Predicting postoperative delirium: Examining the role of preoperative surgical risk and executive function

By

Heidi Lindroth

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Nursing)

at the

University of Wisconsin-Madison

2018

Date of final oral examination: 07/30/2018

The Final Oral Committee:

Lisa Bratzke, Assistant Professor, Nursing
Robert D. Sanders, Assistant Professor, Department of Anesthesiology
Roger Brown, Professor, Nursing
Tonya Roberts, Assistant Professor, Nursing
Kirk Hogan, Professor, Department of Anesthesiology

Copyright Page

Table of Contents

Abstract

Delirium, an acute brain failure, is a significant public health concern because it is independently associated with an increased risk of mortality and morbidity. At least 50% of older adults will experience delirium after their surgery. Identifying high-risk, or vulnerable, individuals, prior to surgery would be advantageous to implementing perioperative delirium prevention measures. The healthcare team is encumbered by an extensive list of predisposing and precipitating risk factors leading to difficulty in selecting important factors. Further, while it is well-known that dementia increases the risk for postoperative delirium, it is not well-known what risk factors, or vulnerabilities, increase the risk in those without pre-existing dementia. Therefore, the present dissertation study was conducted to increase knowledge surrounding delirium vulnerability and prediction.

The first manuscript details the completed systematic review of literature aimed at identifying delirium prediction models in older (≥60) hospitalized adults. Current delirium prediction models are limited with moderate predictive ability, are multi-factorial, largely center on dementia-level risk factors, and only predict delirium incidence. The spectrum of cognitive decline, including specific domains such as executive function, was not evaluated. Lastly, the review did not identify a prediction model for postoperative delirium that incorporated predisposing (preoperative) risk factors with the anticipated precipitating event (surgery). These identified limitations are addressed in the first study of the present dissertation, represented in manuscript two.

A two-factor prediction model for postoperative delirium incidence and severity was developed and is detailed in the second manuscript of the present dissertation. This parsimonious model is composed of the surgical risk score for serious complications from the American College of Surgeons, National Quality Surgical Improvement Program and the Trail Making Test

B, a measure of executive function. Logistic and linear regression using the LASSO and Best Subsets techniques identified these models in a cohort of 97 older adults (≥65yo) undergoing non-cardiac surgery with an estimated length of stay of ≥2 days or greater.

Elucidating distinct risk and vulnerability in older adults without prior impairment is important as delirium may adversely impact their postoperative outcomes. As delirium is bereft of a neuroimaging biomarker, elucidation of regions of cortical atrophy that are associated with high-risk factors, or vulnerabilities to, postoperative delirium may represent a potential biomarker for future studies. Recently, a limitation in cortical thickness measurement, a method used to identify cortical atrophy, was identified. To address this limitation, the third manuscript investigated statistical differences between two methods of cortical thickness measurement and examined whether the statistical results were replicable. The use of multimodal imaging demonstrated higher accuracy and sensitivity in the derivation of cortical thickness measurements within a cohort of 54 older adults (mean age 71, ±SD4.83).

The present dissertation informs future studies. Future studies should validate the developed delirium prediction model in a broad, multisite, prospective perioperative cohort study to confirm predictive ability. Further, future neuroimaging studies should confirm the finding that statistical results between T1-only and T1+T2-FLAIR derived cortical thickness measures are not replicable.

Acknowledgments

I would like to express my deep gratitude to everyone who has nurtured me during my PhD journey as I evolved as a person and a scientist. I am very humbled by the support, encouragement, and mentorship I have received.

I would like to thank all of my committee members for their support; Lisa Bratzke, Tonya Roberts, Roger Brown, Robert D. Sanders, and Kirk Hogan. Each person has had a significant influence on my development as a scientist and I am truly grateful for their mentorship. Truly appreciated!!

I would like to thank all of the participants that consented to be a part of this research study. I am deeply indebted to advance the science of delirium prevention because of them. They were willing to commit a significant amount of their time to help me answer my research questions, sharing their personal details, stories, and experiences. My thinking and perspectives on delirium advanced because of the stories they shared with me. Their thoughts on their experience with delirium will continually inform my future research.

I would like to thank all of the staff at University of Wisconsin Hospital and Clinics that made this research possible. Their support was integral to the success of this research.

I would like to thank all of my friends and colleagues that supported me along this journey, offering bits of sanity and advice whenever I was in need.

A special thank you to my dear friend, Maria Huckleberry, who has been a tremendous support and made this dissertation document shine.

My family has been a tremendous support throughout this entire process and I would not be the person I am today without them. A huge thank you to my family!

And most of all, to the person who planted the seed in my brain that I could get a PhD and has been my anchor throughout this entire journey, I am forever thankful for his love, support, encouragement, advice, and laughter ~ Thank you to my sweetie, Dave Dwyer.

Introduction

Background and Significance

Delirium, an acute brain failure, is a sudden, fluctuating disturbance in consciousness and cognition often precipitated by an acute event such as surgery. The hallmark features of delirium include an abrupt change from an individual's normal state of being, inattention then either a perturbed level of awareness and/or disorganized thinking (American Psychiatric Association, 2013). Additional impairments in memory, perception, and sleep are frequently observed during delirium. All patient care populations experience delirium. The incidence of delirium varies between these populations with at least 29% of older adult medical patients, 50% of surgical patients, and up to 80% of mechanically ventilated critical care patients experience delirium during their hospital stay (S. K. Inouye, Westendorp, & Saczynski, 2014). The substantial morbidity and mortality experienced by those that had delirium reflects a crucial and unresolved public health burden.

