board and/or responsible Technical Authority review the risk evaluation and, if the NDE is to be permitted to be descoped, consider whether alternative requirements to monitor process control should be established.

[3] The Technical Authority process is a part of NASA's system of checks and balances to provide independent oversight of programs and projects in support of safety and mission success through the selection of specific individuals with delegated levels of authority (https://www.nasa.gov/technical-authority/).

[4] The risk matrix is typically used at the system level; however, it is assumed for the purpose of this general discussion that the part in question is the primary driver of the system risk. If there are multiple parts with reliabilities at this level, the system risk will be higher.

[5] Individual programs may have specific risk matrices with different definitions of Likelihood levels.

5.0 Assessment Plan

The original request was to review a proposal from SpaceX for a descope of NDE requirements under NASA-STD-5019A fracture control. Upon the initial meeting, it was made clear that a formal, documented proposal was not available, and SpaceX was asking for assistance in reviewing their existing data to develop risk-evaluation methods that might be acceptable to NASA programs and projects to descope NDE inspections in lieu of NASA-STD-5019A requirements. A revised scope was formulated to develop a probabilistic analysis method for NASA programs and projects to estimate risk associated with descoping NDE. Additionally, the scope would include efforts to demonstrate the method with SpaceX-provided data to perform sensitivity studies as to the future acceptability of specific SpaceX NDE descope requests and to identify additional supporting data that might be required for approval. Acceptance of a request from SpaceX to deviate from NASA-STD-5019A was out-of-scope for this assessment because this is the purview of specific NASA programs/projects.

6.0 Problem Description and Background

This report presents a methodology to evaluate the risk of primary structural failure if NDE flaw screening requirements are relaxed for wrought metallic materials, which is referred to in this report as an NDE descope. The proposed method estimates the probability of a CIFS defect existing in a part based on a historical record of production NDE inspections that are assumed to be predictive of the flaw existence in future parts fabricated under the same time-invariant process control. The methodology conservatively assumes that the CIFS is equal to the NDE detectable flaw size, and it assumes that if a CIFS defect exists it will lead to primary structural failure with a probability of 1.0. These assumptions were necessary because NDE rejections do not include an estimate of flaw size in typical production databases. The primary motivation for SpaceX in descoping NDE is to reduce production cost and schedule. SpaceX also asserts that unnecessary rework of parts to address NDE findings often introduces additional risks. In some cases, NDE descope may be considered necessary to meet a target production rate.

Spaceflight systems include metallic structural parts that can catastrophically fail if a CIFS defect exists and propagates beyond a structural threshold under the anticipated loads and environments within the part's operational lifetime. These parts are referred to as fracture critical,[6] and undergo a flaw/damage size sensitivity analysis to define the CIFS. A fracture-critical part is considered damage tolerant if no flaws greater than the CIFS are present in the part.

NASA-STD-5019A [ref. 1] provides human-rated spaceflight system requirements for establishing a fracture control plan that relies on design, analysis, testing, NDE, and tracking of fracture-critical parts to preclude catastrophic failure through verification of damage tolerance. As an element of the fracture control plan, metallic fracture-critical parts are screened by NDE in

[6] NASA-STD-5019A Fracture Critical: Fracture control classification that identifies a part whose individual failure, caused by the presence of a crack, is a catastrophic hazard and that requires safe-life analysis or other fracture control assessment to be shown acceptable for flight. A part is fracture critical unless it can be shown that there is no credible possibility for a flaw to cause failure during its lifetime or the part failure does not result in a credible catastrophic hazard. Assessments for fracture critical parts include damage tolerance analysis, damage tolerance test, or defined approaches for specific categories. Parts under this classification receive flaw screening by NDE, proof test, or process control and are subjected to traceability, materials selection and usage, documentation, and engineering drawing requirements.

accordance with NASA-STD-5009C "*Nondestructive Evaluation Requirements for Fracture-Critical Metallic Components*" [ref. 3] requirements for establishing the detectable flaw size for an NDE method, metallic material, and application. NASA-STD-5009C defines 'detectable' as the flaw size that has a 90% probability of detection (POD) with 95% confidence.

The method presented in this report is based on the premise that if the probability of CIFS defect existence in a part is considered sufficiently low and shown to support an acceptable risk of primary structural failure, then rationale for descoping NDE may exist. Note that NASA-STD-5019A does not require a specific minimum reliability. However, for structural parts without redundancy in human-rated systems, meeting NASA-STD-5019A requirements notionally corresponds to a Likelihood of 1 with a Consequence of 5 (i.e., 1 x 5 risk posture) for a human-rated system risk matrix. For the NESC risk matrix for crewed missions [ref. 2] (Appendix A), a Likelihood of 1 represents 6 'nines of reliability', or less than a 1-in-a-million chance of structure failure. However, supporting analysis and test are typically not performed to estimate the reliability. Furthermore, it is acknowledged that acceptable risk of primary structural failure is application/system specific, and therefore, a specific part acceptable risk level is not suggested in this report.

The methodology presented in this report is focused on wrought metallic parts fabricated from aerospace alloys that undergo conventional (subtractive) machining processes. Parts being evaluated for NDE descope must satisfy NASA-STD-5019A fracture-control requirements before applying this method.

6.1 Informal Literature Review of Similar Approaches

NASA programs and projects have accepted NDE descoping fracture-critical parts in human-rated spaceflight systems on a case-by-case basis. In discussions with NASA Fracture Control Board members,[7] it was learned that rationale for these descopes were based on multiple factors and not limited to NDE findings during production. NASA-STD-5019A, Section 6.2.5 provides an overview of factors that are considered for low-risk classification. Motivation for past descopes was based on access restrictions for initial and in-service NDE inspections to be performed without 'significant' disassembly. The rationale and approach for accepting NDE descopes was application specific, and acceptance rationale varied. However, common components of the conditional acceptance included augmented process monitoring requirements (i.e., NASA-STD-5019A, Section 8.1.4) and ground and flight fleet leader test and monitoring (i.e., NASA-STD-5019A, Section 7.5.4) [ref. 1]. Similar to the method proposed in this report, the rationale in these cases included a review of past production part NDE inspections. However, none of the reviewed cases provided a quantitative probability of failure or reliability estimate under the NDE descope.

A review of probabilistic structural analysis (PSA) was conducted to identify approaches to quantify risk under reduced NDE flaw screening or NDE descope. In addition, experts from the Federal Aviation Administration (FAA)[8] and Southwest Research Institute (SwRI)[9] were consulted. PSA requires specification of a flaw occurrence distribution in a part. However, examples of how to estimate that flaw distribution were limited, and flaw distributions were

[7] Bence Bartha, Joachim Beek, Jonathan Burkholder, and Gregory Swanson, NASA

[8] Michael Gorelik, Chief Scientist and Technical Advisor for Fatigue and Damage Tolerance, FAA

[9] James Sobotka, Lead Engineer, SwRI

assumed rather than estimated from inspection data. In limited cases where the flaw distribution was estimated, it involved an extensive effort with multiple material suppliers, part manufacturers, and enhanced NDE methods, and the results were applied to specific parts, materials, and features [ref. 4]. None of the PSA examples reviewed considered a reduction of NDE inspections. Conversely, some were reactive to ineffective NDE methods and/or they proposed an increase in NDE inspections.

United States Air Force (USAF) aircraft structural expertise[10] was consulted regarding experience in considering NDE descope for fracture-critical parts. While the USAF has received proposals to consider NDE descope, none were cited as being accepted for operational human-rated spaceflight parts. From the USAF's perspective, sufficient rationale to descope would need to be based on production and flight data with sufficiently large sample sizes, and the rationale would need to demonstrate that it would not degrade the baseline system reliability. It was considered unlikely that sufficient data would exist to allow that baseline risk to be maintained. As shown in this report, any application considered for NDE descope would have to accept a degree of increased risk. Therefore, it was found that these stringent requirements tended to thwart proposals at an early stage of consideration, and therefore a methodology to evaluate the risk of descoping NDE was not developed or evaluated for USAF aircraft.

A limited review of commercial SpaceX systems (i.e., non-NASA) where NDE had been descoped was performed. SpaceX's rationale for descope was based on the observed low number of NDE rejections, expectations of low flaw existence in aerospace grade wrought material processing, and flight history of parts without detected crack initiation and propagation. However, their analysis was largely qualitative, lacked rigorous interrogation of the database, and did not include statistical estimation with confidence bounds that were a function of sample size to account for uncertainty. Furthermore, geometric features that occurred on multiple parts and materials were broadly aggregated and descoped for NDE, but a rationale for aggregating across multiple parts was not provided. Lastly, there was no quantitative assessment of risk associated with descoping NDE, and the risk was qualitatively assumed to remain at its baseline level. SpaceX asserted there was no increase in reliability by implementing NDE, based on the assumption of low probability of CIFS defects existing. In other words, if CIFS defects are not present, then NDE will not find them and, therefore, the inspection does not increase reliability. The findings in this report dispute this assertion. It was found that NDE descope on a fracture-critical structural part that meets NASA-STD-5019A requirements increases the risk of primary structural failure if defects larger than the CIFS can exist.

The literature review and consultations did not identify a methodology to quantify the risk associated with descoping NDE.

