1-27-2017

An Archival Analysis of Stall Warning System Effectiveness During Airborne Icing Encounters

John Michael Maris

AN ARCHIVAL ANALYSIS OF STALL WARNING SYSTEM EFFECTIVENESS DURING AIRBORNE ICING ENCOUNTERS

by

John Michael Maris

A Dissertation Submitted to the College of Aviation
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Aviation

Embry-Riddle Aeronautical University
Daytona Beach, Florida
January 27, 2017

AN ARCHIVAL ANALYSIS OF STALL WARNING SYSTEM EFFECTIVENESS DURING AIRBORNE ICING ENCOUNTERS

by

John Michael Maris

This Dissertation was prepared under the direction of the candidate's Dissertation Committee Chair, Dr. Antonio Cortés, and has been approved by the members of the dissertation committee. It was submitted to the College of Aviation and was accepted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Aviation

Dr. Antonio Cortés, Ph.D.
Committee Chair

Dr. Bruce Conway, Ph.D.
Committee Member

Dr. Alan J. Stolzer, Ph.D.
Department Chair, Doctoral Studies

Dr. Dahai Liu, Ph.D.
Committee Member

Dr. Alan J. Stolzer, Ph.D.
Dean, College of Aviation

Dr. Carolina Anderson, Ph.D.
Committee Member

Dr. Michael P. Hickey, Ph.D.
Dean of Research and Graduate Studies

Dr. Wagdi Habashi, Ph.D.
Committee Member

January 27, 2017

ABSTRACT

Researcher: John Michael Maris

Title: AN ARCHIVAL ANALYSIS OF STALL WARNING SYSTEM EFFECTIVENESS DURING AIRBORNE ICING ENCOUNTERS

Institution: Embry-Riddle Aeronautical University

Degree: Doctor of Philosophy in Aviation

Year: 2017

An archival study was conducted to determine the influence of stall warning system performance on aircrew decision-making outcomes during airborne icing encounters. A Conservative Icing Response Bias (CIRB) model was developed to explain the historical variability in aircrew performance in the face of airframe icing. The model combined Bayes' Theorem with Signal Detection Theory (SDT) concepts to yield testable predictions that were evaluated using a Binary Logistic Regression (BLR) multivariate technique applied to two archives: the NASA Aviation Safety Reporting System (ASRS) incident database, and the National Transportation Safety Board (NTSB) accident databases, both covering the period January 1, 1988 to October 2, 2015.

The CIRB model predicted that aircrew would experience more incorrect response outcomes in the face of missed stall warnings than with stall warning False Alarms. These predicted outcomes were observed at high significance levels in the final sample of 132 NASA/NTSB cases. The CIRB model had high sensitivity and specificity, and explained 71.5% (Nagelkerke R^2) of the variance of aircrew decision-making outcomes during the icing encounters. The reliability and validity metrics derived from

this study suggest indicate that the findings are generalizable to the population of U.S. registered turbine-powered aircraft.

These findings suggest that icing-related stall events could be reduced if the incidence of stall warning Misses could be minimized. Observed stall warning Misses stemmed from three principal causes: aerodynamic icing effects, which reduced the stall angle-of-attack (AoA) to below the stall warning calibration threshold; tail stalls, which are not monitored by contemporary protection systems; and icing-induced system issues (such as frozen pitot tubes), which compromised stall warning system effectiveness and airframe envelope protections. Each of these sources of missed stall warnings could be addressed by Aerodynamic Performance Monitoring (APM) systems that directly measure the boundary layer airflow adjacent to the affected aerodynamic surfaces, independent of other aircraft stall protection, air data, and AoA systems. In addition to investigating APM systems, measures should also be taken to include the CIRB phenomenon in aircrew training to better prepare crews to cope with airborne icing encounters. The SDT/BLR technique would allow the forecast gains from these improved systems and training processes to be evaluated objectively and quantitatively.

The SDT/BLR model developed for this study has broad application outside the realm of airborne icing. The SDT technique has been extensively validated by prior research, and the BLR is a very robust multivariate technique. Combined, they could be applied to evaluate high order constructs (such as stall awareness for this study), in complex and dynamic environments. The union of SDT and BLR reduces the modeling complexities for each variable into the four binary SDT categories of Hit, Miss, False Alarm, and Correct Rejection, which is the optimum format for the BLR. Despite this

reductionist approach to complex situations, the method has demonstrated very high statistical and practical significance, as well as excellent predictive power, when applied to the airborne icing scenario.

DEDICATION

This work is dedicated to “The Captain” whose guidance and limitless encouragement have remained with the author long after the Captain took his final voyage.

ACKNOWLEDGEMENTS

I wish to extend my sincere appreciation to my Committee Chair, and recently appointed Associate Dean of the College of Aviation, Dr. Tony Cortés. I thank Dr. Cortés for his sage guidance, excellent leadership, and boundless encouragement throughout the preparation of this manuscript. My work benefitted immensely from the tireless efforts of my esteemed dissertation committee: Dr. Bruce Conway, Dr. Dahai Liu, Dr. Carolina Anderson, and world-renowned icing expert Dr. Wagdi Habashi. My family's patience and encouragement were also instrumental to my completing my doctoral studies: My son John Michael helped me crystalize the central tenets of this treatise during many hours of late-night discussion; my daughter Stephanie, a professional editor, provided much-needed feedback on the manuscript; and my wife Julia, who issued me a four-year pass to pursue my research. Dr. William (Bill) Tuccio made tireless and invaluable contributions as the Subject Matter Expert for the database encoding. Throughout the program, my good friend and colleague Robert Maxson was a patient and thoughtful sounding board for many of the ideas presented in this treatise. I also wish to acknowledge the staff of the Hunt Library for their invaluable and always-timely assistance, and my editor, Ms. Karen Ostrowski, for her sage inputs throughout this document. Several of the illustrations are courtesy of Ludovic Laberge at Marinvent Corporation. Finally, my thanks go to my Department Chair (and recently appointed Dean of the College of Aviation), Dr. Alan Stolzer, without whose quiet support and ultimate endorsement, this work would neither have begun nor reached its fruition.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Airframe icing has caused severe difficulties for aviators since the dawn of human flight. Aircraft are still lost to icing almost a century after Alcock and Brown's record-setting first trans-Atlantic crossing was nearly thwarted by airframe and engine icing (S. Green, Bettcher, Brachen, & Erickson, 1996). Almost 90 years later, and despite enormous advances in aircraft systems, meteorology, and pilot training, icing continues to cause accidents and loss of life. As a result, the issue remained on the National Transportation Safety Board's (NTSB) Most Wanted List of safety improvements for 14 years (NTSB, 2012).

This study addresses the development and testing of a new theoretical Conservative Icing Response Bias (CIRB) model for evaluating the effect of stall warning system performance on aircrew decision-making outcomes during icing encounters, based on a Signal Detection Theory (SDT) framework. The model was tested via an exploratory archival analysis of NTSB and National Aeronautics and Space Administration (NASA) Aviation Safety Reporting System (ASRS) archives. A non-linear Binary Logistic Regression (BLR) multivariate technique was used to perform the analysis and the related hypothesis tests.

The topic treatment begins with an overview of airframe icing and stall warning systems and the relationship between them. This introduction is followed by a historical review of airframe research, case studies, and regulatory considerations, which provide the operational context for the subsequent theoretical material. The theory sections begin with a review of the basic SDT concepts, vigilance tasks, and Bayes' Theorem. These

collectively form the foundation of the CIRB hypothesis, which is introduced and developed in the next section. The CIRB discussion is followed by an overview of the BLR technique and its application to the evaluation of the CIRB hypothesis using the NTSB and ASRS databases. Chapter I concludes with a review of the delimitations, limitations, and assumptions inherent in the research. The literature review (Chapter II) and methodology section (Chapter III) follow the same general format and sequence as the first chapter. Results and Conclusions are presented in Chapters IV and V, respectively. As airframe icing is the catalyst for every aspect of this proposal, the discussion necessarily begins with a review of icing and its effect on airfoils and the aircraft to which they are attached.

Airframe Icing

Airframe icing adds weight, affects controllability, and, most seriously, can severely compromise an airfoil's ability to create lift "by an unknown amount" (Zeppetelli & Habashi, 2012, p. 612). Innocuous-seeming icing accretions, similar to a thin strip of coarse sandpaper on the critical leading-edge of an airfoil, can result in a 30% loss of lift and a 40% drag increase (Bergrun, 1995). Worse, such icing can cause a *stall*, which represents a significant degradation of airfoil performance, before the aircraft's stall protection systems can alert the crew. The lack of warning arises because current stall warning systems cannot quantify the influence of ice accretions, and the regulations for artificial stall warning systems simply impose fixed, and possibly inadequate, safety margins for flights in icing conditions (FAA, 2011b). Under certain icing conditions, these pre-set margins have proven inadequate, resulting in a spate of

aircraft accidents and Loss of Control (LOC) incidents. The Federal Aviation Administration (FAA) has struggled to address the problem by issuing more than 200 airworthiness directives (AD) and four regulatory amendments (FAA, 2010e), yet LOC incidents and accidents still occur as a result of airborne icing. An analysis of NTSB, FAA, and NASA ASRS data between 1991 and 2010 revealed that LOC was the leading cause of fatalities in the large commercial jet and business jet sectors, accounting for 4,717 lives lost and 44% of all U.S. business aircraft accidents during the period (Veillette, 2012). The same data also showed that icing caused 29% of these fatal LOC events. Weener (2011) noted a continuing safety threat posed by airborne icing, and the icing issue remained on the NTSB's *Most Wanted List* of transportation safety improvements for small and large aircraft from 1997 to 2011 (NTSB, 2012b).

Commuter airlines are even more susceptible to icing effects than the group studied by Veillette because commuter carriers generally operate lower performance aircraft, over shorter sectors, at lower altitudes than the major carriers. These factors confine commuter operators to the atmospheric strata where icing is prevalent. The risk is compounded because commuters are exposed to a greater number of takeoffs and landings than long-range operators, when aircraft performance margins are at their lowest and the stall probability is at its highest.

Airframe icing remains a continuing problem despite the best efforts of the FAA and NTSB. Petty & Floyd (2004) explored NTSB data from 1982 to 2000 and noted that icing accident rates for commuter aircraft operations had not declined during the period. Using U.S. NTSB 2006 – 2010 online accident synopsis data, Appiah-Kubi, Martos, Atuahene, & William (2013) recorded 228 accidents and 30 incidents across all aircraft

categories related to aerodynamic events arising from icing encounters. More recently, the FAA published its 112th icing-related AD, which barred known-icing operations by approximately 4,200 small U.S. registered general aviation (GA) aircraft, based on the ongoing losses caused by icing within this group through a period spanning 30 years (FAA, 2014g; K. Lynch, 2014). Veillette (2012) conducted a study of business aviation LOC accidents between 1991 and 2010. Citing Veillette's results, the National Business Aviation Association (NBAA) found "no cause was nearly as prevalent as aerodynamic stalls." The NBAA also noted that 9 of the 31 stall events (29%) were related to airspeed management in icing conditions (NBAA, 2015). In a study conducted by Boeing of fatal accidents to the worldwide commercial jet fleet, LOC accounted for twice the number of fatalities (1,656) of the second leading cause, Controlled Flight into Terrain (CFIT), which cost 803 lives (Boeing Commercial Airplanes, 2015, p. 22). It is reasonable to assume that the relative incidence of icing accidents in the Boeing study mirrored the 29% observed by Veillette (2012).

In an effort to reverse this trend, there has been a constant evolution of the laws pertaining to icing certification, but Zeppetelli & Habashi (2012) note a number of shortcomings in the icing certification regulations, including subjective and conflicting icing terminology and problems with unrepresentative ice shapes used for certification flight-testing. Furthermore, many of the icing issues cannot easily be addressed by regulatory action alone because they are linked to limitations of Angle of Attack (AoA) - based stall warning systems, to which the discussion now turns.

Stalls, Angle of Attack, and Airfoil Contamination

This section introduces a number of key aerodynamic concepts that have a direct bearing on the icing topic. Figure 1 illustrates several important airfoil definitions used throughout this work. Of particular significance is the geometric *Angle of Attack* (AoA or α_G), which is the angle subtended between the undisturbed free-stream relative airflow and the wing chord line that joins the leading and trailing edges of the airfoil (Figure 1). The AoA can be positive, as shown in the figure, or negative, when the relative wind impinges on the airfoil from above the chord line. Figure 1 also illustrates a number of differing AoA definitions that have application in the fields of Computational Fluid Design (CFD) and aircraft certification. The Absolute AoA (α_{abs}) is measured from the datum at which the airfoil produces zero lift. For positively cambered airfoils, as depicted in Figure 1, α_{zl} occurs at a negative AoA, so:

$$\alpha_{abs} = \alpha_G + \alpha_{zL} \quad (1)$$

The Induced Downwash angle (ε) is characteristic of all 3-dimensional airfoils operating out of ground effect. This downwash is caused by wingtip vortices that reduce the effective AoA (α_ε) experienced by the airfoil, which, in turn, is the root cause of Induced Drag. The downwash and corresponding AoA decrement vary across the span of the airfoil, as a function of the wing's Aspect Ratio. This is relevant to the present study because the changing effective AoA across the span can affect the location and type of icing accumulation.

For commuter and transport category aircraft with an aft-mounted (i.e., conventional) horizontal stabilizer, the main wing almost invariably operates at a positive

AoA, while the tail normally operates at a negative AoA, particularly during slow speed operations with the main wing trailing-edge flaps extended. The relevance of this factor is that icing often accumulates underneath the horizontal stabilizer where it cannot be viewed or monitored by the crew. In addition, the extension of flaps can counter-intuitively exacerbate the potential for a tail stall, even as the flaps increase the stall margin on the main wing. The effect of these important distinctions between wing and tail stalls is explicitly explored in the research design discussed in Chapter III. Figure 1 illustrates the typical location for airborne ice accretions on the leading edge of the airfoil surface at positive angles of attack.

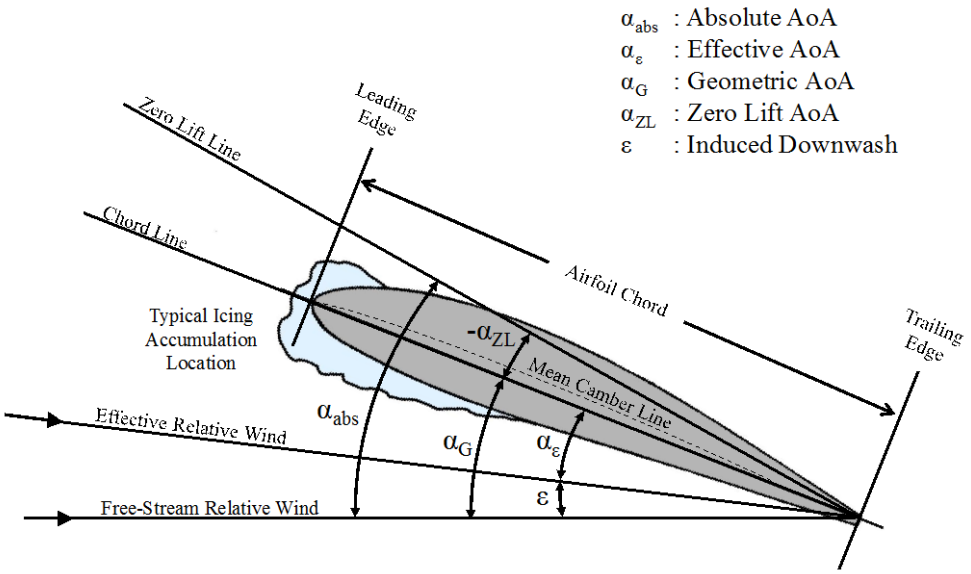

Figure 1. Airfoil definitions. The angle of attack is the angle between the relative wind and the airfoil chord line. A positive AoA is shown in the figure. The icing location and form factor are shown for illustrative purposes only. Actual icing can accrete in an almost limitless number of shapes and coverage extent.

For a given wing planform operating at a fixed Equivalent Airspeed (EAS), the lift force is directly proportional to AoA up to a critical AoA, as shown in Figure 2. The peak of the Lift Coefficient (C_L) vs. α curve, C_{Lmax}, defines the stalling AoA of the airfoil, which is characterized by significant separation of the airflow across the low-pressure side of the airfoil. Stalls result in a significant loss of lift and an increase in drag that can lead to severe performance, stability, and control problems. This study only examines wing and tail stalls and not engine compressor stalls, which are related phenomena that affect the airfoils of the compressor stages of a turbine engine. This distinction between airfoil stalls and engine stalls is not simply semantic; it had important ramifications for the design of the sampling process, in order to avoid the inclusion of the inapplicable engine-related events that would confound the analysis.

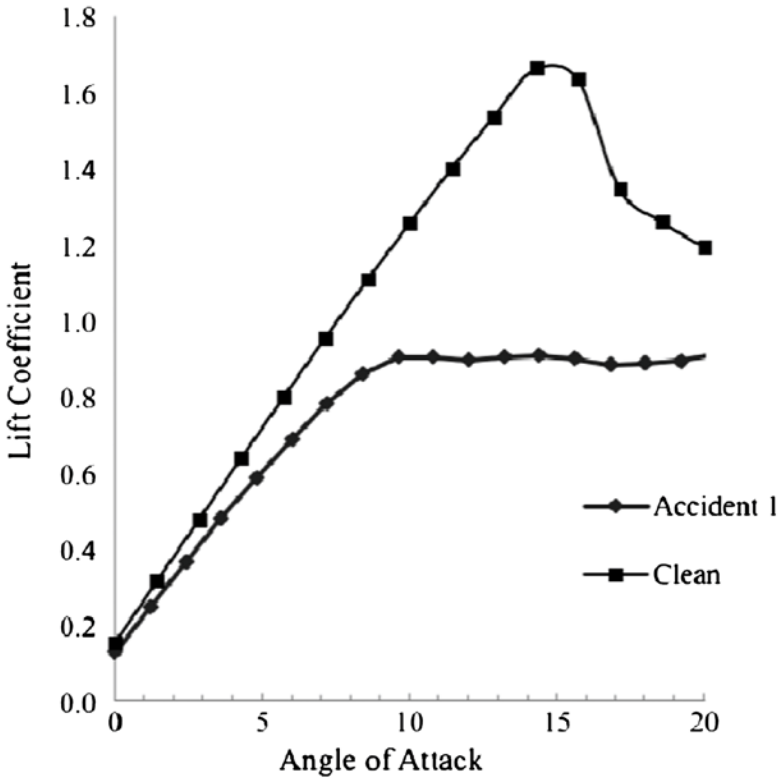

Figure 2. Lift Coefficient (C_L) vs. AoA (α). An airfoil in a *clean* state (i.e., without any icing accumulation) can achieve a higher critical AoA and a greater corresponding lift coefficient than an airfoil contaminated with leading-edge ice, as exemplified by the *Accident 1* trace (Zeppetelli & Habashi, 2012, p. 618). Reprinted with permission.

Stall Warning Systems

Aircraft stall warning systems are designed to alert the crew of an imminent stall by issuing some combination of tactile, visual, or aural cues at a safe margin below the critical AoA (i.e., at a speed margin above the stalling speed). Most contemporary stall warning systems rely on AoA sensors because the critical AoA remains essentially constant, regardless of the aircraft's speed, weight, load-factor, attitude, or other parameters (FAA, 2004, p. 4-3). Unfortunately, these AoA sensors are usually mounted

on the fuselage sides near the nose of the aircraft, where they cannot directly sense the local aerodynamic conditions experienced by the wing or tail surfaces. This is an important shortcoming of AoA-based stall warning systems because even small amounts of airfoil contamination can drastically and unpredictably reduce the stalling AoA from the calibrated warning-threshold value. Current stall warning systems are also unable to provide real-time indication of aircraft performance and controllability margins in icing conditions. The accident record is replete with instances of aircraft control being lost due to the effects of icing, before any warning was presented to the crew. For example, the American Eagle 4184 Roselawn accident was a classic instance of an aircraft departing controlled flight due to airframe icing, prior to the actuation of the stall warning system (NTSB, 1996b). Occasionally, the safety margins built into AoA systems for icing encounters can make the systems too sensitive, generating false stall warning alarms. The Colgan Air 3407 accident sequence began with a 20-knot premature stall warning indication, caused by the incorrect setting of the stall warning reference speed switch by the crew in response to the icing conditions (NTSB, 2010a). This switch is used to recalibrate the stall warning threshold speed to account for the higher approach speeds used in icing conditions. Transport Category aircraft approach speeds are based on a reference landing speed (Vref) (FAA, 2010a, §1.2), which is the minimum acceptable calibrated approach airspeed that must be maintained to a point 50 feet above the landing threshold. Vref must not be less than 1.23 times the stall speed for the selected landing configuration in non-icing conditions (FAA, 2009c, §125). With the datum switch set to the *increase* position, the stall warning system activates at a higher speed to accommodate the greater required speed safety margin. Unfortunately, the Colgan crew

used the standard approach speed, which was 13 knots slower than the stick shaker activation threshold caused by the miss-set switch. This resulted in an unexpected stall warning stick-shaker actuation – an SDT False Alarm (FA) – that startled the crew into a response more suited to a tail stall than a main wing stall. The crew raised the flaps, applied aggressive nose up control inputs, and failed to apply maximum power. These inappropriate inputs started the chain of events that led to a main wing stall, loss of control, and the destruction of the aircraft with the loss of all on board (NTSB, 2010a). Although ice was not a direct contributing aerodynamic factor for the Colgan accident, the FA generated by the stall warning system and the crew's perception of the icing severity were both clearly pivotal to the catastrophic outcome. The American Eagle and Colgan Air case studies clearly illustrate why stall warning FAs and missed warnings in icing conditions were important factors to be considered for the upcoming analysis.

Tail Stalls

Aircraft horizontal and vertical stabilizers (collectively called the *empennage*) are airfoils, like the main wing, and are therefore capable of stalling. In particular, the horizontal stabilizer is more susceptible to icing effects than the main wing because its airfoil has a smaller leading-edge radius (i.e., a sharper leading-edge) than the wing, which increases the stabilizer's efficiency as an ice collector (F. T. Lynch & Khodadoust, 2001, p. 760; Manningham, 1997). As a result, the tail can collect significant ice while the wing remains ice-free. This is a serious problem because the empennage is typically outside the field of view of the aircrew so icing can accumulate undetected, particularly on the horizontal stabilizer's critical underside suction surface. As a complicating factor,

there is no Federal Aviation Regulation (FAR) certification requirement for tail stall warning systems, and none are currently installed on 14 CFR Part 23 or 14 CFR Part 25 aircraft (FAA, 2014e). The lack of a tail stall warning system is important because the initial symptoms of a tail stall can be subtle and hard to distinguish from the normal buffeting and airframe vibrations felt during icing encounters. Unfortunately, tail stalls may be triggered by pilot actions used to *recover* from wing stalls; wing stall recovery generally entails the application of maximum power combined with a firm lowering of the aircraft's nose to reduce the wing's AoA, but these actions may precipitate or aggravate a tail stall (NASA, 1999; Ratvasky, Van Zante, & Riley, 1999). When faced with stall warning indications or an incipient LOC during an icing encounter, the aircrew must determine which type of stall is imminent while under intense time pressure. Bragg (2002) notes that current stall warning systems fail to present aircrew with "processed aircraft performance degradation information" that is vital for the resolution of the wing/tail stall ambiguities during icing encounters. Such incomplete, and possibly misleading, stall warning system information could be expected to influence the outcome of icing encounters, but evaluating this premise required the development of a suitable theoretical framework that could generate testable hypotheses. SDT provided exactly the required framework.

Signal Detection Theory

SDT addresses an observer's ability to discriminate an ambiguous signal from background noise, using a simple binary criterion: The observer decides whether the perceived stimulus indicates the presence or the absence of a target. One of the first

practical applications of SDT related to the challenging task of discriminating real targets (signals) from the ambient noise on early radar displays. There are four permutations of target-state and observer decision states that SDT addresses, as shown in Table 1.

Table 1

SDT Decision Outcome Permutations

	Observer's Decision	
True Signal State	Target Present	No Target Present
Signal present	Hit	Miss
No signal present	False Alarm (FA)	Correct Rejection (CR)

Note. Adapted from "Signal Detection Theory," by H. Abdi (2009), in B. McGaw, P. L. Peterson, & E. Baker (Eds.), Encyclopedia of Education (3rd ed.), pp. 1-10.

In addition to the literal interpretation of physical parameters such as radar target returns, SDT also supports the processing of abstract or metaphorical signals (Abdi, 2009, p. 2). This broader interpretation applies to the ability of a stall warning system to correctly identify an impending stall, based on whatever engineered inputs the system receives. The decisions of the stall warning to activate, in relation to the true stall-state of the aircraft, have the same four outcome permutations as described in Table 1 with *stall warning system* substituted for *observer*. SDT can also be applied at an even higher level, where the system comprises both the aircrew and the stall warning system. This new system has an additional set of inputs that the stall warning alone does not possess: These are the symptoms or indications of an imminent LOC or stall originating from the aircraft, as perceived by the crew. These cues could include buffeting, control anomalies,

performance losses, etc. The combined aircrew / stall warning system processes these available inputs and produces an output, which was characterized for this study as either a correct aircrew response to the aircraft's true stall-state or an incorrect crew response. The system boundaries and relationships for the hypothesized dual-SDT model are shown in Figure 3. An important aspect of the figure is that the output measure for the combined SDT system is a binary choice between a correct aircrew response and an incorrect response (see the *definition of terms* section for the formal definition of these terms for the purposes of this study). This binary crew response property led directly to the selection of the BLR for the research design.

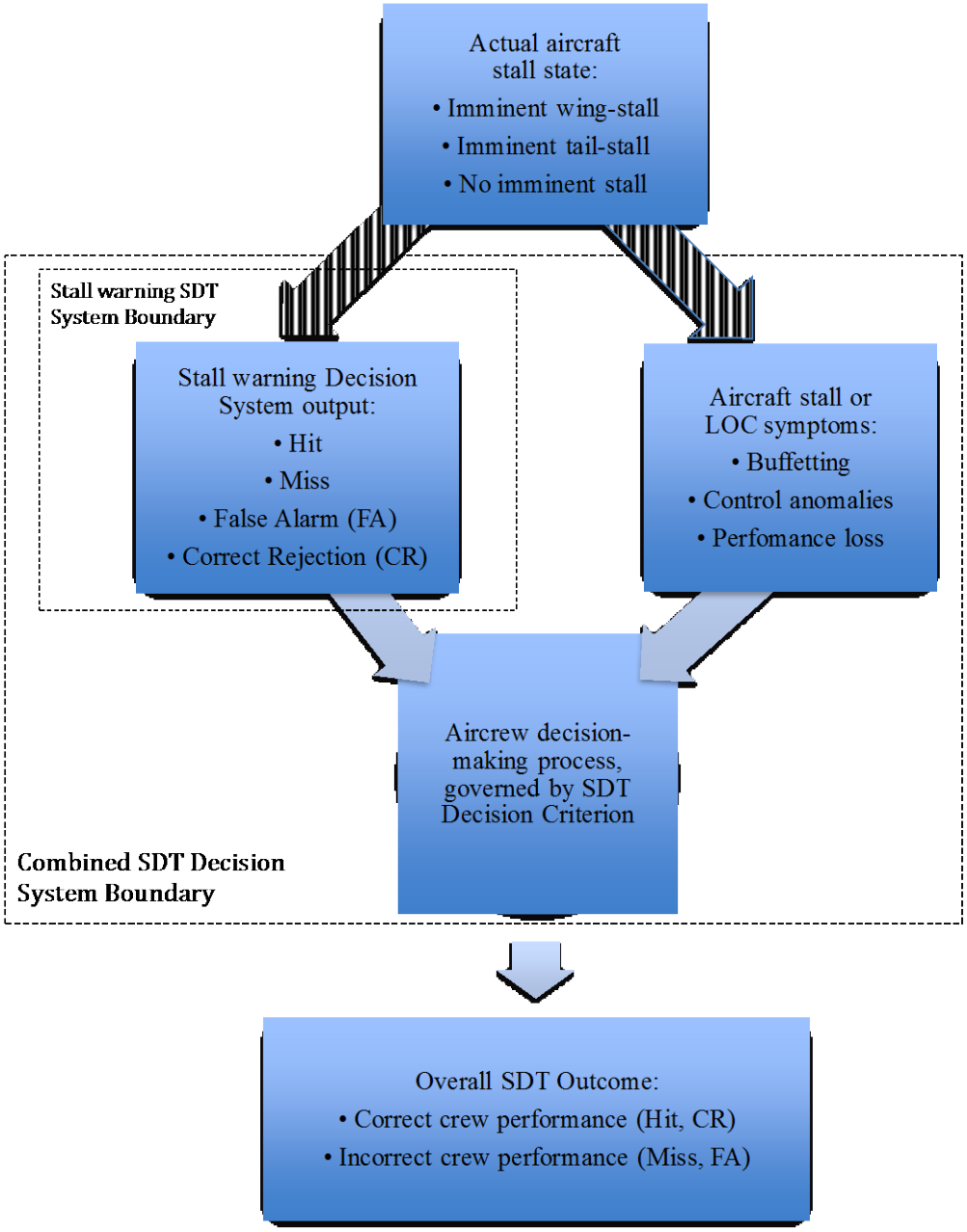

Figure 3. Hypothesized dual SDT system model boundaries. Dashed arrows indicate that the information flow may be incomplete or inaccurate.

The SDT framework includes the related concepts of *system sensitivity, observer response bias, the ideal* (unbiased) *observer*, and the *decision criterion*, which are discussed in detail in Chapter II. This decision criterion sets the observer's minimum threshold for categorizing a stimulus as a signal instead of noise, which directly affects the Hit and FA percentages. SDT defines a hypothetical *ideal observer* as one who sets the decision criterion to minimize the probability of a Miss or an FA. This occurs when the decision criterion is set so that the Miss and FA probabilities are equal (if the signal and noise distributions have identical distributions, as assumed here). If a real observer's decision criterion differs from the ideal observer's, then the former is said to display a conservative or liberal response bias. A conservative response bias is associated with a high criterion, which results in fewer FAs at the expense of fewer Hits and more Misses. Conversely, a liberal response bias results in increased Hits and fewer Misses but more FAs. SDT performance is characterized by this unavoidable tradeoff between Hit and FA rates associated with the setting of the decision criterion. The interplay between these two factors is often shown schematically using families of Receiver Operating Characteristic (ROC) curves for differing decision criteria. (The ROC result for this study is shown in Figure 10 in Chapter IV).

Using the SDT framework, aircrew are assumed to evaluate aircraft cues and stall warning alerts that exceed the decision criterion as a stall condition. Stimuli that fall below the threshold would be categorized as a no-stall condition, and any attendant stall warnings would be judged as FAs. In order for such an analysis to provide useful predictions, the processes that aircrew use to set their decision criterion must be

understood. A brief exploration of Bayes' Theorem and its relationship to vigilance tasks provides an insight into the mechanisms that might be involved.

Vigilance Tasks and Bayes' Theorem

Vigilance tasks may be characterized as those that require operators to monitor systems for long periods while trying to identify phenomena that are infrequent, unpredictable, and possibly insidious in their onset. The majority of icing-related LOC events occur unexpectedly during cruise flight (Appiah-Kubi et al., 2013; Petty & Floyd, 2004), which is normally a low-workload period for the aircrew. Under normal circumstances, the chance of encountering a stall in cruise flight is very small, so the stall monitoring activity can be characterized as a vigilance task. In these circumstances, an operator's decision criterion is calibrated according to the low expected likelihood of the event (e.g., the wing or tail stall). As already discussed, icing can increase the stall probability substantially, which has important ramifications that are quantified by Bayes' Theorem. The theorem is discussed in greater detail in Chapter II, but in its simplest formulation, it indicates that conditional probabilities are not commutative; instead, they depend on the a-priori likelihood of the event, as indicated by the following example: Is the probability of rain, given the occurrence of clouds, the same as the probability of clouds, given the occurrence of rain? Clearly not. Even a cursory analysis reveals that these probabilities could differ, possibly by orders of magnitude. For example, clouds are almost always present when it is raining, so:

$$P(\text{Clouds} \mid \text{Rain}) = 1 \tag{2}$$

Where A | B indicates the conditional probability of A, given the actual occurrence of B. Conversely:

$$0 \leq P\,(\text{Rain} \mid \text{Clouds}) \leq 1 \tag{3}$$

The conditional probability of rain, given the occurrence of clouds, would be near unity during the rainy season in the tropics and near zero for a dry desert region. The conditional probability is therefore strongly determined by the a-priori probability of both the rain and the cloud events, as indicated by Bayes' Theorem:

$$P\,(Rain \mid Clouds) = \frac{P\,(Clouds \mid Rain)\;\; P\,(Rain)}{P\,(Clouds)} \tag{4}$$

As a result:

$$P\,(\text{Clouds} \mid \text{Rain}) \neq P\,(\text{Rain} \mid \text{Clouds}) \tag{5}$$

Bayes' formula is equally applicable to the relationship between stalls and stall warnings:

$$P\,(Stall \mid Stall\ Warning) = \frac{P\,(Stall\ Warning \mid Stall)\;\; P\,(Stall)}{P\,(Stall\ Warning)} \tag{6}$$

This relationship is important because of its implications for the previously discussed SDT decision criterion. During vigilance tasks, aircrew would establish their stall-related decision criterion through years of flight experience, predominantly free of severe icing encounters and their associated LOC situations. The resulting decision

criterion should be near the SDT ideal observer's value for non-icing conditions, otherwise aircrew would experience a preponderance of Misses or FAs, and would fine-tune their criterion to eliminate the imbalance. The criterion would also remain relatively stable because its validity would be reinforced on an almost daily basis.