Postoperative delirium leads to increased risk of adverse events, thereby creating an unsafe environment for patients (Gleason, Schmitt, Kosar, Tabloski, Saczynski, Robinson, Cooper, Rogers, *et al*., 2015). The likelihood of a fall is increased by 3.5-fold during postoperative delirium (Mangusan, Hooper, Denslow, & Travis, 2015). Further, postoperative delirious patients experience significantly more adverse events leading to pulmonary, urological, and neurologic complications (Raats *et al*., 2015). Hospital length of stay is extended on average 2-8 days due to postoperative delirium and 27-33% of these patients will be discharged to an institutional facility instead of returning home (Fimognari *et al*., 2016; Gleason, Schmitt, Kosar, Tabloski, Saczynski, Robinson, Cooper, Rogers Jr, *et al*., 2015; Large *et al*., 2013; Oh, Fong, Hshieh, & Inouye, 2017; Robinson et al., 2009). Complications prolonged length of stay and

disability together contribute to a substantial monetary burden; over $8 U.S. billion per year in hospital costs and $150 U.S. billion in post-hospital costs are attributable to delirium (Leslie & Inouye, 2011; Leslie, Marcantonio, Zhang, Leo-Summers, & Inouye, 2008).

Post-hospitalization morbidity burden is significant with an inability to return to baseline function. Distressing subjective experiences related to the delirious episode are reported along with an increased risk of post-traumatic stress disorder symptoms following hospitalization (Drews *et al*., 2015; Jackson *et al*., 2014; O'Malley, Leonard, Meagher, & O'Keeffe, 2008; Partridge, Martin, Harari, & Dhesi, 2013). Multiple longitudinal studies demonstrate cognitive decline when compared to non-delirious cohorts 1 to 3 years following a delirious episode. Recently, Inouye *et al*., (2016) reported that postoperative delirium changes the cognitive trajectory, resulting in a steeper decline at 3-years when compared to a non-delirious, aging cohort (Sharon K. Inouye *et al*., 2016). Another study found that 1-year following delirium, 34% of individuals demonstrated cognition similar to persons with moderate traumatic brain injury, and 24% had cognitive deficits to those observed in mild Alzheimer's (ALZ) disease (Pandharipande *et al*., 2013). This study also reported executive function scores were below the population norm at both the 3-month and 12- month mark, despite the patient's age. Executive function as a cognitive process is of particular importance to patient outcomes because it allows an individual to continue to function in everyday, normal life (Diamond, 2013). This cognitive process facilities abilities such as reasoning, planning, and problem solving and without this ability, individuals may have difficulty managing their day- to-day life especially those with comorbidities that require self-management, medication management, and sustained functional ability. Prior studies have largely assessed global cognitive function and have not examined how specific cognitive domains are risk factors for, or are impacted by delirium. Assessment of

executive function would aid in identification of patients at particular risk for declines in critical functional outcomes post-discharge.

Delirium has been significantly associated with an increased risk of postoperative mortality (Abelha *et al*., 2013; Ha *et al*., 2018; Raats *et al*., 2015; Veiga, LuÃ-s, Parente, & Abelha, 2013). A recent study noted a 7.35-fold (95%CI: 1.49-36.18) increase in the odds of five-year mortality for those that experienced postoperative delirium following an elective surgery (Moskowitz *et al*., 2017). Postoperative delirium contributes to a significant burden of in-hospital complications and post-hospitalization morbidity and mortality, accordingly, it is imperative to identify those at highest-risk of developing this dangerous syndrome and to work towards prevention.

Vulnerability

The present dissertation defines "vulnerability" as a dynamic process between present risk factors (predisposing) and a set of precipitating factors (surgery, infection, trauma) that, when combined, place the individual "at risk" for delirium. (Nightingale & Fischhoff, 2002) Biologic, cognitive, and social context may be predisposing characteristics to vulnerability. Vulnerabilities are counteracted by biological, personal, and social factors such as coping mechanisms, social support, and resilience. Multicomponent delirium preventative measures currently focus on- counteracting vulnerabilities by promoting early mobility, stimulating the sensory system, maintaining a sleep/wake cycle, hydration and cognitive interaction/re-orientation.

Prevention is paramount

Delirium is preventable. A recent meta-analysis reported that 1 in 3 cases of delirium are preventable when non-pharmacological delirium prevention measures are applied as a bundle in

medical and surgical patients (Hshieh *et al*., 2015; Siddiqi *et al*., 2016). Nonetheless, clinicians struggle with adherence as the bundled intervention programs are time-consuming and it is difficult to identify the patients that would benefit the most from prevention efforts (A. Yevchak *et al.*, 2017; A. M. Yevchak *et al*., 2014; Zaubler *et al*., 2013). Identifying patients who are most vulnerable to delirium upon admission is the first step to implementing non-pharmacological delirium prevention specific to that individual's risk profile. Nurse clinicians are optimally positioned to address delirium risk factors in vulnerable, *i.e*., high risk, individuals allowing for an upstream approach, preventing future decline and cost.

Due to the significant public health burden of delirium, it is crucial to proactively identify risk factors, *i.e*., increased vulnerability, prior to delirium development and implement preventative measures. At present, an extensive list of non-modifiable and modifiable, predisposing and precipitating delirium risk factors constrains clinicians (Table 1), hampering the ability to select the most important, contributing factor (S. K. Inouye *et al*., 2014; Lindroth *et al.*, 2018; Sanders, Pandharipande, Davidson, Ma, & Maze, 2011).

Prediction

The science of prediction models is growing in evidenced-based healthcare due to their improved ability to facilitate early identification of high-risk individuals (Adams & Leveson, 2012; Reilly & Evans, 2006; Ewout W. Steyerberg, 2009; Vickers & Cronin, 2010). Prediction models are statistical models that aid in forecasting which individuals have a higher likelihood of developing a disease (Ewout W. Steyerberg, 2009; E. W. Steyerberg & Vergouwe, 2014).

Table 1. This displays the list of predisposing and precipitating factors that clinicians typically use in practice.