7.0 Analysis

7.1 Methodology

7.1.1 Preliminaries

Damage tolerance is a common approach to ensuring reliability of fracture-critical parts. Under this framework, an undetected flaw that behaves like a crack is assumed to exist and, via analysis and/or test, is shown to not grow to failure during the required service life. Failure in this case

[10] Charles Babish, Technical Advisor, Aircraft Structural Integrity, USAF

could take on a number of forms, such as a surface crack breaking through the exterior surface of a pressure vessel or a fatigue crack growing through a critical structural support resulting in catastrophic failure. Damage tolerance is typically treated as deterministic; an NDE detection threshold is established as a fixed initial flaw size with binary outcome (i.e., flaw exists/does not exist) and failure is based on a worst-case[11] analysis or test with binary result (i.e., pass/fail). However, damage tolerance is rooted in probabilistic concepts and can be generalized.

Assume the following three events can occur: (1) a flaw of a given size exists, (2) failure is predicted to occur by simulation or test, and (3) no flaws were detected by NDE, denoted as A, F, and D_0, respectively. The diagram in Figure 1 illustrates the interplay between these events. Moving clockwise from the top right, the overlapping regions of the event circles represent the following scenarios:

a) *Detectable flaw*: If only events A and F occur, then a flaw of a critical size exists that would result in failure during the anticipated service life, but it would be detected by NDE, resulting in removal of the part from service.
b) *Damage tolerant*: If A and D_0 occur but F does not, then a flaw exists, and it was missed by NDE. However, this flaw is not predicted to fail, meaning the part is tolerant to the existence of that flaw.
c) *Nonexistent flaw*: If F and D_0 occur but A does not, then there is a flaw size that would be life-limiting, but a flaw of that size does not exist in the part.

Any region outside of D_0 for which one or fewer events occur simultaneously would be considered a false positive inspection, resulting in unnecessary part removal or rework. The primary concern is the center triangular region of overlap that defines the existence of a CIFS defect that is not detected. The conditional probability formula represents the failure probability associated with this center region as:

$$P(F, D_0, A) = P(F|D_0, A)P(D_0|A)P(A) \tag{1}$$

where, $P(A)$ is the probability that a flaw of a given size exists, $P(D_0|A)$ is the probability that this flaw will be missed by NDE, and $P(F|D_0, A)$ is the probability that a flaw this size exists and is missed by NDE will fail. In practice, the conditional probability of failure $P(F|D_0, A)$ is independent of whether the flaw is detected, such that $P(F|D_0, A) = P(F|A)$. The multiplication of these terms yields the joint probability $P(F, D_0, A)$, or the part failure probability.

[11] For a worst-case analysis or test, the flaw location, shape, aspect ratio, and orientation are chosen such that they represent an enveloping, worst-case crack growth condition. For additional information, see Reference 1.

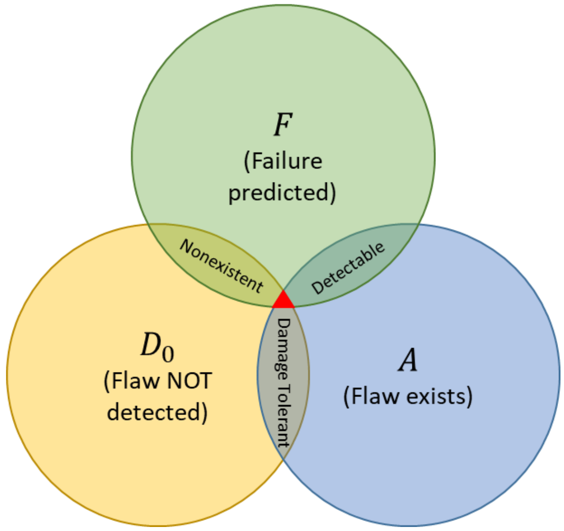


Figure 1. Venn Diagram Representing Three Possible Events Related to Damage Tolerance of a Fracture-Critical Part
Note the red region represents part failure.

Without loss of generality, the concept of flaw size can be reduced to a scalar and represented on the positive real line, as shown in Figure 2. Deterministic damage tolerance as defined in NASA-STD-5019A relies on the following two assumptions:

1. The probability of missing a flaw larger than or equal to the detectability threshold, a_{NDE}, is highly unlikely (i.e., $P(D_0|a \geq a_{NDE}) \approx 0$).[12]
2. The probability of a flaw less than or equal to a CIFS, a_{CIFS}, failing during the service life is highly unlikely (i.e., $P(F|a \leq a_{CIFS}) \approx 0$).

In NASA-STD-5009C and NASA-STD-5019A [refs. 1, 3], the detectability threshold is defined as the flaw size for which POD is 0.90 with 95% confidence. By definition, the CIFS is the largest flaw size for which a worst-case simulation or test does not predict primary structural failure. It is noteworthy that there is typically no assumption regarding the underlying probability of flaw existence. However, it can be assumed that $P(A) = 1.0$ without impacting the analysis outcome. If the regions of zero probability overlap as shown in Figure 2, then Equation (1) evaluates to zero for all flaws, and the part is deemed damage tolerant.

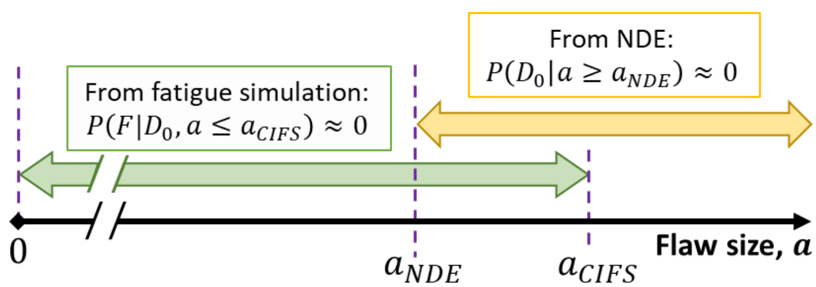


Figure 2. Representation of Flaw Size in One Dimension with Regions of Assumed Probabilities Equal to Zero Highlighted

[12] Crack size has a physical upper bound defined by part geometry that is ignored in this notation for simplicity.

However, there are risks associated with accepting assumptions 1 and 2, and the following mitigation strategies are relied on in the practice of NASA-STD-5019A. First, the potential for non-conservatism in the analysis (i.e., it is possible that $P(F|a \leq a_{CIFS}) > 0$ for an assumed CIFS) is mitigated by making worst-case assumptions and applying scatter factors (e.g., redefining service life as a service factor of 4 × the planned service life [ref. 1]). Second, the potential for missed flaws due to defining a_{NDE} as a flaw size at which POD < 1.0 is mitigated by the fact that the POD is assumed to approach 1.0 as flaw size increases. The larger the margin between a_{CIFS} and a_{NDE} (i.e., Figure 2 overlap regions), the more likely CIFS defects will be detected.

7.1.2 Relative Risk of NDE Descope

The generalization to a probabilistic damage tolerance framework enables a straightforward evaluation of risk associated with descoping NDE. When inspections are removed, the right region of Figure 2 (i.e., the contribution from NDE) is not zero, and the associated probability of failure formula becomes:

$$P(F, A) = P(F|\, A)P(A) \tag{2}$$

The ratio of Equations (1) and (2) provides a relative risk associated with descoping NDE:

$$\frac{P(F, A)}{P(F, D_0, A)} = \frac{P(F|\, A)P(A)}{P(F|D_0, A)P(D_0|A)P(A)} \tag{3}$$

The probability of failure is the same whether the flaw was missed by NDE, or no inspection was conducted, $P(F|A) = P(F|D_0, A)$. Canceling other equal terms and rearranging the equation yields:

$$P(F, A) = \frac{1}{P(D_0|A)} \times P(F, D_0, A) \tag{4}$$

The risk associated with descoping NDE is proportional to the baseline risk multiplied by scale factor $1/P(D_0|A) = 1/(1 - POD)$. In other words, the relative risk increases with increasing detection capability of the NDE method being descoped.

As an example, if the worst-case POD is 0.900 for the region that is not covered by a zero probability of failure in the descoped case (e.g., $A = a > a_{CIFS}$), then the risk of failure in a descoped scenario is at least 10x greater than it was with NDE inspections. For NDE methods that provide POD → 1.0 as $a \to a_{CIFS}$, this results in a relative risk approaching infinity, as illustrated in Table 1. The increase in relative risk is intuitive, and this finding is presented as a warning while acknowledging that the absolute, rather than relative, magnitude of the risk should drive program and project decisions related to NDE descope.