Bayes' Theorem shows that as the stall probability (*P*stall) increases in icing, so does the corresponding conditional probability *P* (Stall | Stall Warning), in direct proportion. This shift would invalidate the aircrew's pre-established decision criterion unless the crew were to adjust the criterion for the new environment. In SDT terms, the negative conditioning exhibited by a failure to shift the decision criterion would represent a conservative crew response bias in icing conditions. Per SDT, the CIRB results in a greater susceptibility to Missed stall detections, albeit with less susceptibility to FA occurrences. A conservative decision-making bias should not be confused with conservative behavior, as the former can actually be very risky if it results in a valid stall warning being ignored. The conservative response bias shift predicted by Bayes' Theorem and SDT in a vigilance task context is the last element required to produce a verifiable model of the aircrew / stall warning interaction in icing conditions. The model is appropriately called the Conservative Icing Response Bias model.

The Conservative Icing Response Bias (CIRB) Model

The CIRB hypothesis combines a dual SDT framework with Bayes' Theorem to create a testable model of the combined aircrew / stall warning system behavior outcomes during icing encounters. The CIRB hypothesis could explain a perplexing aspect of several aviation icing accidents: Why did some aircrew apparently fail to appreciate the

severity of their icing encounter (NTSB, 1996c) while others overreacted to benign conditions (NTSB, 2010a)? Still others obviously grasped the severity of the problem yet failed to take meaningful actions to save the aircraft (Taiwan Aviation Safety Council, 2005, p. ii). The literature gives few insights into the causes of these heterogeneous crew responses, but the CIRB provides a framework for such an analysis. The model begins by coding the stall warning system behavior and aircraft stall-state in SDT terms, as shown in Table 2.

Table 2

Comparison of Stall Warning Outcomes and SDT Constructs

	Stall Warning System Response		
Aircraft State[a]	No Stall Warning	Wing-stall Warning Alert	Tail-stall Warning (Dummy Variable)[b]
Aircraft not stalled	System *CR* No action is required.	System *FA* The wing stall warning must be ignored.	System *FA* The tail stall warning must be ignored.
Incipient or developed wing stall	System *Miss* The wing stall must be identified by other means, and an appropriate recovery executed.	System *Hit* The wing stall warning must be acted upon, for a correct stall recovery to be executed.	System *FA* and *Miss* The tail stall warning must be ignored; the wing stall must be identified by other means. An appropriate wing stall recovery must be executed.
Incipient or developed tail stall	System *Miss* The tail stall must be identified by other means, and an appropriate recovery must be executed.	System *FA* and *Miss* The wing stall warning must be ignored; the tail stall must be identified by other means, and appropriate recovery must be executed.	System *Hit* The tail stall warning must be acted upon, and an appropriate tail stall recovery must be executed.

Note. [a]A simultaneous wing and tail stall is unlikely because a stalled horizontal stabilizer is inherently unable to generate the required down-force to raise the main wing to a stalling AoA. [b]There are currently no 14 CFR Part 23 or 25 airworthiness requirements for systems to provide an artificial warning of an impending tail stall, so the provision of a tail stall warning alert is included as a dummy variable to support the ensuing analysis.

As Table 2 illustrates, a stall warning system could detect (Hit) or fail to detect (Miss) a wing stall. There are currently no stall warning systems that address tail stalls, but these cases still had to be encoded for the STD analysis using a dummy tail stall

warning variable. This is because SDT does not differentiate between Misses due to an absent system and Misses associated with a less-than-perfect installed device; both outcomes would be encoded as stall warning Misses, as shown in the table. In addition to Hits and Misses, stall warning systems can produce FAs by issuing an alert when no stall is imminent or CRs when they remain silent when no threat is present (the nominal situation). Other combinations are less relevant. For example, simultaneously occurring wing and tail stalls are unlikely because a stalled horizontal stabilizer is inherently unable to generate the required down-force to achieve the main wing's critical AoA, unless the aircraft is in an inverted stall as described by Telford (1988). Nevertheless, this combination is so atypical that it is not addressed in the table. The combination of a wing stall warning during a tail stall is also improbable but could not be ruled out until the data were examined, so this combination was retained in Table 2.

Table 2 also illustrates the complex decisions that the crew must make to correctly respond to the numerous permutations of aircraft stall states and stall warning system behaviors during icing encounters. Under these circumstances, an imminent stall or LOC can occur with little or no notice, and the crew must very rapidly discriminate between a wing or tail stall and apply the appropriate recovery techniques in the highly stressful environment associated with an impending loss of control. The crew might encounter severe aircraft buffeting and motions, unusual control forces and deflections, and a multitude of cockpit alarms, all competing for the limited processing capacity of the pilot's working memory (Endsley & Jones, 2012, p. 2.5). All of these factors increase the noise term in the SDT calculations, which should lead to an increase in both types of decision-making errors (Misses and FAs), in the absence of the CIRB hypothesis. In

contrast, CIRB predicts that the negative conditioning associated with vigilance tasks and the resulting conservative decision criterion bias should lead to testable predictions of increased susceptibility to system Misses and a reduced susceptibility to FAs.

In order for these predictions to be tested, SDT requires that the ROC curves be established for a range of decision criteria. This is usually accomplished experimentally by plotting the Hit vs. FA data for differing controlled decision criteria, in order to determine the system's sensitivity and response bias. This information is unfortunately not accessible using archival data, as the decision criterion levels can neither be measured nor systematically varied. This does not imply, however, that the crew response bias effect in icing conditions does not exist. What is needed is a method of gauging the correctness of the crew / stall warning system based on the SDT Hit, Miss, FA, and CR factors. The binary logistic regression provides a mechanism for achieving this integration, as described in the following section.

The Binary Logistic Regression

The BLR is a robust multivariate modeling technique for predicting and explaining a binary categorical (Yes / No) outcome from a combination of metric or non-metric independent variables (Hair, Black, Babin, & Anderson, 2010, p. 317). The mechanics of the method are discussed more fully in Chapter III, but the BLR's importance to this study is that it can test the hypotheses made by the CIRB model. The BLR is the final tool required for the application of SDT techniques to the problem of aircrew interactions with their stall warning systems in icing conditions.

Summary

The literature shows that airframe icing is a longstanding and ongoing problem for aircraft operations. The accident and incident records are replete with inexplicably divergent crew responses to similar situations ranging from apparent nonchalance to overreactions, with both extremes having led to the loss of aircraft. A dual SDT model provides a promising framework for modeling aircrew decision-making performance outcomes, as well as the performance of the stall warning system in icing. As a first step in the model's application, the stall warning system and crew behavioral outcomes are mapped to the standard SDT categories: Hit, Miss, FA, and CR. For the stall warning system, these classifications represent the permutations of the stall warning outputs in the presence or absence of a real stall condition. Similarly, for the combined aircrew / stall warning system, the four categories relate to successful or unsuccessful initial crew decision-making outcomes in response to the perceived aircraft stall cues and / or stall warning alerts.

SDT embodies the concept of a decision criterion that demarcates the threshold above which a stimulus is viewed as a signal. An SDT ideal observer is unbiased and sets the decision criterion in an optimum manner to minimize undesirable Misses and FAs. Any deviation from the ideal threshold represents a conservative or liberal bias. The final CIRB assumption is that aircrew subconsciously set and fine-tune their decision criterion to approximate an ideal observer during routine non-icing operations. This low-intensity stall-monitoring activity during non-icing cruise flight can be characterized as a vigilance task because the stall probabilities are very low and stall warnings are infrequent. Aircrew may become negatively conditioned and therefore carry forward the

same decision criterion to the icing case, where the stall probabilities are much higher. Under these circumstances, Bayes' Theorem predicts that the conditional probability of a stall, subject to a stall warning, increases markedly in icing flight, which invalidates the previously established decision criterion and introduces a conservative aircrew response bias in icing, hence the CIRB nomenclature. CIRB predicts an increased susceptibility to incorrect decision-making outcomes arising from Missed stall warning detections, accompanied by a greater tolerance to FA occurrences during icing conditions, both in relation to the stall warning Hit baseline rates.

Unfortunately, this bias shift hypothesis cannot be directly tested using archival sources because SDT requires a range of decision criterion values and corresponding Hit and FA rates to determine the system sensitivity and response bias parameters. The problem can be solved by incorporating the SDT framework of Hits, Misses, FAs, and CRs as factors in a Binary Logistic Regression, with the crew response outcome as the binary dependent variable. The SDT classifications of the stall warning system behavior and the aircraft wing / tail stall-state were therefore selected as the primary independent variables for the basic CIRB model. The combination of the dual SDT framework, Bayes' Theorem, and the predicted response bias shift constitute the CIRB hypothesis. The BLR allowed the model's bias-shift predictions to be tested using NTSB accident and NASA ASRS archival data.

Statement of the Problem

The literature shows that the majority of airborne icing accidents result from aircraft stalls, yet there is no theoretical model that ties the success or failure of the crew

decision-making in icing to the performance of the stall warning system (Green et al., 1996, p. 2). This knowledge gap imposes reactionary and untargeted regulatory responses that have failed to fully resolve the icing issue, despite decades of effort.

Significance of the Study

The novel application of SDT methodology and Bayes' Theorem has resulted in the CIRB model of the influence of stall warning system performance on crew decision-making outcomes during icing encounters. The CIRB model provides the missing theoretical framework. CIRB builds upon Sarter & Schroeder's (2001) simulator studies of decision-making during high-workload icing encounters, but the model's theoretical basis provides testable predictions regarding the influence of stall warning system Misses and False Alarms in icing. The CIRB hypothesis should be equally applicable to archival research using different databases, and it should be a useful aid for future experimental research designs. If validated by such future research, the underlying theory will provide a new tool that will yield a better understanding of the interaction between the aircrew, the icing environment, and the aircraft's stall protection and warning systems. This research should also yield generalizable system guidelines to aircraft manufacturers, flight test personnel, and certification agencies, relating to the relative significance of correct, misleading, and missing stall warning information for the main wing and horizontal tail surfaces. Some of the processes used in the current research design, such as the correct crew response decision criteria described in Chapter III, should also help provide a standardized methodology for evaluating crew behavior outcomes in complex and dynamic environments, using archival data. If validated, the CIRB hypothesis should

lead to recommendations regarding required levels of stall warning system reliability, selectivity (SDT Hit ratio), and specificity (SDT CR ratio). These recommendations should lead to concrete and verifiable measures for reducing icing-related accidents and for evaluating the relative benefits of installing new protection systems such as Aerodynamic Performance Monitors and tail-stall warning systems. The ultimate and overriding objective of the research is to save lives.

Purpose Statement

The purpose of the research was to evaluate the CIRB SDT model of aircrew decision-making in icing conditions by means of an archival analysis of U.S. aircraft icing incident and accident data. The model was used to determine the influence of stall warning system behavior in cueing aircrew to perform correctly during hazardous airborne icing encounters. For this purpose, *correct aircrew performance* was defined as:

1. Implementing appropriate wing-stall or tail-stall prevention and / or recovery procedures subsequent to a stall warning alarm or other indications of a developing stall condition, such as: controllability and performance issues, airframe buffeting, and abnormal control forces and reactions.
2. Taking no stall prevention or recovery action under False Alarm conditions.

Hypotheses

The fundamental research question pertains to the influence of stall warning system performance on aircrew behavior outcomes during hazardous airborne icing encounters. Using a baseline of normal stall warning system performance (i.e., SDT

Hits), the CIRB hypothesis predicts that aircrew would be more susceptible to making incorrect responses when faced with stall warning system Misses and less likely to make incorrect responses when faced with system False Alarms. The following hypotheses were used to test these predictions. Although the stated hypotheses used two-sided tests for significance, the BLR methodology allows the direction of the relationship to be established:

1. $H_0 1$: There is no significant difference in the crew performance outcome between a valid system stall warning (HIT) and a stall warning Miss.
2. $H_0 2$: There is no significant difference in the crew performance outcome between a valid system stall warning (HIT) and a stall warning False Alarm.

Research Approach

A nomothetic exploratory archival analysis (Babbie, 2013, pp. 91-93; Vogt, Gardner, & Haeffele, 2012, pp. 86 - 95) was conducted using NTSB accident data (NTSB, 2012a) and NASA ASRS reports (NASA, 2014). Because of the relatively small number of icing records, the methodology resembled formalized case study research, which Zotov (2000) deemed the appropriate approach for determining the underlying causes of aircraft accidents. A two-phase research design was employed. The first phase addressed the data sampling, scrubbing, and verification activities, in preparation for the second phase, which entailed the execution of the BLR analysis, along with the associated hypothesis testing. Although the qualitative archival review was a prerequisite for the subsequent quantitative analyses, the process was iterated to fine-tune the sample for the best BLR outcome.

The first phase sampling process began with an exploration of the accident and incident archives to select a sampling frame that contained the airborne icing cases in which the state of the stall warning system, aircraft icing state, aircraft wing / tail stall status, and crew responses could be definitively ascertained. As part of this phase, a pre-test was conducted using both of the target databases to establish the approximate number of cases that would be encountered. The pre-test results indicated that the final number of cases would be between approximately 100 and 500, which was sufficiently small for the case processing and sample selection to be performed manually. In order to pre-screen the ASRS and NTSB database, while ensuring that the required information would be available for the analysis, the samples were constrained to turbine powered non-amateur built aircraft. This subset of the population is generally equipped with cockpit voice recorders (CVR) and / or flight data recorders (FDR), which were essential in many cases for retrieving the required data. The turbine-only constraint had the added benefit of limiting the sample predominantly to the larger turboprop or turbojet aircraft that are the focus of the study.

The case selection for the ASRS and NTSB databases relied on carefully crafted Boolean keyword search queries. The literature contains several useful precedents for searching the NTSB and ASRS databases for icing cases (Appiah-Kubi et al., 2013; Aventin, Morency, & Nadeau, 2015; S. D. Green, 2006). The pre-tests also helped in the refinement of the search terms to be used for the main study.

The sampling phase concluded with the encoding of the selected cases in SDT terms to capture the correctness of the crew response, the stall warning system SDT categorization (Hit, Miss, FA, and CR), and the aircraft stall-state (wing-stall, tail-stall,

no stall). The data were also encoded to allow the individual or collective treatment of the ASRS and NTSB archives. The encoded NTSB and ASRS sample datasets were processed using Microsoft® Excel™ and Access™ software packages, which were used for merging the two databases, data integrity verification, missing data identification, and duplicate data checks. Few data duplications were anticipated or encountered, and the relatively small number of records allowed these records to be identified and merged manually. The scrubbed data and merged data samples were exported into IBM® SPSS™ and SAS® Enterprise Miner™ statistical software packages for the execution of the Binary Logistic Regression in Phase 2. The data sampling and scrubbing processes are described in greater detail in the Treatment of the Data section in Chapter III.

The BLR analysis (Hair et al., 2010, pp. 317-344) was conducted in Phase 2 to determine the influence of stall warning system behavior on aircrew responses during airborne icing encounters, to evaluate the CIRB coefficients, and to test the research hypotheses. A single dichotomous dependent variable (DV), encoded as *correct_response,* was used to record the efficacy of the crew's initial reaction to the perceived imminent stall or LOC event. The primary binary independent variables (IV) for the basic CIRB model were the four SDT permutations related to the stall warning system operation: *Hit, Miss, FA,* and *CR*. The wing stall and tail stall conditions were introduced as secondary independent variables for a comprehensive CIRB model that also included a new system_issue IV, the need for which was identified during the qualitative analysis of the databases. Summary statistics were generated for the ASRS and NTSB data individually and for the combined data ASRS / NTSB dataset. For

reasons that are explained in Chapter IV, the BLR was only conducted on the combined dataset.

Population and Sample

The population was comprised of the fleet of civilian U.S. (*N*-registered), non-amateur built (in NTSB database terminology), turbine-powered airplanes. The sample contained the subset of accidents and incidents of U.S. registered, non-amateur built, turbine-powered airplanes with icing-related NTSB Probable Cause entries and ASRS reports between January 1, 1988, and October 2, 2015, for which the aircraft stall-state, stall warning system performance, and crew performance could all be determined unambiguously by two subject matter experts (SMEs).

The ASRS and NTSB samples were initially encoded independently by the author and a second SME using a structured procedure, an unambiguous codebook, and a formalized checklist, as described in Chapter III and shown in Appendix D. The two SMEs manually selected records containing an imminent onset of an icing-induced wing or tail stall, loss of control situation, or receipt of a stall warning. If the required stall warning, aircraft stall status, and aircrew performance parameters could be definitively determined, these records were encoded with these data and retained in the sample. If one or more of the key data elements were missing, or if the SMEs disagreed on the disposition of an individual record following mutual consultation, then the entire record was deleted from the sample - a *listwise* deletion in IBM® SPSS™ terminology. The pre-tests results and literature review indicated that adequate samples could be obtained from the ASRS and NTSB data to perform the BLR analysis (S. D. Green, 2006; Petty &

Floyd, 2004). Nevertheless, there were a number of challenges associated with the selection, categorization, and encoding of the data samples, as described in the next section.

Delimitations

The research focus was on the subset of the population of U.S. registered aircraft operations that have been recorded in either the NTSB accident database or NASA ASRS databases because of icing encounters. This is an important limitation because the vast majority of icing encounters are successfully negotiated; accordingly, the findings from this study are only generalizable to icing events that lead to accidents or noteworthy incidents, as characterized by their inclusion in the two target databases. The study depended on the availability of first-hand narratives from surviving crewmembers or usable data from installed CVR and / or FDR equipment. In simplified terms, 14 CFR § 91.609 requires flight data recorders and cockpit voice recorders for U.S. civil registered, multiengine, turbine-powered airplane with 10 or more passenger seats and for those with six or more passenger seats for which two pilots are required (FAA, 2010c; FAA, 2010d). The study was therefore limited to turbine aircraft, in order to ensure a reasonable probability of having access to CVR or FDR data. This was particularly important for the NTSB accident cases, where first-hand crew narratives were often unavailable in the absence of a CVR, if the crew did not survive the event. The turbine requirement was applied consistently across the ASRS and NTSB databases, with one relaxation: In some cases, the ASRS database records were missing the explicit entry to the engine type. These records were encoded as turbine events when significant indicators were available

to support this assumption, such as: high cruise altitudes, mention of bleed air, aircraft slats, etc., which predisposed toward a turbine classification. The SMEs did not experience any difficulty or ambiguity in making these rare classification judgment calls.

Overall, the turbine-only limitation should not have a severe impact on the generalizability of the proposed research because the large number of U.S. registered aircraft and their varied operations should ensure sufficient randomization to produce meaningful results. Furthermore, the aircraft classes and configurations that were included in the sampling frame are the most relevant from an icing perspective, as smaller Part 91 and homebuilt aircraft are generally operated in visual flight conditions with far less exposure to icing.

Another delimitation of the study was its reliance on the NTSB probable cause summaries for the initial case selection. The NTSB full narrative was only consulted when the factors of interest were ambiguous in the probable cause statement. This could have led to rare misclassifications in the unlikely event that the full narrative favored a different crew performance outcome or stall warning system behavior than those derived from the probable cause summaries. Within these constraints, the methodology, sample size, and composition should be more than adequate for the hypotheses tests and study outcomes to be reliable and generalizable.

The author, a flight operations and certification SME, performed two important subjective assessments related to the processing of the archival data. This section addresses these activities, as well as the author's qualifications to make the required assessments as an SME. These are important considerations because the validity and

reliability of the study both depend to a large degree on the correct record selection and variable encodings.

The first evaluation was the determination of whether sufficient data were present in each individual NASA ASRS or NTSB record to allow its inclusion into the BLR analysis sample. This required that the aircrew performance, stall warning system operation, and aircraft wing / tail icing status to be determinable unambiguously. The second evaluation was an assessment of the correctness of the crew's initial response to the icing encounter, which constituted the correct / incorrect binary dependent variable in the BLR analysis. Of the two assessments, the first was relatively straightforward, because the required data were either present or not, and there was little subjectivity involved. The second assessment was more challenging because it involved a subjective evaluation of crew performance from a limited archival record. Green succinctly captured this challenge:

> Quantitative analysis was possible in certain data fields; however, with respect to the nature and characteristics of the event sequences, it was necessary to add a carefully considered set of inferences to develop models of the accident morphology. This was rather like reconstructing a number of clay pots from the shards recovered at an archaeological dig. To obtain a sketch of the complete pot, certain gaps must be inferred. An example of such inferences is the addition of a "loss of control" element to a sequence that had concluded with an "uncontrolled descent." A more significant inference that was used extensively, was the addition of a "stall" to the event sequence when a "loss of control" had been identified. The premise used in this case was that a loss of control, if in fact due

> to ice accretion, only occurs when some degree of flow separation takes place. A report, which suggested icing as a cause of an uncontrolled descent, would thus be described in the database as a stall, leading to a loss of control, and concluding with an uncontrolled descent. Analysis of this portion of the data were (sic) therefore more qualitative. (Green, 2006, p. 3)

As an added complication, correct crew performance sometimes resulted in accidents, for example in cases where the icing severity overwhelmed the aircraft and its systems despite appropriate crew action. In other cases, aircraft survived unscathed, despite less than optimal aircrew performance outcomes. For these reasons, a simple successful / unsuccessful outcome categorization was an inadequate proxy for the desired correct crew performance measure. A structured process, described in Chapter III, was developed to minimize the subjective aspects of Green's method. This process incorporated objective checklist criteria for the sample selection and data encoding to minimize subjectivity and bias. The procedure was evaluated during the pre-test of the ASRS data and was found to properly address all the cases with no unresolved ambiguities.

The validity of the proposed research strongly depends on the correct encoding of the crew performance outcomes and stall warning system behavior for the BLR analysis. Incorrect classification of these variables would compromise the internal validity of the BLR, which would in turn severely impact the external validity of any conclusions drawn from the analysis. The required classification rigor and consistency was achieved in two ways: The primary safeguard was the rigorous classification schema, coupled with a robust codebook (Appendix D) for the "exhaustive and mutually exclusive" coding of the

variables of interest, which should ensure the reliability of the data encoding activity (Babbie, 2013, p. 417). The use of a second SME added a level of quality assurance to the author's categorization of the crew performance outcomes and stall warning system behavior factors, as encoded for the BLR. These aspects are discussed in greater depth in Chapters II and III.

Limitations and Assumptions

Accident data were derived from the NTSB accident database from January 1, 1988, to October 2, 2015, inclusive. The former represents the first availability of full NTSB docket data on-line and also corresponded to the start date selected by Green for his (2006) comprehensive study of U.S. inflight icing accidents. The October 2, 2015 cutoff represents the selected end-date for the data extract used in the ASRS pre-test discussed in Chapter III that resulted in 115 usable cases. For consistency, the same end-date was used for the NTSB accident data, which also ensured that probable cause findings had been published for almost all of the NTSB records used in the analysis. The NTSB accident database is comprehensive because of the legal obligation to report all accidents and certain types of incidents (NTSB, 2010b). A pre-test of the NTSB accident database using these parameters (cf. chapter III) retrieved 5,110 records before data scrubbing. This ensured that a viable sample could be obtained for the BLR analysis. The NTSB data were supplemented by the ASRS records to further populate the aircrew performance outcome measure. Unlike the compulsory NTSB accident reporting, participation in ASRS is voluntary and subject to self-reporting (response) bias, as well as manual filtering by data entry personnel. As a result of these factors, a limitation of the

ASRS data is that many incidents go unreported, and even the reported incidents are subject to filtering before being selected for inclusion into the ASRS database. These limitations are clearly outlined in the reference materials published by NASA Ames that are included in Appendix B.

The ASRS selection bias effects were mitigated in a number of ways. First, the combined databases are very large; Green found 9,299 relevant icing events using similar databases and search criteria, albeit using an additional FAA Accidents / Incidents Data System (AIDS) (2006). A second mitigating factor is the anonymous and altruistic nature of incident reporting, as well as the specific ASRS FAA Compliance and Enforcement Program incentive of eliminating noncriminal enforcement action for ASRS respondents (FAA, 2011a). More pragmatically, there are few other sources for the required data, and most alternatives also depend on voluntary reporting.

The crew response outcomes derived from both databases were subject to confounding from uncontrolled external factors such as aircraft type, type of operation, pilot experience, etc. For example, it could be posited that professional crews would be more likely than private pilots to recognize anomalous indications from their stall warning systems during icing encounters. Petty & Floyd (2004) used stratification by type of operation (Part 91, 121, or 135) to minimize these effects, but this was impractical for the current study because the BLR severely constrains the number of possible IVs for a given sample size. Each new stratification variable requires an additional BLR model coefficient to be estimated, and the BLR technique is very sensitive to the minimum number of records in each of these bins. Hair et al. (2010, p. 333) specify that each BLR group should have a minimum sample size of 10 times the number of estimated model

coefficients (which correspond to the number of IVs). This would quickly become problematic if an excessive number of stratification variables were introduced, considering that the ASRS pre-test revealed just 115 total cases. The final data samples were large enough to meet the minimum sample sizes required by the BLR with the limited number of variables proposed in Chapter III. Stratification based on other variables such as pilot experience would have been challenging because of the effect on required group sample sizes, and because the data were not consistently available, particularly in the self-reported ASRS database. Nevertheless, the proposed CVR/FDR requirement tended to exclude the highly heterogeneous private pilot operations and selected towards the more homogeneous professional crew operations, which should reduce the need for stratification.

The impact of these delimitations on the study's validity and reliability was addressed in a number of ways. Vogt, Gardner, & Haeffele encourage the use of triangulation, which entails the application of different methods to explore a single phenomenon in order to add depth of understanding and increase confidence in the outcome of the study (2012, pp. 111 and 113). Triangulation was achieved in this study through the use of combined data from two completely independent databases. Vogt et al. also encourage a *thick* understanding that arises when a phenomenon is examined in close conjunction with its natural context (2012, pp. 71-72). The SME review of the accident narratives provided this level of detailed context-sensitive understanding that should further bolster the construct validity of the study. A final mechanism for improving the study's reliability and validity was the adoption of the BLR multivariate technique, which is robust and resistant to violations of the usual constraints that apply to

most forms of inferential statistical analysis, such as data normality and homoscedasticity. In summary, the proposed application of a mixed-method approach to two large datasets should compensate for the limitations of the study and result in valid and generalizable outcomes of value to future researchers.

Definitions of Terms

Angle of Attack (AoA)	The angle between the chord line (that joins the airfoil's leading and trailing edges) and the relative airflow.
Commuter Category	"Multiengine airplanes that have a seating configuration, excluding pilot seats, of 19 or less, and a maximum certificated takeoff weight of 19,000 pounds or less" (CFR Title 14 §23.3(d)).
Conservative Icing Response Bias (CIRB)	A model based on Signal Detection Theory and Bayes' Theorem that predicts that aircrew are conditioned to exhibit a conservative bias response to stall warnings during icing encounters. This bias would be expected to lead to increase errors in the face of stall warning Misses and reduced errors due to False Alarms.

Correct Crew Performance Outcome	(1) Implementing appropriate wing-stall or tail-stall prevention and / or recovery procedures subsequent to a stall warning alarm or other indications of a developing stall condition, such as: controllability and performance issues; airframe buffeting; and abnormal control forces and reactions; and (2) taking no stall prevention or recovery action under false-alarm conditions. Used interchangeably within the manuscript with *Correct Decision-Making* and Correct Crew Response outcome, dependent on the context. Recorded as the *Correct_Response* measure in the analysis.
Correct Rejection (CR)	An SDT term for the situation when a subject correctly recognizes the absence of a signal during a vigilance task.
Critical Angle of Attack	The angle of attack corresponding to the maximum attainable lift coefficient (C_{Lmax}) of an airfoil.
Decision Criterion (*d'*)	An SDT concept that defines an observer's stimulus

	threshold, above which a perceived stimulus is deemed to represent a target detection (i.e., a *signal*). Below the decision criterion threshold, the perceived stimulus is judged to be *noise*.
Equivalent Airspeed	Indicated Airspeed corrected for position and instrument errors and compressibility effects.
False Alarm (FA)	An SDT term for the situation when a subject incorrectly detects a signal when no signal is present during a vigilance task.
Hit	An SDT term for the situation when a subject correctly detects a signal that actually exists during a vigilance task. *Hit* is capitalized throughout this manuscript when used in the SDT context.
Miss	An SDT term for a subject failing to detect a signal that existed during a vigilance task. *Miss* is capitalized throughout this manuscript when used in the SDT context.

Signal Detection Theory (SDT)	A human factors method for evaluating operator responses to sensory stimuli during vigilance tasks, where the signals of interest occur relatively infrequently and have low amplitudes in relation to the background noise level. These factors result in Missed detections and False Alarms in addition to the desired target detections (Hits) and correct rejections (CR) of noise elements.
Stall	The phenomenon caused by airflow separation beyond an airfoil's critical Angle-of-Attack that results in a loss of lift, increase in drag, and potential stability and control issues for the aircraft. All airfoils (wing, tail, fin) on a conventional aircraft are susceptible to the stall phenomenon.
Stall Warning System	An aircraft system designed to alert the aircrew of an impending aerodynamic stall. A combination of visual, aural, and tactile cues may be employed.
Transport Category	Multi-engine airplanes with more than 19 seats or a maximum takeoff weight greater than 19,000 lbs.

List of Acronyms

14 CFR	Title 14 of the Code of Federal Regulations
α_{abs}	Absolute Angle of Attack
$\alpha\varepsilon$	Effective Angle of Attack
α_{G}	Geometric Angle of Attack
α_{ZL}	Zero Lift Angle of Attack
AC	Advisory Circular
AD	Airworthiness Directive
ADREP	(ICAO) Accident and Incident Data-Reporting Database
AFM	Airplane Flight Manual
AIDS	FAA Accidents/Incidents Data System
AIM	Aeronautical Information Manual
AoA	Angle of Attack
APM	Aerodynamic Performance Monitoring
ASRS	NASA Aviation Safety Reporting System

ASTM	American Society for Testing and Materials
BEA	Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile
BLR	Binary Logistic Regression
CADORS	Civil Aviation Daily Occurrence Reporting System (Transport Canada)
CCR	Correct Crew Response
CFD	Computational Fluid Dynamics
CFIT	Controlled Flight into Terrain
CFR	Code of Federal Regulations
CIRB	Conservative Icing Response Bias (model)
C_L	Lift Coefficient
C_{Lmax}	Maximum Lift Coefficient
CR	Correct Rejection (Signal Detection Theory construct)
CVR	Cockpit Voice Recorder
df	Degrees of Freedom
DSS	Decision Support System
DV	Dependent Variable(s)

ε	Induced Downwash
EAS	Equivalent Airspeed
Exp(B)	Exponentiated Logistic Coefficient(s)
FA	False Alarm (Signal Detection Theory construct)
FAA	Federal Aviation Administration
FAR	Federal Aviation Regulations
FDR	Flight Data Recorder
FOQA	Flight Operational Quality Assurance
GA	General Aviation
ICAO	International Civil Aviation Organization
ICTS	Ice-contaminated-tailplane-stall
IFR	Instrument Flight Rules
IMC	Instrument Meteorological Conditions
IV	Independent Variable(s)
LOC	Loss of Control
NACA	National Advisory Committee for Aeronautics

NASA	National Aeronautics and Space Administration
NBAA	National Business Aircraft Association
NPRM	Notice of Proposed Rulemaking
NPV	Negative Predictive Value (for Binary Logistic Regression)
NTSB	National Transportation Safety Board
Part 23	14 CFR Part 23 (Normal, Utility, Acrobatic, Commuter Category aircraft)
Part 25	14 CFR Part 25 (Transport Category aircraft)
Part 91	14 CFR Part 91 (General operating and flight rules)
Part 121	14 CFR Part 121 (Domestic, flag, and supplemental operations)
Part 125	4 CFR Part 125 (Airplanes that have a seating configuration of 20 or more passengers or a maximum payload capacity of 6,000 pounds or more when common carriage is not involved)
Part 135	14 CFR Part 135 (Commuter and on-demand operating requirements)
PPV	Positive Predictive Value (for Binary Logistic Regression)
*P*stall	Stall Probability
ROC	(SDT) Receiver Operating Characteristic

SAFO	(FAA) Safety Alert for Operators
SDT	Signal Detection Theory
SE	Standard Error
SFAR	Special Federal Aviation Regulation
SLD	Supercooled Large Droplets
SME	Subject Matter Expert
VFR	Visual Flight Rules
VMC	Visual Meteorological Conditions
Vref	Reference Landing Speed (Federal Aviation Administration, 2010a, §1.2)

CHAPTER II

REVIEW OF THE RELEVANT LITERATURE

The purpose of this study was to determine the influence of the stall warning system behavior in cueing aircrew to perform correctly during hazardous airborne icing encounters. A Conservative Icing Response Bias (CIRB) model of aircrew behavior outcomes was proposed, based on an SDT framework and Bayes' Theorem. The model predicts that aircrew should be more error-prone when faced with stall warning Misses and less error-prone in the face of False Alarms (FA), as compared to the baseline condition of correct stall warning system behavior (Hits). Two hypotheses were used to test the significance of Misses and FAs against the baseline crew performance in the face of stall warning Hits.

The literature review begins with an overview of the aerodynamic effects of in-flight icing on airfoils, aircraft, and their stall warning systems. Tail stalls are briefly examined, followed by an overview of several operational factors that make commuter aircraft more susceptible to icing effects than larger airliners. This introduction is followed by a summation of prior archival research on icing accidents and incidents and an examination of the evolution and current status of the applicable certification regulations, including recent initiatives aimed at stemming the number of accidents that have resulted from airborne icing encounters. The regulatory summary is followed by an exploration of the relationship between airframe icing, stall warning systems, and aircrew performance. The theory review concludes with an overview of the SDT framework and the repercussions of Bayes' Theorem in the context of aircrew vigilance tasks, such as monitoring for potential stall situations.