Predisposing factors	Precipitating factors	Delirium-inducing medications
Comorbidities	Acute insults	High risk
Alcoholism	Dehydration	Anticholinergics
Chronic pain	Fracture	Benzodiazepines
History of baseline lung, liver, kidney, heart or brain disease	Hypoxia	Dopamine agonists
Terminal illness	Infection	Meperidine (Demerol)
Demographic factors	Ischemia	Moderate to low risk
Age >65 yrs	Medications	Antibiotics
Male	Metabolic drangement	Anticonvulsants
Geriatric syndromes	Poor nutrition	Antihypertensives
Dementia	Shock	Corticosteroids
Depression	Surgery	Low-potency antihistamines
Falls	Uncontrolled pain	Metoclopramide
History of delirium	Urinary or stool retention	NSAIDs
Malnutrition	Environmental exposures	Tricyclic antidepressants
Polypharmacy	Intensive care unit setting	Sedatives
Pressure ulcers	Sleep deprivation	
Sensory impairment		
Premorbid state		
Inactivity		
Poor functional status		
Social isolation		

These models can take the form of risk stratification tools, decision-trees, as well as algorithms and are referred to in the present dissertation as delirium prediction models. **A precise and timely prediction model would establish the highest impact risk factors into a powerful tool, facilitating early implementation of prevention measures (Lindroth *et al*., 2018).**

The first manuscript of this dissertation, a systematic review of literature completed on delirium prediction models (DPMs) for older hospitalized adults (≥60yo), identified 14 externally validated DPMs that reported moderate predictive ability. Several limitations were identified and are described further in the first manuscript of the present dissertation. The review identified that variables found to be predictive of postoperative delirium such as tobacco use, a surgical risk score, and measures of executive function, were not used in the included delirium prediction models. Further, the prediction of delirium symptoms and duration has not been explored. The present dissertation aims to address these gaps by building a preliminary delirium prediction model using the aforementioned risk factors and examining their use in predicting delirium incidence, severity, and duration.

Gaps in the Literature

Current delirium prediction is limited by the lack of high performing, generalizable models that are clinically applicable. Multi-factor models are cumbersome for clinicians to apply limit the clinical utility of prediction. To optimize the prediction of delirium, a prediction model should be parsimonious, clinically feasible, and ideally identify vulnerability, *i.e.,* risk

factors amendable to interventions before delirium develops. The incorporation of a precipitating factor that provides information on the magnitude of the future insult, leading to the identification of those at high-risk for incident delirium as well as identifying risk modalities that are amenable to interventions may be ideal. The ability to anticipate and address delirium symptom burden would also inform delirium prevention. Delirium symptom severity and duration have been linked to increasing levels of morbidity and mortality (Gunther *et al.,* 2012; Ha *et al.,* 2018; K.-H. Lee, Ha, Lee, Kang, & Koo, 2011; Morandi *et al.,* 2012; Racine, Fong, Gou, *et al.,* 2017; Vasunilashorn *et al.,* 2016). Yet there is a lack of research on which vulnerabilities are predictive of delirium symptom severity and duration.

The application of a surgical risk score has not been explored in current delirium prediction models. This has the potential to improve postoperative delirium prediction by estimating the magnitude of insult from the planned surgical procedure. Albeit exclusive to surgical patients, a surgical risk score allows the incorporation of several premorbid factors and the planned surgery into one score, thus achieving the combination of both premorbid and future precipitating factors and simplifying use statistically and practically. Prior delirium prediction models in older adults (≥60yo) have used scores such as the American Society of Anesthesiologists (ASA) score (J. L. Rudolph *et al.,* 2009), the Acute Physiology and Chronic Health Evaluation II (APACHE II) score (Kalisvaart *et al.,* 2006), and the Systemic Inflammatory Response Score (SIRS).(Sarah T. Pendlebury *et al.,* 2016; Pendlebury, Lovett, Smith, Wharton, & Rothwell, 2016) While these do provide an estimation on the severity of illness, the ASA score does not consider the pending surgical procedure, the APACHE II score use physiological information gathered during the 1st 24-hours of admission when delirium may already be present and the SIRS more applicable to medical patients as it measures physiologic

response to an already present stimulus. A surgical risk score that incorporates baseline characteristics with the planned procedure may be a more accurate predictor of risk. The American College of Surgeons developed an online risk calculator from their National Surgical Quality Improvement Program (NSQIP) database (Bilimoria *et al.,* 2013; Guillamondegui *et al.,* 2012; Shiloach *et al.,* 2010). This online tool uses age, gender, functional status, several vascular comorbidities including smoking, BMI and the planned surgical procedure to estimate the potential risk for serious complications, readmission, and mortality among several other scores for each individual. Taking less than 3 minutes to complete, it has been widely validated in several surgical populations including orthopedic, abdominal, cardiac, and vascular surgeries (Bilimoria *et al.,* 2013; Bohnen *et al.,* 2017; Guillamondegui *et al.,* 2012; Helman *et al.,* 2017; Mogal *et al.,* 2017; Shiloach *et al.,* 2010). As previous studies have shown age and vascular burden to both significantly associated with impaired cognitive function as well as an increased risk for delirium incidence (Galyfos, Geropapas, Sianou, Sigala, & Filis, 2017; Oldroyd *et al.,* 2017; Scholz, Oldroyd, McCarthy, Quinn, & Hewitt, 2016; Sheline *et al.,* 2006; Williams- Russo, Sharrock, Mattis, Szatrowski, & Charlson, 1995), a risk score that combines these factors into one score is important. Nonetheless, it has not been used in delirium prediction. This dissertation proposes to examine how surgical risk predicts delirium incidence, severity and duration.