Table 1. Relative Risk Scaling Factor Associated with NDE Descope

$POD = 1 - P(D_0\|A)$	0.900	0.990	0.999	1.000
Risk Factor (lower bound)	10x	100x	1000x	∞

7.1.3 Absolute Risk of NDE Descope

As discussed, the basis for damage tolerance relies on coverage of all possible flaw sizes through NDE- and test/simulation-based screening of flaws to compensate for the lack of knowledge regarding $P(A)$. In contrast, estimating the absolute risk associated with NDE descope requires knowledge of $P(A)$. Characterizing $P(A)$ over the entire flaw space requires the estimation of the probability density function ($p(a)$) over flaw sizes, as shown in blue in Figure 3. Accurately characterizing this distribution is difficult since it requires high-sensitivity flaw characterization methods (e.g., high-resolution radiography, ultrasound testing, or destructive serial sectioning and microscopy). Even with state-of-the-art flaw characterization methods, there is a practical limit in describing the flaw size distribution as it approaches zero. In addition, a large sample size is needed to choose an appropriate distributional model. The NDE detection capability for a particular method is often defined in two or three dimensions (e.g., surface crack detectability depends on crack length and depth), further complicating the distributional modeling task by introducing a multi-dimensional probability of flaw existence. It is assumed the estimation of $p(a)$ is intractable in practice due to these complexities and required resources. In contrast, the proposed method avoids the use of $p(a)$ and multi-dimensional complexities and, instead, treats flaw size as a scalar.

If NDE is descoped, the assumption that $P(F|\ a \leq a_{CIFS}) = 0$ for flaws equal to or smaller than the CIFS is still valid. From Equation (2), the failure probability due to NDE descope can be then expressed as:

$$P(F, A) = P(F, a > a_{CIFS}) = P(F|a > a_{CIFS})P(a > a_{CIFS}) \quad (5)$$

Noting that $P(F|a > a_{CIFS}) = 1.0$ by definition,[13] Equation (5) can be simplified to:

$$P(F, A) = P(a > a_{CIFS}) \quad (6)$$

Therefore, the risk of NDE descope is equal to the probability of a flaw size exceeding the CIFS, which is a more feasible quantity to estimate than the full flaw size existence distribution. However, challenges remain since computing this probability requires NDE methods that can accurately estimate flaw size and, potentially, aspect ratio and/or orientation to determine if the flaw size exceeds the CIFS. Again, sample size and cost of such an approach is expected to limit its application.

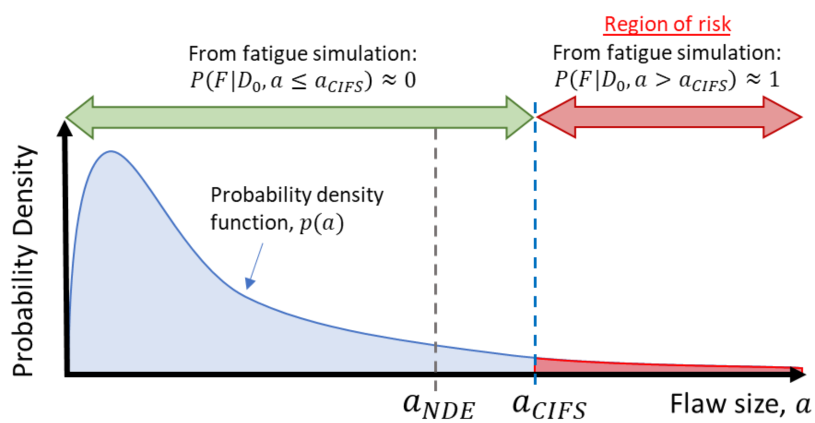


Figure 3. Illustration of the Region of Risk when Descoping NDE
Note the highlighted region of probability density function indicates probability of existence for flaws with size exceeding the CIFS.

---

[13] Previously, the CIFS was defined such that flaws with size less than or equal to the CIFS survive with probability 1.0. If true, then it is likely not true that flaws above fail 100% of the time unless there is zero analysis uncertainty. A conservative definition is assumed to explore potential consequences of descoping NDE.

It is more likely that flaw exceedance data could be inferred from existing, historical NDE databases. However, utilizing such flaw size data is problematic as NDE methods typically only provide hit/miss information (i.e., the damage size is not quantified). In this case, the NDE data would enable evaluation of $P(a \geq a_{NDE})$, but not $P(a > a_{CIFS})$. Assuming that a descope request is initiated for a part that meets damage tolerance requirements in NASA-STD-5019A prior to descope, all flaws satisfy $a_{NDE} \leq a_{CIFS}$ by definition. Thus, the probability of flaws exceeding the detectability threshold is an acceptable, albeit conservative, substitute for the CIFS exceedance probability as:

$$P(a > a_{CIFS}) \leq P(a \geq a_{NDE}) \quad (7)$$

Apart from conservatism, simplicity is an additional benefit in that the complexities of sizing multi-dimensional flaws can be ignored because the data extraction task involves counting the number of NDE rejections in a database of NDE inspection results. A downside of this conservatism for the requestor of the NDE descope is the potential overestimation of risk leading to rejection of the request, especially if there is significant separation between a_{NDE} and a_{CIFS}.

7.1.4 Proposed Approach

This section proposes a framework for evaluating NDE descope risk for human-rated spaceflight parts subject to fracture-control requirements under NASA-STD-5019A. The framework can be applied to a single part or family of parts defined through a principled aggregation approach. Output of this framework is not a decision on whether to descope but is intended to be a tool for the responsible Technical Authority to evaluate expected risk increase associated with an NDE descope request based on available time-invariant process control and NDE data. Further, the framework is not intended to be an NDE capability assessment or advancement and is agnostic to the NDE approach being descoped. A flowchart of the method is provided in Figure 4.

The framework has the following requirements:

1. The part being proposed for NDE descope satisfies fracture control requirements under NASA-STD-5019A.
2. Data are available that comprise binary (i.e., hit/miss) or quantitative sizing results from historical NDE inspections of a production/flight part, or other flaw characterization efforts relevant to the part (e.g., higher-resolution NDE methods than used in production inspections).
3. The wrought material processing, fabrication, and inspection technique being descoped was under verifiable time-invariant process control for the period over which the NDE data were acquired.
4. Historical NDE data are deemed predictive of future flaw existence probabilities, which implies that changes to wrought material processing or fabrication method that could reasonably alter the probability of CIFS exceedance would invalidate risk calculations related to NDE descope.
5. The frequency of process escapes detected only by the NDE method proposed for descope is captured/represented by the historical data.

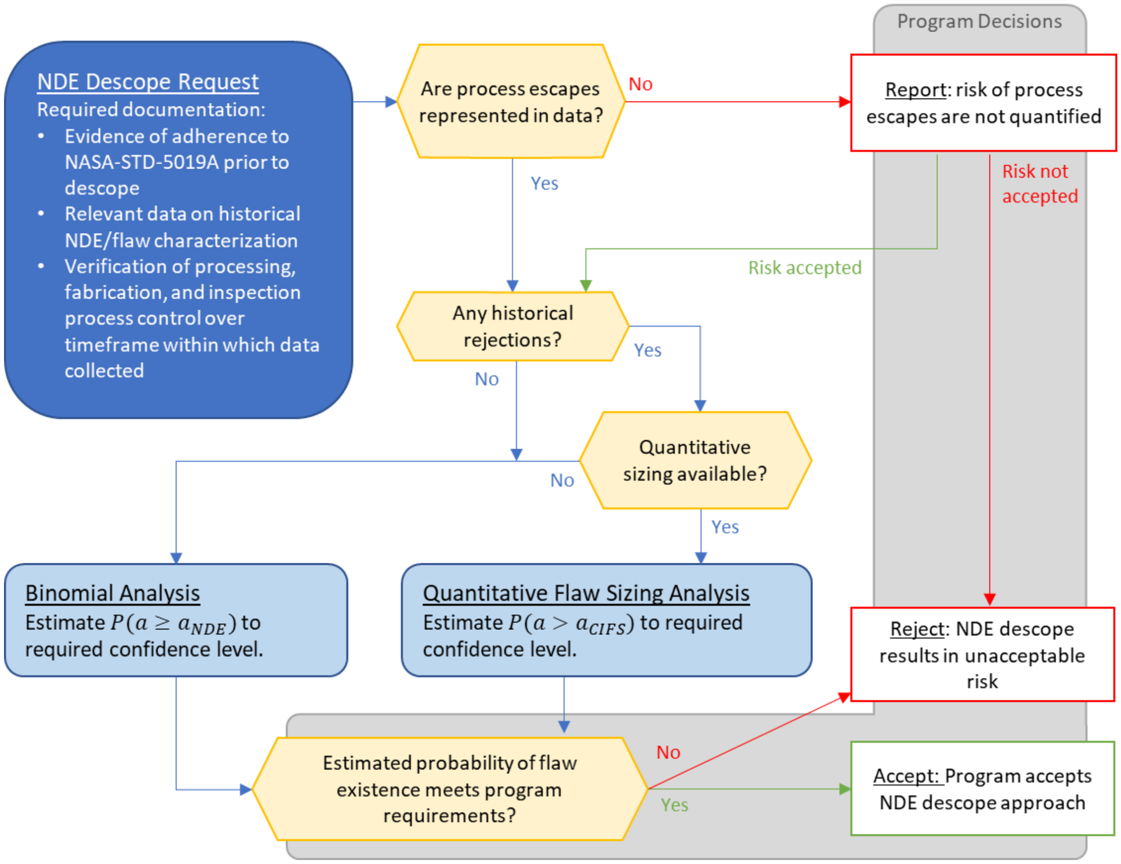


Figure 4. Flowchart of the Proposed Approach for Assessing Risk Associated with NDE Descope

Process escapes are defined as off-nominal processing conditions or events that lead to the introduction of unexpected, potentially life-limiting flaws into the part. The potential for missing such escapes after NDE has been descoped was cited as a primary concern during this NESC assessment. In some cases, process escapes may be detected with methods beside that being considered for descope (e.g., during visual or dimensional inspections). However, if NDE is the only reliable means for identifying flaws, then an assessment of risk associated with descope must include the effect of the undetected process escapes. Since process escapes are expected to be rare, none may be observed when building the NDE database. However, process escapes are expected to be documented when they do occur. For example, evidence should be provided that NDE-based rejections are included in the database prior to the part being reworked, scrapped, or excessed. Time-invariant process control must be maintained such that this frequency of occurrence can be relied upon in future production. The responsible Technical Authority must determine if process escapes are adequately considered.