Icing Effects on Airfoils and Aircraft

Icing adversely affects an airfoil's maximum lift coefficient and stall Angle of Attack (AoA), which can cause severe performance losses and seriously degrade aircraft stability and controllability. The impact of airframe icing is difficult to predict because numerous factors influence the effect of icing contamination on airfoil performance, and individual airfoils differ markedly in their sensitivity to icing. Thurber cites a Canadian study that noted that an icing density equivalent to a salt-grain sized ice crystal per square centimeter of the critical airfoil leading edge led to a 33% C_{Lmax} reduction when out of ground effect (2008). Veillette noted that heavy rain alone could cause an 18% reduction in C_{Lmax} and a 40% drag increase, characterized by Veillette as "an exceptionally difficult process to understand" (2009, p. 26). Lynch & Khodadoust (2001, p. 760) observed the same C_{Lmax} reduction in their comprehensive archival study of experimental aerodynamic research.

The Transportation Safety Board of Canada (2006) cited several studies noting that the National Advisory Committee for Aeronautics (NACA) 23000 series airfoil used in a very popular light utility aircraft "has been found to be very sensitive to leading edge ice accretions. Compared to other general aviation airfoils, the NACA 23012 has the most severe performance loss" (p. 15). Zeppetelli & Habashi (2012) observed that the popular NACA 23012 airfoil accounted for 25% of the events in the International Civil Aviation Organization's (ICAO) accident and incident data-reporting database (ADREP), and cited Abbott and von Doenhoff's findings that contamination greatly affects the lift capability of the 230XX series airfoils (1959, p. 705). Broeren, Bragg, & Addy (2004) performed tests at the Goodrich Icing Wind Tunnel of a representative NACA 23012

general aviation airfoil with simulated leading-edge ice shapes that were cast from actual ice formations obtained during the ice-tunnel testing. The authors were interested in evaluating the effects of residual *intercycle* ice that forms in-between the activation cycles of leading-edge pneumatic deicing boots. The study revealed very severe aerodynamic degradations from this relatively small icing accumulation, including a 60% loss in maximum lift coefficient and an 8-degree drop in the airfoil's stall AoA.

Icing Effects on Stall Warning Systems

It has long been recognized that AoA-based stall warning systems cannot inherently adapt to the changing aerodynamic characteristics caused by icing. In 1944, Klemin reported on the development of three new stall warning systems, two of which would "function successfully under all flight conditions except when ice has accumulated on the wings" (1944, p. 195). The third system described by Klemin was intended to operate correctly even with contaminated airfoils because its diaphragm-operated sensor was vented directly to the suction surface of the airfoil. Unfortunately, Klemin failed to elaborate on the aerodynamic criteria for triggering such a system, and this promising development appears to have stalled. Luers reviewed the airborne icing literature and noted "dramatic decreases in maximum lift…lead(ing) to premature stall… destroy(ing) the safety margin of an aircraft approaching stall" (1983, p. 54). Garvey cautioned that there is no "rule of thumb" to estimate or quantify the performance impacts of airborne icing, even though these can far exceed the customarily applied safety margins (2010). The situation did not change appreciably over the next 40 years, and the provision of a satisfactory stall warning solution in icing remained elusive. This situation would be

exacerbated when another aspect of airfoil icing rose to the forefront with the dramatic accidents that befell American Eagle 4184 (NTSB, 1996a) and Comair Flight 3272 (NTSB, 1998a): icing-induced roll upsets.

Roll Upsets

In addition to the loss of lift and increased drag, another undesirable effect of in-flight icing is its potential to adversely influence control forces and moments on aircraft with unpowered (reversible) flight controls, such as the ailerons and elevators. Airflow degradation and flow separation bubbles caused by airfoil icing can drive a control surface to a full deflection (Bragg, 1996), as experienced by American Eagle 4184 (NTSB, 1996a) and Comair Flight 3272 (NTSB, 1998a), which caused both aircraft to rapidly roll to extreme attitudes following their icing encounters. The problem is exacerbated following a sudden disconnect of the autopilot, which can occur when the icing-induced control forces become excessive (Flight Safety Foundation, 1996; Flight Safety Foundation, 2008; Manningham, 1997; NTSB, 1996b; NTSB, 1998b). This phenomenon can affect the pitch axis as well as the roll axis, and longitudinal stick forces as high as 195 lbs. have been recorded during flight test investigations into icing-induced tail stalls (B&CA Staff, 1997).

Icing and Tail-Plane Stalls

Unlike the main wing, conventional aft-mounted tail-planes operate as inverted airfoils. Much of the following information on this phenomenon stems from a 4-year, multi-disciplinary, flight-testing, and wind-tunnel study led by NASA Lewis, with

participation from Ohio State University (Ratvasky et al., 1999). The flight trials were performed using the Lewis de Havilland DHC-6 Twin Otter turboprop.

Horizontal stabilizers generally produce a down-force to compensate for the nose-down pitching moment of the main wing, particularly at slow aircraft speeds. Large main wing flap deflections move the center of pressure aft, which further increases the aircraft's nose-down pitching tendency that must be overcome by the tail. The main wing flaps also increase the downwash impinging on the horizontal tail, increasing the tail's negative AoA and putting the stabilizer closer to a stall situation. Compounding this effect, the horizontal stabilizer may be six times more susceptible to icing accumulations than the main wing because the tail airfoil's sharper leading-edge radius makes it a more efficient ice collector (F. T. Lynch & Khodadoust, 2001, p. 760; Manningham, 1997). All of these factors can lead to, or exacerbate, an ice-contaminated-tail-plane-stall (ICTS).

A tail stall recovery entails a reduction in power, pulling back on the control column, and raising the flaps (Detwiler, 2015; North, 1998), which are the opposite inputs from those required to recover from a wing stall. The recovery must be executed under the extreme time pressure and stress of an imminent stall, and the choice of technique depends on subtle cues, such as detecting control wheel vibration in the absence of airframe buffet. The wrong decision has resulted in at least 16 crashes (Carlisle, 2006; North, 1998) and at least 139 fatalities (B&CA Staff, 1997, p. 80).

Commuter Aircraft Icing Exposure

The 1970s saw the start of the proliferation of commuter category turboprops and small regional jets. These aircraft typically fly numerous short sectors at mid-altitudes where icing is prevalent, which increases the icing exposure frequency, severity, and duration for commuter types compared to larger transport category aircraft. Conversely, larger airliners tend to spend most of their flight time at cruise altitudes well above the freezing level, climbing and descending quickly through the freezing layer, which limits their exposure to the icing environment. Commuter aircraft have therefore experienced a disproportionate amount of icing accidents, and despite FAA efforts, the U.S. commuter icing accident rate did not reduce appreciably during the 1982 to 2000 period studied by Petty and Floyd (2004). A number of high-profile commuter aircraft icing accidents prompted an intense research focus on airborne icing, as demonstrated by the emergence of numerous archival studies of the icing phenomenon.

Icing Accident and Incident Archival Studies

The following discussion summarizes a number of archival studies of icing-related accidents and incidents to the U.S. civil aircraft population in approximately chronological order. Cole & Sand reviewed NTSB accident data from 1975 to 1988, examining a number of variables, including type of operation, phase of flight, and seasonal distribution, before concluding that airborne icing encounters can be “extremely hazardous” and “a significant number of larger commercial aircraft…have been involved in icing related accidents” (1991, pp. 9 and 10). Petty & Floyd (2004) explored NTSB data from 1982 to 2000 and added a useful stratification by type of operation: 14 CFR

Part 91 (general aviation (GA)), Part 121 (air carrier), Part 135 scheduled (commuter), and Part 135 non-scheduled (air taxi). The authors noted that 39.8% of the icing-related accidents and 50.6% of the fatalities occurred during the cruise phase of flight. Despite a gradually declining icing accident rate throughout the period, icing still accounted for 819 deaths, and Petty & Floyd concluded their review by stating that airframe icing remained a "serious aviation hazard."

Green (2006) broadened the investigation by merging the FAA Accidents / Incidents Data System (AIDS) with the NASA ASRS and NTSB accident data in a comprehensive study of 5,604 reports covering the 1978 to 2002 period. Green reduced the data to 693 "aerodynamically significant" events and noted that the prevalent outcome from an icing encounter was a stall followed by a LOC event. Hard landings caused by unexpected stalls were another leading contributor to the recorded accidents and incidents. Green also highlighted instances of icing overwhelming pitot-static system heaters, particularly on smaller GA aircraft, leading to a loss of cockpit airspeed indications. This phenomenon would prove to be equally dangerous for much larger aircraft, with the catastrophic loss of an Airbus 330 on June 1, 2009, just three years after Green's study (Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (BEA), 2012). Air France 447 experienced pitot-tube blockage due to icing that led to the temporary failure of the aircraft's air data systems. This led to a degradation of the flight control laws that resulted in the loss of the sophisticated flight envelope and stall protections normally provided to the crew. The aircraft was then unintentionally held into a stall for several minutes, despite the aircraft's aural stall warning sounding 75

times during the fully stalled descent into the Atlantic Ocean, with the loss of all 228 aboard (BEA, 2012).

Veillette's analysis of NTSB, FAA, and ASRS data between 1991 and 2010 revealed that LOC was the leading cause of fatalities in the large commercial jet and business jet sectors (2012). Veillette reported that LOC accounted for 4,717 lives lost, and 44% of all U.S. business aircraft accidents during the period. Of particular note, Veillette's data showed that icing caused 29% of these fatal LOC events. A worldwide Boeing study of large (> 60,000 lbs.) jet aircraft accidents between 2005 and 2014 corroborated Veillette's findings: LOC was the leading cause of fatal accidents, accounting for 1,706 deaths – more than double the 804 fatalities resulting from CFIT, the next leading cause (Boeing Commercial Airplanes, 2015). Although not specifically broken out, the proportion of icing-induced LOC accidents in the Boeing data should mirror the 29% observed by Veillette (2012), based on the very similar aircraft classifications used in the two studies. In recognition of these trends, the NTSB added LOC to its Most Wanted List (NTSB, 2016a), effectively absorbing and embracing the icing issue that was removed as a stand-alone item in 2012.

Small aircraft have also been well represented in the icing accident statistics. Aarons (1995) cites an internal NTSB study covering the period between 1986 and 1995, which records 154 icing-related accidents experienced by 14 CFR Part 23 aircraft operating under 14 CFR Parts 91 (private) and 135 (air taxi) operations. In a follow-on study for the period 1989 to 1997, the NTSB noted that icing was a factor in 11% of weather-related GA accidents, 6% of air taxi and commuter mishaps, and 2% of air carrier accidents (National Aviation Weather Program Council, 1999, p. 7-1). FAA

Aviation Safety Information Analysis and Sharing analysts reviewed the NTSB accident and incident databases for the period 2003 – 2007; of the 1,740 weather related accidents, 64 (3.7%) were related to icing encounters. Of the icing total, 57 pertained to 14 CFR Part 91 (private) operations and the remainder to Part 135 (air taxi) operators (FAA, 2010f). No Part 121 carriers experienced any icing-related losses during the period, although this respite would end in February 2009 with the high-profile loss of Colgan Air Flight 3407 over Buffalo, NY (NTSB, 2010a).

Appiah-Kubi, Martos, Atuahene, & William (2013) examined NTSB and ASRS data from 2006 to 2010 in a study that yielded 228 accidents and 30 incidents related to aerodynamic events arising from airborne icing encounters. Similar to Petty and Floyd's (2004) study, Appiah et al. (2013) noted that the majority of the accidents (40%) occurred following the first detection of icing during the cruise phase of flight. The majority of the accidents (53%) also resulted from a stall and subsequent loss of control after such encounters, with only two aircraft out of 40 successfully recovering from the resulting stall. This finding was borne out by the previously cited National Business Aircraft Association (NBAA) study of business aviation LOC accidents between 1991 and 2010 that found "no cause was nearly as prevalent as aerodynamic stalls" (NBAA, 2015, p. 1). The NBAA noted that 9 of the 31 stall events (29%) were related to airspeed management in icing conditions.

Aventin, Morency, & Nadeau (2015) examined data from the Transport Canada Civil Aviation Daily Occurrence Reporting System (CADORS) to evaluate the influence of on-ground de-icing / anti-icing on icing accidents. The small size of the final sample (19 events) limited the causal inferences that could be drawn, but the pattern was

consistent with the earlier research: Small GA proved more vulnerable to icing than larger aircraft, with turboprop aircraft accounting for 58% of the accidents in the study.

Zeppetelli & Habashi (2012) researched the ICAO ADREP database that comprises mandatory reports of aviation accidents to aircraft over 2,250 kg in accordance with Annex 13, Chapter 7 of the Chicago Convention. The authors identified 323 airborne icing occurrences since the database's inception in 1970. Once again, the cruise phase of flight accounted for the largest percentage of these accidents (approximately 33%), closely followed by the approach phase (32%). A worldwide Boeing study of commercial jet aircraft accidents between 1959 and 2014 noted that LOC accounted for 17 of 72 total accidents and 1,656 of the 3,946 total on-board fatalities, although icing accidents were not broken out specifically (2015).

Lynch & Khodadoust (2001) performed a comprehensive review of public-domain flight test and wind tunnel data pertaining to icing effects on airfoil surfaces of fixed-wing aircraft. The authors considered four principal icing formations: (1) small initial leading-edge ice accumulations; (2) runback icing, which is characteristic of Supercooled Large Droplet (SLD) conditions; (3) large, irregular ice formations resulting from extended exposures with inadequate or failed ice-protection capability; and (4) ground frost on the upper wing surfaces, typically caused by chilled fuel in the wings during a quick-turnaround following a cold-soak at altitude. Lynch & Khodadoust's findings corroborated many of the icing hazards already discussed, but the authors grouped their conclusions into four insightful icing threat categories:

1. "Dangerous because of (the) possibility of being under-estimated and / or misunderstood;"

2. “Dangerous because of (the) potential for catastrophic reductions in aerodynamic effectiveness;”
3. “Dangerous because (the) upper limits of potential aerodynamic consequences are not really defined;” and
4. “Dangerous because of (the) portion of flight operation envelope involved” (F. T. Lynch & Khodadoust, 2001, pp. 759, 760).

These four hazards of severe inflight icing were tragically and simultaneously demonstrated by a LOC accident recounted by Telford (1988). Ironically, the crash occurred to his organization’s B-26 atmospheric research aircraft in the course of an investigation into SLD icing. The SLD phenomenon has been the subject of many regulatory changes in recent years and remains a difficult challenge for the certification and safe operation of contemporary aircraft. For example, the FAA issued “Airworthiness Directive (AD) 96-09-25 requiring crews of de Havilland Model DHC-7 and DHC-8 Series Airplanes to avoid or exit SLD conditions as a matter of urgency” (FAA, 1996a). In contrast, the Telford accident aircraft had deliberately been exposed to SLD and accumulated substantial airframe ice. The accident sequence began during a descent in clear air as the ice started to melt and run back on the wings and tail surfaces. The resulting residual ice ridges apparently caused the inboard wing sections to stall, which led to a loss of pitch and roll stability. The aircraft executed a series of unstable pitch and roll excursions before entering into an inverted stall and diving into the ground at a 60-degree angle. Telford conjectured that the tail was also fully stalled during the final negative-g maneuver, precluding any possibility of recovery. This single accident, to a fully-instrumented icing research aircraft, exhibited almost every characteristic of

severe icing that has already been discussed, including: performance degradation; premature stalling with a lack of stall warning indications; loss of pitch stability; violent roll divergence; and, finally, an unrecoverable tail stall (Telford, 1988). Lynch & Khodadoust observe that this type of accident could have been avoided if decades of lessons learned had not been ignored or forgotten (2001, p. 761).

Aarons (1999) critiqued the historical certification approach to stall warnings in icing conditions, which allowed approval based on a mixture of artificial and natural stall cues (such as airframe buffeting). Veillette (2006) noted the flaws in this approach, as exemplified by six premature stall events on a single commuter aircraft type that occurred with little or no warning from the aircraft's stall protection systems, and which failed to alert the crews by other means. Zeppetelli & Habashi (2012) also noted a number of shortcomings in the icing certification regulations, ranging from problems with the icing terminology to unrepresentativeness of the simulated ice shapes used for certification flight-testing. These certification issues are central to the current investigation, so they merit a dedicated discussion.

Operational Approvals for Known Icing Flight

Sections of Title 14 of the Code of Federal Regulations (CFR), popularly known as the Federal Aviation Regulations (FAR), govern the approval for flight in known icing conditions for all U.S. registered civilian aircraft. Known icing authorization has two elements: airworthiness certification of the aircraft and its systems and operational approval for flight in known icing, which is the subject of this section. The type of operation dictates the applicable standards: 14 CFR Part 91 (General operating and flight

rules); Part 135 (Commuter and on-demand operating requirements); Part 125 (airplanes that have a seating configuration of 20 or more passengers or a maximum payload capacity of 6,000 pounds or more when common carriage is not involved); and Part 121 (Domestic, flag, and supplemental operations). The broadest requirements are contained in 14 CFR 91.527(c) *operating in icing conditions* that states:

> Except for an airplane that has ice protection provisions that meet the requirements in section 34 of Special Federal Aviation Regulation No. 23, or those for transport category airplane type certification, no pilot may fly an airplane into known or forecast severe icing conditions. (FAA, 2009d)

This apparently clear-cut regulation has proved difficult to interpret and implement. Jeck (2001) captured the evolution of the FARs related to operations in icing conditions and notes the conflicts that have arisen because of the continued use of stale terminology that has not kept pace with the changing regulations. The problem partially stems from the definition of *severe icing* contained in the FAA Aeronautical Information Manual (AIM), which defines the phraseology to be used for pilot reporting purposes: "The rate of accumulation is such that deicing / anti-icing equipment fails to reduce or control the hazard. Immediate flight diversion is necessary" (FAA, 2014a, p. 7-1-45). Under the AIM definition, flight in severe icing should never be countenanced, regardless of the installed aircraft equipment. This conflicts with the wording in 14 CFR 91.527(c) that implies that a certain level of equipage should allow such operations. Zeppetelli & Habashi (2012, p. 612) also clearly highlight this contradiction between theory and practice.

Airworthiness Approvals for Known Icing Flight

In addition to the operational approval for flight in known icing, 14 CFR also contains standards for the airworthiness approvals of aircraft and equipment required for flight in known icing conditions. 14 CFR Part 23 lists the airworthiness standards applicable to normal, utility, aerobatic, and commuter category airplanes. These standards generally apply to small aircraft with a maximum weight of 12,500 lbs., with the exception of the commuter category that is limited to 19,000 lbs. and 19 passenger seats. Large aircraft are governed by 14 CFR Part 25, which encompasses the *Transport Category* airworthiness standards; these are generally more stringent than their Part 23 equivalents. Collectively, 14 CFR Parts 23 and 25 contain the regulations applicable to aircraft stall behavior, stall warning systems, and anti-icing / de-icing equipment. Appendix C to Part 25 also embodies a Special Federal Aviation Regulation (SFAR) that defines the precipitation and icing envelopes for known icing certifications (FAA, 2009a). Pending the enactment of 2014 amendments to this SFAR discussed in a later section, certain weather phenomena, such as SLD precipitation, were outside the Appendix C envelope. As a result, SLD encounters have occasionally overwhelmed the anti-icing or deicing capabilities of known-icing certified aircraft, leading to a number of accidents (NTSB, 1996b; NTSB, 1996c; Taiwan Aviation Safety Council, 2005).

In addition to the SLD problem, the stall warning provisions of 14 CFR 25.207(e) raise additional issues (the following paragraph numbering mirrors the regulation):

> (e) In icing conditions, the stall warning margin in straight and turning flight must be sufficient to allow the pilot to prevent stalling (as defined in §25.201(d)) when the

pilot starts a recovery maneuver not less than three seconds after the onset of stall warning. When demonstrating compliance with this paragraph, the pilot must perform the recovery maneuver in the same way as for the airplane in non-icing conditions. Compliance with this requirement must be demonstrated in flight with the speed reduced at rates not exceeding one knot per second, with -

(1) The most critical of the takeoff ice and final takeoff ice accretions defined in Appendices C and O of this part, as applicable, in accordance with §25.21(g), for each configuration used in the takeoff phase of flight;

(2) The most critical of the en route ice accretion(s) defined in Appendices C and O of this part, as applicable, in accordance with §25.21(g), for the en route configuration;

(3) The most critical of the holding ice accretion(s) defined in Appendices C and O of this part, as applicable, in accordance with §25.21(g), for the holding configuration(s);

(4) The most critical of the approach ice accretion(s) defined in Appendices C and O of this part, as applicable, in accordance with §25.21(g), for the approach configuration(s); and

(5) The most critical of the landing ice accretion(s) defined in Appendices C and O of this part, as applicable, in accordance with §25.21(g), for the landing and go-around configuration(s).

(f) The stall warning margin must be sufficient in both non-icing and icing conditions to allow the pilot to prevent stalling when the pilot starts a recovery maneuver not less than one second after the onset of stall warning in slow-down turns with at least 1.5 g load factor normal to the flight path and airspeed deceleration rates of at least 2

> knots per second. When demonstrating compliance with this paragraph for icing conditions, the pilot must perform the recovery maneuver in the same way as for the airplane in non-icing conditions. (FAA, 2014e)

The problem with these apparently stringent requirements is that no existing stall warning system can determine the correct stall speed under the multitude of icing conditions that could be encountered. Bragg et al. (1998) proposed a sophisticated ice management system that would use advanced sensors to predict the performance and controllability effects of airborne icing, but such a system has yet to be certified. Considerable research has been conducted into stall warning methods based on direct airflow measurements (Catlin, 1992; Lerner, 1985; Maris, 1991; Maris, 1996; Pederson, 2003), but Bragg et al. note "no processed aircraft performance degradation information is available to the pilot" (2002, p. 1). Instead, a common solution is to apply a fixed safety increment to the stall warning margin when ice is detected via an on-board sensor or via a pilot-selectable speed reference switch that changes the airspeed thresholds for the activation of the stall warning and protection systems. The latter approach was implemented in the Dash-8 aircraft that was involved in the Colgan Air icing accident (NTSB, 2010a, p. 18). Neither solution can guarantee an adequate warning margin under all flight conditions, and there may also be circumstances when the system provides too much warning, particularly if the aircrew misinterpret or forget the position of the icing reference switch, as was the case in the Colgan Air accident (NTSB, 2010a, p. 151). As a result, the use of fixed or even dual stall warning sensitivity thresholds inevitably results

in some Missed stall warnings or False Alarms in icing conditions that have important ramifications for this study.

The Evolution of Icing Legislation

The push for icing legislation reform goes back many years, but a watershed event occurred in January 1997 when a commuter twin-turboprop experienced an un-commanded icing-induced roll excursion on approach to the Detroit Metropolitan / Wayne County Airport (DTW). The aircraft dove steeply into the ground, partially inverted, with no survivors (NTSB, 1998b). Reehorst, Chung, Potapczuk, & Choo (2000) analyzed the accident scenario using Computational Fluid Dynamics (CFD) techniques and determined that as little as five minutes of ice accretion were sufficient to produce the aerodynamic degradations that led to the complete loss of control. As a result of the Detroit accident, the NTSB made 21 sweeping recommendations to the FAA, including a reiteration of two of its previously issued safety recommendations to the Agency:

> Revise the icing criteria published in 14 Code of Federal Regulations Parts 23 and 25, in light of both recent research into aircraft ice accretion under varying conditions of liquid water content, drop size distribution and temperature, and recent development in both the design and use of aircraft. Also, expand the Part 25 Appendix C icing certification envelope to include freezing drizzle / freezing rain and mixed water / ice crystal conditions as necessary (A-96-54).

> Revise the icing certification testing regulation to ensure that airplanes are properly tested for all conditions in which they are authorized to operate, or are otherwise shown to be capable of safe flight into such conditions. If safe operations cannot be demonstrated by the manufacturer, operational limitations should be imposed to prohibit flight in such conditions and flightcrews should be provided with the means to positively determine when they are in icing conditions that exceed the limits for aircraft certification (A-96-56). (NTSB, 1998b, p. 185)

The FAA had already begun to act in 1997 by publishing its Inflight Icing Plan, overseen by the FAA Icing Steering Committee, which detailed the FAA's intended activities and milestones for improving flight safety in icing conditions. The plan incorporated sweeping changes to the certification regulations, including an overhaul of the Appendix C icing envelopes to address the SLD and ice crystal phenomena (FAA, 1997). In parallel, the FAA began publishing a long run of ADs targeting the vulnerable commuter turboprop category, and imposing operational restrictions in certain types of icing. In addition, the AD instructed aircrew on procedures for recognizing and escaping from dangerous SLD ice (Flightglobal, 1996). Also in 1997, the NTSB decided to incorporate the airframe icing issue into its Most Wanted List of Safety Improvements for the first time. Airframe icing would remain on the list until 2011 (Weener, 2011). The FAA responded in 1999 by proposing a sweeping series of ADs affecting the operation in icing conditions of the "Beechjet 400-series, Cessna T303 Crusader, de Havilland Dash 6 Twin Otter, Embraer EMB-110 Bandeirante, Jetstream 31-series, and Nihon Aeroplane YS11" (B&CA, 1997).

Throughout this period, the FAA continued to update the ice protection regulations applicable to Part 23 aircraft. Prior to 1993, the regulations required no flight evaluations in real icing conditions for GA known-ice approvals. This changed when Amendment 23-43 was issued in 1993, required these aircraft to meet the same performance and flying quality criteria in icing as outside it, which was beyond the capability of most applicants. The evolving regulations also required tail contamination effects to be evaluated in flight for all GA known-icing approvals. Newton (2006) observed that 13 years after the new rules came into effect, only one aircraft type, the Extra 400, had been certified to Amendment 23-43. With this one exception, known icing certification was not even being attempted for the most vulnerable segment of the GA fleet.

The FAA also tried to address the icing issue for transport aircraft via a string of regulatory amendments to the ice protection standard (14 CFR § 25.1419) in 1970, 1990, 2007, and 2009 (FAA, 2009b). Of note, the 2007 "Activation of Ice Protection" Notice of Proposed Rule-making (NPRM), and the subsequent regulatory amendment, introduced the concept of a "primary ice detection system that automatically activates or alerts the flightcrew to activate the airframe ice protection system" (FAA, 2007). These changes, along with the ever-growing list of ADs, were well intended but reactive measures that did not constitute a long-term strategic solution to the icing problem. This shortcoming was exemplified by the crash of a small business jet short of the runway at Pueblo, CO, in 2005, where Fiorino noted "the (NTSB) board found the FAA's failure to establish adequate certification requirements for flight into icing conditions was a contributing factor in the crash. It resulted in the failure of the aircraft's stall warning

system to provide an adequate warning margin" (2007, p. 47). Fiorino also cited the NTSB's recommendation to "modify the Cessna 560 stall warning system to require a warning margin that takes into account the size, type, and distribution of ice" (2007, p. 47). The FAA had already published an Advisory Circular (AC) on Aircraft Ice Protection (AC 20-73A) that addressed this issue and stated the need to "provide acceptable stall warning margins and to prevent a stall during flight in icing conditions" (FAA, 2006, p. 27). Unfortunately, like all ACs, AC 20-73A is not a regulation; it simply describes "an acceptable means, but not the only means of showing compliance with the …Regulations" (FAA, 2006, p. i).

On February 24, 2010, NTSB Chair Deborah Hersman, testifying to the House Aviation Subcommittee of the Committee on Transportation and Infrastructure, elected to keep the icing issue on its Most Wanted List, due in part to more than 50 accidents and 200 associated deaths resulting from ice encounters (NTSB, 2010c). The same day, the FAA issued a press release "Fact sheet – flying in icing conditions" that listed the Agency's historical and planned efforts at addressing the problem (FAA, 2010e).

Table *3* summarizes this fact sheet and includes updates to the FAA activities that occurred after the press release was issued.

Table 3

FAA Icing Activities 1996 - 2014

Date	FAA Action	Description[a]
1996	AD 96-09-25	*Airplane Flight Manual (AFM) revision to limit or prohibit the use of various flight control devices and provide flight crews with recognition queues and procedures for exiting from severe icing conditions.*
1999	AD 99-19-18	*Mandated revisions to AFM to advise flight crews to activate airframe pneumatic de-icing boots at the first sign of ice accumulation.* Applicable to aircraft with history of icing issues.
Mar 29, 2006	Safety Alert for Operators (SAFO) 06002	*Ground deicing practices for turbine airplanes in nonscheduled Part 135 and Part 91 service.*
Oct 6, 2006	SAFO 06014	*Hazards posed by polished frost.*
Nov 11, 2006	SAFO 06016	Aimed to *increase awareness of in-flight icing dangers for pilots flying turbo-propeller powered airplanes.*
Aug 8, 2007	Final rule: icing certification standards.	*New airworthiness standards for the performance and handling characteristics of transport airplanes in icing conditions… Harmonize(s) the U.S. and European airworthiness standards for flight in icing conditions. Comprehensive set of airworthiness requirements that manufacturers must meet to receive approval for flight in icing conditions, including specific performance and handling qualities requirements, and the ice accretion (size, shape, location,*

Date	FAA Action	Description[a]
		and texture of ice) that must be considered for each phase of flight. These revisions will ensure that minimum operating speeds determined during the certification of all future transport airplanes will provide adequate maneuvering capability in icing conditions for all phases of flight.
Nov 30, 2007	SAFO 07009	*Inform(s) owners, operators, and FAA entities of training requirements for pilots of CE-208 (Cessna Caravan 1) and CE-208B (Cessna Grand Caravan) airplanes for flight into icing conditions.*
Dec 2007	AC 91-74A	*Affect* (sic) *of ice crystals on turbine engines.*
May 8, 2008	NPRM: Polished frost	*Remove(d) language from its regulations that allowed some operators – not commercial airplanes – to operate with polished frost. Unlike commercial airplanes which must have a clean wing, corporate aircraft were permitted to fly with smooth or "polished frost." That practice has been deemed unsafe.*
May 20, 2008	SAFO 0812	*Aircraft taxi operations during snow and ice conditions.*
Feb 11, 2009	SAFO 09004 SAFO 0812 elaborated	*Emphasize preflight and in-flight planning for winter airport operations for taxi, takeoff, and landing.*
Aug 3, 2009	Final Rule: icing certification standards for transport category airplanes.	Rule requires *either the automatic activation of ice protection systems or a method to tell pilots when they should be activated. The rule applies to new transport aircraft*

Date	FAA Action	Description[a]
		designs and significant changes to current designs that affect the safety of flight in icing conditions… *The standards further require that after initial activation, the ice protection system must operate continuously, automatically turn on and off, or alert the pilots when the system should be cycled.*
Nov 29, 2009	NPRM: ice detectors for air carrier airplanes	Proposed rule w*ould require either the installation of ice detection equipment or changes to the procedures for activating the ice-protection system to ensure timely activation of the ice-protection system. This proposed rule would apply to all current and future airplanes in service with air carriers whose maximum takeoff weight is less than 60,000 pounds.*
Mar 16, 2010	SAFO 10006	*In-Flight Icing Operations and Training Recommendations…* encouraging *directors of safety and directors of operations (part 121 and 135); and training managers for all operators…to review and amend, if required, flight crewmember and dispatcher training programs.*
June 29, 2010	NPRM: SLD	*The proposed rule would improve safety by taking into account supercooled large droplet (SLD) icing conditions for transport category airplanes most affected by these icing conditions, mixed-phase and ice-crystal conditions for all transport category airplanes, and supercooled large droplet, mixed phase, and ice-crystal icing conditions for all turbine engines.*

Date	FAA Action	Description[a]
Nov 4, 2014	Final Rule: Part 25 and 33 icing certification standards.	Appendix C icing envelope expanded to accommodate SLD and ice crystal icing. Added performance and handling requirements for transport aircraft in SLD. New Part 25 icing certification requirements for airspeed and AoA indicating systems.

Note. Adapted from FAA Fact Sheet "Flying in Icing Conditions" FAA (2010e), retrieved from http://www.faa.gov/news/fact_sheets/news_story.cfm?newsid=10398
[a]Italicized text is quoted verbatim from the source document.

In June 2010, the FAA took the important step of codifying the mass of advisory material into a broad NPRM that would add significant legislative weight to tackle the SLD and aircraft systems problems in icing (FAA, 2010b). The NPRM was finally enacted on November 4, 2014, and came into effect on January 5, 2015, as the "Airplane and engine certification requirements in supercooled large drop, mixed phase, and ice crystal icing conditions" (FAR, 2014d). The updated regulation addressed a number of previous problem areas, notably including an update to the Appendix C icing envelope to accommodate SLD and ice crystal icing, as well as adding performance and handling requirements for transport aircraft operations in SLD. The new regulation also introduced updated system requirements, including the need for AoA sensors and airspeed indicators, "to perform in freezing rain, freezing drizzle, mixed phase, and ice crystal conditions" (FAA, 2014g, p. 1). Two advisory circulars supported the revised airworthiness standards. The first, AC 25-25A, addressed compliance demonstration with the performance and handling requirements for the new Appendix C SLD regulation

(FAA, 2014f). The second circular, AC 25-28, addressed the broad new certification requirements for transport category aircraft operations in icing conditions (FAA, 2014c). Unfortunately, it proved impractical to make the updated certification requirements retroactive, so the more stringent standards only apply to certifications commencing after January 5, 2015, when the new regulations came into force (FAA, 2014g).

In parallel with its legislative activities, the FAA and other groups were heavily engaged with industry. The FAA hosted an international conference on aircraft icing in 1996 (FAA, 1996b), while SAE International addressed the icing issue with a pair of working groups: the AC–9 Aircraft Icing Technology sub-committee and the G-12 committee, which was given the mission "to improve worldwide safety in matters related to aircraft ground deicing" (SAE International, 2014, p. 1). In 1997, the University of Illinois constituted an interdisciplinary research center for aircraft icing, which embarked on the Smart Icing Systems Project in conjunction with the NASA Lewis Research Center. The project team implemented a multidisciplinary combination of basic and applied research to address the issue of icing safety (Bragg et al., 1998). As a result of these ongoing efforts and an increasing awareness of the serious and unpredictable consequences of airfoil ice contamination, certification authorities such as the FAA and Transport Canada adopted the *Clean Aircraft Concept* that precludes attempted takeoffs with any ice or frost adhering to the critical surfaces of the aircraft, including its wings, propellers, and stabilizers (Transport Canada, 2004).