Tobacco use history has shown to be predictive of delirium incidence (Hessler *et al.,* 2015; Hsieh, Shum, Lee, Hasselmark, & Gong, 2013; James L. Rudolph et al., 2007). The NSQIP risk calculator uses “Current smoker within 1-year”, however, does not incorporate pack years or past smoking history. As shown by the recent systematic review of delirium prediction models, tobacco use, past, current or pack years, are not used in current models. Including total pack years and history of past tobacco use may contribute important information to a prediction model and these variables will be included as covariates in these analyses. Depression has been

identified as a significant risk factor for delirium incidence in older adults (Koskderelioglu *et al.,* 2017; O'Sullivan, Inouye, & Meagher, 2014; K. S. Radinovic *et al.,* 2014; Smith, Attix, Weldon, & Monk, 2016). Psychological symptom burden has been associated with increased delirium incidence and duration (O'Sullivan *et al.,* 2014). Further, depression has demonstrated a significant association with executive function (Sheline *et al.,* 2006; Yoon, Shin, & Han, 2017). Depression often coexists with cognitive decline as depressed individuals show deficits in processing speed, selective attention and response inhibition; several key features of executive function (O'Sullivan *et al.,* 2014; Sheline *et al.,* 2006; Steffens & Potter, 2008; Yoon *et al.,* 2017). Depression may be a confounder with impaired cognition and executive function, and is included in these analyses.

The current DPMs are limited by the paucity of distinct premorbid cognitive domain measurements. Pre-existing global cognitive impairment is a known risk factor for delirium development and is the variable most frequently used in DPMs (Davis *et al.,* 2015; Hshieh, Inouye, & Oh, 2018; S. K. Inouye *et al.,* 2014; Jones *et al.,* 2016; Lindroth *et al.,* 2018; Oh *et al.,* 2017). These models employ measures that capture cognitive impairment at the level of dementia, largely overlooking discrete changes in cognition. The risk for delirium steadily increases as an individual's cognitive ability declines, even for those whom are considered unimpaired and do not meet diagnostic criteria for mild cognitive impairment or dementia. Jones et al. (2016) demonstrated this effect as each half standard deviation decline in an individual's Global Cognitive Performance score, resulted in a two-fold increase in the risk of delirium incidence following surgery, regardless of age, gender, estimated IQ and existing comorbidities (Jones *et al.,* 2016). Even though mild and moderate cognitive impairment (Franco *et al.,* 2010; Kazmierski *et al.,* 2014; Racine, Fong, Gou, *et al.,* 2017; Veliz- Reissmuller, Aguero Torres, van

der Linden, Lindblom, & Eriksdotter Jonhagen, 2007) are both established risk factors for delirium development, the impact of decline in specific cognitive domains (*e.g.*, executive function) has not been well established. Given the lack of high-performing, generalizable models that are clinically relevant and inclusive to the spectrum of cognition, future models should explore incorporating distinct measures of cognition, such as executive function. Selecting a cognitive measure that is sensitive and specific to gradual decline as well as feasible to administer in a clinical setting may improve delirium prediction.

Current DPMs lack a measure of premorbid executive function. Inhibitory control, working memory, and cognitive flexibility are core domains of executive functions (Diamond, 2013). These domains are particularly sensitive, or vulnerable, to adverse conditions such as increased stress, depression, sleep deprivation, physical inactivity, and loneliness and as such, are the first to show signs of dysfunction under such circumstances (Brandt *et al.,* 2009; Diamond, 2013; Glisky, 2007; Mansouri, Tanaka, & Buckley, 2009; Sheline *et al.,* 2006; Smith, Attix, Weldon, Greene, & Monk, 2009; Yoon *et al.,* 2017). Perturbed executive functions results in inattention, disinhibition, inability to self-control or correct, and impaired logic, reasoning and problem solving. Each of these disturbances is a key symptom of delirium. Further, as delirium may result from an acute, stressful situation resulting in increased sleep deprivation and physical inactivity, it is evident that executive dysfunction and delirium are intricately interconnected. Preoperative executive function has demonstrated a significant association with postoperative delirium incidence in several studies supporting this proposed relationship (Greene *et al.,* 2009; James L. Rudolph *et al.,* 2006; Smith *et al.,* 2009). The incorporation of a valid, and reliable, executive function test into a delirium prediction model may provide a clinically feasible marker of vulnerability prior to surgery.

Investigating patterns of cortical atrophy and their association with preoperative executive function may elucidate potential areas of vulnerability within the brain, predisposing an individual to delirium. As hypothesized by the *Cognitive Disintegration Model*, an individual with higher vulnerability prior to the precipitating insult, *i.e.,* a high-risk individual, will require less of a stimulus to cross over the "delirium threshold" (Sanders, 2011). Examining the biological underpinnings of executive function in the brain may elucidate a biomarker of vulnerability for delirium. It is not known how vulnerability to delirium is reflected within the brain, specifically in patterns of cortical atrophy. Patterns of cortical atrophy have been identified as a sensitive and accurate biological marker to indicate conversion from MCI to Alzheimer's disease (AD), this pattern is known as the AD Signature (Bakkour, Morris, & Dickerson, 2009; M. Zhou, Zhang, Zhao, Qian, & Dong, 2016; Q. Zhou et al., 2014; Y. Zhou & Lui, 2013). Regions within the AD Signature are shown to significantly correlate with postoperative delirium severity (Racine, Fong, Travison, *et al.,* 2017). Nonetheless, it is not known if other regions of atrophy are significantly associated with postoperative delirium and how vulnerability is represented within those regions of atrophy. Individuals that share similar patterns of atrophy may have a distinct set of clinical characteristics that could serve as a proxy for their cortical atrophy signature, *i.e.*, biomarker of vulnerability. Most cortical thickness research to-date has focused on variable-oriented analyses, examining the relationships between variables at aggregate levels. This type of analysis leads to an understanding of associations between specific clinical variables and areas of atrophy. As an example, thinner cortex in the frontoparietal regions has correlated with poor performance on a number of different executive function measures (MacPherson *et al.,* 2017; Nowrangi, Lyketsos, Rao, & Munro, 2014; Yuan & Raz, 2014). These findings are scientifically informative, yet are specific to the type of executive