If initial requirements are met, then the flow of the calculation proceeds as follows. If there are no NDE rejections in the database, then the only option for estimating the probability of detectable flaws existing is to use zero-failure binomial analysis. Otherwise, the analysis approach depends on whether the NDE data are binary (i.e., hit/miss) or quantitative (e.g., flaw sizes can be estimated). For the former, binomial analysis employing a one-sided 95% upper confidence bound accounts for statistical uncertainty associated with the sample size. For the latter, the quantitative NDE data are used to reduce uncertainty in the estimate of $P(a \geq a_{NDE})$, assuming the data are conducive to the application of extreme value theory methods (e.g., peaks over threshold [ref. 5]). Both methods are elaborated on in the following subsections with a brief discussion on data aggregation.

7.1.2.1 Binomial Analysis for Zero NDE Rejections or Hit/Miss Data

The binomial analysis proceeds in the same manner regardless of the number of NDE rejections, including zero. The $(1-\alpha) \times 100\%$ upper confidence bound from Reference 6 is used:

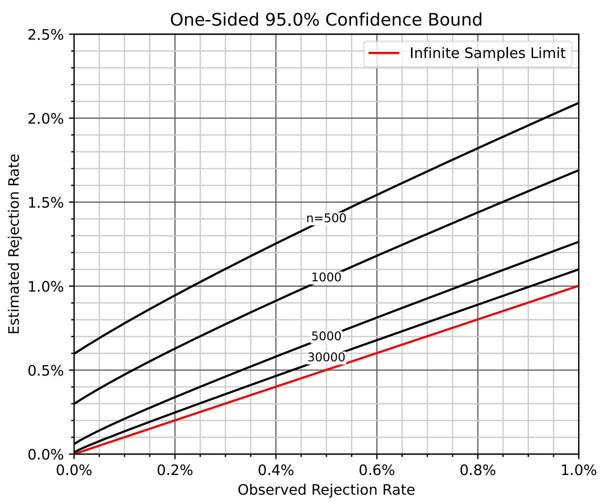
$$\pi_u = qbeta(1-\alpha;\ x+1,\ n-x). \tag{8}$$

where α is the confidence level, n is the total number of inspections in the database, and x is the number of rejections. The function $qbeta$ is the beta distribution quantile function, which can be evaluated using most modern statistical software packages.[14] It is proposed that an appropriate choice for α is 0.05, which corresponds to a 95% upper confidence bound. For intuition about this bound, if the procedure of gathering the historical data over time and calculating the bound is repeated an infinite number of times, then the resulting bound covers the NDE rejection rate 95% of the time.

The upper bound is plotted as a function of observed rejection rate (i.e., the point estimate x/n) and total number of inspections shown in Figure 5. The left plot shows the upper bound, π_u, as a rejection rate, or the estimated probability $P(a \geq a_{NDE})$ as a percentage. The right plot shows the same data, but the ordinate is converted to the number of nines associated with the reliability, defined as:

$$R = 1 - P(F, A) = 1 - P(a \geq a_{NDE}) \tag{9}$$

where number of nines is calculated as $-\log_{10}(R)$. For example, a reliability of 0.999 has three nines and is equivalent to a probability of failure of 0.001. Reliability as a number of nines can be a convenient way to present risk for decision makers.



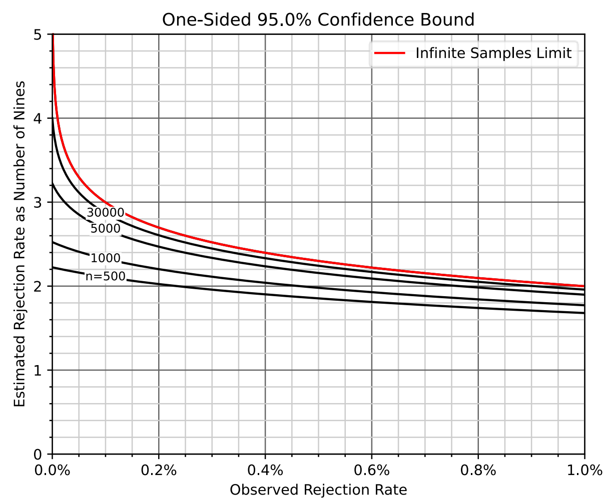


Figure 5. 95% Confidence Upper Bound on Rejection Rate as a Function of Total Inspections and Proportion of Rejections
Note the infinite limit (red) asymptotes at 0.0%, and the limit at five nines is simply a plotting artifact.

[14] For example, the Microsoft® Excel® formula for Equation (8) is BETA.INV(1- α, x+1, $n-x$).

For each plot, there are curves for a given number of total inspections in the historical database. As shown in red, there is a bounding infinite sample limit where the proportion of rejections is equal to $P(a \geq a_{NDE})$ (i.e., there is no uncertainty due to finite sample size). There are diminishing returns in terms of uncertainty reduction as the total number of samples increases toward infinity. These charts can be used to determine if an acceptable risk level has been met, or whether obtaining additional data or performing aggregation could potentially result in a lower risk level.

7.1.2.2 Quantitative Flaw Sizing Analysis

The binomial analysis approach is generally applicable, regardless of whether the NDE method is inherently hit/miss or signal response and is expected to be the common approach assuming most databases only report NDE rejections and not flaw sizing information. However, it is acknowledged that dichotomizing continuous signal-response NDE method to hit/miss reduces information content. Assuming an NDE method has been calibrated[15] to report flaw sizing, a signal-response method could provide additional information on flaw sizes detected to estimate $P(a > a_{CIFS})$. This could reduce the conservatism of assuming the NDE detectable flaw size is equal to the CIFS in Equation (7).

While a binomial approach could be used based on flaw sizing information, other avenues (e.g., extreme value theory [ref. 5]) may be preferred if there is measurable separation between a_{NDE} and a_{CIFS}, and if the number of observed flaws with size exceeding a_{CIFS} is rare compared to the total number of rejections. Here the peaks-over-threshold method [ref. 5] might be used to fit a generalized Pareto distribution to the exceedances, $\delta = a_{observed} - a_{NDE}$, obtained from the quantitative inspection data. While additional details can be found in the provided references, under certain assumptions, the distribution of these exceedances has asymptotic properties that enable a direct estimate of $P(a > a_{CIFS})$ with limited observations of these large flaws. Confidence bounds on the fitted distribution parameters enable estimation of $P(a > a_{CIFS})$ at a desired confidence level.

Even though there are potential advantages (e.g., reducing conservatism), a methodology that incorporates quantitative flaw sizing is anticipated to be problematic to apply in practice due to (i) the need for a signal-response NDE method that has been calibrated to estimate flaw size over the applicable range of flaw sizes and, potentially, (ii) the need to perform a multi-dimensional extreme value theory analysis [ref. 7]. Using hit/miss data to estimate $P(a \geq a_{NDE})$ accounts for multi-dimensional flaws (i.e., it makes the problem one-dimensional) whereas an estimate of $P(a > a_{CIFS})$ may require a probability estimate for a range of CIFS (e.g., at different aspect ratios and flaw depths in the case of surface cracks). These challenges would be exacerbated if attempting to aggregate data from parts/features with varying CIFS, whereas this task is straightforward if using a fixed NDE method and detectability threshold.

---

[15] A calibrated NDE method provides a signal response related to flaw size and uncertainty based on testing with independently characterized flaws (e.g., an estimated 0.060-inch-long crack may an uncertainty of ±0.005 inch). Boundary conditions on the flaw size estimation capability are defined (e.g., a lower threshold as the flaw signal approaches the system noise level and/or an upper threshold based on signal saturation and/or sensor design). An NDE-based flaw sizing calibration is not typically performed in a POD demonstration.

7.1.2.3 Data Aggregation

The proposed framework can be applied to descope a single NDE inspection of a feature on a specific part. However, it may be desirable and/or advantageous to aggregate datasets over multiple parts as this (i) increases value by expanding the number of inspections that are descoped, and (ii) decreases uncertainty in estimates of $P(a \geq a_{NDE})$ by increasing sample size.