Icing was finally removed from the NTSB's Most Wanted List in 2012, but the problem was far from conquered. In March 2014, the FAA published a new AD (its 112th related to icing) that barred known icing operations by approximately 4,200 small

GA aircraft, based on 52 mishaps and 36 fatalities attributed to in-flight icing within this group in the past 30 years (FAA, 2014g; Lynch, 2014). Two years after the NTSB removed icing from the Most Wanted List, an Embraer Phenom crashed following an icing encounter because the pilot failed to activate the aircraft's ice-protection systems. As a result of this accident, the NTSB issued a recommendation to the FAA and the General Aviation Manufacturer's Association that a system be developed to automatically alert pilots of certain aircraft when the ice protection systems should be activated (NTSB, 2016b).

In parallel with the FAA and NTSB efforts, the American Society for Testing and Materials (ASTM) convened Technical Committee F44 on general aviation aircraft, with a primary mandate to streamline the Part 23 regulations for the certification of light aircraft. As a result of the committee's work, the FAA issued an NPRM "Revision of Airworthiness Standards for Normal, Utility, Acrobatic, and Commuter Category Airplanes" that included a major emphasis on LOC and icing, including SLD (FAA, 2016). In the interim, the FAA issued a revised policy to substantially ease the certification burden of installing AoA systems in Part 23 aircraft "to provide precise information to the pilot (that) could help avoid needless accidents" (FAA, 2014b).

Despite these efforts, the NTSB's recommendation for improvements in cockpit stall warning systems still remain unmet by the new regulations for icing scenarios, and in the absence of such a capability, aircrew continue to experience FAs and Missed stall warnings during severe airborne icing encounters. The following discussion examines the issue of crew decision-making and the interaction between the aircrew and the

aircraft's stall warning system under sub-optimal conditions such as severe icing encounters.

Aircrew Performance During Icing Encounters

Advani wryly observed "Aerodynamic Stall Can Prompt 'Brain Stall.'" (2014, p. 58), and icing-induced stalls are no exception. According to Sarter and Schroeder (2001), surprisingly few studies have been performed on the effectiveness of Decision Support Systems (DSS), under a combination of time pressure and incorrect cueing from the DSS. Sarter and Schroeder's study used a scripted simulator exercise to examine the impact of different levels of DSS cueing and reliability on pilot response times and error rates in a multiple-task, highly dynamic (icing) environment (2001). The researchers observed a positive correlation between the aircrew's successful handling of the encounter and the level of cueing provided to the pilots. They also noted that incorrect DSS command information led to more stall recovery errors than systems that provided simple status information, even when the latter were in error. Inaccurate command display information was especially problematic in *unfamiliar* icing conditions (the authors' term for tailplane icing), and DSS errors compromised the operator's ability to evaluate and respond to other valid cues that were presented (Sarter & Schroeder, 2001, p. 8). As the literature shows, contemporary stall warning systems, which can be characterized as DSS tools in icing, sometimes experience Misses and FAs in icing conditions. This violates one of Billings' key principles for human-centered automated systems: machine processes must be predictable if the automation is to help, rather than hinder, the human operator's situational awareness (1997).

Green et al. observe that the majority of icing encounters are successfully negotiated, but they also caution that "remarkably, (there is) little pragmatic understanding of what made those few unsuccessful and how they might be avoided" (1996, p. 2). The authors contrasted the state of icing management with thunderstorm avoidance. They noted that thunderstorms are well understood meteorologically, and that useful tools, such as airborne and ground weather radar, are available to manage interactions with thunderstorms. In contrast, icing encounters tend to be very spatially and temporally localized, which makes accurate forecasting difficult, even though the potential value of such forecasts has been demonstrated experimentally by Vigeant-Langlois, & Hansman (2000). Green et al. also added their voices to the body of critiques concerning the highly subjective and inconsistent nature of current icing reporting terminology and recommended the adoption of an objective, quantitative, graduated parametric (icing) severity index. In the absence of such a forecasting tool, the authors observed that wing-mounted aerodynamic performance monitoring (APM) technologies could give aircrew an "objective indication of the wing's performance… that would allow the pilot to make tactical decisions in a timely and informed matter" (1996, p. 4). As the literature shows, contemporary AoA-based stall warning systems do not monitor airfoil performance and therefore occasionally manifest stall-detection errors in icing conditions. Under these literal Hit and Miss circumstances, Signal Detection Theory provides a useful framework for evaluating the crew / stall warning interactions during airborne icing encounters.

Signal Detection Theory

Signal Detection Theory is widely used in the human factors field. The theory was initially formulated by Peterson, Birdsall, and Fox (1954) and extended in important works by Tanner and Swets (1954) and Green and Swets (1966). SDT provides a formal framework for modeling the outcome of a binary (Yes / No) decision task when an observer attempts to discriminate a signal from the background noise (Figure 4).

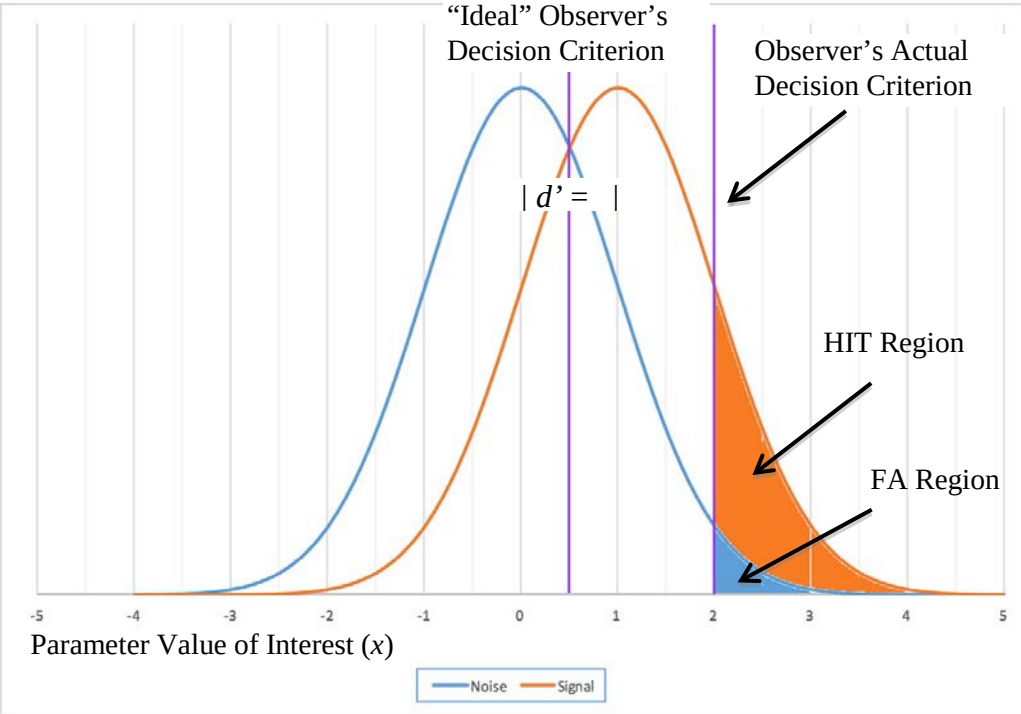

Figure 4. Signal Detection Theory concepts.

The horizontal axis represents the value of the stimulus parameter or *decision variable* (*x*), such as the target's brightness on the radar display. The two curves

represent the probability distributions of the noise and the combined (signal + noise) stimuli. The noise is assumed to be random and normally distributed. For mathematical convenience, this Gaussian distribution is usually re-expressed in non-dimensional terms, with a mean of zero and a variance of one. The signal adds to the noise, so the combined (signal + noise) stimulus is also assumed to have a Gaussian shape but shifted to the right by a distance d' ('*d*-prime'), which represents the mean value for the signal distribution (recalling that the mean of the noise distribution is zero). The quantity d' is a measure of the *sensitivity* of the system. In the commonest SDT models, the variances of the two Gaussian distributions are assumed equal (Lee, 2008, p. 450), as shown by the two identically shaped curves in Figure 4.

The observer must determine which curve a stimulus belongs to, based on its perceived strength. The task is simple if the signal is strong relative to the noise, making the curves widely separated (i.e., d' is large). The required discrimination is much more difficult when the curves have a significant overlap, as shown in Figure 4. SDT addresses the four possible outcome permutations under these circumstances: the presence or absence of a signal, and the response or absence of response from the operator. In SDT terms, these permutations are self-evidently labeled Hit, Miss, FA and Correct Rejection (CR). The FA region is equivalent to a Type I error and alpha level in statistical hypothesis testing, while a Miss corresponds to a Type II error and beta level.

SDT assumes that the operator discriminates between the signal and noise by setting an internal *decision criterion*, above which the stimulus would be categorized as a target (signal + noise), and below which the stimulus would be classified as noise. In Figure 4, the selected decision criterion is $x = 2$, and the area of the shaded regions

indicate the Hit and FA probabilities corresponding to this decision criterion. If the threshold is increased, there is a reduction in FAs at the cost of increased Misses; conversely, a reduced threshold would result in more Hits, at the cost of more FAs. By convention, SDT defines an ideal observer as one who chooses a zero-bias decision criterion threshold that exactly balances the probabilities of the undesirable Misses and FAs. This occurs where the two curves intersect for the equal variance model shown in Figure 4.

Non-ideal observers exhibit *response bias* that is deemed *liberal* when their selected decision criterion is below the ideal observer's and *conservative* when it is above the ideal observer's. The selected decision criterion and the corresponding response bias are affected by the *costs* of making the wrong decision (Miss, FA) and the *benefits* (or *payoffs*) of achieving a correct outcome (Hit, CR). For example, a radar operator trying to avoid two aircraft colliding would set a relatively low (liberal) decision criterion, resulting in high Hits and high FAs. This is because the consequences of an undetected target and subsequent collision could be a major loss of life. Conversely, a radar operator operating an anti-aircraft battery in peacetime would set a very high decision criterion in order to avoid the risk of shooting down a non-threatening target. The reduction in FAs would come at the risk of an elevated Miss probability against a real threat, but this would be an acceptable compromise in peacetime. As these examples show, the ideal observer is a theoretical construct for evaluating bias, not an individual person.

The preceding discussion applies SDT to model an observer's reactions, based on known signal and noise distribution parameters. SDT is more commonly applied in the converse sense: to estimate the SDT parameters, such as system sensitivity and response

bias, based on observed Hit and FA rates. The resulting data are often plotted on Receiver Operating Characteristic (ROC) curves, which allow predictions of one parameter (e.g., response bias) if the other two are known (e.g., Hit and FA rates). Sheriden & Parasuraman (2000) investigated the use of SDT Miss and False Alarm rates in this manner to determine the optimum balance between human operators and automation. This research extends Sheriden & Parasuraman's work by treating the crew and stall warning as interacting SDT systems, in order to investigate the influence of the degraded stall warning system operation under icing conditions on the eventual crew performance outcomes.

For this study, archival icing accident and incident data were analyzed using the SDT framework introduced in Table 2. Accidents and incidents were classified based on the stall warning system performance and the resulting influence on the crew performance outcomes. For example, in the case of the American Eagle 4184 accident (NTSB, 1996b), the aircraft departed controlled flight with no prior stall warning - an SDT *Miss* by the stall warning system, and an incorrect crew response because the precipitating event was not avoided. Conversely, in the Colgan Air 3407 crash (NTSB, 2010a), a premature stall warning led to an incorrect crew response resembling a tail-stall recovery, which led to a main-wing stall. The initial stall warning in the Colgan case would therefore be characterized as an SDT False Alarm. An SDT Hit is exemplified by the crash of Air Florida Flight 90, where the stick-shaker stall warning actuated immediately after the aircraft became airborne and continued until impact into the Potomac river 30 seconds later (NTSB, 1982). CRs represent the null case of no-pending-stall or LOC and no stall warning.

The classification of aircrew responses and the associated stall warning SDT categories form the foundation of the CIRB model of the stall warning system's impact on aircrew performance during icing encounters. Unfortunately, the aircrew's SDT decision criterion cannot be directly measured or manipulated using archival data. The analysis requires a new construct related to the available Hit and FA data. Bayes' Theorem provides this missing link and yields two testable predictions.

Bayes' Theorem

Bayes' Theorem addresses the conditional probability of an event happening subject to the occurrence of another event, as described in the rain and clouds example in Chapter I. Lee (2008) cites numerous advantages of using Bayesian methods with SDT, including their complete representation of uncertainty, but Bayes' Theorem is also useful for the evaluation of the crew / stall warning interactions because of its possible influence on the SDT decision criterion, which is the observer's threshold for identifying a stimulus as a signal rather than noise.

Stalls and stall warnings are both rare occurrences during normal flight operations. In the absence of any overriding factors, the likelihood of a stall increases during icing encounters due to the reduced critical angle of attack with leading edge ice contamination, as discussed in Chapter I. Bayes' Theorem predicts the probability of a stall, given that a stall warning event has occurred, as follows:

$$P\ (Stall \mid Stall\ Warning) = \frac{P\ (Stall\ Warning \mid Stall)\ \ P\ (Stall)}{P\ (Stall\ Warning)} \qquad (7)$$

As equation 7 shows, the probability of a stall, given that a stall warning has occurred, is directly proportional to the a priori probability of a stall occurring (P_{Stall}), which increases with icing. The stall probability is also inversely related to the probability of a stall warning occurring ($P_{\text{Stall Warning}}$). The latter is influenced by the sensitivity and FA rates of the stall warning system that may or may not be influenced by the icing. Bayes' Theorem therefore predicts that the FA rate should increase as the a priori probability of the precipitating event (an actual stall) decreases. In the extreme, the FA rate could become unacceptable when the precipitating event is extremely unlikely (Sheridan & Parasuraman, 2000). This phenomenon may condition aircrew to treat stall warnings in cruise flight as nuisance FAs, even though their perception should change drastically in icing conditions, where P_{Stall} can increase markedly. Aircrew are unlikely to be aware of these Bayesian consequences and are likely to use the decision criterion that they established over many years of uneventful flying when assessing the risk of a stall occurring in icing. During an icing encounter, when the Bayesian probability of a stall increases, aircrew should lower their decision criterion because a given stimulus is more likely to represent a true stall event and less likely to be a FA. This is easily understood if taken to the extreme: There is some level of ice accretion that would produce a 100% probability of a stall, so the crew should treat any signal under these circumstances as a stall. Conversely, a clean aircraft in un-accelerated high-speed flight is unlikely to stall, so a high decision criterion should be set to avoid excessive FAs under these circumstances.

If aircrew fail to adjust their decision criterion and therefore treat all stall warnings equally, then SDT predicts that incorrect crew responses arising from stall

warning Misses would increase during icing encounters while FAs would lead to fewer errors than the baseline non-icing condition (Figure 4). This phenomenon should be more noticeable for vigilance tasks that are characterized by long periods of inactivity, such as flight operations in the cruise segment, because the decision criterion would likely be set at a higher threshold for such tasks. Operators perceive vigilance tasks as fatiguing and stressful (Warm, Parasuraman, & Matthews, 2008) and often fail to respond appropriately when the stimulus appears. In the icing environment, these difficulties could be compounded by the crew's conservative decision criterion, which arises from the conditioning that takes place during routine (non-icing) cruise operations. The situation is exacerbated because stalls do not always present with consistent symptoms in a well-defined sequence, particularly with contaminated airfoils, which makes stall identification and warning interpretation even more difficult (NTSB, 2010a). Flottau (2012) quotes the Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (BEA) report into the loss of Air France 447 following a temporary failure of the aircraft's stall protection features and primary air data indications: "the occurrence of the failure in the context of flight in cruise completely surprised the crew of flight AF447.... The startle effect played a major role in the destabilization of the flight path and in the two pilots' understanding of the situation" (BEA, 2012, p. 209).

The failure of the crew to adjust their decision criterion in icing corresponds to a conservative SDT response bias. The CIRB theory predicts an increase in incorrect crew performances when exposed to stall warning Misses but reduced errors in the face of FAs when compared to the stall warning Hit baseline. Unlike the theoretical SDT analysis

discussed previously, Bayes' Theorem allows this prediction to be tested using archival data.

Summary

The literature shows that airborne icing has a clearly demonstrated potential to cause serious and sometimes catastrophic degradations in aircraft performance, stability, handling qualities, and system performance. Current aircraft stall warning systems are unable to provide consistent and reliable warnings during icing encounters. This is because no existing stall warning directly samples the airflow over the wing where the flow separation occurs, so the stall warning margin provided to the crew is, at best, an informed estimate. The NTSB accident reports and ASRS pre-test data contain numerous instances of aircraft stalling before the receipt of any warning by the crew. In some cases, this has happened during the flare, and the result has been a hard landing; in many others, such as the loss of American Eagle Flight 4184 and Comair Flight 3272, the result has been the tragic loss of the aircraft and all on board (NTSB, 1996b; NTSB, 1998a). Furthermore, no certified system monitors or warns against tail stalls, which significantly increases the challenges faced by the crew when trying to differentiate between a wing or tail stall. This is a critical shortcoming, as the recoveries for the two types of stall are almost diametrically opposed. The application of an inappropriate recovery technique undoubtedly exacerbates the situation and may be unrecoverable, as demonstrated by the loss of Colgan Air Flight 3407 (NTSB, 2010a).

Airframe icing remained on the NTSB's Most Wanted List of transportation safety improvements for small and large aircraft from 1997 to 2011 (NTSB 2012b).

Airframe icing is a continuing problem, despite the FAA's publication of more than 200 Airworthiness Directives (2010e) and multiple regulatory amendments affecting almost every class of aircraft and type of operation. Aside from a small number of simulation studies (Sarter & Schroeder, 2001), there has been very limited research into the interaction between the aircraft's wing-stall or tail-stall status, the operation of the stall warning system, and the outcome of the aircrew's decision-making process. As a result, current efforts at addressing the loss of control problem in icing are reactive rather than proactive, and the expensive, complex measures that have been adopted have not prevented a number of major icing-related accidents.

The current research is based on a modification of the Signal Detection Theory framework (Abdi, 2009) that treats the crew and the stall warning as two interacting SDT systems that can be analyzed using classical SDT methods: The aircraft is either about to experience a wing stall, a tail stall, or no stall at all. The aircraft's stall warning attempts to determine whether a warning should be issued, based on incomplete and possibly erroneous sensor information. These factors result in four permutations of stall warning behavior, self-evidently labeled *Hits*, *Misses*, *False Alarms,* and *Correct Rejections* in SDT terms, which form the IVs for the CIRB model. In turn, the crew uses basic cues from the aircraft (such as buffeting, vibration, and altered control responses) and its stall warning system to determine the stall state of the vehicle. The crew response is either correct or incorrect, and this simple binary measure is the dependent variable for the analysis. The CIRB model suggests that aircrew establish a decision criterion for reacting to a stall warning indication. The crew would treat the combined stimuli from the aircraft's behavior and stall warning system as an actual stall when total stimulus

exceeds the decision criterion, and ignore stimuli below this threshold. Aircrew would establish their individual stall warning decision criterion from extensive exposure to event-free flight, characterized in SDT terms as a vigilance task, which would lead to a very conservative setting in order to avoid overreactions to the relatively high number of False Alarms that are characteristic of such low probability events (Sheridan & Parasuraman, 2000).

Bayes' Theorem indicates that conditional probabilities are directly affected by the a priori event in question, and these conditional probabilities are not commutative (Lee, 2008). This was demonstrated in Chapter I with the rain vs. cloud example. Unfortunately, aircrew probably instinctively treat stall and stall warning probabilities as commutative, so they may fail to adjust their decision criterion sufficiently when encountering serious icing conditions. This results in a strong conservative SDT decision-making bias, which should lead to increased incorrect aircrew responses in the face of stall warning Misses, but reduced errors in the face of stall warning False Alarms. For convenience, this model has been titled the Conservative Icing Decision Bias model. The CIRB predications are testable under experimental conditions if the Hit and False Alarm data could be generated for a range of decision criteria. Under these circumstances, the system's sensitivity and bias could also be determined. Alternatively, if the SDT model parameters are known, any one of the SDT properties (Hit rate, False Alarm rate, and operator sensitivity and bias) could be inferred from the others.

Two comprehensive archival databases were selected to test the conservative bias decision criterion theory: NASA's self-reported Aviation Safety Reporting System (ASRS) incident repository (NASA, 2014) and the NTSB online accident report database

(NTSB, 2012a). Unfortunately, the decision criterion cannot be manipulated experimentally using archival data, so an additional step was required in order to test the hypotheses stemming from the CIRB theory. The solution lay in the application of the Binary Logistic Regression to the archival data. The BLR is a powerful non-linear multivariate method that is used to predict and explain binary categorical (Yes / No) outcomes from a combination of metric or non-metric independent variables (Hair et al., 2010, p. 317). The BLR is extremely resilient to violations of normality and to heteroscedasticity, and its robustness is well suited to the proposed exploratory analysis. The aircraft (wing-stall / tail-stall / no-stall) state and the stall warning SDT classifications were the factors (IVs) evaluated with the BLR analysis. The single dichotomous dependent variable was the *Correct_Response* measure of the crew's initial response to the stall warning (or lack of one). The BLR tied together the factors and measures required to evaluate the two CIRB hypotheses derived from SDT and Bayes' Theorem: CIRB predicts that icing encounters should lead to a significant reduction in aircrew correct responses for the stall warning Miss cases compared to the stall warning Hits. Conversely, the theory predicts that stall warning False Alarms should not result in a significant increase in incorrect aircrew responses compared to the baseline Hit condition. These predictions differ from an equilibrium situation where either error would be expected to have the same influence on crew behavior outcomes. The next chapter details the application of these concepts for analyzing the NASA ASRS and NTSB database archives.

CHAPTER III

METHODOLOGY

Research Approach

This chapter describes the research approach, population and sampling methods, data sources, and data treatment that were applied to test the predictions stemming from the Conservative Icing Response Bias (CIRB) model. A nomothetic exploratory archival analysis (Babbie, 2013, pp. 91-93; Vogt et al., 2012, pp. 86-95) was used to determine the influence of stall warning system behavior on aircrew performance outcomes during airborne icing encounters. Two research hypotheses were evaluated using a Binary Logistic Regression (BLR) analysis of archival NTSB accident data (NTSB, 2012) and NASA ASRS incident data (FAA, 2011a). Records from these two databases were examined and encoded by two subject matter experts (SMEs) on flight operations and certification processes, as discussed in the Delimitation section of Chapter I. The following sections detail the population and sampling methods that were applied to the ASRS and NTSB databases.

Population and Sample Overview

The population was comprised of the fleet of civilian U.S. (*N*-registered), non-amateur built, turbine-powered airplanes. The sample contained two subsets from the population: (1) U.S. registered, non-amateur built, turbine-powered airplanes with icing-related probable cause entries in the NTSB accident database, and (2) icing-related Loss of Control (LOC) ASRS events obtained using the query syntax shown in Appendix A,

which was developed in close cooperation with the ASRS database specialists. The sample date range used for both the ASRS and NTSB queries was January 1, 1988, to October 2, 2015, inclusive. Sub-samples were generated from the ASRS and NTSB samples for which the aircraft stall-state, stall warning system performance, and crew performance outcome could all be determined unambiguously, as described in the following sections. To aid subsequent researchers, the NASA ASRS and NTSB records included in the final sub-samples are tabulated at Appendix C2 and C3, respectively.

Data Encoding

A large amount of archival NTSB and NASA ASRS narrative data was processed for the quantitative BLR evaluation. Two SMEs were employed for this activity. The author drew upon substantial experience with icing stall research in making the judgments necessary for the proposed research (FAA, 1996b; Lerner, 1985; Maris, 1996; Maris, 2009). The literature identifies the benefit of using a second SME to provide independence and quality assurance of the work of the Principal Investigator: Bazeley advocates the use of an independent observer to validate researchers' encoding decisions and stresses the importance of a robust audit trail for achieving the sought-after reliability (2013, p. 151). Babbie also advocates for "some verification" of the researcher's encoding decisions, but notes the overriding importance of "exhaustive and mutually exclusive" code categories (2013, pp. 416, 417). This study adopted both of these recommended safeguards. A highly structured encoding process was employed, using a second SME to provide a quality assurance check on the author's sample selection and data encoding of the ASRS and NTSB databases. The appointed SME is a highly

experienced pilot with 7,500 hours total flight time, and is a certified multi-engine and instrument flight instructor with an FAA Airline Transport Pilot certificate and type ratings for the ATR42/72. In order to familiarize the SME with the protocols involved in this study, the SME was briefed using the checklists and procedures included in Appendix D, which also includes the data collection worksheet. The following sequence was followed for the data sampling and encoding processes:

1. Both SMEs queried the on-line NTSB accident archive using the criteria shown in Table 7. The NASA-provided custom ASRS extract already conformed to the specifications shown in the table, without additional processing.
2. The SMEs compared their recorded ASRS and NTSB sample sizes and addressed any discrepancies in order to achieve identical sample sets.
3. The SMEs examined the ASRS and NTSB samples and rejected records for which the aircraft stall state, stall warning system performance, and crew performance outcomes could not be unequivocally determined. Any discrepancies were addressed to ensure that the resulting sub-sample sets from the two SMEs were identical.
4. The SMEs independently encoded the sub-samples regarding stall warning system performance and aircrew performance outcomes using the criteria discussed in the next section and summarized in Table 4. Discrepancies were discussed openly, and each SME presented the reasoning behind their encoding to the other SME. Any unresolved disagreement in these assessments resulted in the rejection of the entire record, based on its failing to meet the overwhelming evidence threshold listed in the table.

5. The selected sub-samples were recorded in the Appendix D worksheet and subsequently encoded into the software used for the BLR analysis and hypothesis tests.

The preceding steps are described in more detail in the following paragraphs. The stall warning status and aircraft stall state were both relatively unambiguous parameters: They were either present in the record, leading to the record's inclusion in the sample, or they were absent, and the entire record was eliminated from the sample. An ASRS pre-test (described below) comprising 18,214 records, resulted in a 115-case sample with no ambiguities encountered that required a subjective assessment of the stall warning operation and stall warning status. This lack of ambiguity does not imply that the required information was present in all the records; it only indicates the availability of the required data and that resultant treatment of the individual record was easy to determine without guesswork.

The evaluation of the crew responses was a more subjective exercise because it involved an assessment of the interplay between human behavior; a complex environment; and a framework of rules, procedures, and industry norms. For these reasons, a structured methodology was employed to add rigor to the process and to minimize any subjective biases. The author and the independent SME each classified the crew performance outcomes as correct or incorrect based on the sequential application of the criteria shown in Table 4, as evinced by the crew's initial reactions during the onset of the stall, stall warning, or LOC event.

Table 4

Crew Performance Outcome Evaluation Criteria

	Incorrect Crew Performance	Correct Crew Performance
1.	The NTSB probable cause or contributing factor in a factual or final report indicates that the crew's initial response was inappropriate (e.g. BEA, 2012; NTSB, 1996b; NTSB, 2010a).	The NTSB probable cause or contributing factor in a factual or final report indicates that the crew's initial response was appropriate (e.g. BEA, 2012; NTSB, 1996b; NTSB, 2010a).
2.	The ASRS submitter indicated that the crew response was inappropriate.	The ASRS submitter indicated that the crew response was appropriate.
3.	The crew first became aware of the impending stall or loss of control after their onset (i.e., the crew allowed the situation to degrade to the point where control was lost before recognizing this fact).	The crew first became aware of the impending stall or loss of control before their onset and made positive efforts to avoid the event, regardless of the success of the outcome.
4.	An appropriate stall warning was not acted upon in time to avoid a true aerodynamic stall or loss of control.	An appropriate stall warning was acted upon in a timely fashion, regardless of the success of the outcome.
5.	The crew response was markedly different from an accepted norm (i.e., adding power and firmly lowering the nose to prevent a wing stall) (e.g. NTSB, 2010a).	The crew response conformed to the accepted norm, regardless of the success of the outcome.
6.	The crew appeared to be unaware of the stall-state of the aircraft or misdiagnosed its state (e.g. BEA, 2012).	The crew appeared to be aware of the stall-state of the aircraft.
7.	There is *overwhelming evidence* from a subjective review of the record that the crew's initial response was inappropriate.	There is *overwhelming evidence* from a subjective review of the record that the crew's initial response was appropriate.

Only the sixth criterion listed in Table 4 was subjective, and this was the reason for the strict imposition of the overwhelming evidence threshold (criterion 7). If the crew performance could not be determined using these criteria, or if the two SMEs differed on the coding of a particular record, then the de-facto overwhelming evidence criterion was not met, and the entire record was rejected, listwise. The heuristics in Table 4 were evaluated and found usable in the NASA ASRS pre-test, which gave considerable confidence for their application to the NTSB data. Nevertheless, the ASRS and NTSB database sampling methods differed because ASRS data are always de-identified and do not contain CVR or FDR data, but usually contain first-person narratives. Conversely, the NTSB cases often contained CVR and FDR information that made the assessments easier for these cases. The following sections address the individual sampling approaches used for the ASRS and NTSB databases, after a short discussion regarding inter-rater reliability between the two SMEs.

Inter-Rater Reliability

There are several precedents for the use of two raters in peer reviewed works and dissertations. Joslin (2013) employed two SME raters in a comparative study of Runway Incursion Models and cited numerous other works where two raters were used, including two aerospace studies (Hendriksen & Holewijn, 1999; Zuschlag, 2005). Bazeley mentioned the use of “a second person” to check coding reliability but notably did not extrapolate the concept to greater numbers of coders (2013, p. 150). Instead, Bazeley placed her main emphasis for improving reliability on “the strength of your argument and clarity and comprehensiveness of your evidence” (2013, p. 151). For these reasons, and

as a result of the procedural precautions taken for the data encoding process, two raters were deemed appropriate for this study, one of whom was the author acting in the capacity of an SME, while the second provided a quality assurance function. The two raters interacted regularly to converge on a common outcome, so the second SME should not be considered as an independent rater for statistical purposes.

Inter-rater reliability between the author and the second SME was not quantified because the methodology required complete consensus on all retained records, yielding a de-facto inter-rater reliability of 100%. More sophisticated measures, such as Cohen's kappa (*k*), could have been employed to contrast the observed agreements between raters with chance outcomes (Cohen, 1960), but the kappa statistic can seriously underestimate the inter-rater reliability when the contingency table is skewed by a prevalent response (Feinstein & Cicchetti, 1990). Based on the ASRS and NTSB pilot studies, such skewing was anticipated, so the use of the kappa statistic would have added little insight to the inter-rater reliability, negating its usefulness.

ASRS Database, Sample, and Pre-test

The NASA ASRS sample was obtained via a customized data extract in XLSX format for subsequent processing using a combination of MS® Excel™, MS® Access™, IBM® SPSS™ Statistics and SAS® Enterprise Miner™ software applications. The search query was developed via personal communications between the author and the ASRS data specialist with the objective of being sufficiently broad to avoid the accidental elimination of relevant cases. This conservative approach resulted in the inclusion of superfluous records that had to be scrubbed before the BLR analysis. The final query

syntax is listed in Appendix A. Important ASRS limitations are included in Appendix B, and the ASRS imported data structure is shown in Appendix C1.

The ASRS data complemented the picture of the icing phenomenon provided by the NTSB database because the accident data, unsurprisingly, yielded very few correct crew responses to the stall warning cues. The ASRS cases of interest were expected to contain a higher proportion of crews responding correctly to the icing exigencies, thereby avoiding an appearance in the NTSB's database.

ASRS pre-test. A pre-test was performed to validate this assumption and to determine if the quality and quantity of available ASRS icing encounter data would support the proposed BLR analysis. The sample was comprised of 115 ASRS records from the inception of ASRS in January 1988 to October 2, 2015. ASRS data were obtained via a customized extract in XLSX format from the publicly available database of 182,214 records. The pre-test query string included in Appendix A was developed and refined in the course of a number of personal communications between the author and the ASRS database Project Manager. Only five CRs were observed in the resulting sample, which is unsurprising as these represent the null case of no-pending-stall or LOC and no stall warning. These cases were excluded from the analysis because an accepted minimum bin size for a BLR is 10 observations per estimated parameter (Hair et al., 2010, p. 322). Table 5 details the steps in the refinement of the sample used for the ASRS pre-test.

Table 5

ASRS Pre-test Population and Sample Summary

Group	Size
ASRS sample date range	January 1, 1988 - October 2, 2015
Full ASRS dataset, since inception	182,214 records at October 2, 2015
Initial sample, per tailored request #7212[b]	381 cases
Scrubbed sample for further analysis	200 cases
Final sample size for further analysis	**115 cases**[a]

Note. [a]All five "Correct Rejection" (CR) cases were eliminated because of Logistic Regression sample-size restrictions. [b]The search string used to generate the 381 sample cases from the complete 182,214 ASRS data record is included in Appendix A.

ASRS pre-test findings. Aircrew responded correctly to the icing encounter in 45% of the pre-test cases. The logistic regression model was statistically significant, $\chi^2(2) = 24.615$, $p < .0005$. The Hosmer and Lemeshow test was not statistically significant ($p = 1.0$), indicating that the model was a satisfactory fit despite the small sample size, and no multivariate outliers were noted using a two-standard deviation cutoff. The model explained 33.6% (Nagelkerke R^2) of the variance in crew performance outcomes and correctly classified 75.7% of the cases. Sensitivity was 54.0%, specificity was 92.3 percent, positive predictive value was 15.6%, and negative predictive value was 72.3%. Of the three predictor SDT-state variables, two were statistically significant: STD_MISS and STD_FALSE_ALARM. Compared to a missing stall warning, aircrew had 16.54 times greater odds of performing correctly when faced with a stall warning False Alarm. The BLR pre-test indicated that a significant proportion of successful crew

performance outcomes could be predicted using the SDT model, paving the way for the full study and its associated hypothesis tests. An unexpected finding from the pre-test was the need for the inclusion of a new *System_Issue* variable to address incidents caused by systems failures directly related to the icing conditions that were not initially related to stalls or LOC. The loss of Air France 447 (BEA, 2012) due to a stall resulting from the aircrew's response to pitot icing exemplifies the need for this new variable.