function test used and covariates applied within the statistical analysis. As delirium is a heterogeneous syndrome with multiple risk factors, a gestalt analysis that facilitates the examination for unknown risk factors is warranted. This alternate form of analysis is referred to as person-oriented because it investigates how unknown, or latent, subgroups are represented within a population then examines the descriptive (biologic, social, and cognitive) differences within and between subgroups (Bergman, Magnusson, & El Khouri, 2002). It has not been widely employed in cortical thickness research and it is suggested as a method to apply as this approach provides an opportunity to discern new, or unknown clusters of variables within-like individuals, and facilitates the creation of biologic biomarkers (Malpas, 2016). Building knowledge on the biological underpinnings of a particular risk factor, *i.e.,* developing a biomarker of vulnerability to delirium, may inform future therapeutic interventions. The biomarker could then be identified by clinical characteristics instead of an MRI image as it is not clinically or financially feasible to image every older adult. A recently identified limitation in cortical thickness measurement is the decreased sensitivity of anatomical T1-weighted images to measure the signal intensity in aging gray and white matter (Salat *et al.,* 2009). This is an important limitation that may impact the ability for cortical thickness measures to identify accurate sub-groups of individuals. Given that cortical thickness measures may be improved with the addition of a second anatomical scan, this present dissertation aimed to investigate the statistical differences between cortical thickness measures derived from T1-only and T1+T2-FLAIR anatomical images and examine whether T1+T2-FLAIR data replicates the findings of T1- only. Further, the present dissertation applies the statistical method latent profile analysis to identify sub-groups of cortical atrophy then examines risk factors including performance on measures of executive function, vascular risk, and global cognition in relation to the identified

sub-groups to develop a preliminary biomarker of vulnerability to future delirium. Future work will build on these findings by using the preliminary biomarker to examine risk factors within groups and measuring each sub-group's association with delirium incidence, severity, and duration.

Conceptual Framework

The *Cognitive Disintegration Model (Sanders, 2011)* provides the conceptual framework for this investigation because it considers an individual's baseline status/vulnerability and the nature/caliber of the precipitating insult. Two factors are suggested that determine an individual's vulnerability to delirium. The first is their baseline neural network connectivity. This represents the integrity of their existing neural network, facilitating communication between various neurons and brain regions and is a reflection of their age, chronic illness and level of cognition among other factors such as depression and dementia. The second factor is their level of inhibitory tone in their brain. Neurons within the brain receive constant input from GABAergic neurons, setting an inhibitory tone. Precipitating events such as surgery, inflammation and medications may increase the level of inhibitory tone within the brain, impacting and altering connectivity between different brain regions. Individuals with higher levels of baseline network connectivity may require a greater change and therefore an increase in the level of inhibitory tone to precipitate exceeding the "delirium threshold". In summary, the *Cognitive Disintegration Model (Sanders, 2011)* focuses on neural network connectivity and brain inhibitory tone as potential etiological factors related to delirium. The aim of this project is not to test or explore potential etiological factors of delirium, rather, this dissertation aims to complement the model by investigating preoperative vulnerabilities, *i.e.,* risk factors, *i.e.,* risk factors, which may contribute to delirium incidence, severity, and duration. This investigation

epitomizes a holistic nursing perspective of delirium and considers a patient's clinical picture.

Study Design

This proposed study enrolling 100 older adults (≥ 65yo) having elective non-cardiac surgery investigates 1) the ability of an executive function score and a surgical risk score to predict postoperative delirium incidence, severity, and duration, and 2) the development of a preliminary phenotype, *i.e.,* biomarker, of vulnerability to future delirium risk and cognitive decline.

Specific Aim 1: To examine the ability of executive function and surgical risk to predict postoperative delirium incidence, severity, and duration.

Specific Aim 2: To examine the statistical differences in cortical thickness measurements between two neuroimaging methods and investigate the replicability of findings. Further, to use latent profile analysis to identify subgroups of cortical atrophy then examine cognitive and vascular risk factors within those subgroups.

This is a perioperative, prospective cohort study. Preoperative and postoperative measures were administered to older adults (≥65yo) whom underwent non-cardiac surgery at University Hospital. The data for this study were collected in parallel with an ongoing parent study (Sanders, R.D., PI) at University Hospital. The primary aim of the parent study is to evaluate the Role of Preoperative Cingulate Cortex Functional Connectivity (FC) as a Predictor of the Risk of Postoperative Delirium and is funded by the Department of Anesthesiology and a K23 NIH grant. This study has IRB approval (2015- 0374).

Setting

Participants were recruited from six surgical specialty clinics at University Hospital in the areas of vascular, spine, general surgery, urology and ear, nose, and throat. Participants underwent

procedures at University Hospital and then were followed postoperatively as an inpatient in that facility. Permission and full written support was granted from the participating surgeons, clinics, unit managers and the Department of Nursing at University Hospital.

Sample

100 participants were recruited following the inclusion and exclusion criteria in Table 2.

Table 2

Inclusion Criteria	**Exclusion Criteria**
65 years of age or older	Previous documented history of dementia, CAM+
Undergoing non-cardiac surgery and receiving general or regional anesthetic	Unable to communicate with research staff to complete neuropsychological testing (language barrier, significant impairments in hearing/vision)
Approximate length of stay-≥2 days	Contraindication to MRI

Recruitment

Participating surgeons and clinic staff identified potentially eligible participants and introduced the study during the pre-operative consultation visit. If interested, the study team obtained full informed written consent following a thorough explanation of the study premise and procedures. Participant confidentiality was maintained by ensuring a private room was used for obtaining full informed consent and data collection. The preoperative study visit included the following measures; consent (15-min.), neuropsychological battery (1-hour), and a MRI brain scan (1-hour).

Methods

Timing of Measures

I. **Preoperative (preop) Data Collection.** Demographics, clinical variables, MRI images.

   A. Clinical variables: Preoperative cognitive function with Trail Making Test A&B (TMT-A, TMT-B), the Repeatable Battery for the Assessment of

Neuropsychological status (RBANS), the Controlled Oral Word Association Test (COWAT), the Boston Naming Test (BNT), geriatric depression score (GDS-15), tobacco use history, past medical history for surgical risk calculation, years of education, and delirium (CAM).