Care must be taken when aggregating datasets to ensure the multiple aspects of similitude. If sub-populations have different flaw size exceedance probabilities, then higher probabilities could be masked through aggregation with lower-risk sub-populations, resulting in non-conservative estimates across the aggregated population. To mitigate this, a number of constraints are identified. First, it is suggested that the framework be applied to aerospace-grade materials, which have more consistent processing over an established production history. Furthermore, it is suggested that aggregation be conducted over parts constructed from a single wrought material. Additionally, all supporting data are assumed to be produced via the same NDE inspection method proposed for descope (i.e., sub-population data are gathered with the same probability of detection). The following three elements of rationale are proposed for aggregating over parts, features, and fabrication methods:

- Engineering similitude: The part fabrication processes, features, and materials to be aggregated are expected to result in the same underlying probability of flaws exceeding the NDE detectability threshold based on documented expert elicitation that includes assumptions, limitations, and boundary conditions.
- Statistical similitude: The sub-populations being aggregated have similar NDE rejection rates when accounting for different sample sizes and associated uncertainty. This can be demonstrated using hypothesis testing, or comparison of rejection rate point estimates if the sample sizes are similar.
- Time-invariant process control: Part fabrication and the inspection technique are empirically demonstrated to be consistent throughout the data gathering period and across all sub-populations.

Evidence supporting this rationale may be quantitative (i.e., available data) and/or qualitative (i.e., engineering/fabrication expert elicitation).

7.2 Case Study

To demonstrate the risk-evaluation framework, SpaceX provided access to their production part NDE database and engineer support to help interpret and interrogate this dataset. The goal of the case study was to gain practical experience using the risk-evaluation framework and to document lessons learned. This case study is for demonstration purposes and did not result in formal engineering decisions or recommendations regarding NDE inspection descope under NASA-STD-5019A.

The NESC team was given access to SpaceX's proprietary database that comprised manufacturing, fabrication, and quality-assurance data and included NDE results for NASA-STD-5019A compliant production parts. To scope the effort for this exploratory study, a subset of the hardware was selected, and the production time was limited to a period spanning 5 years. There were six unique part numbers in the subset considered for NDE descope with a varying number of bolt holes (on the order of 100 per part) drilled using computer numerical control (CNC). The parts were fabricated from a common wrought aerospace alloy. After drilling, holes were inspected using Eddy current testing (ET) according to a single NDE specification, and POD was assumed by the team to be consistent across inspections. ET inspections were proposed to be omitted on the selected subset of future hardware, which constitutes an NDE descope proposal. Visual inspections were proposed to remain in production, and thus the term NDE descope does not imply that all NDE would be omitted. Inspections performed by the wrought material supplier remain unchanged in the NDE descope considered in this study. Features with higher variance (e.g., hand-drilled, at-assembly holes) were excluded from NDE descope consideration. This case study analysis evaluated a hypothetical ET descope proposal. It was assumed that the future parts without ET would be under the same time-invariant process control as the historical database. Process control verification is required, and, for this case study, it was performed by SpaceX, but not independently reviewed by the NESC team.

The ET inspections can result in one of three outcomes: pass, report,[16] or reject. When an inspection results in a rejection, an issue ticket is created, and the part is subsequently dispositioned. Corrective action is prescribed when feasible (e.g., a hole might be oversized to remove a crack-like indication). Under descoped NDE, this corrective action would not be taken, resulting in a potential life-limiting flaw entering service. The goal of the case study was to identify the rate of NDE rejections:

$$r_f = x/n, \tag{10}$$

where r_f is the rejection rate, x is the number of NDE rejections, and n is the total number of inspections. The rejection rate was used to estimate the risk of the proposed ET descope using Equation (8). The case study was used as an opportunity to develop and evaluate rationale for aggregating multiple parts and assessing time-invariant process control during the data-gathering period.

The NDE inspection database used in this case study was developed to meet SpaceX's internal process/quality-control goals and not specifically for estimating flaw size exceedance probabilities. In general, archiving NDE historical data is considered best-practice, but it may not be in a format that is searchable or retrievable when applying this risk-evaluation methodology. The raw data format considered in this study was not conducive to computing the probability of flaws exceeding the NDE detectable flaw size in an automated fashion, which is essential when processing inspections numbering on the order of 10^3 to 10^4. For example, when flaws were identified, a decision was made to take corrective action, scrap, or attempt to salvage the affected part. However, the decision process was found to be ad-hoc or unique to the specific case and sometimes in the form of a detailed write-up rather than a database entry that can be queried. This required a case-by-case review of each rejection. Therefore, accurate interpretation and processing of this information was challenging, time consuming, and required specialized expertise. To efficiently apply the risk-evaluation framework in the future, modifications to the design of the production database should be considered. It is expected that this experience is representative of most historical databases that would be used for justifying NDE descope.

[16] Excessive noise or indications that are less than 50% of the calibration reference signal amplitude were considered reportable in the NDE specification but were not considered indications of rejectable crack-like features in this case study. However, the rate of occurrence of reportable indications could be estimated under a similar framework.

7.2.1 Initial Results

Despite reducing scope to a subset of parts over a specific time period, obtaining the necessary data in a usable format was tedious and its interpretation was challenging. A web-browser-based graphical user interface to the database was made available to the NESC team, but accessing the SpaceX database via a more sophisticated programming language-based interface[17] was required to extract the desired dataset. A SpaceX reliability engineer wrote the necessary queries to extract the data requested by the NESC team.

An initial execution of the database query and application of the risk-evaluation framework resulted in 54,807 total inspections with a rejection rate $> 2.0\%$. Through extensive collaboration with the cognizant SpaceX reliability and design engineers, a number of additional filters were developed and applied to the initial data query. For example, pass-through holes were not being proposed for NDE descope, but were inadvertently included in the dataset. Additionally, some data were incomplete or duplicated for certain parts. The cause of the latter issue was inspectors partially filling out or abandoning empty forms after their creation in the reporting system. While inconsequential for the original use case of the database, these issues could artificially inflate or deflate the estimated rejection rate. Appropriate filters were developed to identify and remove these entries from the dataset.

Based on SpaceX's knowledge of the process control evolution from development to production, it was found that the inspection reporting requirements were not standardized during the first 2 years of the 5-year period. This is reflected in Figure 6, which is a graphical representation of the evolution of field names in the ET reporting form completed by the NDE technicians. In this figure, the abscissa represents versions of the form (i.e., starting at version 1.0 and ending at version 5.0). The ordinate represents unique field names used to record the ET inspection results. While the quantities being recorded were consistent over time, the field name changes made it difficult to automate the data extraction and aggregation across multiple form versions. An example of this is the change from "Number of Holes Required" to "Number of Features Requiring Inspection." Upon further review, SpaceX noted the initial 2 years coincided with more frequent changes to part drawings. Therefore, only the final 3 years of data were used to assure time-invariant process control over the data gathering period (i.e., Figure 6 highlighted region).[18]

[17] Structured Query Language (SQL)

[18] The final version number appears to change significantly in Figure 6. However, the field name modifications were easily mapped to those from earlier versions as these were purely name modifications made for clarity plus a few added fields that were not critical to the analysis.