NTSB Accident Database, Sample, and Pre-test

NTSB accident data were downloaded for the period January 1, 1988, to October 2, 2015, inclusive, from the public on-line query page: http://www.ntsb.gov/_layouts/ntsb.aviation/index.aspx. The start date represents the first availability of full NTSB docket data on-line and also corresponds to the start date of Green's (2006) study of U.S. inflight icing accidents that is discussed below. The October 2, 2015, cutoff corresponded with the end-date of the NASA ASRS data extract used in the pre-test; it also ensured the majority of the NTSB probable causes had been established in time for the data reduction, based on a six-month buffer between the record retrieval date and the time of the data processing.

NTSB accident archive processing. The NTSB archive was imported in delimited text format directly into Excel™ and IBM® SPSS™, using the " | " (vertical bar) symbol as the delimiter. Figure 5 shows the pre-test results of importing the NTSB accident database into SPSS™; identical results were achieved with the Excel™ import.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns
1	EventId	String	15	0		None	None	15
2	InvestigationType	String	10	0		None	None	10
3	AccidentNumber	String	13	0		None	None	13
4	EventDate	String	12	0		None	None	12
5	Location	String	39	0		None	None	39
6	Country	String	15	0		None	None	15
7	Latitude	Numeric	12	6		None	None	12
8	Longitude	Numeric	13	6		None	None	13
9	AirportCode	String	6	0		None	None	6
10	AirportName	String	32	0		None	None	32
11	InjurySeverity	String	13	0		None	None	13
12	AircraftDamage	String	13	0		None	None	13
13	AircraftCategory	String	14	0		None	None	14
14	RegistrationNumber	String	8	0		None	None	8
15	Make	String	32	0		None	None	32
16	Model	String	19	0		None	None	19
17	AmateurBuilt	String	5	0		None	None	5
18	NumberofEngines	Numeric	3	0		None	None	8
19	EngineType	String	15	0		None	None	15
20	FARDescription	String	31	0		None	None	31
21	Schedule	String	6	0		None	None	6
22	PurposeofFlight	String	25	0		None	None	25
23	AirCarrier	String	37	0		None	None	37
24	TotalFatalInjuries	Numeric	4	0		None	None	8
25	TotalSeriousInjuries	Numeric	3	0		None	None	8
26	TotalMinorInjuries	Numeric	3	0		None	None	8
27	TotalUninjured	Numeric	5	0		None	None	8
28	WeatherCondition	String	5	0		None	None	5
29	BroadPhaseofFlight	String	13	0		None	None	13
30	ReportStatus	String	16	0		None	None	16
31	PublicationDate	String	12	0		None	None	12
32	V1	Numeric	1	0		None	None	8

Figure 5. SPSS import of NTSB accident database variables.

NTSB accident archive sample. The sampling method was patterned after Green's (2006) comprehensive archival study of U.S. inflight icing accidents and incidents between 1978 and 2002 using NTSB, FAA, and ASRS data. Green's initial

NTSB sample contained 11,174 cases, which was reduced by shortening the review period and by careful selection of the Boolean search terms. Green also eliminated a number of the retrieved reports relating to engine icing or icing system anomalies by adjusting the search terms and by manual inspection. The final search string employed by Green ("icing | freezing | rime | glaze | sleet | frost") yielded a more manageable 2,212 cases, which were then manually reduced to a working sample of 693 accidents for Green's detailed analysis. Based on the similarity between the scope of the present study and Green's work, Green's strategy and search string were reused for the current research. This will facilitate meaningful comparisons with Green's earlier findings. As indicated in the delimitations section (Chapter I), the analysis required a level of inference regarding icing-induced loss of control and stall events, which sometimes presented solely as "uncontrolled descents" in the accident data (Green, 2006, p. 3). A similar approach to Green's was adopted to address this issue, although additional formalized structure was imposed, as previously discussed.

The NTSB accident database is comprehensive because of the legal obligation to report all accidents (NTSB, 2010b), so a large number of icing-related records were anticipated, in line with Green's findings. Unlike the ASRS database, many of the NTSB cases related to accidents without survivors, making it impossible to determine the stall warning performance and crew outcomes without access to a CVR or FDR. The NTSB archive was therefore screened for aircraft that were likely to be equipped with CVR and / or FDR equipment in order to facilitate the extraction of the factors and measures required for the BLR analysis. As the FDR and CVR requirements have evolved with time and contain grandfather clauses, it was not possible to specify which records would

meet the selection requirement before a detailed examination of the data. Based on the current FARs at the time of publication, the following aircraft classes were expected to have FDR and / or CVR equipment installed: U.S. civil registered, multiengine, turbine-powered airplane with 10 or more passenger seats (FAA, 2010d) and "U.S. civil registered, multiengine, turbine-powered airplanes… (with) six passengers or more and for which two pilots are required by type certification or operating rule" (FAA, 2010d). In order to capture these classes of aircraft as simply as possible using the NTSB database search criteria, the NTSB sampling frame was limited to non-amateur built turbine (turbojet, turbofan, or turboprop) airplanes. The turbine limitation was not strictly necessary for the ASRS data because the ASRS pre-test indicated that the sample size would remain manageable, even without this filtering, and also because crew narratives are almost always available in the ASRS records, which obviates the need for CVR or FDR equipage. Nevertheless, the NTSB turbine-only limitation was also applied to the ASRS data in order to maximize the similarities between the sample sets.

The filtered NTSB turbine aircraft accident *Probable Cause* synopses were manually examined to select cases where an aerodynamic stall or loss of control were encountered, with the remaining records discarded. The crew performance outcomes and stall warning system behavior were then encoded for the retained records. For those records where the crew and stall parameters could not be determined from the NTSB on-line synopses, the NTSB full narratives were consulted, where available, to determine the missing parameters. If the NTSB full narratives were unavailable or failed to include the required parameters, then the affected records were discarded. The full narratives for the

remaining records were used to encode the stall warning and crew outcomes for the subsequent BLR investigation.

NTSB pre-test. A pre-test was performed to evaluate the potential sample size of the icing-related occurrences contained in the NTSB database. NTSB accident data for the period January 1, 1978, to October 2, 2015, were downloaded in text format directly from the publicly available archive via the on-line query page: http://www.ntsb.gov/_layouts/ntsb.aviation/index.aspx. The pre-test sample was obtained using Green's (2006) search string with the following syntax: "icing" or "freezing" or "rime" or "glaze" or "sleet" or "frost." The sample was constrained to non-amateur built airplanes with turbine engines (i.e., turboprop, turbofan, and turbojet classifications in the NTSB database). An additional extract was performed with piston-powered aircraft in order to gauge the size of this sub-group, in case it was required in order to achieve an adequate overall sample size, but this subset was not used in the final analysis. For reference, before the application of the keywords and engine-type delimiters, 5,110 records were retrieved for the sample period, of which 855 met the search string criteria. The resulting sampling frames were not examined further to determine the actual usable sample size, but approximately 30% of the sample frame was expected to result in usable sample data, based on the ASRS pre-test results (Table 5). The NTSB pre-test results and initial estimates of the sample sizes for the BLR analysis are shown in Table 6.

Table 6

NTSB Pre-test Sample Size Findings

Engine Type	Pre-test Sample Size (records)	Estimated Final Sample Size (records)[a]
Turboprop	106	32
Turbofan	42	14
Turbojet	13	4
Total Turbine Airplane Sample Size	**161**	**50**
Piston	694	208
Maximum Sample Size	**855**	**258**

Note. Retrieved from http://www.ntsb.gov/_layouts/ntsb.aviation/index.aspx based on Instrument Flight Rules (IFR) operations between January 1, 1978, and October 2, 2015, with the following query syntax: "icing" or "freezing" or "rime" or "glaze" or "sleet" or "frost." [a]Estimated sample size is 30% of sampling frame size, based on ASRS pre-test findings.

Hair et al. (2010, p. 333) indicate that each BLR group should have a minimum bin size of 10 times the number of estimated model coefficients. Based on the IV factors under consideration (stall warning system state and aircraft stall state), a minimum sample size of 20 was desired. The findings from the NTSB database pre-test indicated that the BLR analysis should be viable using turbine-engine aircraft records alone, although the option of including the piston-engine airplanes was retained, if needed to achieve an acceptable sample size.

Database Summary

The key characteristics of the ASRS and NTSB databases are summarized in Table 7.

Table 7

Key Database Characteristics

Characteristic	ASRS Data	NTSB Data
Database source	Custom extract provided by ASRS data specialist	Downloaded from NTSB accident database website[a]
Data format	MS XLSX file	Delimited TXT File
Extract start date	January 1, 1988	January 1, 1988
Extract end date	October 2, 2015	October 2, 2015
Primary Filter	Customized Boolean search query (Appendix A)	Turbine Engine Type
AircraftCategory Filter	Airplane	Airplane
AmateurBuilt Filter	N/A	No
EngineType Filter	N/A	Turbine aircraft (turbojet, turboprop, turbofan)
WeatherConditions Filter	N/A	All
Suitable Information[b] Filter	SME evaluation	SME evaluation

Note. [a] Retrieved from http://www.ntsb.gov/_layouts/ntsb.aviation/index.aspx. [b]Suitable information is defined as adequate to determine aircraft stall state, stall warning system state, and crew response.

Data Reliability

Field defines reliability as "the ability of a measure to produce consistent results when the same entities are measured under different conditions" (2009, p. 792). Three techniques were used to enhance the reliability of the analysis. The first entailed the application of the Hosmer and Lemeshow test, which has specific application to the BLR method. A model is deemed not to have a poor fit when the Hosmer and Lemeshow test value does not achieve statistical significance at the desired level. The second method attempted to take advantage of the two completely independent datasets used for the analysis. The BLR was run independently on each of the ASRS and NTSB data samples, with the objective of validating the results across the two analyses. Unfortunately, sample size limitations precluded this approach, as discussed in Chapter IV, which required the use of a merged database for the BLR. The reliability of the merged database was therefore established by partitioning the sample into training and holdout (validation) samples. This required the BLR to be run a number of times while varying the relative training / holdout proportion in order to attain the best balance between the model specification (which required a large training sample) while reducing variance in the validation dataset (which required a large holdout sample). A satisfactory compromise was achieved with a 50 – 50 split between the two partitions, as discussed in Chapter IV.

A second important reliability consideration stemmed from the self-reporting nature of the ASRS source data, which introduced the potential for significant bias. Although there are definite benefits of filing an ASRS report, such as indemnification from FAA prosecution for non-criminal violations (FAA, 2011a), this enticement could

not completely overcome the self-selection bias inherent in the program. Appendix B contains an excellent summary of the limitations of the ASRS data provided by the program office. Another source of ASRS bias arises because a noteworthy event has already occurred for a record to appear in the ASRS database, so every ASRS record likely corresponds to a large number of uneventful icing encounters that would never appear in the database. This is undoubtedly the reason for the relatively poor (45%) aircrew correct response outcome rate observed in the ASRS pre-test. For many of these records, the first indication reported by the crew in the ASRS data was often an LOC, which automatically constituted an incorrect aircrew response based on the criteria in Table 4. These incorrect crew response classifications were not intended to impute blame, because there were several instances where no warning was given to the crew before the LOC. The terminology should be understood simply as a category for the dependent variable, with no blame attribution. Conversely, an event would be categorized as an incorrect crew response if the crew's initial responses were inappropriate, even if the event did not result in an accident. Despite these shortcomings, the ASRS data provided an important perspective that would have been missing had the NTSB accident data been analyzed in isolation. In summary, the ASRS and NTSB databases should offset each other's weaknesses to some degree, thereby increasing the reliability of the study.

Data Validity

Field defines validity as "evidence that a study allows correct inferences about the question it was aimed to answer" (2009, p. 795). A number of validities were considered

for this exploratory design, including ecological validity, content validity, face validity, and external validity. *Ecological validity* relates to the absence of bias that could be caused by the researcher's presence (Field, 2009, p. 12). As there was no researcher observing the targeted archived events as they unfolded, the ecological validity of the research should be sound. Babbie defines *content validity* as the degree to which a measure covers the range of meanings included within a concept (2013, p.152). The factors and measures used in this study were robust and fully encompassing, either due to their binary nature (for the IV and several of the DVs) or because their definition comprehensively covered all the available outcomes (i.e., the SDT Hit, Miss, FA, and CR permutations). For these reasons, the content validity should also be high. Babbie describes *face validity* as the quality of an indicator that makes it seem to be a reasonable measure of some variable (2013, p. 151). Once again, the characterization of the stall warning system's performance in SDT terms (Hit, Miss, FA, and C.) had high face validity and very little subjectivity. The face validity of the Correct_Response measure is much harder to establish because it can be very difficult to divine the crew's thoughts as they responded to icing challenges. As previously noted, anti-ice and deice systems have occasionally been overwhelmed by severe icing resulting in the loss of the aircraft, despite textbook crew responses. Conversely, some crews have survived unscathed, despite their executing a series of ill-advised actions. The structured approach employed for evaluating crew responses (Table 4) should mitigate bias from these characteristics, thereby improving the external validity of the research.

Shadish, Cook, & Campbell define *external validity* as "…whether the cause-effect relationship holds over variation in persons, settings, treatment variables, and

measurement variable" (2003, p. 38). External validity is strongly influenced by the quality of the research design and by the robustness of the statistical tools used in the analysis. Although it is difficult to control the validity of the archives used in any exploratory research, a number of techniques were used to increase the validity of results derived from them. These included *triangulation* (Vogt et al., 2012, p. 113), whereby conclusions were drawn using independent paths and the development of a thick understanding of the situation by examining the phenomena in the most representative possible real environment (Vogt et al., 2012, pp. 71, 72). This study attempted to implement both of these techniques by sampling from independent pools of highly pertinent data. The large number of U.S. registered aircraft and their varied operations should ensure sufficient randomization to produce meaningful results. The NTSB accident data were comprehensive due to the mandatory nature of aircraft accident reporting. In contrast, the ASRS reports were voluntary and subject to self-reporting and response biases, but the NTSB accident data should mitigate these effects, as should the anonymous and altruistic nature of ASRS reporting, with its associated incentive of eliminating noncriminal event enforcement action. The robustness of the statistical tools employed in the study is another major factor in the achievement of good external validity: The BLR was selected primarily because of its inherent resilience to heteroscedasticity and to violations of normality (Hair et al., 2010, pp. 317, 321). In combination, these considerations should help achieve good external validity and the consequent generalizability of the conclusions to be drawn from the research.

Treatment of the Data

Data processing was accomplished using a phased approach with the application of four software packages: Microsoft® Excel™, MS Access™, IBM® SPSS™ Statistics, and SAS® Enterprise Miner™. These applications were used collectively to perform the four data treatment activities: importation, scrubbing, variable encoding, and BLR execution. The ASRS and NTSB data samples were initially imported into Excel™ and SPSS™, where the two datasets were merged while retaining the identification of the source database for stratification purposes. The combined record was manually scrubbed for duplicates based on event dates, and records with missing values for any of the key variables were rejected on a listwise basis. The scrubbed data were then manually encoded for the selected BLR variables and imported into Access™ to facilitate review and comparison between the two SMEs. These processes are described in the following sections.

Data importation and scrubbing. The ASRS and NTSB pre-test findings indicated that the proposed Boolean keyword searches would return a considerable number of unwanted records that were unrelated to airframe icing, such as engine compressor stalls and carburetor icing events. These confounding records were eliminated by inspection of the imported data files using the Microsoft® Excel™ and MS Access™ applications. Bazeley's "describe, compare, and relate" (2013) strategy was used for the selection, scrubbing, and encoding of candidate cases, and to facilitate the manual filtering of these extraneous records. The scrubbed data were encoded for the

factors and measures to be used in the subsequent BLR evaluation of the CIRB hypothesis.

Variable encoding. The ASRS and NTSB databases did not contain several of the factors and measures required for the BLR analysis, so these variables were manually encoded into the data files. The data operationalization was accomplished by adding an additional column into the Excel™ spreadsheet for each desired variable shown in Table 8. The SME hand-encoded the appropriate values for each new variable based on scrutiny of the NTSB and ASRS narratives using the heuristics shown in Table 4.

Table 8

Primary BLR Variables

Variable Name	Function	Attribute
Correct_Response	DV	Binary Y/N
SDT_Class	IV	Categorical: Hit, Miss, FA, CR
Pending_Wing_Stall	IV	Binary Y/N
Tail_Stall	IV	Binary Y/N
Stall_Warning	IV	Binary Y/N
System_Issue	IV	Binary Y/N

Missing data. The CIRB model postulates that the *correct aircrew performance outcome* construct can be modeled through the application of Bayes' Theorem to an SDT theoretical structure. This relationship is pivotal, because the Conservative Icing Response Bias at the heart of the CIRB model represents a cognitive shift that eludes direct measurement, so the BLR analysis of the aircrew responses was a vital construct for quantifying the intangible CIRB effect. For this reason, it was essential to avoid extraneous assumptions about the data in order to avoid skewing the outcome. Accordingly, the BLR required an unambiguous response to three questions in order for a record to be included in the analysis. The first consideration was whether the crew's initial response to the incipient stall, loss of control, or FA could be determined, regardless of a successful or failed outcome. A correct response entailed implementing appropriate wing-stall or tail-stall prevention and / or recovery procedures subsequent to a stall warning alarm or other indications of an incipient stall condition. A correct response also required that no stall prevention or recovery action be undertaken under FA conditions. The second consideration related to the SME's ability to characterize the stall warning system's performance in SDT terms (Hit, Miss, FA, or CR). The final consideration required the unambiguous determination of the aircraft's actual stall status (wing stall, tail stall, or no imminent stall). If any of these characteristics could not be unambiguously determined, no imputation was attempted, and the entire record was deleted from subsequent analysis (listwise, in SPSS™ terminology). This approach was essential to avoid skewing the BLR analysis with erroneously categorized records, which was an important consideration given the relatively small samples under consideration.

Duplicate data. The ASRS and NTSB data records each carried unique identifiers, so few, if any, duplicate records were anticipated within either dataset. It was, however, anticipated that some ASRS reports could be filed for events that led to accidents, in which case some duplication was to be expected. As ASRS data are fully de-identified, automated methods could not be used that depend on unique identifiers, such as the aircraft registration, for de-duplication. Accordingly, duplicate records were manually screened, identified, and merged using a combination of unique field data, such as occurrence date, aircraft class, etc. The BLR was executed once the data had been properly scrubbed with suitable attention to duplicate records and missing values.

Binary Logistic Regression Overview

This section describes the mechanics of the BLR as outlined by Hair et al. (2010, p. 317-344) and also addresses the specific application of the method for the testing of the CIRB hypothesis. The BLR was the appropriate multivariate technique for the proposed investigation based on decision-tree classification methodology developed by Hair et al. (2010, pp. 12-13). Table 9 outlines the process and decision nodes used to arrive at the BLR methodology.

Table 9

Multivariate Method Selection Decision Tree

Decision Node	ASRS Attribute
What type of relationship is being examined?	Dependence
How many variables are being predicted?	One dependent variable in a single relationship (i.e., Correct_Response)
What is the measurement scale of the dependent variable?	Nonmetric (i.e., binary)
Appropriate Multivariate Method	**Linear Probability Model: Binary Logistic Regression**

Note. Adapted from "Multivariate data analysis" (7th ed.), by Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E., (2010), Upper Saddle River, NJ: Prentice Hall, pp. 12 and 13.

The BLR is a subset of logistic regression methods "formulated to predict and explain a binary (two-group) categorical variable" (Hair et al., 2010, p. 317). For the present study, the binary dependent variable (DV) was the Correct_Response measure, which categorizes initial aircrew responses to icing-induced stall, stall warning, or loss of control events, as either correct or incorrect. Candidate independent variables for a BLR can be any combination of nonmetric or metric independent variables (IV). The primary IVs for this investigation were the four Signal Detection Theory (SDT) classifications of the stall warning system response to wing or tail icing: Hit, Miss, FA, or CR. Indicator Coding was used to convert the four levels of SDT IVs into dichotomous values through the use of dummy variables. The SDT Hit IV was used as the reference variable for the indicator coding, and was therefore excluded from the regression analysis. The BLR

processed these IVs in a manner that maximized the discrimination between the two DV states by using a logistic or logit curve shown in Figure 6 (Hair et al., 2010, pp. 321-323). As the figure illustrates, the logit curve is a continuous sigmoid function that asymptotically approaches the values zero and one, while represent the two possible states of the binomial dependent variable. The figure highlights examples of correctly classified and misclassified data, based on the logit model.

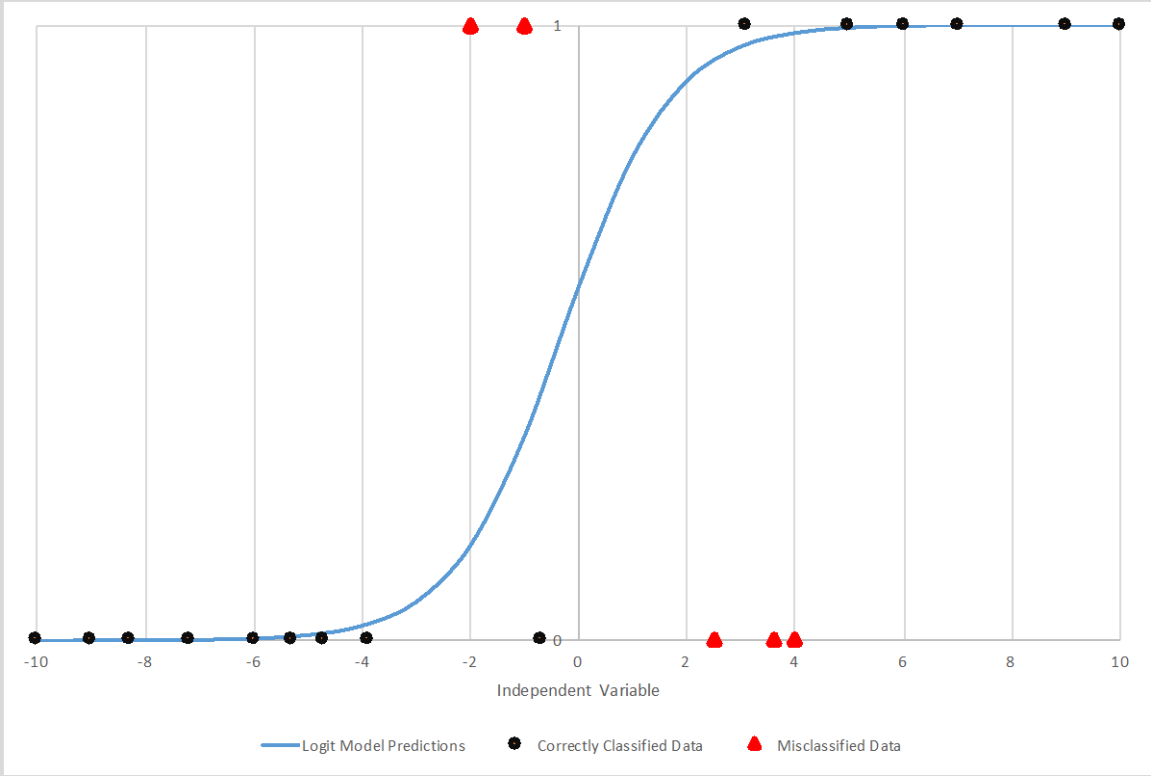

Figure 6. Sample binary logistic regression logit function. The figure shows the source data and a sample logit model with correctly classified and misclassified data identified. Derived from Hair (2010, p. 322).

Hair et al. showed that the coefficients (b_i) for the variables (X_i) that define the logistic curve can be calculated using the mathematically equivalent logit (Equation 8) and exponentiated / Odds Ratio (Equation 9) formulae (2010, p. 326):

$$Logit_i = \ln\left(\frac{prob_{event}}{1 - prob_{event}}\right) = b_0 + b_1X_1 + \cdots + b_nX_n \quad (8)$$

$$Odds_i = \left(\frac{prob_{event}}{1 - prob_{event}}\right) = e^{b_0 + b_1X_1 + \cdots + b_nX_n} \quad (9)$$

These equations highlight the inherently non-linear relationship between the IVs and DV in a BLR, unlike the standard multiple regression techniques, which require a linear relationship. Further, the error terms of the discrete BLR dependent variable are binomially (i.e., not normally) distributed and the error variances are not constant across the IV values (heteroscedasticity). Although these properties violate the statistical requirements of linear regression methods, the BLR technique is not affected by such violations.

BLR model fit. The distinctive nature of the logistic curve and its underlying assumptions require a different approach to model fit from traditional regression methods. BLR fit is evaluated using statistical measures that are unique to the BLR technique. These tests evaluate overall model fit, differences between models, and the significance of the parameters within a model. Overall fit is evaluated using *maximum likelihood estimation* to derive the *-2 log likelihood* (- *2LL*) value for the model, which is

analogous to the sum of square errors obtained in regression analysis (Hair, et al., 2010, p. 327). Lower likelihood values indicate a better model fit, and a perfect fit is indicated by a zero -2LL value.

A second statistical measure of the overall BLR model fit is provided by the Hosmer and Lemeshow test that evaluates the significance of differences between the actual DV values and the expected values derived from the model. Smaller differences are desired, and an acceptable model fit is indicated by a non-significant Hosmer and Lemeshow result. Good model fit is not necessarily a measure of practical significance, so the -2LL and Hosmer and Lemeshow tests are supplemented by the *Pseudo* R^2 statistic that is equivalent to the multiple regression coefficient of determination. Pseudo R^2 is a measure of the statistical difference between two models, evaluated using a chi-square significance test of the difference between their respective –2LL values. Pseudo R^2 is interpreted in a similar manner to coefficients of determination and has a range of zero to one, with one representing a perfect fit. The Cox and Snell R^2 and Nagelkerke R^2 are refinements to the pseudo R^2 test and are assessed identically (Hair et al., 2010, p. 339). The results of this study are presented in terms of the Nagelkerke R^2. The basic pseudo R^2 statistic is calculated as follows (Equation 10):

$$R^2{}_{LOGIT} = \frac{-2\,LL_{null} - (-2\,LL_{model})}{-2\,LL_{null}} \qquad (10)$$

where:

R^2 = pseudo R^2

LL = Log Likelihood

BLR classification accuracy. Once the BLR model overall fit and practical significance are established using the preceding tests, the model's practicality as a predictive tool is assessed by means of a classification matrix, as shown in Table 11. The classification matrix captures a number of key statistics from the BLR model, including the *Hit Ratio,* which is the percentage of combined Hit and CR outcomes successfully predicated by the model, as well as the model's *sensitivity* and *specificity* that correspond to the model's individual SDT Hit and CR ratios, respectively.

BLR coefficient weights. BLR coefficients are equivalent to those for a multiple regression, but the former are logarithmic when the DV is expressed using the logit function. BLR logit coefficients represent the "change in the ratio of the probabilities (the odds)" (Hair et al., 2010, p. 329) that reflect the relative weights of each IV. Logit coefficients are real numbers, and a zero coefficient indicates an odds ratio of 1.0 and a corresponding probability of 0.5. Negative logit coefficients indicate lower odds ratios and corresponding probabilities less than 0.5. Positive logit coefficients indicate the converse. The logit coefficients are useful in determining the strength of the coefficient relationships with the DV.

Coefficient weights can be expressed using an alternative but equivalent format to facilitate the interpretation of the direction of the relationship between the DV and IVs: *Exponentiated logistic coefficient*s are the antilogs of their equivalent logit coefficients. These exponentiated coefficients are positive real numbers, where 1.0 corresponds to a relationship with no direction, and values above and below zero reflect positive and

negative directional relationships between the DV and the selected IV, respectively. The exponentiated format is also useful for determining the relative weights of the coefficients, as shown in Equation 11 (Hair et al., 2010, p. 331):

$$\text{Percentage change in odds} = (\text{Exponentiated Coefficient}_i - 1.0) \times 100 \quad (11)$$

.

Categorical IVs, such as the SDT classifications used in this study, entailed the use of dummy (indicator) variables. In such cases, the calculated percentage change in odds is in relation to the *reference category* chosen for the analysis: SDT Hits for this study. The odds ratios quantify the relative weights of Misses, FAs, and CRs on the crew performance outcomes, in relation to the Hit baseline crew performance.

The Wald statistic. The Wald statistic is used to test the significance of the coefficients derived in a BLR analysis. The statistic is applied and interpreted in the identical manner to the *t* value significance test of multiple regression coefficients. The preceding concepts are summarized in Table 10, which contrasts the BLR parameters with their more familiar multiple regression equivalents.

Table 10

Comparison of BLR and Multiple Regression Parameters

Multiple Regression Property	Equivalent BLR Property	BLR Parameter Range and Interpretation
Total sum of squares, Error sum of squares	– 2LL of base model	Smaller is better: 0 = perfect model fit.
Regression sum of squares	Difference of – 2LL for between models	The model with the smaller -2LL value is the better fit.
F test of model fit	Chi-square test of – 2LL difference	Standard Chi-square significance test for each evaluated model.
F test of model fit	Hosmer and Lemeshow Chi-square test fit test	Non-significant outcomes indicate an acceptable model fit.
Coefficient of determination (R^2)	Pseudo R^2 Nagelkerke R^2	$0 \leq R^2 \leq 1.0$
Coefficient significance (*t*-value)	Wald statistic	Interpreted similarly to the *F* and *t* values used in significance testing of regression coefficients.
Coefficient weight	Logit coefficients	Real numbers. A zero coefficient indicates an odds ratio of 1.0 and a corresponding probability of 0.5. Negative coefficients indicate lower odds ratios and corresponding probabilities less than 0.5. Positive coefficients indicate the converse.
Coefficient weight (alternative formulation)	Exponentiated logistic coefficient	Positive real numbers. An exponentiated coefficient of 1.0 corresponds to a relationship with no direction. Values < 1.0 indicate a negative relationship direction, while values > 1.0 indicate a positive relationship direction.

Note. Adapted from Hair et al. (2010, p. 328).

BLR Execution

The BLR was initially performed on the sample data using IBM® SPSS™ and SAS® Enterprise Miner™ statistical software, as detailed in the BLR and data treatment sections. Each package has differing strengths and weaknesses, and the use of both in parallel provided a useful cross-check of the outcomes. Enterprise Miner™ is a modular application that provides a unified and expandable interface for conducting advanced statistical analyses including linear and non-linear modeling. The program has flexible import and export capabilities, including the import and export to the Excel™ XLSX format that was used for the data scrubbing. The BLR was executed using the GLM function in Enterprise Miner™ and the *Analyze > Regression > Binary Logistic* in SPSS™, acting on the data exported and encoded from Excel™ during the data selection and scrubbing phase. Both packages offered several alternative BLR methodologies that are discussed in the next section.

BLR Models

Three CIRB BLR models, each with four variations, were constructed to perform data mining, hypothesis testing, and validity and reliability evaluations. The models are referenced as the *basic*, *comprehensive*, and *validation* CIRB models. In order to compare the CIRB outcomes with raw stall warning data, the three CIRB models were contrasted with a BLR baseline analysis based solely on the activation of the stall warning system, as shown in Equation 12:

$$\text{Correct Crew Response (Y/N)} = f\,(\text{Stall Warning (Y/N)} + \text{error}) \quad (12)$$

With the exception of the validity and reliability evaluations, each BLR model was evaluated using the entire sample dataset in order to achieve the best fit. The CIRB BLR models were evaluated using five alternative BLR methodologies: the baseline *all-in* technique, with all variables retained as the model iterates, as well as forward, backward, and *stepwise* methods, in which the variables were either added or eliminated sequentially, based on defined criteria such as the Wald statistic. The ASRS pre-test evinced very little difference among these alternate methods, but all four methods were applied and contrasted using the Enterprise Miner™ to evaluate the BLR sensitivity to the specific methodology. Figure 7 illustrates the structure developed in the Enterprise Miner™ software for the evaluation of each of the BLR models described below.

Basic CIRB model. The basic CIRB model was comprised of a constant term and the SDT stall warning performance outcome as the sole independent variables and the aircrew performance outcome as the sole binomial dependent variable. When the BLR incorporates nonmetric (dummy) variables, the resulting Odds Ratios are referenced to a selected baseline category, which is subsequently excluded from the logistic equation. For this study, SDT Hits were selected as the baseline category because Hits reflect the intended functioning of the stall warning system. The remaining SDT parameters (Miss, FA, and CR) were coded as dummy categorical variables, resulting in a BLR relationship of the form shown in Equation 13:

$$Logit_{ccr} = \ln\left(\frac{prob_{ccr}}{1 - prob_{ccr}}\right) = b_0 + b_1 CR + b_2 FA + b_3 Miss + error \quad (13)$$

where:

CCR = Correct Crew Response outcome

Comprehensive CIRB model. The comprehensive model included the same independent variables used in the basic model, with the addition of Wing Stall, Tail Stall, and System Issue IVs (Equation 14):

$$Logit_{ccr} = b_0 + b_1 CR + b_2 FA + b_3 Miss + b_4 Wing_{Stall} + b_5 Tail_{Stall} + b_6 System_{Issue} + error \quad (14)$$

where:

CCR = Correct Crew Response outcome

CIRB validation model. The CIRB validation model was identical to the comprehensive model, except the sample was split into training and holdout (validation) sub-samples to evaluate the BLR reliability.