B. MIR Images: High-resolution MRI images were captured using a 1.5T or 3T scanner Eachscan took approximately 45 minutes, as the parent study obtains additional sequences. Head motion was restricted using a pillow and foam, and earplugs used to attenuate scanner noise.

II. **Postoperative Data Collection**. Delirium incidence, severity, and duration (CAM, CAM-ICU, DRS-98-R). Participants were assessed for delirium on post-operative (POD) 1-4 between the hours of 0500-1000 and 1600-2200. Each assessment time point was 0.5 days (12-hours).

Description of Measures

a) **Demographics and clinical variables (preop):** Age, gender, educational level, tobacco use, comorbidities, and surgical procedures were collected from interview and chart review.

b) **Anatomical MRI images (preop)**: T1-weighted MPRAGE (TR: 2,530ms, TE: 3.09ms, Flip Angle: 10˚, 256 x 256 matrix, 208 coronal slices, 1 mm isotropic resolution) and T2-fluid attenuated inversion recovery (FLAIR, T2*-weighted gradient-echo echo planar imaging pulse sequence sensitive to BOLD contrast: field of view 224mm, matrix 64 x 64, TR: 2600ms, TE: 22ms, Flip Angle 60˚, 40 axial plane slices of 3.5mm thickness with 3.5mm spacing between slices) images were obtained on 1.5-telsa and 3-telsa General Electric scanners using an 8-channel head coil at the University of Wisconsin Hospital and the Wisconsin Institute for Medical Research center.

c) **Depression (preop):** The GDS-15 evaluated depression with 15 yes/no questions. A score >10 indicates the presence of depression. Validity and reliability are supported in clinical and research.

d) **Executive Function (preop):** TMT-A & TMT-B measured executive function, it is widely validated and tests attention and task-switching ability. Participants are scored through time completion at their ability to connect 25 circled numbers (TMT-A) and 23 circled numbers and letters (TMT-B) in sequential order.

e) **Neuropsychological Battery (preop):** The RBANS assessed global cognitive function in a 30-minute comprehensive cognitive instrument testing the domains of immediate memory, visuospatial/constructional, language, attention, and delayed memory. This is well-validated in the older adult population and scores are adjusted for age (Randolph, Tierney, Mohr, & Chase, 1998; Strauss, 2006). BNT assessed object naming ability and COWAT tested verbal fluency. COWAT is adjusted for age and education (Strauss, 2006).

f) **Surgical Risk (precipitating):** The copyrighted NSQIP Surgical Risk Calculator incorporates an individual's baseline demographics (age, gender) and comorbidities (functional status, emergency case, ASA class, steroid use for chronic condition, ascites within 30 days prior to surgery, systemic sepsis within 48-hours prior to surgery, ventilator dependent, disseminated cancer, diabetes, hypertension requiring medication, congestive heart failure in 30 days prior to surgery, dyspnea, current smoker within 1 year, history of severe COPD, dialysis, acute renal failure and BMI) along with their procedure into an estimated risk score. The risk of "serious complications" (NSQIP-SC) and "risk of death" (NSQIP-D) were used in this study.

g) **Delirium, Outcome Measure 1**: Confusion Assessment Method (CAM) measured the outcome variable, delirium. Inouye et al. (1990) developed this detection tool based on the definition in the Diagnostic and Statistical Manual of Mental Disorders, third edition (S. K. Inouye, Viscoli, Horwitz, Hurst, & Tinetti, 1993). Well validated in various adult populations with a sensitivity and specificity greater than 90%, the CAM is a four-step algorithm utilizing clinician observation and patient interview. The CAM-ICU was used when appropriate (patient is intubated, unable to verbalize due to tracheostomy or respiratory support). The CAM-ICU is derived from the CAM and uses the same 4 features and algorithm to diagnose delirium (Ely, Inouye, *et al.,* 2001; Ely, Margolin, *et al.,* 2001). Any delirium occurring on postoperative days 1-4 was used in the analysis.

h) **Delirium Outcome Measure 2**: Delirium symptoms and symptom severity were concurrently assessed with the Delirium Rating Scale-98-R (DRS-98-R), which is a widely validated and reliable 16-item scale assessment tool. Thirteen-items are specific to assessing the symptom severity and 3-items are specific to diagnostic criteria including the temporal nature and fluctuation of delirium. (Trzepacz *et al.,* 2001) The max score is 44-points.

i) **Delirium Outcome Measure 3**: If delirium persisted on POD4-PM assessment, the participant was followed until delirium resolved to provide information on *delirium duration*.

Data Analysis

Analysis was conducted using FreeSurfer Mplus, R, NCSS, and STATA. Statistical significance was $p \leq 0.05$. Patient characteristics were described using means $\pm$ standard deviations for continuous variables and frequency counts with percentages for categorical variables. Dependent on the distribution of the data confirmed by statistical plots, continuous variables were compared using Student's t-test or Mann-Whitney U-test. Categorical variables were

compared using *X2*. Relative risks and 95% confidence intervals were calculated and adjusted using standard errors and parameter assessments. Missing data was assessed through Little's test for MCAR (missing completely at random). Contingent upon either MCAR missing data, imputation was considered to respond to the level of missingness and to be able to use the optimal amount of information for analysis.