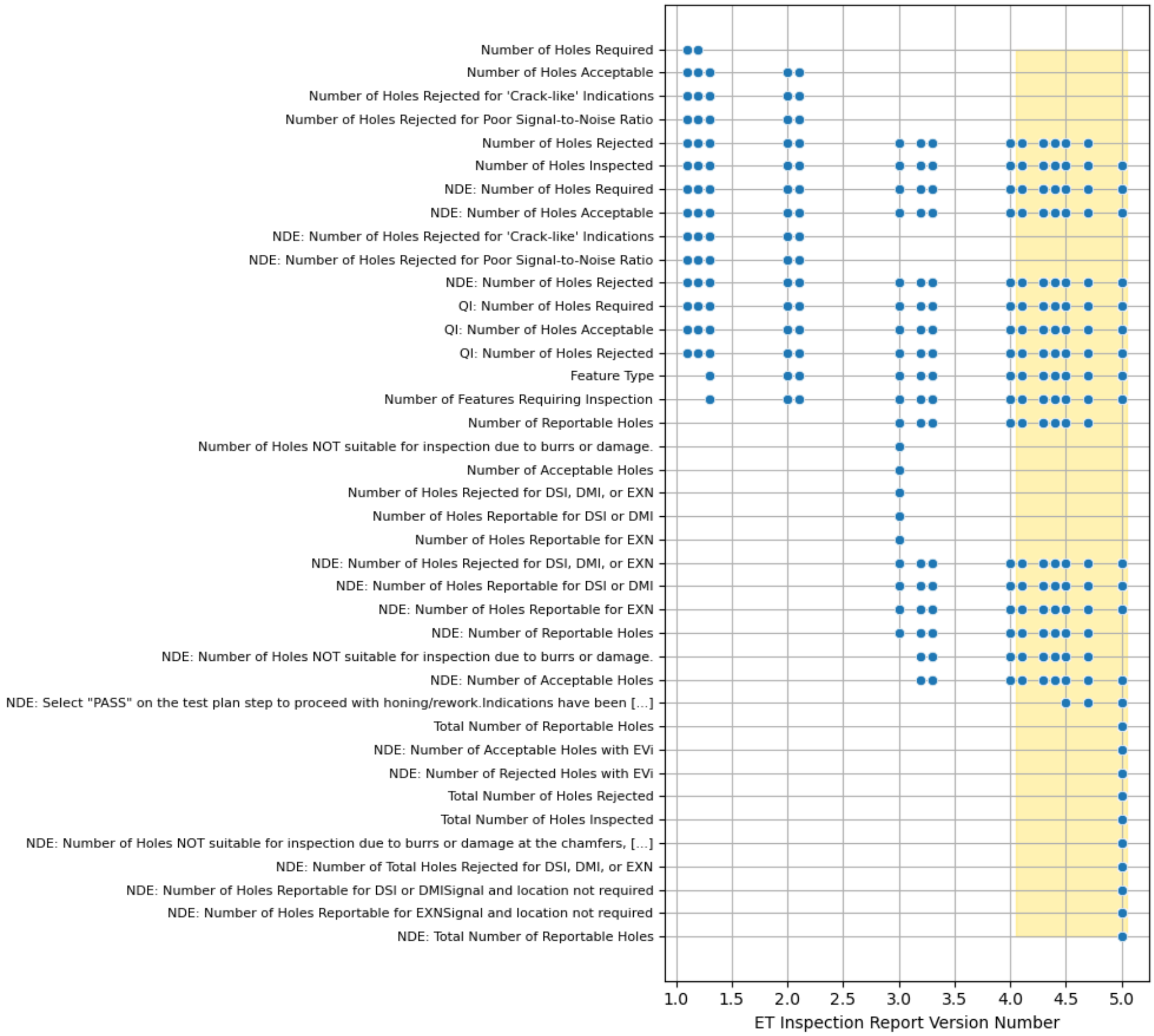


Figure 6. Visualization of the Evolution of Field Names in ET inspection Reports Over a 5-year Period
Note the highlighted region represents the versions used for the case study.

As part of an internal SpaceX evaluation, it was found that initial ET after drilling may reject holes due to surface anomalies when no crack-like flaw existed. As the goal of the risk-evaluation framework was to identify the proportion of NDE inspections that identified life-limiting flaws, it was deemed important to remove these 'false positives' from consideration. A SpaceX fabrication protocol was developed to avoid unnecessary corrective action in these cases. If a hole was rejected on the initial ET inspection, then the hole was honed to maintain the drawing dimensional allowances but remove surface anomalies. After honing, reinspection was conducted via ET. If the hole passed, or the indication changed from rejectable to reportable, then it was assumed the original indication was due to surface imperfections or debris rather than a crack-like flaw. However, if the hole was rejected, then it was assumed that a crack-like flaw existed, and corrective action was required. Utilizing this verification step, rejections on the second ET inspection were considered when evaluating Equation (10) and the subsequent estimation of $P(a \geq a_{NDE})$.

7.2.2 Final Results

The refined dataset contained a total of 33,630 bolt hole inspections over a 3-year period on three parts. Of these, 85 holes were rejected upon initial inspection. Five holes were rejected after honing. However, upon review of the disposition documentation, it was found that one of the passing holes had excess material removed during the honing process (i.e., oversizing the hole from the drawing allowable dimension). Determining whether a crack-like flaw originally existed was not possible and, therefore, it was decided that it was conservative to count this case as a rejectable crack-like flaw. This example highlights the challenges associated with database interpretation.

The final result was 6 crack-like features from 33,630 inspections, resulting in the point estimate $r_f = 0.02\%$. Accounting for uncertainty based on sample size (i.e., Equation (8)) yields a 95% confidence upper bound rejection rate of 0.04% or $P(a \geq a_{NDE}) = 0.0004$ for each hole, which can be expressed as approximately 3.4 nines of reliability. Note whole numbers of nines of reliability are typically used to bin risk. Therefore, this result would typically be reported as 3 nines of reliability.

To add context to these results, the predicted rejection rate and number of nines of reliability associated with $n = 33{,}630$ are plotted as functions of the observed rejection rate in Figure 7. The red square represents the $x = 6$ rejections. The infinite sample size limit is plotted in red, and a hypothetical $n = 100{,}000$ line is plotted in black to provide insight into the diminishing returns of increased sample size. In this particular case, increasing the number of inspections in the dataset to 100,000 (i.e., multiplying by a factor of 3) marginally increases the number of nines to 3.5. However, 3 nines would be reported. At this observed rejection rate (i.e., 0.02%), 4 nines of reliability are not achievable with infinite samples. Even if the rejection rate were 0.01%, 4 nines would require an impractical infinite sample size. It is expected that the rejection rates/sample sizes in this case study are on the order of magnitude of what would be observed/available in practice. Since 2 nines or less would equate to a significant increase relative to baseline risk for NASA Human Spaceflight Programs, a minimum sample size of 5,000 inspections is recommended.

These quantitative results based on historical data provide the responsible Technical Authority with information to evaluate risk of an NDE descope, under the identified assumptions. Other factors that may influence the risk determination include the amount of conservatism associated with assuming that the CIFS is equal to the NDE detectable flaw size (i.e., assuming that there is no margin between a_{NDE} and a_{CIFS} as shown in Equation (7)), an assessment that the process may not produce flaws greater than the CIFS, and/or an expectation that flaws greater than CIFS would be detectable with visual quality inspections. The aggregation rationale should be assessed per requirements outlined in Section 7.1. If the risk is deemed unacceptable, then Figure 7 can be used to determine if increasing sample size is beneficial. Alternatively, expanding aggregation or making process improvements to reduce the rejection rate could be considered. Note that the latter option would require data used to justify future NDE descope be gathered *after* process modifications are implemented and process controls verified (i.e., changing process controls resets the historical database).

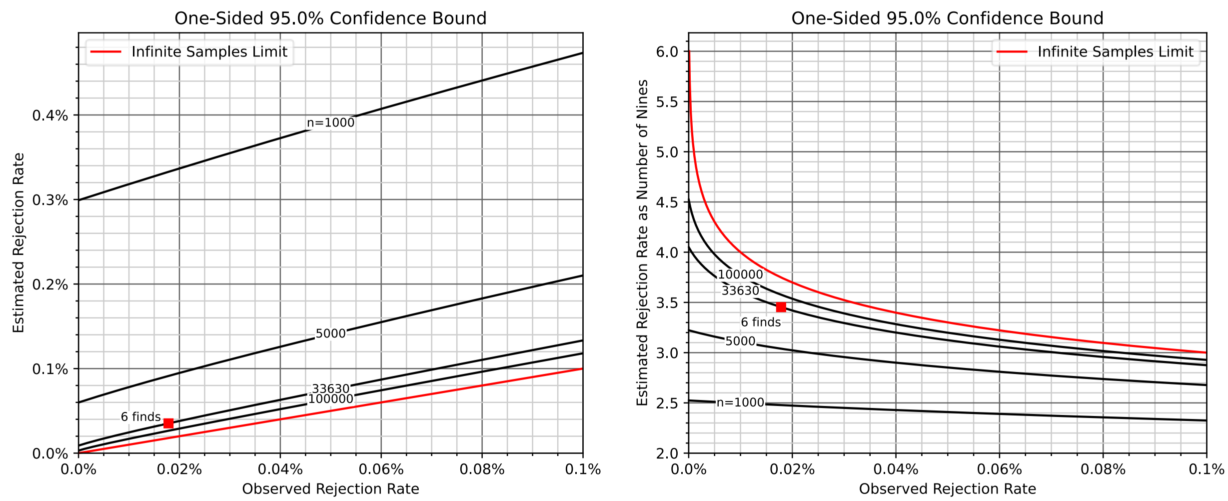


Figure 7. 95% Confidence Upper Bound on Rejection Rate as a Function of Total Inspections and Proportion of Rejections
Note the axis ranges are set based on the results of the case study. Red squares illustrate the rejection rate point estimate.

7.2.3 Discussion on Time-Invariant Process Control

The proposed risk-evaluation framework relies on time-invariant process control to ensure that estimated exceedance probabilities from historical inspections are predictive of future probabilities after a potential NDE descope. Ensuring a consistent process during the data collection period is an important first step in verifying existing controls.

The top panel of Figure 8 shows total bolt hole inspections over the time period considered in this case study. Each bar represents a 10-day window/bin. The middle panel shows the number of bolt holes that were honed after an initial ET rejection, with the same window/bin size. This middle panel represents holes that were initially rejected and honed per the standard operating procedure. Of these honed holes, 6 were rejected, representing the crack-like indications reported in this study and shown in the bottom panel.

It is suggested, at a minimum, a time history of inspection data similar to that shown in Figure 8 be included in a descope request. Such a plot provides insight into whether NDE rejections are uniformly distributed throughout the time period (i.e., consistent processing), or clustered (i.e., failure of process controls). Furthermore, statistical process monitoring methods [ref. 8] can also be employed to quantitatively evaluate whether the rejection rate remains with expected levels over time or indicates a deterioration in process control. It should be noted that this type of information is considered necessary, but not sufficient, for time-invariant process control verification and should be a component of a broader supporting information package (e.g., manufacturing procedures, technician qualifications, etc.).