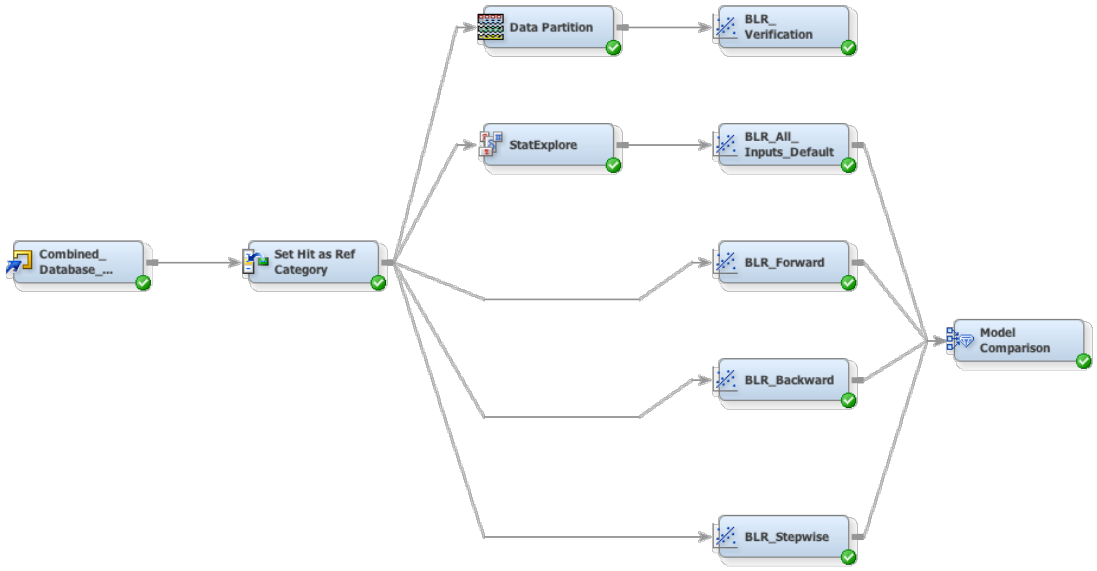

Figure 7. Enterprise Miner™ BLR model structure.

BLR Descriptive Statistics

The computed BLR outcomes were computed and presented using a Classification Matrix (Table 11) and a BLR Outcome Matrix (Table 12). The terminology used in these matrices is explained in Table 13.

Table 11

Sample Binary Logistic Regression Classification Matrix

	Incorrect Crew Response Predicted[a]	Correct Crew Response Predicted[a]	Percentage of Correct Predictions
Incorrect response observed	AA[b]	BB	Specificity[c]
Correct response observed	CC	DD[b]	Sensitivity[d]
Figures of Merit	NPV	PPV	Correctly Classified[e]

Note. [a]Cut value: .50. [b]Lead-diagonal elements of matrix represent correct model predictions. [c]*Specificity* = (AA / (AA + BB))% [d]*Sensitivity* = (DD / (CC + DD))% *Positive predictive value* (PPV) = (DD / (BB + DD))% ; Negative *predictive value* (NPV) = (AA / (AA + CC))% .

Table 12

Sample Binary Logistic Regression Outcome Format

							95% C.I. for EXP(B)	
	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
SDT_MISS (Reference)			24.615	2	.000			
SDT HIT	.667	.563	1.406	1	.236	1.949	.647	5.872
SDT FA	2.806	.566	24.603	1	.000	16.537	5.458	50.112
Constant	-1.119	.288	15.109	1	.000	.327		

Table 13

Binary Logistic Regression Calculated Parameters

Statistic	Symbol	Meaning[a]
Degrees of Freedom	df	Degrees of Freedom
Exponentiated Logistic Coefficient	Exp(B)	An alternative expression of the Logistic Coefficient. Always positive. Values > 1 indicate positive relationships; values < 1 indicate negative relationships between the IVs and DV.
Hosmer and Lemeshow significance	N/A	A test for BLR model fit. A good fit is indicated when the Hosmer and Lemeshow test is not statistically significant.
Logistic Coefficient (logit) and 95% confidence interval	B	Weighting factor for each IV in relation to its discriminatory power. A zero indicates 50 / 50 odds. Negative numbers indicate probabilities < 50%, positive numbers indicate probabilities > 50%.
Nagelkerke R^2	R^2	A *Pseudo* R^2 measure applicable to the BLR technique that is analogous to the coefficient of determination ($R^{2)}$ in a multiple linear regression.
Negative Predictive Value	N/A	Percentage of incorrect crew performance outcomes correctly predicted by the model related to the total number of incorrect crew performance outcomes predicated by the model. 100% is ideal.
Percentage of Correct Aircrew Responses	N/A	The number of correct crew performance outcomes divided by the total number of records, after scrubbing.

Statistic	Symbol	Meaning[a]
Percentage Correctly Classified (*Hit Ratio*)	N/A	The percentage of crew performance outcomes correctly classified by the BLR model.
Positive Predictive Value	N/A	Percentage of correct crew performance outcomes correctly predicted by the model related to the total number of correct crew performances predicated by the model. 100% is ideal.
Sensitivity	N/A	Percentage of correct crew performance outcomes correctly predicted by the model related to the total number of correct crew performance outcomes (i.e., true positives). 100% is ideal.
Specificity	N/A	Percentage of incorrect crew performance outcomes correctly predicted by the model related to the total number of incorrect crew performance outcomes (i.e., true negatives). 100% is ideal.
Standard Error	SE	Standard Error of the BLR coefficient.
Statistical Significance (χ^2 *p* value)	Sig.	Significance level corresponding to the Wald statistic. Analogous to the *t*-test significance level in a multiple regression.
Wald Statistic	Wald	Statistical significance of each BLR coefficient. Analogous to the *t* value in a multiple regression.

Note. [a]Definitions from Hair et al. (2010, pp. 318, 319, 331). Quotation marks omitted from embedded verbatim text for clarity.

Reliability testing. BLR reliability was established by partitioning the combined ASRS and NTSB database into training and holdout (validation) samples (Hair et al., 2010, p. 341). The BLR was computed using the training partition, and the reliability was established using the validation partition. The two samples were compared on the basis of their misclassification rates and average squared error. The BLR was repeated with differing training / holdout proportions in order to optimize the model definition in the training sample while reducing the variance in the holdout sample. The selected partition ratio is presented in Chapter IV.

Hypothesis testing. The purpose of the study was to determine the influence of the stall warning system behavior in cueing aircrew to perform correctly during hazardous airborne icing encounters, as evaluated by the following hypotheses:

1. $H_0 1$: There is no significant difference in the crew performance outcome between a valid system stall warning (HIT) and a stall warning Miss.
2. $H_0 2$: There is no significant difference in the crew performance outcome between a valid system stall warning (HIT) and a stall warning False Alarm.

Although these hypotheses used 2-sided tests for significance, the BLR methodology allows the direction of the relationship to be established. In each case, the significance threshold for rejecting the null hypothesis was established as a Wald statistic below .05 ($p < .05$).

Qualitative data. The qualitative analysis resulted from the manual examination of the databases while classifying the action of the stall warning system in SDT terms

(Hit, Miss, FA, or CR) and the *Correct_Response* outcomes, as defined in Chapter 1. The purpose of the qualitative evaluation was primarily to perform these classification functions and to note pertinent data for the BLR. No inferential statistical analyses were performed on the qualitative data.

Summary

Two hypotheses derived from the Conservative Icing Response Bias theory were evaluated using an archival study of NASA ASRS icing incident data and NTSB accident data pertaining to non-amateur built turbine-powered aircraft. A Binary Logistic Regression non-linear multivariate technique was used to test the CIRB predictions concerning the stall warning system's impact on the crew performance outcomes. In particular, the combination of Bayes' Theorem and SDT applied to the crew stall monitoring vigilance task suggested that stall warning Misses would have a greater impact in icing conditions, while the impact of False Alarms would be reduced, both when evaluated against the baseline Hit condition. These predictions formed the basis of the two hypotheses that were tested using the BLR.

Two SMEs, including the author, scrubbed, merged, and encoded the ASRS and NTSB archives with the following information: aircraft icing state; wing or tail stall state; the action of the stall warning system, icing related system issues, and the reaction of the crew to the potential stall or loss of control situation. Aircrew responses were categorized as either correct or incorrect, using a structured process, and this classification was used as the dependent variable for the basic, comprehensive, and validation BLR models. The independent variables for the basic CIRB model were the

Signal Detection Theory classifications (Hit, Miss, False Alarm, and Correct Rejection) of the stall warning system performance, as manually derived from each data record. The comprehensive model included additional IVs including the wing and tail stall state and the *system issue* variable. Key elements of the investigation are illustrated in Figure 8.

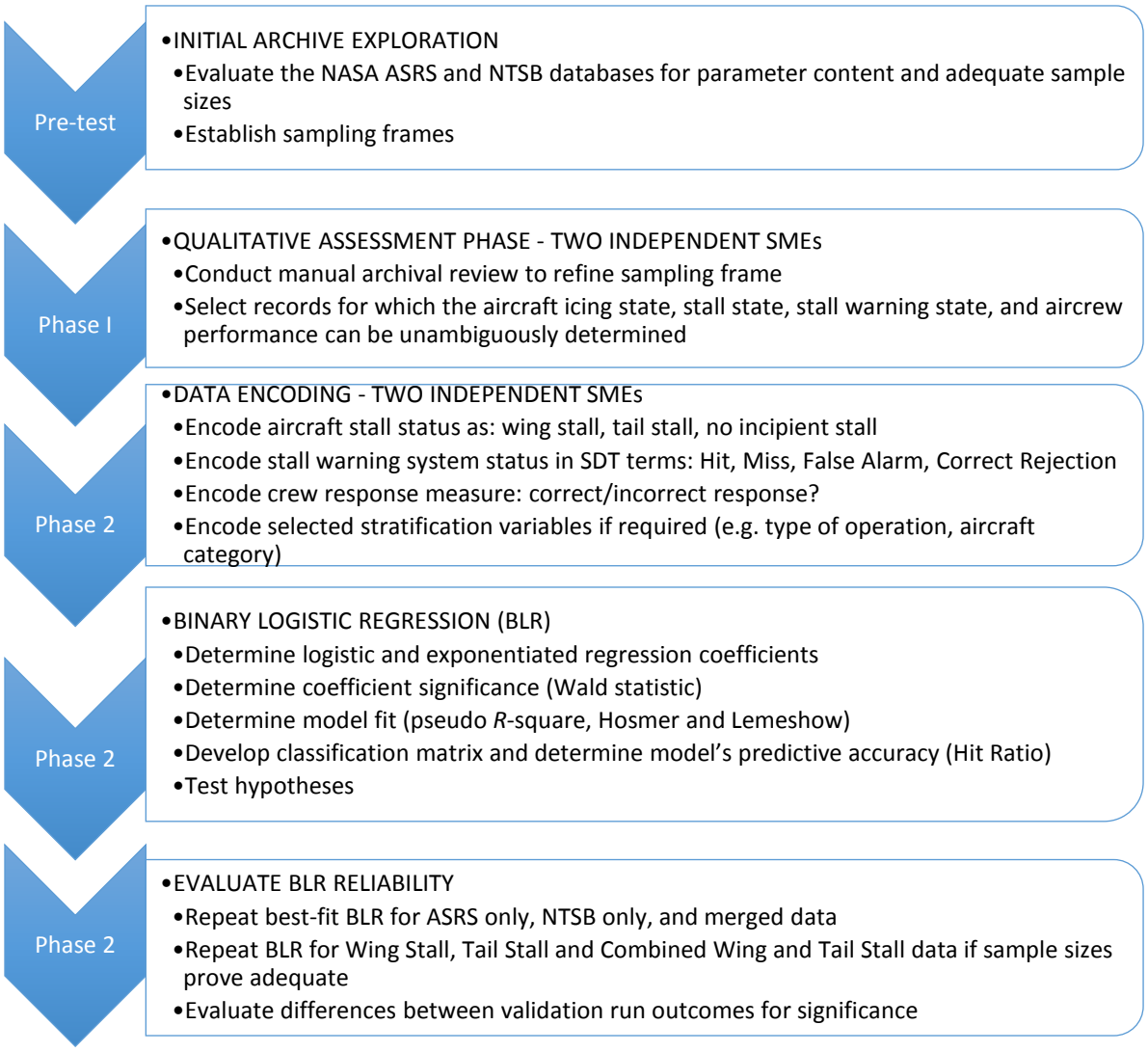

Figure 8. Research activity flowchart.

CHAPTER IV

RESULTS

The results of the analysis are presented incrementally, beginning with generalized descriptive statistics and concluding with the hypothesis test results. The first section summarizes the descriptive statistics for the ASRS, NTSB, and the combined samples. This is followed by summary statistics for key parameters, such as: correct crew response ratios; wing stall vs. tail stall vs. system issues; and the Signal Detection Theory (SDT) stall warning classifications. The next section addresses the Binary Logistic Regression (BLR) test outcomes and BLR model validity tests. The chapter concludes with the outcomes of the two hypothesis tests that were postulated in Chapter I.

Sample Descriptive Statistics

Table 14 summarizes the ASRS and NTSB sampling outcomes based on the criteria expounded in Chapter III. The table reflects the 2-phase scrubbing that was used to identify candidate icing events. Records that lacked the required parameters for the BLR were excluded listwise, with no imputation for missing values. Six duplicate NTSB cases were identified and merged. In addition, three ambiguous cases were rejected in the final down-sampling because of differing interpretations between the author and the external Subject Matter Expert (SME) relating to the crew performance outcome or Signal Detection Theory (SDT) classifications.

Table 14

ASRS and NTSB Database Sample Summary

Sample Group	ASRS Database	NTSB Database	Combined Databases
Unfiltered database size at October 2, 2015 cutoff	18,214	77,544	95,758 records
First-scrub sample size, per tailored icing-event extracts[a]	381	3,039[b]	3,420 records
Second-scrub sample size[c]	126	108[d]	234 records
Final sample for BLR (without imputation)[d]	79	53	132 records

Note. [a]The first-scrub sample sets contained all eligible icing-related incidents or accidents; the numbers are approximate because of the iterative nature of the down-sampling process. [b]Filtered for non-amateur built airplanes. [c]The second-scrub samples excluded records with unknown crew performance outcomes; the numbers are approximate because of the iterative nature of the down-sampling process. [d]Net of six duplicate Visual Meteorological Conditions (VMC) / Instrument Meteorological Conditions (IMC) cases. [d]The final sample excluded cases with indeterminate crew or SDT outcomes.

Figure 9 illustrates the distribution of the final subset of ASRS and NTSB icing events by year, based on the event date. The figure approximates the relative incidence of such events, but it should not be used to evaluate the absolute occurrence rates because the table excludes legitimate records that were rejected from the study due to incomplete data. In addition, the figure has not been normalized to account for annual flight hour exposure, cyclical weather variations, or other confounding factors.

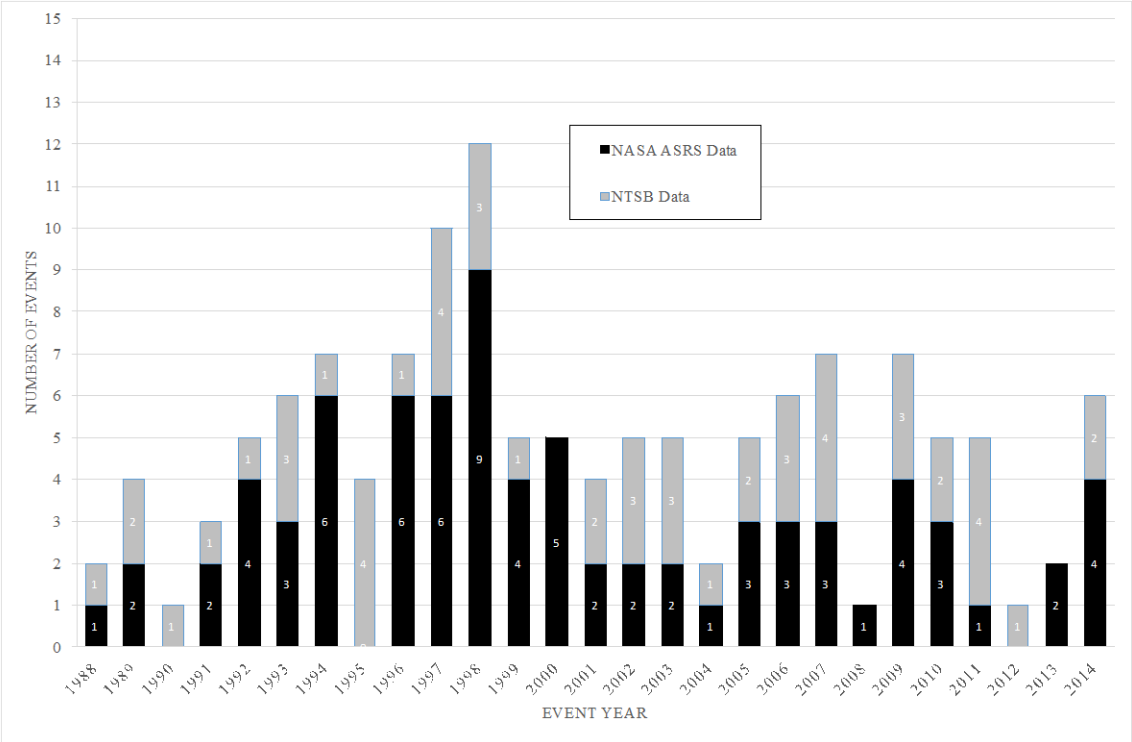

Figure 9. Annual incidence of icing events in study, by event year.

Icing Event Summary Statistics

Table 15 summarizes key icing event statistics for each of the databases independently and for the final combined ASRS / NTSB data sample of 132 records. As expected, wing stalls accounted for most of the events, but tail stalls and system issues collectively accounted for more than 37% of the sample, which required the inclusion of these categories in the subsequent BLR analysis. Each of the event categories (wing stall, tail stall, and system issue) exceeded the minimum desired count of 10 observations for the combined dataset, but the NTSB data failed to achieve the minimum in three of the four categories, and the ASRS dataset had only 14 tail stall records. As will be shown,

these small sample counts for the identified IV categories had consequences on the BLR execution. Note that the totals in Table 15 exceed 100% because of some overlap between the icing event categories, such as a simultaneous occurrence of a stall warning with a system issue.

Table 15

Icing Event Classification Frequencies

Icing Event Class	ASRS	NTSB	Combined
Wing Stall	41 (51.9%)	47 (88.7%)	88 (66.7%)
Tail Stall	14 (17.7%)	5 (9.4%)	19 (14.4%)
Stall Warning	30 (38.0%)	6 (11.3%)	36 (27.3%)
System Issue	23 (29.1%)	7 (13.2%)	30 (22.7%)
Total	108[a]	65[a]	173[a]

Note. [a]Column totals exceed 100% because some of the icing event categories, such as stall warning and system issues, can occur simultaneously and therefore overlap in the statistics.

Tail Stall Identification

The SMEs identified tail stall events conservatively during the data-encoding process. Tail stall events were never inferred; records were only encoded as tail stalls when the NTSB or ASRS narratives specifically alluded to a tail stall event, as in the following example. The abbreviations and text contractions are as they appear in the source record (emphasis added):

> We were noticing very light rime ice, but it was not accumulating on the wings... Upon initiating the clb, the acft pitched down, buffeting. The PIC

> attempted to regain pitch ctl and the stall horn went off. At this point, the PIC pushed up the pwr, and I contacted ctr to ask for the nearest arpt and advise that we had a 'vibration' and were experiencing difficulty… Looking back, I believe that, *although the wings were relatively free from ice, there may have been an accumulation on the tail*. The acft was equipped with pneumatic boots, but there was never enough ice present (that I observed, and I was watching) to cycle the boots effectively. *Somehow, the flow around the empennage was disturbed,* and the pitch up to initiate the clb to 13000 ft from 11000 ft *must have stalled the horiz stabilizer*. There was no perceptible trim change to alert us to this condition, and the subsequent events were very rapid. (NASA 2014, extract ACN #265218)

Stall Warning System SDT Performance

As discussed in Chapter III, the CIRB model is comprised of two interacting SDT systems that each produce a binary output based on incomplete input data. The first system contains the aircraft's stall warning system, which performs its function subject to limited data available from the aircraft's sensors and air data systems. Table 16 summarizes the SDT outcome measures of the stall warning system's performance, as determined from the final database sample. Correct stall warning behavior, represented by the Hit and Correct Rejection entries in the table, totaled 21.2% of the sample. The remaining 78.8% represent the undesirable SDT Miss and FA outcomes. These stall warning system SDT outcomes were hypothesized to be critical inputs to the second

CIRB SDT system: the crew decision-making element, which is discussed in the next section.

Table 16

Stall Warning SDT Classification Frequencies

SDT Classification	ASRS	NTSB	Combined
Hit	12 (15.2%)	5 (9.4%)	17 (12.9%)
Correct Rejection	8 (10.1%)	3 (5.7%)	11 (8.3%)
False Alarm	17 (21.5%)	1 (1.9%)	18 (13.6%)
Miss	42 (53.2%)	44 (83.0%)	86 (65.2%)
Total	79 (100%)	53 (100%)	132 (100%)

Aircrew Performance Outcomes

As described in Chapter III, aircrew must process imperfect stall warning, environmental, and aircraft cues to produce an aircrew performance outcome. The stall warning SDT outputs shown in Table 16 were hypothesized to have a significant effect on these aircrew performance outcomes. For the combined sample, correct crew responses were observed in 35 of the 132 cases (26.5%), and incorrect responses were recorded in the remaining 97 cases (73.5%), so the aircrew performance outcomes clearly exhibited some of the Hit and Miss traits that characterize SDT systems, as discussed in Chapter II. The BLR was used to evaluate two hypotheses that linked the crew decision-making outcomes to the SDT performance of the stall warning system.

Binary Logistic Regression

As presented in Chapter III, the BLR technique was used to evaluate three CIRB models: a basic CIRB model, a comprehensive model, and a validation model. These models were compared to a baseline stall warning model that contained the stall warning actuation as the sole IV. The binary aircrew response outcome was adopted as the sole dependent variable for the three CIRB models and the baseline stall warning model. The results are presented in order of increasing model complexity, beginning with the baseline stall warning model and proceeding through the basic CIRB model to the comprehensive CIRB model.

Baseline stall warning BLR. The stall warning BLR was used as the baseline for evaluating the three subsequent CIRB models. Aside from the model constant, the sole IV was the stall warning actuation (Y / N). Results from the baseline stall warning model are shown in Table 17, which forms the basis for the subsequent comparison with the CIRB models.

Table 17

Stall Warning Baseline Logistic Regression Outcomes

	Incorrect Crew Response Predicted	Correct Crew Response Predicted	Model Classification Accuracy
Incorrect Crew Response Observed	82	15	84.5% Specificity
Correct Crew Response Observed	14	21	60.0% Sensitivity
Model Predictive Values	85.4%	58.3%	78.0% Overall

CIRB models. The basic and comprehensive CIRB models were run independently for each of the ASRS and NTSB sample sets, with the objective of comparing the results across the two databases as described in Chapter III. Unfortunately, the individual databases did not meet the minimum BLR sample size requirements of 10 observations per category, as indicated by the single-digit frequencies in Tables 15 and 16, particularly for the NTSB data. This precluded the BLR from converging to a solution for either database in isolation, even after 20 iterations using the standard .50 cutoff value. In contrast, the BLR converged in only six iterations with the combined dataset, with a parameter change of less than .001 at the last cycle. The remainder of the analysis was therefore confined to the combined ASRS / NTSB dataset of 132 samples, using the conventional test and holdout methodology to establish the reliability and validity of the merged data.

Basic CIRB model. The basic CIRB model was of the form shown in equation 15:

$$Logit_{ccr} = \ln\left(\frac{prob_{ccr}}{1 - prob_{ccr}}\right) = b_0 + b_1 CR + b_2 FA + b_3 Miss + error \qquad (15)$$

where:

CCR = Correct Crew Response outcome

Four variants of the BLR were examined to achieve the best model fit: the default method, in which all BLR variables were retained as the model iterated, and three alternatives (stepwise, forward, and backward), which used different criteria for selectively excluding IVs as the model iterated. All four methods produced identical results for the basic CIRB model. The BLR was statistically significant in every case: ($\chi 2(3) = 85.328$ $p < .0005$). The model explained 69.5% (Nagelkerke R^2) of the variance in crew performance outcomes and correctly classified 90.9% of the cases. Only 12 of the 132 cases were misclassified by the model (three false positives and nine false negatives). The Hosmer and Lemeshow test was not statistically significant for the basic CIRB model ($\chi 2(2) = 0.000$ $p = 1.000$), indicating that the model was a satisfactory fit. Other parameters for the basic CIRB model are shown in the crew response classification matrix (Table 18), and the corresponding BLR outcomes are shown in Table 19.

Table 18

Basic CIRB Model Crew Response Classification Matrix

	Incorrect Crew Response Predicted	Correct Crew Response Predicted	Model Classification Accuracy
Incorrect Crew Response Observed	94[b]	3	96.9% Specificity
Correct Crew Response Observed	9	26[b]	74.3% Sensitivity
Model Predictive Values	91.3%[c]	89.7%[d]	90.9%[b] Overall

Note. [a]$n = 132$, cut value = .50. [b]Lead-diagonal elements of matrix represent correct model predictions. [c]Negative Predictive Value. [d]Positive Predictive Value.

Table 19

Basic CIRB Model Outcomes

SDT Class	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for Exp(B) Lower	Upper
Hit (Reference)			41.821	3	.000			
Correct Rejection	2.909	1.165	6.232	1	.013	18.333	1.868	179.895
False Alarm	2.686	.906	8.795	1	.003	14.667	2.486	86.529
Miss	-2.714	.777	12.217	1	.000	.066	.014	.304
Constant	-.606	.508	1.426	1	.232	.545		

Comprehensive model. The comprehensive CIRB model was of the form shown in Equation 16:

$$Logit_{ccr} = b_0 + b_1 CR + b_2 FA + b_3 Miss + b_4 Wing_{Stall} + b_5 Tail_{Stall} + b_6 System_{Issue} + error \quad (16)$$

where:

CCR = Correct Crew Response outcome

For the comprehensive CIRB model, the four BLR iteration alternatives produced slightly different results, with the all-variables-included method producing a slightly better fit, as shown in Table 20. Accordingly, the comprehensive BLR model analyses, and the results that follow, are based on the all-in regression technique.

Table 20

Comprehensive Model: BLR Method Comparison

Model	Misclassification Rate	Squared Error
BLR_All_Inputs (Default)[a]	0.083333	0.067738
BLR_Forward	0.083333	0.068831
BLR_Stepwise	0.090909	0.071701
BLR_Backward	0.090909	0.071701

Note. [a]The selected model for the BLR, based on lowest misclassification rate and least squared error criteria.

Comprehensive Model Outliers. Six multivariate outliers were noted using a two-standard deviation cutoff. These outliers are listed in Table 21. As the table shows, the outliers reflected diverse and random permutations of the independent and dependent variables, so they were deemed unlikely to skew the analysis. Further, their elimination reduced some of the sample counts, such as the tail stall category, below the minimum group sizes required for the BLR execution. Accordingly, the outliers were retained in the analysis.

Table 21

Comprehensive CIRB Model Outliers

Case	Source	Source ID	Wing Stall?	Tail Stall?	Stall Warn?	Correct Response[a]	System Issue?	SDT Class	ZResid.[b]
24	NTSB	CHI97FA047	Y	N	N	Y	N	M	6.464
48	NTSB	NYC05MA083	N	N	N	N	Y	C	-4.533
50	NTSB	NYC97FA045	N	N	Y	N	Y	F	-2.791
82	ASRS	389483	N	Y	N	Y	N	M	3.098
96	ASRS	456868	N	Y	N	Y	N	M	3.098
101	ASRS	495957	N	N	Y	N	Y	F	-2.791

Note. [a]All of the cases in the correct response column were misclassified by the comprehensive BLR model. [b]Outlier threshold: 2 standard deviations.

Comprehensive model statistics. The comprehensive model was statistically significant ($\chi^2(6) = 88.911$ $p < .0005$). The model explained 71.5% (Nagelkerke R^2) of the variance in crew performance outcomes and correctly classified 91.7% of the cases. The Hosmer and Lemeshow test was not statistically significant ($\chi^2(5) = 1.279$ $p = .937$), indicating that the model was a satisfactory fit. Other parameters for the comprehensive CIRB model are shown in the crew response classification matrix (Table 22), the outcome matrix (Table 23) and the correlation matrix (Table 24). Table 23 highlights an important outcome: The SDT Hit and Miss parameters were the only significant IVs in the comprehensive SDT model. No other parameter (i.e., Wing Stall, Tail Stall, and System Issues) approached statistical significance. Also of note, the incidence of any of these three issues resulted in a decreased chance of a correct crew performance outcome, as shown by the negative sign of the B coefficients in Table 23.

Table 22

Comprehensive CIRB Model Crew Response Classification Matrix

	Incorrect Crew Response Predicted	Correct Crew Response Predicted	Model Classification Accuracy
Incorrect Crew Response Observed	93[b]	4	95.9% Specificity
Correct Crew Response Observed	7	28[b]	80.0% Sensitivity
Model Predictive Values	93.0%[c]	87.5%[d]	91.7%[b] Overall

Note. [a]*n* = 132, cut value = .50. [b]Lead-diagonal elements of matrix represent correct model predictions. [c]Negative Predictive Value. [d]Positive Predictive Value.

Table 23

Comprehensive CIRB Model Outcomes

							95% C.I. for Exp(B)	
SDT Class	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Hit (Reference)			15.938	3	**.001**			
Correct Rejection	2.339	1.275	3.364	1	.067	10.375	.852	126.367
False Alarm	1.449	1.505	.927	1	.336	4.258	.223	81.330
Miss	-2.887	.911	10.040	1	**.002**	**.056**	.009	.332
Wing Stall	-1.802	1.351	1.778	1	.182	.165	.012	2.332
Tail Stall	-.357	1.649	.047	1	.828	.699	.028	17.717
System Issue	-.358	1.057	.114	1	.735	.699	.088	5.555
Constant	.953	1.380	.478	1	.490	2.594		

Table 24

Comprehensive CIRB Model Correlation Matrix

	Constant	SDT CR	SDT FA	SDT Miss	Wing Stall	Tail Stall	System Issue
Constant	1.000	-.288	-.643	-.144	-.906	-.749	-.426
SDT_CR	-.288	1.000	.391	.174	.186	.136	-.197
SDT_FA	-.643	.391	1.000	.038	**.639**	**.504**	-.252
SDT_Miss	-.144	.174	.038	1.000	-.102	-.337	.147
Wing_Stall	-.906	.186	**.639**	-.102	1.000	.784	.298
Tail_Stall	-.749	.136	**.504**	-.337	.784	1.000	.285
System_Issue	-.426	-.197	-.252	.147	.298	.285	1.000

CIRB Model Comparison

Table 25 contrasts the basic and comprehensive CIRB models with the ASRS pre-test outcomes and the baseline stall warning model. Both CIRB models were superior to the pre-test model that was developed using a sub-sample of NASA ASRS data. This is attributed to the larger overall sample size of the combined ASRS and NTSB databases that reduced the model variance.

Table 25

CIRB Model and Pre-Test Comparisons

	Nagelkerke R^2	Correctly Classified	Sensitivity %	Specificity %	PPV[a] %	NPV[b] %
ASRS pre-test results	.336	75.7	54.0	92.3	15.6	72.3
Stall warning baseline model	.243	78.0	60.0	84.5	58.3	85.4
CIRB basic model	.695	90.9	74.3	96.9	89.7	91.3
CIRB comprehensive model	.715	91.7	80.0	95.9	87.5	93.0

Note. [a]Positive Predictive Value. [b]Negative Predictive Value.

The sensitivities of the basic and comprehensive CIRB models (74.3% and 80.0%, respectively) were also markedly better than the baseline stall warning model (60.0% sensitivity). Overall, the CIRB basic and comprehensive models correctly classified 90.9% and 91.7% of the sample cases, respectively, compared to 78% for the baseline stall warning model. Although the comprehensive model was somewhat more sensitive than the basic model, the latter performed almost as well in almost every evaluated parameter, including model fit, Nagelkerke $R^{2,}$ cases correctly classified, and specificity. This is an important finding because the basic CIRB model outcomes indicate that a simple SDT analysis of stall warning system performance explains 69.5% of the variance in aircrew decision-making outcomes, while correctly classifying 90.9% of the cases in the final data sample. These findings must now be examined in the context of the BLR reliability assessments.

BLR Reliability

The reliability and validity of the BLR analysis were evaluated with several tests. As a pre-requisite, the combined NTSB and ASRS database sample sizes were deemed adequate for the successful application of the BLR method, as evinced by the rapid convergence to a unique solution. The Hosmer and Lemeshow test was not statistically significant for the basic CIRB model ($\chi^2(2) = 0.000$; $p = 1.000$) and for the comprehensive model ($\chi^2(5) = 1.279$; $p = .937$), indicating that the fit of both models was satisfactory. The close results achieved using the four differing BLR approaches with both models also attested to their robustness and reliability.

The final test of model reliability entailed the partitioning of the combined data sample into training and holdout (validation) sub-samples. The BLR was run several times while varying the relative proportion of these two partitions. The objective was to attain the best balance between the model specification (which required a large training sample), while minimizing the variance in the validation dataset (which required a large holdout sample). An even split ($n = 66/66$) between the two categories was eventually selected that yielded a good balance between these competing influences. Table 26 compares the training and holdout (validation) misclassification rates and average square errors.

Table 26

Training Vs. Holdout Samples

	Training Partition[a]	Validation Partition[a]	Difference (Percent)
Misclassification Rate	0.075758	0.075758	0
Average Squared Error	0.059682	0.072321	21.2

Note. [a]Basis: $n = 66$ for both the training and holdout samples.