Specific Aim 1: **To examine the ability of executive function and surgical risk to predict postoperative delirium incidence, severity and duration.**

First, to evaluate the predictive ability of NSQIP-SC over NSQIP-D, logistic (DELYN) regression model were completed. ASA classification and Framingham risk were also evaluated in this manner. Second, a delirium prediction model was developed. We did not employ univariate statistics to select candidate predictors as this may lead to poor performing predictors and overfitting (Ogundimu, Altman, & Collins, 2016). To counter the effects of small sample sizes and reduce bias within data, we employed a statistical shrinkage regression technique, using Least Absolute Shrinkage and Selection Operator (LASSO) (Lockhart, Taylor, Tibshirani, & Tibshirani, 2014; Tibshirani, 1997; Tibshirani *et al.,* 2012). This technique reduces the noise within the data, allowing true signals to be detected and avoids common problems such as model overfitting. Candidate variables demonstrating the smallest Mallow's C*p* value (Mallows, 1973), indicating precise predictors, were then applied in Best Subsets regression. Best Subsets regression is an automated regression approach that evaluates all possible combinations of candidate predictors (Hosmer, Jovanovic, & Lemeshow, 1989; King, 2003). The output provides a set models with model fit statistics. Model selection was based on assessment of model fit using Akaike information criteria (AIC), Bayesian information criteria (BIC), and McKelvey and Zavoina's R2 (E. W. Steyerberg *et al.,* 2010). The area under the

receiver operating characteristic curve (AUROC) with 95% CI was calculated. Calibration was assessed through goodness-of-fit tests calculated by the Hosmer-Lemeshow statistic. Sensitivity, specificity, positive predictive and negative predictive values were calculated and reported. The peak delirium severity (DRS) score was transformed using the Box-Cox Method (Arzideh, Wosniok, & Haeckel, 2011) with the optimal Lambda value due to the positive skew, please refer to supplementary data for raw and transformed plots. The regression modeling procedures outlined in the paragraph above for logistic regression were repeated for the linear regression model. Model selection was based on assessment of model fit using Akaike information criteria (AIC), Bayesian information criteria (BIC), and adjusted R2 (E. W. Steyerberg *et al.,* 2010).

Specific Aim 2: To examine the statistical differences in cortical thickness measurementsbetween two neuroimaging methods and investigate the replicability of findings. Further, to use latent profile analysis to identify subgroups of cortical atrophy then examine cognitive and vascular risk factors within those subgroups.

MRI Image Processing. Cortical thickness (mm) measures were obtained using a publicly available software package, FreeSurfer v6.0 (https://surfer.nmr.mgh.harvard.edu/). First T1-weighted MPRAGE images were processed using the well-documented recon-all processing stream which includes motion correction, skull-stripping, registration, segmentation, smoothing, and parcellation mapping. Coronal slices were manually inspected and ranked on quality level by two independent observers. The quality rankings were as follows: 0) 0-1 segmentation error; 1) $\leq$2 segmentation errors; 2) $\leq$5 segmentation errors; 3) moderate segmentation error with dura overlap; 4) poor quality, discard. Seven scans were discarded due to motion artifact, poor segmentation, or anatomical abnormalities. All scans were then re-processed using T1+T2-

FLAIR multimodal recon-all processing stream. Cortical parcellation statistics were extracted using the Desikan- Killiany Atlas (DKT), which contains 68 regions, 34 per hemisphere. Means, standard deviations, and data distribution were then evaluated. Following re-processing, manual segmentation was completed by two independent observers and scans were re-ranked on their level of quality Statistical analysis was performed to evaluate the differences between the T1-weighted recon-all images versus the T1-weighted/T2-FLAIR recon-all images (n=54).

Cortical thickness measures were adjusted for scanner-type and intracranial volume (ICV). The ComBat harmonization correction tool was applied to account for inter-site scanner variation (Fortin *et al.,* 2018). This tool performs well in small sample sizes and is able to maintain biologic variability while correcting for inter-site scanner variation. Latent class analyses were performed using the mixture-modeling package in M*plus*, v8.1. Each hemisphere (DKT atlas, 34 regions per hemisphere) was loaded into M*plus* as classification variables (Desikan *et al.,* 2006). The presence and dependency of a latent variable is assumed by LCA and the latent variable signifies the unobserved sub-classification of the data. Model parameters were estimated with an expectation- maximization algorithm. The optimal number of classes was selected based on the following selection criteria: 1) interpretability; 2) theoretical justification; 3) parsimony; 4) Entropy >0.75; 5) lowest adjusted Bayesian Information Criteria (BICa) score; 7) average posterior probability in each class >0.75 and no more than 10% overlap between non-contiguous clusters; 8) at least 2.5% of total count in each group; and 9) no significant improvement as assessed by likelihood ratio test (Lo-Mendall-Rubin). Univariate entropy signifying the quality of a latent class qualifier will be reported for the selected classes (Muthen, 2014). Auxiliary analysis was performed to evaluate statistically significant differences between groups. False Discovery Rate (FDR) and effect sizes with 95% confidence intervals were

calculated for the auxiliary variables. To examine if class assignment predicted regions of cortical atrophy, FreeSurfers gui QDEC was used. Posterior probabilities generated by MPlus for LCA class assignment per participant were loaded into QDEC. Both unadjusted and scanner-adjusted (scanner type as a nuisance factor) models were evaluated.

C. Limitations

1. Small sample size leading to potential overfitting of prediction model and latent profile models. Recommendations to minimize the creation of an overly-optimistic model will be followed including statistical shrinkage of candidate predictor variables, the use of penalized or Lasso regression and internal validation procedures such as bootstrapping.
2. The results of this may not be generalizable to different surgical populations and is not generalizable to medical populations.

Introduction to Three Manuscripts

The following three manuscripts were produced from this dissertation work.

1. Manuscript 1: Systematic review of prediction models for delirium in the older adult inpatient
    a. This manuscript details delirium prediction models in older hospitalized adults (≥60yo) In both medical and surgical populations. BMJ Open, PMID: 29705752.
2. Manuscript 2: The role of a surgical risk calculator and executive function in predicting postoperative delirium incidence and severity.
    a. This manuscript examines the ability of a surgical risk calculator (NSQIP-SC and NSQIP-D) and executive function (TMTA, TMTB) to predict postoperative delirium incidence, severity and duration. Anticipated journal: The British Journal of Anaesthesia.

3. Manuscript 3: Examines the statistical differences in cortical thickness measurements between two neuroimaging methods and investigates the replicability of findings.
    a. Cortical thickness measurements derived from anatomical T1-only MRI images are compared to data from T1+T2-FLAIR MRI images. Models using both forms of data are built using latent profile analysis and compared. Anticipated journal: Frontiers.