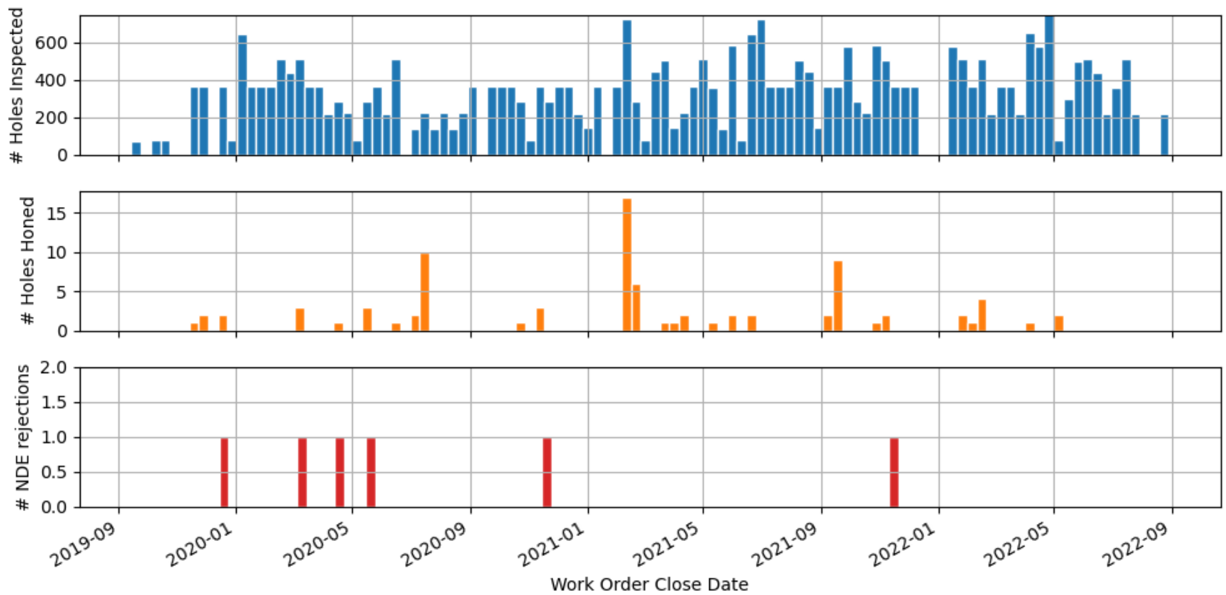


Figure 8. Time History of Inspection Data
Note the top frame shows the number of inspected bolt holes over time with a bin size of 10 days, the middle frame shows the number of honed bolt holes over time with the same bin size, and the bottom frame shows the 6 NDE rejections.

7.2.4 Discussion on Aggregation

Aggregating data across multiple parts can increase value by expanding the number of inspections that are descoped and/or decreasing uncertainty in estimated rejection rates. Ensuring appropriate rationale for data aggregation is a key component to the proposed risk-evaluation framework. A baseline review of aggregation rationale for this case study is provided in this section. It should be noted that a more rigorous review is necessary when making NDE descope decisions, but it was considered out of scope for this case study.

The data were aggregated across three different part numbers. All parts were geometrically similar, made of the same wrought material, and did not include welds. Each part had a different number of bolt holes, and the hole sizes varied within and across parts. However, fabrication protocols and tooling were the same for all holes with the exception of changes to drill bit size. The ET NDE method was the same for every hole inspected. The same ET specification was used for all holes in the database, so it was assumed POD was consistent over this population.

From an engineering perspective, there was no reason to believe that the NDE rejection rate would be different across sub-populations. As previously noted, the original 5-year period was reduced to 3 years to ensure time-invariant process control.[19] From a statistical perspective, low rejection rates and varying numbers of inspections per part made it problematic to quantitatively evaluate the effect of aggregation on the final predicted risk. Figure 9 shows a breakdown of the total number of inspections by the unique part identifier (i.e., “A” through “C”). Parts A, B, and C accounted for similar proportions of holes inspected (left). However, the six rejections were attributed to Part B (middle). In other words, parts A and C had a rejection rate point estimate of 0%, while Part B had $r_f = 0.05\%$ over 12,118 total inspections. In an attempt to account for

[19] Limited evidence was requested to verify this claim due to case study time constraints.

variability in sample size and associated uncertainty, a 95% confidence upper bound was calculated for each sub-population and plotted as number of nines of reliability (right). While Part B results in a lower reliability estimate than A and C, the upper bound on rejection rate is within one order of magnitude (i.e., 1 nine of reliability) across all parts and all would be reported as 3 nines. From this information, it is difficult to determine whether there is an issue with aggregation rationale. In practice, all rejections attributed to a single part may be a warning and necessitate a more rigorous engineering and statistical examination of similitude.

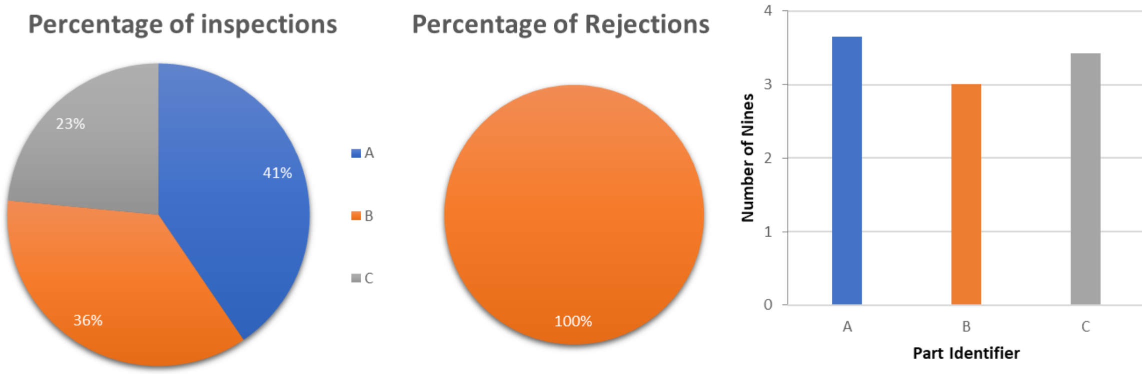


Figure 9. Attribution of Inspections and Rejections to Individual Parts in the Dataset Along with Per-part Calculated 95% Confidence Upper Bounds Presented as Number of Nines of Reliability

8.0 Findings and NESC Recommendations

8.1 Findings

F-1. Descoping NDE on a fracture-critical structural part that meets NASA-STD-5019A requirements increases the risk of primary structural failure if defects larger than the CIFS can exist.

a. The failure risk of a fracture-critical structural part meeting NASA-STD-5019A requirements is proportional to the probability of defects larger than the CIFS prior to initial use.

F-2. A quantitative primary structural failure risk-evaluation framework was developed that relies on an estimate of the probability of NDE-detectable flaw existence for assessing requests to descope NDE for fracture critical wrought structural parts satisfying NASA-STD-5019A requirements.

F-3. The risk evaluation uniquely applies to a wrought part population (family) that uses a single NDE method, is under verified, measurable/monitored time-invariant process control, and demonstrated similarity in probability of NDE detectable flaw existence.

a. Changes to the fabrication process are assumed to change the probability of an NDE detectable flaw and invalidate an existing risk evaluation.

F-4. Similarity rationale to aggregate multiple parts considered for NDE descope relies on qualitative and quantitative engineering/fabrication expert elicitation and time-invariant process control data.

F-5. The risk-evaluation framework does not quantify the effect of rare process escape defects unless their rate of occurrence is overtly captured in the inspection database used to estimate the probability of NDE detectable flaw existence.

F-6. Descoping NDE removes one means of monitoring process control.

8.3 NESC Recommendations

The following NESC recommendations are directed to the NASA programs/projects responding to a request for NDE descope for fracture critical parts:

R-1. Implement the risk-evaluation methodology if NDE descope is proposed on a wrought structural part that meets NASA-STD-5019A. ***(F-2, F-3, F-4, F-5)***

a. Requires rationale to aggregate data from multiple parts based on qualitative and/or quantitative engineering/fabrication expert elicitation and time-invariant process control. ***(F-4)***

b. Requires evidence and/or rationale that the risk-evaluation database has sufficient sample size to support risk posture and is representative (predictive) of future parts and under time-invariant process control. ***(F-4)***

c. Requires evidence that process escape defects, which are assumed to be rare, would be captured in the risk-evaluation database. ***(F-5)***

R-2. Review NDE descope risk evaluation as a component of an overall fracture control evaluation by the NASA Fracture Control Board and/or responsible Technical Authority. ***(F-1, F-6)***

a. Specify alternative requirements in the fracture control plan to ensure time-invariant process control under descoped NDE. ***(F-6)***

9.0 Alternate Technical Opinion(s)

No alternate technical opinions were identified during the course of this assessment by the NESC assessment team or the NESC Review Board (NRB).

10.0 Definition of Terms

Finding	A relevant factual conclusion and/or issue that is within the assessment scope and that the team has rigorously based on data from their independent analyses, tests, inspections, and/or reviews of technical documentation.
Lesson Learned	Knowledge, understanding, or conclusive insight gained by experience that may benefit other current or future NASA programs and projects. The experience may be positive, such as a successful test or mission, or negative, as in a mishap or failure.
Observation	A noteworthy fact, issue, and/or risk, which is not directly within the assessment scope, but could generate a separate issue or concern if not addressed. Alternatively, an observation can be a positive acknowledgement of a Center/Program/Project/Organization’s operational structure, tools, and/or support.