Figure 10 contrasts the SDT Receiver Operating Characteristic (ROC) curves obtained for the training and holdout samples. The similar ROC curves confirm the training / holdout reliability depicted in Table 26. The ROC curves also approached the ideal forms for maximum sensitivity and specificity, which requires them to be asymptotic to both axes, and convex towards the upper left quadrant. Collectively, these outcomes indicate that the results of the BLR methodology should be reliable and valid when applied to the target U.S. registered, turbine, non-amateur-built aircraft population addressed in this study.

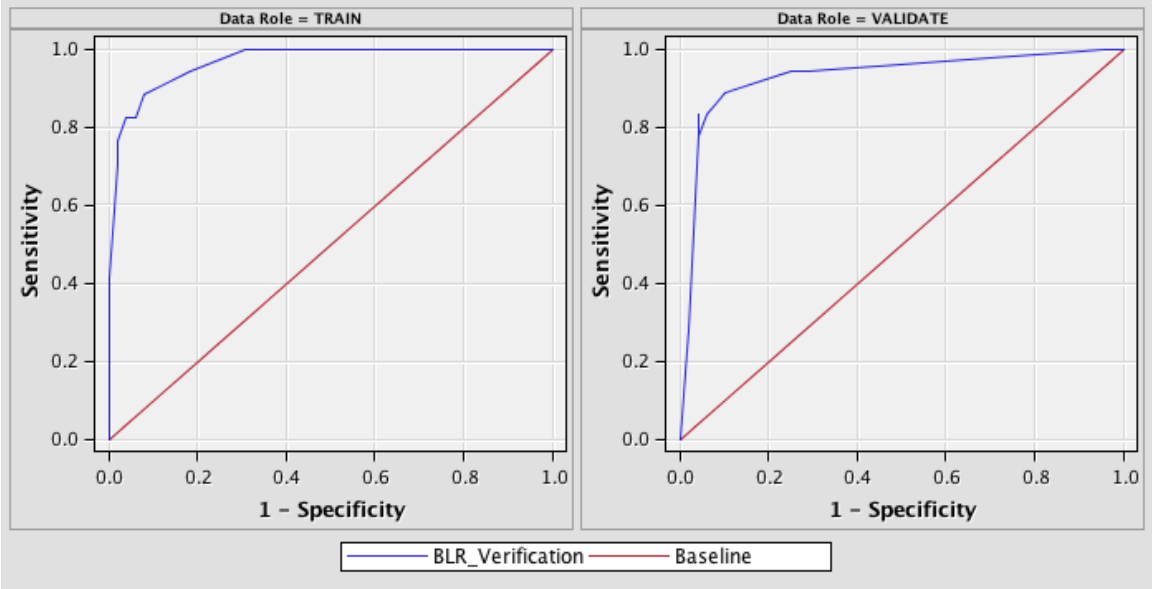

Figure 10. Baseline BLR model crew response receiver operating characteristics. Training and validation sample comparison.

Hypothesis Tests

Table 27 contains key statistics for the Hit, CR, Miss, and FA SDT predictor variables for the basic and comprehensive CIRB models. The SDT Hit and Miss variables and coefficients were significant and similar for both models, but the significance of the FA predictor was notably different between them. This is due to the strong negative correlation between FA and wing stalls (-.639) and between FA and tail stalls (-.504), as shown in Table 24.

Table 27

CIRB Model Comparison

SDT Class	B	S.E.	Wald	df	Sig.	Exp(B)
Basic CIRB Model						
Hit (Reference)			41.821	3	.000	
Correct Rejection	2.909	1.165	6.232	1	.013	18.333
False Alarm	2.686	.906	8.795	1	**.003**	**14.667**
Miss	-2.714	.777	12.217	1	**.000**	**.066**
Comprehensive Model						
Hit (Reference)			15.938	3	.001	
Correct Rejection	2.339	1.275	3.364	1	.067	10.375
False Alarm	1.449	1.505	.927	1	**.336**	**4.258**
Miss	-2.887	.911	10.040	1	**.002**	**.056**

Based on the comprehensive CIRB model exponential coefficients in Table 27, aircrew had approximately 4.3 times greater odds of performing correctly when faced with a stall warning FA than with a stall warning Hit in icing conditions. The difference in crew performance outcomes between stall warning Misses and Hits was statistically significant ($p < .0002$). Conversely, a stall warning Miss was 17.9 times (1 / 0.056) more likely to result in an incorrect response than a Hit. Although the Miss IV was not significant in the comprehensive model ($p < .336$), the statistic was significant ($p < .003$) in the basic model, which was based solely on the stall warning SDT categories. Combining the Hit, Miss, and FA ratios, a Miss was 76.0 times more likely to lead to an incorrect outcome than an FA with the comprehensive CIRB model. The Miss: FA ratio

for the basic CIRB model was even greater (222.2:1). In the absence of the CIRB model, these probabilities would be expected to be approximately equal. Accordingly:

H01: There is no significant difference in the crew performance outcome between a valid system stall warning (HIT) and a stall warning Miss, therefore the hypothesis is rejected.

H02: There is no significant difference in the crew performance outcome between a valid system stall warning (HIT) and a stall warning False Alarm, therefore the hypothesis is rejected.

CHAPTER V

DISCUSSION, CONCLUSIONS, & RECOMMENDATIONS

Discussion

This research was predicated on the assumption that a previously unrecognized link exists between the action of an aircraft's stall warning system and the successful or unsuccessful negotiation of an airborne icing encounter by the aircrew. Signal Detection Theory (SDT) concepts were used to define two interacting SDT decision-making systems under these circumstances: the aircraft's stall warning system, operating with incomplete sensor information, and the aircrew, also operating with incomplete and conflicting stall cues, including those from the stall warning system. It was postulated that the stall warning system performance, in terms of Hits (i.e., valid warnings), Misses (required warnings that the system didn't issue), False Alarms (FA), and Correctly Rejected (CR) warnings could influence the aircrew SDT system in a unidirectional relationship. The application of Bayes' Theorem to these interacting SDT models led to a predicted shift of the aircrew's stall detection decision criterion in icing conditions that has been termed the Conservative Icing Response Bias (CIRB) model in this study. The CIRB led to testable relationships between stall warning Misses, FAs, and crew performance outcomes. This relationship was evaluated by the application of a Binary Logistic Regression (BLR) technique to an archival analysis of NASA Aviation Safety Reporting System (ASRS) incident data and National Transportation Safety Board (NTSB) accident data. The discussion begins with an assessment of the summary

statistics obtained from the archival analysis and concludes with a discussion of the implications of the CIRB hypothesis test outcomes.

Wing Stalls, Tail Stalls, and System Issues

Descriptive statistics were used to assess the proportion of different icing impacts (wing / tail / system), the relative incidence of correct and incorrect crew response outcomes, and the performance of the stall warning system in SDT terms (Hit, Miss, FA, or CR). A total of 132 cases met the criteria for the application of the BLR technique. Of these, the majority (66.7%) related to wing stalls, with tails stalls accounting for 14.4% of the cases. System issues, pertaining to the loss of air data capability due to the freezing of the pitot-static systems or AoA probes, figured relatively prominently (22.7%) in the final sample of 132 cases. The inclusion of the System Issue category was not initially envisaged, but it was added as a new independent variable (IV) for the comprehensive BLR analysis based on its prominence in the pilot study, coupled with the severe consequences observed for this type of failure in the archives.

The relative incidence of tail stalls and system issues may have important repercussions for flight operations, system design, and airworthiness certification requirements. For example, despite an observed 14.4% incidence, none of the aircraft appearing in the archives were equipped with any form of tail stall detection or prevention system because current airworthiness certification requirements do not require such a system. The qualitative review of the NTSB and Archives indicated that tail stalls generally caught aircrew completely by surprise, with the expected undesirable outcomes.

Similarly, icing-induced system failures such as frozen pitot-static ports and angle-of-attack vanes represented 22.7% of the sampled events. These failures sometimes severely degraded the primary flight instrument systems and compromised important flight envelope protections and stall warning functionality. The archival narratives even contained some instances of system issues giving rise to simultaneous stall and over-speed warnings, and this combination of overwhelming and erroneous cues clearly exacerbated the potential CIRB effect. Unfortunately, contemporary stall warning systems are often rendered inoperable when these types of failures arise, disabling the stall protections when they are most needed. Similarly, the literature has shown that Angle-of-Attack (AoA)-based stall warning systems are incapable of differentiating between wing stalls, tail stalls, and system issues, leaving the crew with an extremely challenging analysis task under very difficult conditions. A simple, direct, indication of the aircraft's wing and tail stall margins would go a long way to mitigating the limitations of AoA-based stall warning systems. An ideal system would also operate independently from the aircraft's highly integrated air data and AoA systems that have proven vulnerable to failure during severe icing encounters. These stall warning system shortcomings were clearly revealed in the SDT outcomes derived during the study, as discussed next.

Stall Warning System Effectiveness

One of the two basic assumptions of the CIRB model was that aircraft stall warning systems operate with incomplete sensor information, particularly in an airborne icing context, and therefore perform imperfectly. In SDT terms, such systems can fail to

function when they should operate (a Miss), or they can activate when they should not (a False Alarm). The statistics confirmed this effect in a very significant manner for Missed stall warnings. As noted in Table 23, the only IVs that achieved statistical significance in the comprehensive CIRB model were the SDT Hit or Miss stall warning parameters. No other parameter (Wing Stall, Tail Stall, or System Issues) approached statistical significance. Counterintuitively, the actual wing-stall or tail-stall state did *not* approach significance in relation to aircrew performance outcomes. These findings strongly support the interacting crew / stall warning SDT system basis of the CIRB model, which explains 70% of the observed variability in crew response outcomes to airborne icing encounters. As predicted by CIRB, these outcomes were strongly and negatively biased by poor stall warning system performance (Misses and FAs), whether the aircraft was in a stalled condition or not.

In terms of overall stall warning system performance, correct operation (i.e., Hits and CRs during non-stall conditions) accounted for only 21.2% of the final 132 cases. False Alarms accounted for 13.6% of the cases, and Misses accounted for the majority (65.2%) of the stall warning system SDT outcomes. These statistics reinforce the shortcomings of conventional stall warning systems that cannot directly respond to the aerodynamic degradations caused by icing or monitor for tail stalls, as discussed at length in Chapter II. The repercussions of these findings are two-fold. First, the basic assumptions of the CIRB model are validated. Second, stall warning systems need to be developed with better Miss: FA ratios for both wing and tail stalls. The literature shows that aerodynamic performance monitors, which direct measure the boundary-layer

separation that is always associated with a stalled condition, could help achieve these objectives.

It could be argued that the apparent failures of current stall warning systems are simply the result of self-selection bias, such that the research sample was necessarily derived from those cases that had already resulted in icing incidents and accidents. Arguably, this process selected those very rare occurrences of poor stall warning system behavior that led to poor outcomes, and which did not otherwise occur in the general population. There are two counterarguments to this line of reasoning: First, and most importantly, efforts to reduce accidents must focus on the unsuccessful outcomes and their causes, so the selection bias is a strength, not a weakness of the study. For all safety endeavors, it is the successful outcomes that form the baseline to which incremental safety improvements must be added by addressing the failures.

The second counterargument is that the literature is replete with aerodynamic explanations for poor stall warning performance, particularly Misses, in icing conditions. It is therefore no more valid to assume that stall warnings perform properly in the absence of an accident than it is to assume the inverse. In other words, it would be equally valid to argue that several successfully negotiated icing encounters (that did not result in entries into the NTSB or ASRS archives) resulted *despite* the erroneous performance of the stall warning systems, not because of their excellent performance.

For these reasons, the conclusion stands that tangible safety benefits would be achieved in icing operations if stall warning system design and certification addressed the CIRB effect. This would be achieved by deploying warning systems that reduce the incidence of stall warning Misses – that led to the poorest outcomes – even at the cost of

some increase in FAs, to which crews proved to be relatively resilient. The data revealed three principal causes of stall warning Misses during icing encounters: incorrect stall warning trigger thresholds, tail stalls, and system issues. Incorrect stall warning trigger thresholds resulted from the reduced critical angle of attack (AoA_{crit}) caused by icing-induced aerodynamic degradation. Contemporary stall warning systems cannot adjust for the highly variable effects of ace accumulations in real-time, despite the stall-margin allowances that are made to accommodate the ice shapes used for certification demonstrations. In consequence, there will continue to be occasional icing encounters that result in airfoil stalls before the activation of the aircraft's stall warning system. Tail stalls also led to Missed stall warnings, because no system currently monitors the empennage to provide tail stall warning or alerting. Tails stalls therefore constitute stall warning Misses, almost by definition, unless a simultaneous wing stall resulted in the activation of the stall warning system when the tail stalled. The final category of stall warning Misses related to icing-induced system failures that were observed to compromise both the aircraft's stall warning and envelope-protection functions. These complex failures represented some of the most challenging stall warning scenarios, as they sometimes resulted in near-simultaneous presentation of stall warning Hits, Misses, and False Alarms to the crew.

Aerodynamic performance measurement (APM) systems address the three causes of stall warning Misses identified above. APM systems directly sample and respond to the degraded aerodynamics associated with airfoil icing, in contrast to current Angle-of-Attack based systems which tend to underestimate the stall threat in icing conditions, as amply supported by the literature review and the findings from this study. APM systems

can be used to monitor the empennage to provide appropriate warnings in the event of impending tail stalls. Finally, APM systems operate completely independently from the air data and AoA systems, which currently provide the stall protection functions. For these reasons, APM-based stall warning systems would be less susceptible to the icing-induced sensor degradations that were observed in this study.

Crew Performance

The descriptive statistics obtained from the analysis of the NASA ASRS and NTSB accident archives confirmed the findings from the literature review concerning the previously unexplained variability of crew responses during icing encounters. For the combined NASA ASRS and NTSB database archive, correct crew responses were observed in only 35 of the 132 cases (26.5%), and incorrect responses were recorded in the remaining 97 cases (73.5%). These statistics could be attributed to self-selection bias causing the final sample to capture only those crews that performed incorrectly, while the majority successfully negotiated their icing encounters.

The counterargument is the same as was raised for the stall warning discussion: Even if the results stemmed from self-selection, these same cases would still need to be addressed to improve the safety record. Further, it could be argued that many cases were excluded from the final sample despite an incorrect performance outcome from the crew because an accident was avoided (and hence went unreported). Several records that might have further supported the CIRB hypothesis were also excluded because one or more parameters of interest could not be explicitly determined using the overwhelming evidence threshold set for this study. These considerations support the conclusion that

aircrew performed imperfectly during icing encounters largely due to the influence of the stall warning system, which is addressed in the following section.

Stall Warning System Influence on Crew Performance Outcomes

As predicted by the CIRB model, the BLR analysis indicated that Missed stall warnings had a significant and adverse influence on the outcome of airborne icing encounters for both the basic and comprehensive CIRB models. More surprisingly, the actual stall state of the wing or tail did not prove statistically significant as a predictor of crew performance outcomes. This implies that crews *did* react appropriately to stalls when they were correctly identified by the stall warning system. Otherwise, statistically significant degradations in correct crew response outcomes should have been observed in the presence of actual stall conditions. Nevertheless, stalls and loss of control (LOC) events were often accompanied by system issues and adverse environmental influences, such as airframe vibration and buffeting, which undoubtedly added an increased noise component to the aircrew's stall-detection task, in SDT terms. If the aircrew did not adjust their decision-making criterion appropriately, the increased SDT noise would further complicate the task of detecting a stall. This SDT noise effect would compound the overly-conservative SDT decision-making criterion bias predicted by the CIRB model, which explains 70% of the variability in the crew performance outcomes. This finding also supports Advani's assertion, first noted in Chapter II, that "Aerodynamic Stall Can Prompt 'Brain Stall'" (2014, p. 58).

An incorrect assertion could be made that False Alarms caused fewer poor outcomes because of their scarcity, but this view is not supported by the data. FAs

accounted for 13.6% of the final data sample, second in prevalence only to Misses. More importantly, the CIRB model predicts that the increased aircrew vulnerability to stall warning Misses is a direct result of the Bayesian and SDT origins of the model. This counterargument is further validated by the almost identical outcomes obtained with the basic and comprehensive CIRB models. The basic model, which incorporated the stall warning system SDT performance as the only IV, correctly classified 90.9% of cases, with only 12 of 132 cases being misclassified by this relatively simple model. The comprehensive SDT model, which added the wing stall, tail stall, and system issue IVs, correctly classified only one additional case. The stall warning SDT was therefore the dominant factor in determining the aircrew performance outcomes in this study.

Research Design Lessons Learned

The BLR technique was applied successfully to achieve the objectives of this study, but several lessons-learned arose from the application of the method. As anticipated, the BLR proved resilient to violations of traditional statistical requirements related to normality and heteroscedasticity, but the tradeoff for these benefits was the BLR's requirement for a larger sample size than other multivariate techniques, such as multiple linear regression. Successful BLR execution was highly dependent on minimum sample size constraints being met for the overall sample, as well as for the number of observations within each variable category grouping. As indicated in Chapter III, the desired sample size for this analysis was 10 cases per estimated parameter or category, but the NTSB data had less than 10 samples each for the Tail_Stall, Stall_Warning, and System_Issue variables. The decision made at the outset of the research design to use a

combination of ASRS data and NTSB archival data proved fortuitous. The BLR would not run successfully with either archive in isolation due to one or more violations of the minimum sample-size requirements.

Although the BLR executed properly using the combined ASRS / NTSB database, the final sample of 132 cases was still near the lower acceptable group sample size bounds for the successful execution of the BLR. This was evinced when the BLR analysis was attempted with six identified outliers removed from the data sample: The BLR failed to converge to a solution because the lower sample size limits had been violated. This sensitivity of the BLR to the overall sample size and the sample sizes within each independent variable group should be considered carefully if the technique is to be reattempted because it is unlikely that the BLR would run successfully with a sample any smaller than was used for this study. Similarly, the addition of more independent variables would significantly raise the minimum sample size that would also likely preclude the successful execution of the BLR. Future researchers contemplating the application of this method should therefore carefully consider the tradeoff between minimum required sample size and the number of variables during the early stages of their experimental design because other analytical methods might prove more suitable than the BLR if these constraints cannot be met.

A second research-design lesson-learned relates to the importance of making a very clear distinction between *crew performance* outcomes and adverse or satisfactory *event* outcomes. Extensive efforts were made to isolate the crew responses from the event outcomes when viewed in an icing / stall context. This is because False Alarms are inherently associated with nonthreatening (no stall) situations, while Misses correspond

to threatening stall or LOC conditions. An incorrect response to an FA would therefore likely correspond to a benign outcome, despite a crew error, whereas a correct crew response to a Miss might still result in an accident. If proper account were not taken of this phenomenon, the data would simply correlate with the potential seriousness of a stall in icing, rather than the intended SDT independent variables. Efforts to replicate this study should therefore take similar precautions to properly isolate the SDT phenomenon of interest to avoid seriously confounding the analysis.

A third lesson-learned related to the relative usefulness of the ASRS and NTSB databases. It was anticipated that the NTSB accident archives would yield a higher percentage of usable records than the ASRS incident data because of the presumed availability of flight data recorders (FDR) and cockpit voice recorders (CVR) in the tailored sample of turbine aircraft that normally carry this equipment. This did not turn out to be the case. It proved difficult to determine the exact crew responses and stall warning system status from the NTSB accident data because most of the aircraft in the sample frame lacked an FDR. Conversely, many of the ASRS records contained detailed and useful pilot narratives, often with explicit declarations concerning the items of interest. This unexpected windfall proved the value, once again, of using two different archival sources, as planned from the inception of the research design.

The final lesson-learned related to the benefits of building on prior research, particularly with the regard to the generation of appropriate search strings. For example, the use of Green's (2006) search string, and some of Green's associated methodology, significantly streamlined the processing and down-sampling of the massive ASRS data

archive. As an added benefit, this standardization should facilitate future research synthesis and meta-analyses related to the topic.

Conclusions

This study applied the Binary Logistic Regression technique to a hypothesized Conservative Icing Response Bias model of aircrew performance outcomes during airborne icing encounters. The evaluation of the CIRB hypothesis resulted in a convergent BLR solution with adequate combined sample sizes and positive analytical measures of reliability and validity. Accordingly, the following conclusions should be generalizable to the target population of U.S. registered, non-amateur-built, turbine aircraft, as intended.

The research demonstrated a significant adverse effect of Missed stall warnings (that accounted for 65.2% of the events studied) on aircrew performance outcomes during airborne icing encounters. Conversely, aircrew proved far less susceptible to FAs, as predicted by the CIRB hypothesis. The fundamental CIRB assumptions concerning two interacting Signal Detection Theory Systems were therefore validated.

CIRB provides a much-needed theoretical model that explains the apparently heterogeneous aircrew reactions to icing encounters. The model uses only four SDT parameters (Hit, Miss, FA, and CR) related to the aircraft's stall warning system's performance to predict the aircrew performance outcomes during airborne icing encounters. CIRB produces quantitative predictions, so it can be formally applied and tested with any dataset that meets the sample size requirements for a BLR analysis. By extension, these results highlight the potential application of SDT to a much broader

context than airborne icing encounters. Although the current study was predicated on the SDT modelling of crew response outcomes, these binary outcomes are manifestations of a higher-level Situational Awareness construct, which could be termed *stall awareness*. This extension of SDT methodology beyond simple perceptual tasks associated with physical stimuli into higher-level abstract or metaphorical signals was mentioned in the introduction (Abdi, 2009, p. 2). This potential was realized in the findings from the current study, where the crew decision-making outcomes were the result of complex interactions between physical stimuli, the environment, training, workload, aerodynamics, and numerous other factors. The application of the combined SDT / BLR method reduced these numerous and sometimes unknown or unquantifiable factors to the four basic binary elements of SDT: Hits, Misses, FAs, and CRs. Despite this significant simplification, the resulting model yielded impressive sensitivity, specificity, and correct classification statistics. If the methods of this study can be successfully replicated, then the reductionism achieved by combining the SDT method and the BLR technique has important ramifications for the modelling of other complex human-in-the loop processes. Previously intractable problems can be reframed in SDT terms, with the dependent variable representing the output of a high-level construct such as stall awareness. The resulting models could then be quantitatively evaluated for statistical and practical significance using the SDT / BLR technique, as applied in this study. Once successfully modeled, the same methods would allow the quantitative evaluation of the effects of changed parameters on the model output. In the case of the current study, the CIRB SDT / BLR model could be applied to the development and evaluation of aircraft stall warning systems and their associated certification regulations. Ongoing data mining

methods could then be applied to determine if the anticipated benefits of APM systems or revised aircrew training are being realized as expected.

Recommendations

This section addresses three broad topic areas: stall warning system design and certification, aircrew training, and future research directions. The following recommendations are intended to explicitly address the problem statement developed in Chapter I, and repeated below:

> The literature shows that the majority of airborne icing accidents result from aircraft stalls, yet there is no theoretical model that ties the success or failure of the crew decision-making in icing to the performance of the stall warning system. This knowledge gap imposes reactionary and untargeted regulatory responses that have failed to fully resolve the icing issue, despite decades of effort.

Stall warning system design. The CIRB phenomenon has shown that it would be advantageous to modify stall warning system certification regulations (e.g., 14 CFR §23.207 and §25.207 *Stall Warning*) to achieve systems with improved Miss: FA ratios. This could be accomplished in several ways: Stall warning systems should be developed that maintain the correct warning margins in the face of airfoil icing. Warning systems should also monitor for tail stalls, which accounted for 14.4% of the sampled events and which often led to stall warning Misses with unfavorable consequences. Finally, stall warning systems should be developed and certified that continue to function

independently and correctly in the face of icing-induced system issues, such as frozen pitot-static ports and angle-of-attack vanes, which represented 22.7% of the sampled events. The literature has shown that aerodynamic performance monitors have promise in fulfilling all three of these requirements.

Aircrew training. Aircrew training programs for airborne icing operations should explicitly address the response-bias phenomenon and its attendant dangers. The CIRB phenomenon stems from the failure of aircrew to adjust their SDT decision criterion appropriately to the more conservative value required during airborne icing encounters. This response bias is a direct consequence of Bayes' Theorem, which predisposes aircrew to underestimate the possibility of a stall or loss of control during icing encounters, based on stall expectations derived during extensive exposure to non-icing flight. The incorporation of specific training objectives to familiarize aircrew with the CIRB effect would strongly complement the current emphasis on the meteorological and technical aspects of airframe icing. For example, crews should be conditioned to respond to all performance and flying quality degradations in icing as aerodynamic stalls, unless proven otherwise by overwhelming evidence. Although this recommendation seems self-evident, there were numerous cases in the archives and literature where the crew failed to make the connection between their control difficulties and impending icing-induced aerodynamic stalls. In some cases, the crew forced the aircraft into a stall, and then maintained inappropriate control inputs, sometimes for several minutes, which precluded any chance of recovery. Conflicting cues, an unfamiliar environment, and high

stress levels undoubtedly contributed to these unfortunate outcomes, and better training to recognize and address such situations would be a very important mitigation.

In addition to the preceding training recommendations, the CIRB / BLR methodology could be applied to Flight Operational Quality Assurance (FOQA) data obtained during routine operations and from simulator training data. The application of comprehensive data mining techniques, using the same variables examined in this study as a baseline, would yield important benefits. Most importantly, data mining would generate a much larger sample than was achieved by this archival study, because FOQA data are routinely collected during every flight, not just the operations that resulted in incidents or accidents, as was the case in this study. The larger sample size would allow additional stratification variables to be incorporated in the CIRB model, which was not possible with this research because of the minimum sample size limitations of the BLR. FOQA and simulator data would also capture all the successful crew performance outcomes, which would minimize the self-selection bias towards unsuccessful outcomes inherent in accident and incident archival database research. Finally, the application of the CIRB / BLR method to FOQA and simulator data would facilitate the objective quantitative evaluation of the benefits achieved from updated training practices or improvements in aircraft equipage.

Future research directions. The CIRB model should be further validated and extended using databases that were not incorporated into this study. Domestic examples include the FAA Accidents/Incidents Data System (AIDS), airline Flight Operational Quality Assurance (FOQA) data, and ASRS / NTSB icing data for reciprocating engine aircraft that were excluded from the BLR analysis. The model could also be further

validated and expanded using non-U.S. data sources such as the ICAO Accident and Incident Data-Reporting Database (ADREP). ADREP contains mandatory reports of all aviation accidents to aircraft over 2,250 kg, in accordance with Annex 13, Chapter 7 of the Chicago Convention. Zeppetelli & Habashi (2012) identified 323 relevant airborne icing occurrences since the ADREP's inception in 1970. Incorporation of ADREP data would therefore substantially expand the sample size that could be evaluated using the techniques presented in this study. Transport Canada operates the Civil Aviation Daily Occurrence Reporting System (CADORS) which could further supplement the sample size, particularly in light of the prevalence of icing encounters in Canadian airspace due to the country's climate and geography.

There is also a wealth of proprietary untapped data that is not in the public domain to which the model could be applied and tested. The CIRB / BLR methodology should be considered for applications outside the narrowly defined scope of this study. Examples include evaluations of the factors leading to non-icing related loss of control, runway incursions, or even applications in unrelated industries, such as nuclear power plant operational safety.

Future researchers are encouraged to duplicate the use of dissimilar database archives, as was done for this study. This beneficially increases the overall sample size to meet the demands of the BLR, and should also help to ensure that minimum group membership levels are achieved when the number of variable categories is increased. There is a risk of the BLR analysis failing to converge if these measures are not taken. Future research could be performed using additional stratification variables, such as crew qualifications, type of flight, etc., that could not be evaluated with the sample size

available for this study. The use of dissimilar databases also increases the triangulation that can be achieved, with beneficial consequences to the reliability and external validity of the research.

Given the analytically demonstrated validity and reliability of the basic and comprehensive CIRB models, the BLR coefficients derived in this study could be used for predictive purposes, and the preceding conclusions and recommendations should be valid and generalizable to the target population of U.S. turbine-powered, non-amateur-built aircraft. In a broader context, the SDT/BLR model can be generalized and applied to other human-in-the loop tasks, where its potential to simplify complex relationships would allow quantitative evaluations to be performed on otherwise intractable behavioral constructs, such as judgment and decision-making. Collectively, the successful implementation of these recommendations should achieve the ultimate and overriding objective stated in the first chapter of this treatise: to save lives.

REFERENCES

Aarons, R. N. (1995, Jan). NTSB icing warning. *Business & Commercial Aviation, 76*(1), 80.

Aarons, R. N. (1999, Oct). New Approaches to flight in icing conditions. *Business & Commercial Aviation, 85*(4), 126.

Abbott, I. H., & von Doenhoff, A. E. (1959). *Theory of wing sections*. New York, NY: Dover Publications.

Abdi, H. (2009). Signal detection theory. In B. McGaw, P. L. Peterson & E. Baker (Eds.), *Encyclopedia of Education* (3rd ed., pp. 1-10). New York, NY: Elsevier.

Advani, S. (2014, Mar). Opinion: Aerodynamic stall can prompt 'brain stall.' *Aviation Week & Space Technology, 176*(11), 58.

Appiah-Kubi, P., Martos, B., Atuahene, I., & William, S. (2013). U.S. inflight icing accidents and incidents, 2006 to 2010 Abstract ID: 32. *IIE Annual Conference Proceedings,* 1-15.

Aventin, A., Morency, F., & Nadeau, S. (2015). Statistical study of aircraft accidents and incidents related to de-icing / anti-icing procedures in Canada between 2009 and 2014. *Canadian Aeronautics and Space Institute Aero 2015,* Montreal QC, Canada.

Babbie, E. (2013). *The practice of social research* (13th ed.). Belmont, CA: Wadsworth, Cengage Learning.

Bazeley, P. (2013). *Qualitative data analysis: Practical strategies*. Thousand Oaks, CA: Sage.

Bergrun, N. (1995). A warming trend for icing research. *Aerospace America, 33*(8), 22.

Billings, C. E. (1997, Feb). Issues concerning human-centered intelligent systems. *National Science Foundation Workshop on Human-Centered Systems: Information, Interactivity, and Intelligence*, Arlington, VA.

Boeing Commercial Airplanes. (2015). *Statistical summary of commercial jet airplane accidents, worldwide operations: 1959 – 2014.* Seattle, WA: Boeing Commercial Airplanes.

Bragg, M. B. (1996). Aircraft aerodynamic effects due to large droplet ice accretions. *34th Space Sciences Meeting & Exhibit,* Reno, NV.

Bragg, M. B., Basar, T., Perkins, W. R., Selig, M. S., Voulgaris, P. G., Melody, J. W., & Sarter, N. B. (2002). Smart icing systems for aircraft icing safety. *AIAA 2002-0813: 40th Aerospace Sciences Meeting & Exhibition,* Reno, NV.

Bragg, M. B., Perkins, W. R., Sarter, N. B., Basar, T., Voulgaris, P. G., Gurbacki, H. M., McCray, S. A. (1998). An interdisciplinary approach to inflight aircraft icing safety. *36th Aerospace Sciences Meeting & Exhibit,* Reno, NV.

Broeren, A. P., Bragg, M. B., & Addy, H. E. (2004). Effect of intercycle ice accretions on airfoil performance. *Journal of Aircraft, 41*(1), 165-174.

Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (BEA). (2012). *Final report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro - Paris.* (Accident Report - Final). France: BEA.

Carlisle, D. (2006, Oct). Ice-contaminated tailplane stall. *Business & Commercial Aviation, 99*(4), 103-106.

Catlin, P. (1992). *Stall warning using contamination detecting aerodynamics* (SAE Technical Paper 922010). doi:10.4271/922010

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37-46. doi:10.1177/001316446002000104

Cole, J., & Sand, W. (1991). Statistical study of aircraft icing accidents. Paper presented at the *29th AIAA Aerospace Sciences Meeting*, Reno, NV. doi:10.2514/6.1991-558

Detwiler, R. (2015). *How to recover from tailplane icing*. Retrieved from http://aviationweek.com/business-aviation/how-recover-tailplane-icing

Endsley, M. R., & Jones, D. G. (2012). *Designing for situation awareness. An approach to user-centered design* (2nd ed.). Boca Raton, FL: CRC Press.

Federal Aviation Administration. (1996a). *Airworthiness Directives; de Havilland Model DHC-7 and DHC-8 Series Airplanes,* 49 U.S.C. 14 C.F.R. §39.

Federal Aviation Administration. (1996b). *Proceedings of the FAA international conference on aircraft in flight icing, volume 1: Plenary sessions,* Springfield, VA.

Federal Aviation Administration. (1997). *FAA inflight aircraft icing plan.* Retrieved from http://lessonslearned.faa.gov/AmericanEagle/FAA_Icing_Plan.pdf

Federal Aviation Administration. (2004). *Airplane flying handbook.* Oklahoma City, OK: FAA. Retrieved from https://www.faa.gov/regulations_policies/handbooks_manuals/aircraft/airplane_handbook/media/FAA-H-8083-3B.pdf

Federal Aviation Administration. (2006). *Aircraft ice protection*. (Advisory Circular No. 20-73A). Washington, DC: Department of Transportation.

Federal Aviation Administration. (2007). *Activation of ice protection*, 49 U.S.C. 72 Fed. Reg. 20924 (proposed Apr. 26, 2007, to be codified at 14 CFR pt. 25).

Federal Aviation Administration. (2009a). *Ice Protection Rule*, 14 CFR § 25.1419.

Federal Aviation Administration. (2009b). *Landing*, 49 U.S.C. 14 C.F.R. § 25.125.

Federal Aviation Administration. (2009c). *Operating in icing conditions rule*, 14 CFR § 91.527.

Federal Aviation Administration. (2010a). *Abbreviations and symbols*. 14 CFR § 1.2.

Federal Aviation Administration. (2010b) *Airplane and engine certification requirements in supercooled large drop, mixed phase, and ice crystal icing conditions*, 49 U.S.C.

75 Fed. Reg. 37311 (proposed June 29, 2010, to be codified in 14 C.F.R. pts. 25 and 33).

Federal Aviation Administration. (2010c). *Cockpit voice recorders*, 49 U.S.C. 14 C.F.R. § 121.359.

Federal Aviation Administration. (2010d). *Flight data recorders and cockpit voice recorders*, 49 U.S.C. 14 C.F.R. § 91 609.

Federal Aviation Administration. (2010e). *Press release: Fact sheet – flying in icing conditions*. Retrieved from http://www.faa.gov/news/fact_sheets/news_story.cfm?newsid=10398

Federal Aviation Administration. (2010f). *Weather-related aviation accident study 2003–2007*. Washington, DC: Federal Aviation Administration.

Federal Aviation Administration. (2011a). *Aviation safety reporting program.* (AC 00-46E). Washington, DC: Federal Aviation Administration.

Federal Aviation Administration. (2011b). *Flight test guide for certification of transport category airplanes: Chapter 8 - airworthiness: miscellaneous items 228. Design and function of artificial stall warning and identification systems.* (AC 25-7B). Washington, DC: Federal Aviation Administration.

Federal Aviation Administration. (2014a). *Aeronautical information manual*. Washington, DC: Department of Transportation.

Federal Aviation Administration. (2014b). *Approval of non-required angle of attack (AoA) indicator systems.* Washington, DC: Federal Aviation Administration.

Federal Aviation Administration. (2014c). *Compliance of transport category airplanes with certification requirements for flight in icing conditions.* (AC 25-28). Washington, DC: Federal Aviation Administration. (2014d). *Airplane and engine certification requirements in supercooled large drop, mixed phase, and ice crystal icing conditions*, 49 U.S.C. 14 C.F.R. pts. 25 and 33.

Federal Aviation Administration. (2014e). *Stall warning rule*, 14 CFR § 25.207.

Federal Aviation Administration. (2014f). *Performance and handling characteristics in icing conditions.* (AC 25-25A).

Federal Aviation Administration. (2014g). *Press release: FAA issues final rule updating aircraft icing standards.* Retrieved from https://www.faa.gov/news/press _releases/news_story.cfm?newsId=17674

Federal Aviation Administration. (2016). *Revision of airworthiness standards for normal, utility, acrobatic, and commuter category airplanes*, 49 U.S.C. 81 Fed. Reg. 13452 (proposed Mar. 14, 2016, to be codified at 14 C.F.R. pts. 21, 23, 35, 43, 91, 121 and 135).

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low Kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology, 43*(6), 543-548. doi:10.1016/0895-4356(90)90158-L

Field, A. (2009). *Discovering statistics using SPSS* (third ed.). Thousand Oaks, CA: Sage Publications.

Fiorino, F. (2007, Jan). NTSB aims to reduce hazards of icing. *Aviation Week & Space Technology, 166*(5), 47.

Flight Safety Foundation. (1996). Pilots can minimize the likelihood of roll upset in severe icing. *Flight Safety Digest, 15*(1), 1-9.

Flight Safety Foundation. (2008). *Supplement #1 to the airplane upset recovery training aid*. Alexandria, VA: Flight Safety Foundation.

Flightglobal. (1996). *FAA icing rules change.* Retrieved from https://www.flightglobal.com/news/articles/faa-icing-rules-change-16988/

Flottau, J. (2012, Jul 09). AF447 crash report cites 'startle effect.' *Aviation Week & Space Technology, 389*(4), 54.

Garvey, W. (2010, Oct 01). Icing insights from Glenn's gurus. *Business & Commercial Aviation, 106*(10), 76.

George, Fred. (1997, Dec). Ice-contaminated-tailplane-stall. *Business & Commercial Aviation, 81*(6), 80.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York, NY: Wiley.

Green, S. D. (2006). A study of U.S. inflight icing accidents and incidents, 1978 to 2002. In *AIAA 2006-82* from 44th AIAA Aerospace Sciences Meeting and Exhibit, Reno, NV.

Green, S., Bettcher, J., Brachen, J., & Erickson, S. (1996). *Tools for the operational management of inflight icing in the twenty-first century.* American Institute of Aeronautics and Astronautics. doi:10.2514/6.1996-136

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Hendriksen, I., & Holewijn, M. (1999). Degenerative changes of the spine of fighter pilots of the Royal Netherlands Air Force (RNLAF). *Aviation Space Environmental Medicine,* 70, 1057–1063.

Jeck, R. K. (2001). *A history and interpretation of aircraft icing intensity definitions and FAA rules for operating in icing conditions.* (Final No. DOT/FAA/AR-01/91). Atlantic City, NJ: FAA Office of Aviation Research.

Joslin, R. E. (2013). *Validation of new technology using legacy metrics: Examination of surf-IA alerting for runway incursion incidents* (Doctoral dissertation). Retrieved from ProQuest Dissertation Database. (3594575)

Klemin, A. (1944). The I.Ae.S. annual meeting. *Aircraft Engineering and Aerospace Technology, 16*(7), 195-202.

Lee, M. D. (2008). Bayes SDT: Software for Bayesian inference with signal detection theory. *Behavior Research Methods, 40*(2), 450-456.

Lerner, E. J. (1985). Stall warning - catching it early. *Aerospace America, 23*, 42-43.

Luers, J. K. (1983). Wing contamination - threat to safe flight. *Astronautics and Aeronautics (0004-6213), 21*, 54-59.

Lynch, K. (2014, March 6). FAA bans 4,200 twin piston Cessnas from flight into known icing. *Aviation Week Intelligence Network.* Retrieved from http://aviationweek.com/awin/faa-bans-4200-twin-piston-cessnas-flight-known-icing

Lynch, F. T., & Khodadoust, A. (2001). Effects of ice accretions on aircraft aerodynamics. *Progress in Aerospace Sciences*, *37*(8), 669-767.

Manningham, D. (1997, Feb 01). New problems with ice. *Business & Commercial Aviation, 80*(2), 72.

Maris, J. M. (1991). The turbulence intensity parameter as a basis for stall prediction. *Canadian Aeronautics and Space Institute Flight Test Section,* Montreal, QC.

Maris, J. M. (1996). Airfoil performance monitoring using the Turbulence Intensity Parameter. *Proceedings of the FAA International Conference on Aircraft Inflight Icing, Volume II 601-607,* Springfield, VA.

Maris, J. M. (2009). Avionics flight test principles and practices. *Transport Canada 2009 national aircraft certification - Delegates conference,* Ottawa, ON.

NASA. (1999). Tail plane icing. *Scientific and Technical Aerospace Reports (STAR), 37.*

NASA. (2014). *Aviation safety database reporting system (ASRS) database report sets: Commuter and GA Icing incidents.* Retrieved from http://asrs.arc.nasa.gov/search/reportsets.html

National Aviation Weather Program Council. (1999). *National aviation weather initiatives.* (No. FCM-P34-1999). Washington, DC: National Aviation Weather Program Council.

NBAA. (2015). *Loss of control in flight.* Retrieved from https://www.nbaa.org/ops/safety/in-flight-safety/loss-of-control-in-flight/

Newton, D. (2006, Sep). What's ahead in icing certification. *Business & Commercial Aviation, 99*(3), 42.

North, D. M. (1998, Feb). NASA safety study focuses on tailplane icing stalls. *Aviation Week & Space Technology, 148*(6), 81.

NTSB. (1982). *Air Florida, Inc. Boeing 737-222, N62AF collision with 14th Street Bridge, near Washington National Airport, Washington, D.C. January 13, 1982* (Final report No. NTSB-AAR-82-8). Springfield, VA: NTSB.

NTSB. (1996a). *In-flight icing encounter and loss of control, Simmons Airlines, d.b.a. American Eagle flight 4184, Avions de Transport Regional (ATR) model 72-212,*

N401AM, Roselawn, Indiana, October 31, 1994. Washington, D.C.; Springfield, VA: NTSB.

NTSB. (1996b). *Volume I. In-flight icing encounter and loss of control, Simmons Airlines, d.b.a. American Eagle Flight 4184, Avions de Transport Regional (ATR) Model 72-212, N401AM. Roselawn, Indiana, October 31, 1994* (No. 2013).

NTSB. (1996c). *Volume II: In-flight icing encounter and loss of control, Simmons Airlines, d.b.a. American Eagle Flight 4184, Avions de Transport Regional (ATR) Model 72-212, N401AM. Roselawn, Indiana, October 31, 1994. Response of Bureau Enquêtes-Accidents to safety board's daft report* (No. 2013). Retrieved from http://purl.access.gpo.gov/GPO/LPS109375

NTSB. (1998a). *Comair flight 3272 Embraer EMB-120RT, N265CA Monroe, Michigan, January 9, 1997* (Final report No. 6997A/B/C). Springfield, VA: NTSB.

NTSB. (1998b). *In-flight icing encounter and uncontrolled collision with terrain* (Aircraft accident report No. PB98-910404). Washington, DC: NTSB.

NTSB. (2010a). *Loss of control on approach, Colgan Air, Inc., operating as Continental Connection Flight 3407, Bombardier DHC-8-400, N200WQ, Clarence Center, New York, February 12, 2009*. Washington, DC: NTSB.

NTSB Initial notification of aircraft accidents, incidents, and overdue aircraft, 49 U.S.C. Subtitle B § 830.5 (2010b).

NTSB. (2010c). *Press release: NTSB Chairman Hersman testifies on aircraft icing.* Retrieved from http://www.ntsb.gov/news/press-releases/Pages/NTSB_Chairman_Hersman_Testifies_on_Aircraft_Icing.aspx

NTSB. (2012a). *Aviation accident database.* Retrieved from http://www.ntsb.gov/aviationquery/index.aspx

NTSB. (2012b). *Most wanted list: General aviation safety*. Retrieved from http://www.ntsb.gov/safety/mwl-2.html

NTSB. (2016a). *Most wanted list of transportation safety improvements: prevent loss of control in General Aviation.* (Fact Sheet). Retrieved from https://www.ntsb.gov/safety/mwl/Documents/2017-18/2017MWL-FctSht-LossControl-A.pdf

NTSB. (2016b). *Press release: NTSB finds failure to use de-ice system caused 2014 crash of Embraer jet.* Retrieved from http://www.ntsb.gov/news/events/Pages/2016_gaithersburg_BMG.aspx

Pederson, E. T. (2003). *Investigation of the iced flowfield characteristics related to the stall margin instrumentation used in icing conditions* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (305294529)

Peterson, W. W., Birdsall, T. G., & Fox, W. (1954). The theory of signal detectability. *Information Theory, Transactions of the IRE Professional Group On, 4*(4), 171-212. doi:10.1109/TIT.1954.1057460

Petty, K., & Floyd, D. (2004). A statistical review of aviation airframe icing accidents in the U.S. *11th Conference on Aviation, Range, and Aerospace,* Hyannis, MA., 81425-1 11.2.

Ratvasky, T. P., Van Zante, J. F., & Riley, J. T. (1999). NASA/FAA tailplane icing program overview: AIAA-99-0370. *37th Aerospace Sciences Meeting & Exhibit,* Reno, Nevada.

Reehorst, A., Chung, J., Potapczuk, M., & Choo, Y. (2000). Study of icing effects on performance and controllability of an accident aircraft. *Journal of Aircraft, 37*(2), 253-259.

SAE International. (2014). *SAE G-12 aircraft ground deicing committee charter and operating procedures*. Retrieved from https://www.sae.org/works/committeeResources.do?resourceID=395984

Sarter, N., & Schroeder, B. (2001). Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human Factors, 43*(4), 573.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2003). *Experimental and quasi-experimental designs for generalized causal inference* (2nd ed.). Boston, MA: Houghton Mifflin.

Sheridan, T. B., & Parasuraman, R. (2000). Human versus automation in responding to failures - an expected-value analysis. *Human Factors, 42*(3), 403-407.

Taiwan Aviation Safety Council. (2005). *In-flight icing encounter and crash into the sea TransAsia Airways flight 791 ATR72-300, B-22708 17 kilometers southwest of Makung City, Penghu Islands, Taiwan, December 21, 2002.* (Occurrence Investigation Report No. ASC-AOR-05-04-001). Taipei, Taiwan: Aviation Safety Council.

Tanner, W. P., Jr., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review, 61*(6), 401-409. doi:10.1037/h0058700

Telford, J. W. (1988). An example of the behavior of an aircraft with accumulated ice - latent instability. *Journal of Applied Meteorology, 27,* 1093-1108.

Thurber, M. (2008, February 7). Winter's here, time to brush up on training for ground icing. *AINonline.* Retrieved from http://www.ainonline.com/aviation-news/aviation-international-news/2008-02-07/winters-here-time-brush-training-ground-icing

Transportation Safety Board of Canada. (2006). *Loss of control Georgian Express Ltd. Cessna 208B Caravan C-FAGA Pelee Island, Ontario, 17 January 2004.* (No. A04H0001). Ottawa, ON Canada: Minister of Public Works and Government Services. Retrieved from http://www.tsb.gc.ca/eng/rapports-reports/aviation/2004/a04h0001/a04h0001.pdf

Transport Canada. (2004). *TP10643 When in doubt. Small and large aircraft critical surface contamination training for aircrew and groundcrew* (7th ed.). Ottawa, ON: Transport Canada.

Veillette, P. R. (2006, Nov). Residual ice: It's dangerous and overlooked. *Business & Commercial Aviation, 99*(5), 36.

Veillette, P. R. (2009, Aug). The hazard of heavy rain. *Business & Commercial Aviation, 105*(2), 26.

Veillette, P. R. (2012, Jul-Sep). Investigating and preventing the loss of control accident, Part 1. *ISASI Forum, 45*(3), 5-9.

Vigeant-Langlois, L., & Hansman, R. J., Jr. (2000). Influence of icing information on pilot strategies for operating in icing conditions. *Journal of Aircraft, 37*(6), 937-946.

Vogt, W. P., Gardner, D. C., & Haeffele, L. M. (2012). *When to use what research design*. New York, NY: The Guilford Press.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors, 50(3),* 433-441.

Weener, E. (2011). NTSB focus area - inflight icing. *International Winter Operations Conference,* Montreal, Quebec, Canada.

Zeppetelli, D., & Habashi, W. G. (2012). In-flight icing risk management through computational fluid dynamics-icing analysis. *Journal of Aircraft, 49*(2), 611.

Zotov, D. (2000). *Scientific methods for accident investigation*. Australian Society of Air Safety Investigators (ASASI) 2000 Regional Seminar, Christchurch, NZ.

Zuschlag, M. (2005). *Violations of temporary flight restrictions and air defense identification zones: An analysis of airspace violations and pilot report data* (NASA/CR-2005-213923). Cambridge, MA: VOLPE.

APPENDIX A

ASRS QUERY STRING

```
SELECT DISTINCT AL1.ITEM_ID FROM QWPUBLIC.ALL_ITEMS AL1,
QWPUBLIC.TEXT AL2, QWPUBLIC.ENVIRONMENT AL3, QWPUBLIC.TEXT
AL4 WHERE (AL2.ITEM_ID=AL1.ITEM_ID AND AL3.ITEM_ID=AL1.ITEM_ID
AND AL4.ITEM_ID=AL1.ITEM_ID)  AND (CONTAINS(AL2.TEXT,'STALL%')>0
AND AL3.ATTRIBUTE='Weather Elements / Visibility' AND AL3.VALUE='Icing')
UNION  SELECT AL1.ITEM_ID FROM QWPUBLIC.ALL_ITEMS AL1,
QWPUBLIC.TEXT AL2, QWPUBLIC.ENVIRONMENT AL3, QWPUBLIC.TEXT
AL4 WHERE (AL2.ITEM_ID=AL1.ITEM_ID AND AL3.ITEM_ID=AL1.ITEM_ID
AND AL4.ITEM_ID=AL1.ITEM_ID)  AND (CONTAINS(AL2.TEXT,'STALL%')>0
AND CONTAINS(AL4.TEXT,'ICE OR ICING OR ICED OR RIME OR RHIME OR
FROZE OR FREEZ%')>0) UNION  SELECT AL1.ITEM_ID FROM
QWPUBLIC.ALL_ITEMS AL1, QWPUBLIC.TEXT AL2,
QWPUBLIC.ENVIRONMENT AL3, QWPUBLIC.TEXT AL4 WHERE
(AL2.ITEM_ID=AL1.ITEM_ID AND AL3.ITEM_ID=AL1.ITEM_ID AND
AL4.ITEM_ID=AL1.ITEM_ID)  AND (AL1.ENTITY='Component' AND
AL1.ATTRIBUTE='Aircraft Component' AND AL1.VALUE IN ('Stall Barrier System',
'Stall Protection System', 'Stall Warning System') AND AL3.ATTRIBUTE='Weather
Elements / Visibility' AND AL3.VALUE='Icing') UNION  SELECT AL1.ITEM_ID
FROM QWPUBLIC.ALL_ITEMS AL1, QWPUBLIC.TEXT AL2,
QWPUBLIC.ENVIRONMENT AL3, QWPUBLIC.TEXT AL4 WHERE
(AL2.ITEM_ID=AL1.ITEM_ID AND AL3.ITEM_ID=AL1.ITEM_ID AND
AL4.ITEM_ID=AL1.ITEM_ID)  AND (AL1.ENTITY='Component' AND
AL1.ATTRIBUTE='Aircraft Component' AND AL1.VALUE IN ('Stall Barrier System',
'Stall Protection System', 'Stall Warning System') AND CONTAINS(AL4.TEXT,'ICE
OR ICING OR ICED OR RIME OR RHIME OR FROZE OR FREEZ%')>0)
```

Figure 11. ASRS Query String.

APPENDIX B

ASRS LIMITATIONS

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, CA 94035-1000

TH: 262-7

MEMORANDUM FOR: Recipients of Aviation Safety Reporting System Data

SUBJECT: Data Derived from ASRS Reports

The attached material is furnished pursuant to a request for data from the NASA Aviation Safety Reporting System (ASRS). Recipients of this material are reminded when evaluating these data of the following points.

ASRS reports are submitted voluntarily. The existence in the ASRS database of reports concerning a specific topic cannot, therefore, be used to infer the prevalence of that problem within the National Airspace System.

Information contained in reports submitted to ASRS may be amplified by further contact with the individual who submitted them, but the information provided by the reporter is not investigated further. Such information represents the perspective of the specific individual who is describing their experience and perception of a safety related event.

After preliminary processing, all ASRS reports are de-identified and the identity of the individual who submitted the report is permanently eliminated. All ASRS report processing systems are designed to protect identifying information submitted by reporters; including names, company affiliations, and specific times of incident occurrence. After a report has been de-identified, any verification of information submitted to ASRS would be limited.

The National Aeronautics and Space Administration and its ASRS current contractor, Booz Allen Hamilton, specifically disclaim any responsibility for any interpretation which may be made by others of any material or data furnished by NASA in response to queries of the ASRS database and related materials.

Linda J. Connell

Linda J. Connell, Director
NASA Aviation Safety Reporting System

CAVEAT REGARDING USE OF ASRS DATA

Certain caveats apply to the use of ASRS data. All ASRS reports are voluntarily submitted, and thus cannot be considered a measured random sample of the full population of like events. For example, we receive several thousand altitude deviation reports each year. This number may comprise over half of all the altitude deviations that occur, or it may be just a small fraction of total occurrences.

Moreover, not all pilots, controllers, mechanics, flight attendants, dispatchers or other participants in the aviation system are equally aware of the ASRS or may be equally willing to report. Thus, the data can reflect **reporting biases**. These biases, which are not fully known or measurable, may influence ASRS information. A safety problem such as near midair collisions (NMACs) may appear to be more highly concentrated in area "A" than area "B" simply because the airmen who operate in area "A" are more aware of the ASRS program and more inclined to report should an NMAC occur. Any type of subjective, voluntary reporting will have these limitations related to quantitative statistical analysis.

One thing that can be known from ASRS data is that the number of reports received concerning specific event types represents the **lower measure** of the true number of such events that are occurring. For example, if ASRS receives 881 reports of track deviations in 2010 (this number is purely hypothetical), then it can be known with some certainty that at least 881 such events have occurred in 2010. With these statistical limitations in mind, we believe that the **real power** of ASRS data is the **qualitative information** contained in **report narratives**. The pilots, controllers, and others who report tell us about aviation safety incidents and situations in detail – explaining what happened, and more importantly, **why** it happened. Using report narratives effectively requires an extra measure of study, but the knowledge derived is well worth the added effort.

APPENDIX C

Tables

Table C1

ASRS Data Structure

ASRS Variable	ASRS Variable (cont.)
1. ACN	2. Date
3. Local Time Of Day	4. Locale Reference
5. State Reference	6. Relative Position.Angle.Radial
7. Relative Position.Distance.Nautical Miles	8. Altitude.AGL.Single Value
9. Altitude.MSL.Single Value	10. Flight Conditions
11. Weather Elements / Visibility	12. Work Environment Factor
13. Light	14. Ceiling
15. RVR.Single Value	16. ATC / Advisory
17. Aircraft Operator	18. Make Model Name
19. Propulsion	20. Aircraft Zone
21. Crew Size	22. Operating Under FAR Part
23. Flight Plan	24. Mission
25. Nav In Use	26. Flight Phase
27. Route In Use	28. Airspace
29. Maintenance Status.Maintenance Deferred	30. Maintenance Status.Records Complete
31. Maintenance Status.Released For Service	32. Maintenance Status.Required / Correct Doc On Board
33. Maintenance Status.Maintenance Type	34. Maintenance Status.Maintenance Items Involved
35. Cabin Lighting	36. Number Of Seats.Number
37. Passengers On Board.Number	38. Crew Size Flight Attendant.Number

ASRS Variable	ASRS Variable (cont.)
	Of Crew
39. Aircraft Component	40. Manufacturer
41. Aircraft Reference	42. Problem
43. ATC / Advisory	44. Aircraft Operator
45. Make Model Name	46. Aircraft Zone
47. Crew Size	48. Operating Under FAR Part
49. Flight Plan	50. Mission
51. Nav In Use	52. Flight Phase
53. Route In Use	54. Airspace
55. Maintenance Status.Maintenance Deferred	56. Maintenance Status.Records Complete
57. Maintenance Status.Released For Service	58. Maintenance Status.Required / Correct Doc On Board
59. Maintenance Status.Maintenance Type	60. Maintenance Status.Maintenance Items Involved
61. Cabin Lighting	62. Number Of Seats.Number
63. Passengers On Board.Number	64. Crew Size Flight Attendant.Number Of Crew
65. Location Of Person	66. Location In Aircraft
67. Reporter Organization	68. Function
69. Qualification	70. Experience
71. Cabin Activity	72. Human Factors
73. Communication Breakdown	74. ASRS Report Number.Accession Number
75. Location Of Person	76. Location In Aircraft
77. Reporter Organization	78. Function

ASRS Variable	ASRS Variable (cont.)
79. Qualification	80. Experience
81. Cabin Activity	82. Human Factors
83. Communication Breakdown	84. ASRS Report Number.Accession Number
85. Anomaly	86. Miss Distance
87. Were Passengers Involved In Event	88. Detector
89. When Detected	90. Result
91. Contributing Factors / Situations	92. Primary Problem
93. Narrative	94. Callback
95. Narrative	96. Callback
97. Synopsys	98. –

Table C2

NASA ASRS Final Sample Case Listing

ASRS ACN	Event Year	ASRS ACN	Event Year	ASRS ACN	Event Year
100792	1988	386766	1997	642483	2005
104442	1989	389483	1997	643904	2004
115422	1989	390641	1997	665350	2005
189745	1991	391550	1998	682246	2005
191028	1991	393189	1998	684037	2006
200004	1992	393446	1998	692028	2006
202249	1992	395823	1998	714794	2006
211430	1992	403299	1998	760888	2007
225830	1992	418260	1998	765665	2007
231194	1993	419839	1998	765691	2007
235939	1993	423056	1998	774091	2008
250881	1993	423333	1998	832021	2009
260890	1994	425239	1999	845030	2009
264355	1994	441448	1999	849667	2009
265218	1994	452162	1999	852531	2009
268036	1994	456868	1999	881246	2010
282950	1994	463853	2000	881955	2010
286127	1994	470303	2000	924002	2010
326726	1996	476789	2000	925811	2011
327563	1996	479942	2000	1090560	2013
327661	1996	495957	2000	1128912	2013
327877	1996	519723	2001	1147583	2014
330391	1996	522830	2001	1152737	2014
357096	1996	541639	2002	1168045	2014
357245	1997	565131	2002	1227048	2014
366589	1997	589618	2003		
376201	1997	601072	2003		

Table C3

NTSB Final Sample Case Listing

NTSB #	Event Year	NTSB #	Event Year
ANC00LA017	1999	DCA09MA027	2009
ANC02FA020	2002	DCA15MA029	2014
ANC05CA040	2005	DCA90MA011	1989
ANC08LA027	2007	DCA91MA019	1991
ANC10LA019	2010	DCA92MA025	1992
ANC11TA031	2011	DCA93IA027	1993
ANC98LA018	1998	DCA95MA001	1994
ANC98MA008	1997	DCA97MA017	1997
CEN09MA142	2009	ERA11LA344	2011
CEN10LA090	2010	FTW03FA089	2003
CEN11FA144	2011	FTW93MA143	1993
CEN12LA095	2011	FTW95FA094	1995
CEN12LA153	2012	FTW95FA129	1995
CEN15LA091	2014	LAX02FA108	2002
CHI02IA151	2002	LAX02LA030	2001
CHI04IA056	2004	LAX06IA076	2006
CHI06IA127	2006	LAX95IA128	1995
CHI06LA058	2006	MIA98LA061	1998
CHI07LA059	2007	NYC04LA044	2003
CHI89IA034	1988	NYC04LA050	2003
CHI89MA057	1989	NYC05MA083	2005
CHI90IA106	1990	NYC07LA081	2007
CHI93MA061	1993	NYC97FA045	1997
CHI97FA047	1996	NYC98LA028	1997
CHI98LA084	1998	SEA08FA042	2007
DCA01MA031	2001	SEA95LA059	1995
DCA09IA064	2009		

APPENDIX D

SME DATA ENCODING CHECKLIST

Introduction

You have graciously agreed to participate in a doctoral dissertation research study that examines the effect of stall warning system performance on crew behavioral outcomes during airborne icing encounters. Airborne icing has been a major cause of aircraft accidents and loss of life since the earliest days of aviation and continues to cause aircraft accidents and incidents. Airframe icing has been featured in the National Transportation Safety Board's (NTSB) *Most Wanted List* of transportation safety improvements for 14 years (NTSB, 2012b). The objective of this research program is to help increase the understanding of the complex interactions between aircrew, the icing environment, aircraft dynamics, and the aircraft systems during icing encounters. The findings should result in improved guidelines for the design, certification, and operation of stall protection systems, with the overall objective of reducing accidents and saving lives.

The research entails the selection and encoding of icing-encounter cases that meet strict criteria from the NTSB Aviation Accident database and the NASA Aviation Safety Reporting System (ASRS) database. This activity addresses both the selection and the encoding of the candidate cases for further study. For a case to be included in the study, the record must contain unambiguous information about a number of critical variables. These include: the aircraft's wing and / or tail stall state; the activation state of stall warning and / or stall prevention equipment (e.g. aural stall alerts, stick shaker, and stick

pusher systems); and whether the initial crew response was *correct*, according to specific criteria defined for the study. The *correct response* criterion is used in a very specific manner in the study: It does not relate to the success or failure of the crew interventions nor is it intended to judge the competency of the crew. In order for the study to be valid and reliable, very detailed procedures and criteria must be applied to ensure that the appropriate records are selected and that they are consistently encoded without introducing observer bias. Any inclination to assess the accident or incident situations subjectively must be studiously avoided. The next sections address the two principal archive processing activities: record selection and data encoding.

Record Selection

The ASRS records to be analyzed are contained in a tailored extract of 381 cases provided by NASA ASRS analysts, and the sample is ready for encoding as presented. The NTSB sample records must be downloaded from the NTSB on-line query page: http://www.ntsb.gov/_layouts/ntsb.aviation/index.aspx using the criteria shown in Table D1. Unlisted parameters must be left at their default values. A separate extract will be required for each of the turbojet, turboprop, and turbofan engine types.

Table D1

NTSB Database Search Criteria

Characteristic	**Value**
Event start date	January 1, 1988
Event end date	October 2, 2015
AircraftCategory Filter	Airplane
AmateurBuilt Filter	No
EngineType Filter	Turbine aircraft (turbojet, turboprop, turbofan). Separate extracts will be required for each of these.
WeatherConditions Filter	All
Word search string	"icing" or "freezing" or "rime" or "glaze" or "sleet" or "frost"

Archive Encoding

The ASRS and NTSB extracts must be evaluated for records containing the required aircraft, stall warning, and crew response information. The resulting sub-samples of conforming records must then be encoded for these variables. Both processes are accomplished using the procedure shown in Table D2.

Table D2

NTSB and ASRS Database Encoding Checklist

Step	Action	Record
1	Record Initial ASRS and NTSB sample sizes:	
1a	Initial ASRS sample size	381 Cases
1b	Initial NTSB turbojet sample size	____Cases
1c	Initial NTSB turboprop sample size	____Cases
1d	Initial NTSB turbofan sample size	____Cases
2	Perform ASRS and NTSB sub-sampling: Reject all records that were not related to *airframe* icing (e.g. engine issues related to ice crystal ingestion, runway overruns due to landing surface contamination, etc.). Reject all records for which the aircraft stall-state, stall warning system performance, and crew performance outcomes[a] cannot be unequivocally determined. Knowledge of the stall warning system state entails explicit evidence of aural stall warnings, stick-shaker, or stick pusher activation, with one exception: if a detailed crew narrative of the accident or incident is available and no indication or mention is made of a stall warning actuation, then the stall warning system state will be encoded *None*.	
2a	Record ASRS sub-sample size	____Cases
2b	Record NTSB turbojet sub-sample size	____Cases
2c	Record NTSB turboprop sub-sample size	____Cases
2d	Record NTSB turbofan sub-sample size	____Cases

Step	Action	Record
	For each remaining sub-sample:	
3	Encode the aircraft stall state (no stall, imminent or actual wing stall, imminent or actual tail stall, or system issue[b]).	
4	Encode the crew response outcome using the criteria in Table D3.	
5	Encode the stall warning system outcome into one of four states: *Hit, Miss, False Alarm, or Correct Rejection* using the criteria in Table D4.	
6	Record encoded cases in worksheet for further processing in the ASRS and NTSB Sample Data Entry Worksheet. Duplicate and use as many sheets as necessary to record all of the sub-sample data. Number each of the sheets and record your name at the bottom of each completed sheet. Cross out and initial any incorrect entries and any unused rows on the last sheet.	Complete Table D5

Note. [a]Crew performance outcomes are defined using the criteria in Table D3. [b]A system issue is one that resulted in an icing stall event that was *not* caused by airframe icing. An example would be loss of primary airspeed information due to a frozen pitot-static system.

Table D3

Crew Performance Outcome Evaluation Criteria for SME

	Incorrect Crew Performance	**Correct Crew Performance**
1.	The NTSB probable cause or contributing factor in a factual or final report indicates that the crew's initial response was inappropriate (e.g. BEA, 2012; NTSB, 1996b; NTSB, 2010a).	The NTSB probable cause or contributing factor in a factual or final report indicates that the crew's initial response was appropriate.
2.	The ASRS submitter indicated that the crew response was inappropriate.	The ASRS submitter indicated that the crew response was appropriate.
3.	The crew first became aware of the impending stall or loss of control after their onset (i.e., the crew allowed the situation to degrade to the point where control was lost before recognizing this fact).	The crew first became aware of the impending stall or loss of control before their onset and made positive efforts to avoid the event, regardless of the success of the outcome.
4.	An appropriate stall warning was not acted upon in time to avoid a true aerodynamic stall or loss of control.	An appropriate stall warning was acted upon in a timely fashion, regardless of the success of the outcome.
5.	The crew response was markedly different from an accepted norm (i.e., adding power and firmly lowering the nose to prevent a wing stall) (e.g. NTSB, 2010a).	The crew response conformed to the accepted norm, regardless of the success of the outcome.
6.	The crew appeared to be unaware of the stall-state of the aircraft or misdiagnosed its state (e.g. BEA, 2012).	The crew appeared to be aware of the stall-state of the aircraft.
7.	There is *overwhelming evidence* from a subjective review of the record that the crew's initial response was inappropriate.	There is *overwhelming evidence* from a subjective review of the record that the crew's initial response was appropriate.

Table D4

Stall Warning System State Encoding

Aircraft Wing or Tail Stall State	**Stall Warning System State *preceding* the aircraft event**	**SDT[a] Classification**
No imminent or actual wing or tail stall[b]	None	Correct Rejection (CR)
No imminent or actual wing or tail stall	Alert	False Alarm (FA)
Imminent or actual wing stall	None	Miss (M)
Imminent or actual wing stall	Alert	Hit (H)
Imminent or actual tail stall	None	Miss (M)
Imminent or actual tail stall	Alert[b]	Hit (H)

Note. [a]Signal Detection Theory. [b]Imminent or actual stall can be inferred from airframe buffeting or vibration, wing or nose drop, marked control difficulties, or an inability to stop a descent, such as uncontrolled sink in the landing flare. [c]There are currently no artificial stall warning systems capable of detecting a tail stall, so this is a placeholder category only.

Table D5

ASRS and NTSB Sample Data Entry Worksheet

1. Duplicate and use as many sheets as necessary to record all of the sub-sample data. **2. Number each of the sheets and record your name at the bottom of each completed sheet.** **3. Cross out and initial any incorrect entries and any unused rows on the last sheet.**								
	ASRS ACN or NTSB ID #	**Event Date YYYYMM**	**Pending Wing Stall (Y/N)**	**Pending Tail Stall (Y/N)**	**Stall Warning (Y/N)**	**Correct Crew Response (Y/N)**	**System Issue (Y/N)**	**SDT Class (H, M, FA, CR)**
e.g.:	12345	200112	**Y**	**N**	**N**	**N**	**N**	**M**
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								
23								
24								
25								
SHEET ____ of ____ **SME Name:_____**								