Recommendation	A proposed measurable stakeholder action directly supported by specific Finding(s) and/or Observation(s) that will correct or mitigate an identified issue or risk.

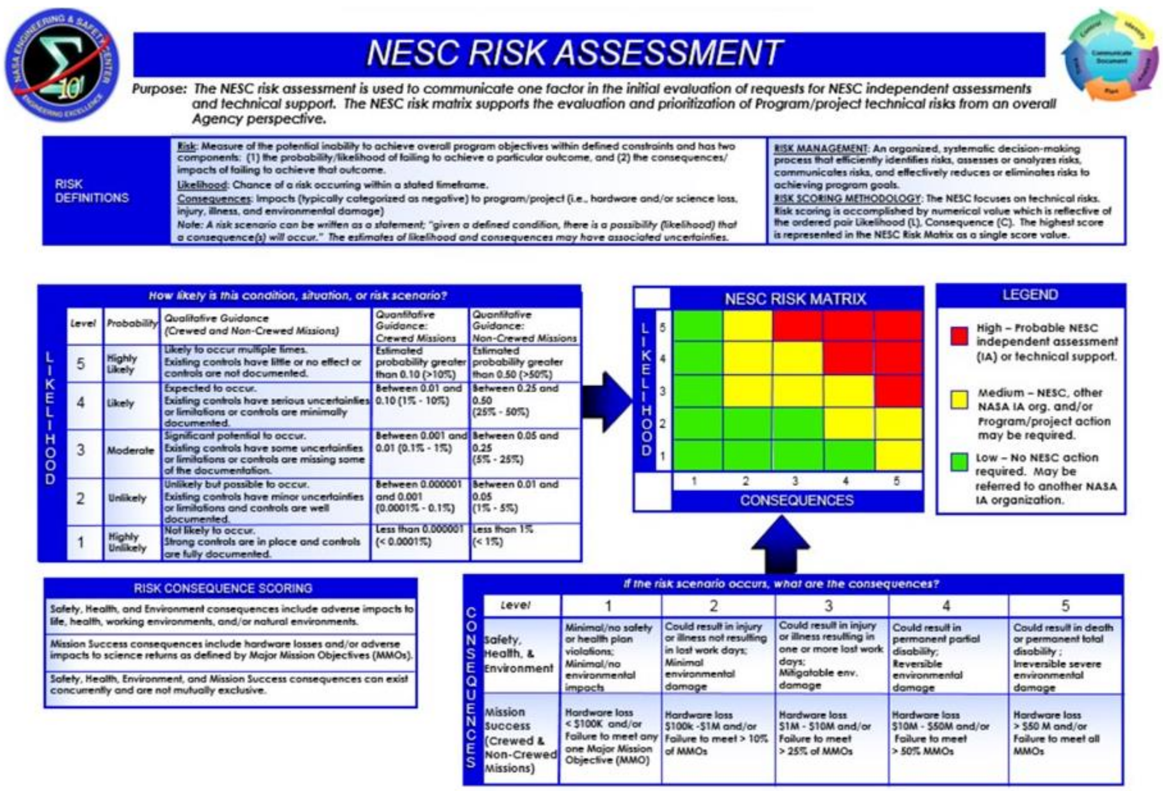
11.0 Acronyms and Nomenclature List

CIFS	Critical Initial Flaw Size
CNC	Computer Numerical Control
ET	Eddy Current Testing
FAA	Federal Aviation Administration
NDE	Nondestructive Evaluation
NESC	NASA Engineering and Safety Center
NRB	NESC Review Board
OCE	Office of Chief Engineer
POD	Probability of Detection
PSA	Probabilistic Structural Analysis
SQL	Structured Query Language
SwRI	Southwest Research Institute
USAF	United States Air Force

12.0 References

1. NASA-STD-5019A: *Fracture Control Requirements for Spaceflight Hardware*. National Aeronautics and Space Administration, 2016.
2. NESC Risk Assessment, July 2016 "NESC Risk Matrix Initial Evaluation Template.ppt," Slide 13, NESC-PR-011-TP-02, Version 2.1.
3. NASA-STD-5009C: *Nondestructive Evaluation Requirements for Fracture-Critical Metallic Components*. National Aeronautics and Space Administration, 2023.
4. Aerospace Industries Association Rotor Integrity Sub-Committee: "The development of anomaly distributions for aircraft engine titanium disk alloys." *Proceedings of 38th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. 1997.
5. Coles, Stuart, et al.: "An introduction to statistical modeling of extreme values." Vol. 208. London: Springer, 2001.
6. Meeker, William Q., Gerald J. Hahn, and Luis A. Escobar: "Statistical Intervals: A Guide for Practitioners and Researchers." 2nd ed., Wiley, 2017, pg. 103.
7. McNeil, Alexander J.: "Extreme value theory for risk managers." *Departement Mathematik ETH Zentrum* 12.5 (1999): pp 217-237.
8. Montgomery, D. C.: "Introduction to Statistical Quality Control," 8th ed., Wiley, 2019.

APPENDIX A. NESC Risk Assessment Matrix



NESC RISK ASSESSMENT

Purpose: *The NESC risk assessment is used to communicate one factor in the initial evaluation of requests for NESC independent assessments and technical support. The NESC risk matrix supports the evaluation and prioritization of Program/project technical risks from an overall Agency perspective.*

RISK DEFINITIONS

Risk: Measure of the potential inability to achieve overall program objectives within defined constraints and has two components: (1) the probability/likelihood of failing to achieve a particular outcome, and (2) the consequences/impacts of failing to achieve that outcome.

Likelihood: Chance of a risk occurring within a stated timeframe.

Consequences: Impacts (typically categorized as negative) to program/project (i.e., hardware and/or science loss, injury, illness, and environmental damage)

Note: A risk scenario can be written as a statement; "given a defined condition, there is a possibility (likelihood) that a consequence(s) will occur." The estimates of likelihood and consequences may have associated uncertainties.

RISK MANAGEMENT: An organized, systematic decision-making process that efficiently identifies risks, assesses or analyzes risks, communicates risks, and effectively reduces or eliminates risks to achieving program goals.

RISK SCORING METHODOLOGY: The NESC focuses on technical risks. Risk scoring is accomplished by numerical value which is reflective of the ordered pair Likelihood (L), Consequence (C). The highest score is represented in the NESC Risk Matrix as a single score value.

How likely is this condition, situation, or risk scenario? (LIKELIHOOD)

Level	Probability	Qualitative Guidance (Crewed and Non-Crewed Missions)	Quantitative Guidance: Crewed Missions	Quantitative Guidance: Non-Crewed Missions
5	Highly Likely	Likely to occur multiple times. Existing controls have little or no effect or controls are not documented.	Estimated probability greater than 0.10 (>10%)	Estimated probability greater than 0.50 (>50%)
4	Likely	Expected to occur. Existing controls have serious uncertainties or limitations or controls are minimally documented.	Between 0.01 and 0.10 (1% - 10%)	Between 0.25 and 0.50 (25% - 50%)
3	Moderate	Significant potential to occur. Existing controls have some uncertainties or limitations or controls are missing some of the documentation.	Between 0.001 and 0.01 (0.1% - 1%)	Between 0.05 and 0.25 (5% - 25%)
2	Unlikely	Unlikely but possible to occur. Existing controls have minor uncertainties or limitations and controls are well documented.	Between 0.000001 and 0.001 (0.0001% - 0.1%)	Between 0.01 and 0.05 (1% - 5%)
1	Highly Unlikely	Not likely to occur. Strong controls are in place and controls are fully documented.	Less than 0.000001 (< 0.0001%)	Less than 1% (< 1%)

NESC RISK MATRIX

LEGEND

High – Probable NESC independent assessment (IA) or technical support.

Medium – NESC, other NASA IA org. and/or Program/project action may be required.

Low – No NESC action required. May be referred to another NASA IA organization.

RISK CONSEQUENCE SCORING

Safety, Health, and Environment consequences include adverse impacts to life, health, working environments, and/or natural environments.

Mission Success consequences include hardware losses and/or adverse impacts to science returns as defined by Major Mission Objectives (MMOs).

Safety, Health, Environment, and Mission Success consequences can exist concurrently and are not mutually exclusive.

If the risk scenario occurs, what are the consequences? (CONSEQUENCES)

Level	1	2	3	4	5
Safety, Health, & Environment	Minimal/no safety or health plan violations; Minimal/no environmental impacts	Could result in injury or illness not resulting in lost work days; Minimal environmental damage	Could result in injury or illness resulting in one or more lost work days; Mitigatable env. damage	Could result in permanent partial disability; Reversible environmental damage	Could result in death or permanent total disability ; Irreversible severe environmental damage
Mission Success (Crewed & Non-Crewed Missions)	Hardware loss < $100K and/or Failure to meet any one Major Mission Objective (MMO)	Hardware loss $100k -$1M and/or Failure to meet > 10% of MMOs	Hardware loss $1M - $10M and/or Failure to meet > 25% of MMOs	Hardware loss $10M - $50M and/or Failure to meet > 50% MMOs	Hardware loss > $50 M and/or Failure to meet all MMOs