Manuscript 1

Title: Systematic Review of Prediction Models for Delirium in the Older Adult Inpatient

Heidi Lindroth, Lisa Bratzke, Suzanne Purvis, Roger Brown, Mark Coburn, Marko Mrkobrada, Matthew TV Chan, Daniel Davis, Pratik Pandharipande, Cynthia M. Carlsson, Robert D. Sanders

University of Wisconsin School of Medicine and Public Health, Department of Anesthesiology, School of Nursing, 600 Highland Avenue, Madison, WI 53792 Heidi Lindroth BSN PhD Candidate[1]

University of Wisconsin School of Nursing, 701 Highland Avenue, Madison, WI 53705 Lisa Bratzke PhD Assistant Professor[2]

University Hospital, 600 Highland Avenue, Madison, WI 53792 Suzanne Purvis DNP, Clincial Nurse Specialist[3]

University of Wisconsin School of Nursing, 701 Highland Avenue, Madison, WI 53705 Roger L. Brown PhD, Professor[4]

University Hospital RWTH Aachen, Department of Anesthesiology, Pauwelsstrasse 30, 52074 Aachen, Mark Coburn MD, Professor[5]

Western University, Schulich School of Medicine and Dentistry, 339 Windermere

Road, Room B9-100, London ON N5X 3W6, Marko Mrkobrada MD Associate Professor[6]

The Chinese University of Hong Kong, Department of Anaesthesia and Intensive Care, 4/F Main Clinical Bock and Trauma Center, Prince of Wales Hospital, Hong Kong Special Administrative Region, China, Matthew TV Chan, MB, BS, PhD, Professor[7]

MRC Unit for Lifelong Health and Ageing at UCL, 33 Bedford Place, London, WC1B 5JU, Daniel Davis PhD MRCP, Senior Clinical Researcher[8]

Vanderbilt University Medical Center, Department of Anesthesiology, 1211 Medical Center Drive, Nashville TN 37322, Pratik Pandharipande MD, Chief, Division of Anesthesiology Critical Care Medicine[9]

University of Wisconsin School of Medicine and Public Health, Madison VA Geriatric Research Education and Clinical Center (GRECC) (D4211), 2500 Overlook Terrace, Madison, WI, 53705, Cynthia M. Carlsson, MD, MS, Associate Professor of Medicine, Division of Geriatrics[10]

University of Wisconsin School of Medicine and Public Health, Department of Anesthesiology, 600 Highland Avenue, Madison, WI 53792, Robert D. Sanders MBBS PhD FRCA, Assistant Professor[11]

[+]**Corresponding author**: *Heidi Lindroth, Department of Anesthesiology, University of Wisconsin School of Medicine and Public Health, 600 Highland Avenue, B6/319 CSC Madison, WI 53792-3272 Telephone: 608- 262-6469 Fax: 608-263-0575 Madison, USA. Email:* hlindroth@wisc.edu

Word count: 4033

Objective

To identify existing prognostic delirium prediction models and evaluate their validity and statistical methodology in the older adult (≥60yo) acute hospital population.

Design

Systematic review

Data Sources and methods

PubMed, CINAHL, PsychINFO, SocINFO, Cochrane, ISI and EMBASE were searched from 1990/1/1 to 2016/12/31. The PRISMA Statement guided protocol development. Inclusion criteria: age ≥60, inpatient, developed/validated a prognostic delirium prediction model. Exclusion criteria: alcohol-related delirium, sample size ≤ 50. The primary performance measures were calibration and discrimination statistics. Two authors independently conducted search and extracted data. First author synthesized data. Mentoring author resolved disagreement.

Eligibility Criteria

Inclusion criteria: Age ≥ 60, inpatient, developing or validating an existing prognostic delirium prediction model. Exclusion criteria: alcohol-related delirium, sample size ≤ 50. Data were extracted from published studies. The primary performance measures were calibration and discrimination statistics. Secondary measures included applied statistical methodology.

Results

The initial search resulted in 7,502 studies. Following full-text review of 192 studies, 33 were excluded based on age criteria (<60yrs) and 27 met the defined criteria. Twenty-three delirium prediction models were identified, thirteen were externally validated and three were internally validated. The following populations were represented: 11-medical, 3-medical/surgical, and 13-

surgical. The assessment of delirium was often non- systematic resulting in varied incidence. Five models demonstrated an AUROC >0.75, indicating moderate predictive ability. Limitations in design, data collection methods, and calibration statistics were identified.

Conclusions

Delirium prediction models for older adults show variable and typically inadequate predictive capabilities. Our review highlights the need for development of robust models to predict delirium in older inpatients. We provide recommendations for the development of such models.

Funding

Department of Anesthesiology, University of Wisconsin-Madison.

R01 NHLBI (HL111111) and research grant from Hospira Inc.,

NIH collaboration. All authors completed ICMJE disclosure forms

and no conflicts of interest declared.

Keywords

Delirium. Aging. Cognition. Prediction. Statistical Models.

Strengths and Limitations of this Study

- The PRISMA Statement and CHARMS checklist were used to develop the protocol for this systematic review.
- Interprofessional authorship providing different perspectives on delirium prediction models.
- Comprehensive search using multiple databases and search terms
- Limited by age (≥60yo)
- Limited to studies developing or validating predictive models, did not include predictive risk factors

INTRODUCTION

Delirium is an acute disturbance of consciousness and cognition precipitated by an acute event such as sudden illness, infection, or surgery. This syndrome is a serious public health concern, as up to 50% of hospitalised older adults will experience delirium in medical and surgical populations.[1-3] Delirium has been independently associated with increased mortality, morbidity in terms of impaired cognition and functional disability along with an estimated annual expenditure of $164 billion.[4-9] Prediction models allow clinicians to forecast which individuals are at a higher risk for the development of a particular disease process and target specific interventions at the identified risk profile.[10-13] At present, an extensive list of modifiable and non-modifiable, predisposing, and precipitating delirium risk factors encumbers clinicians, hindering the ability to select the most important or contributing risk factor.[1 14] An accurate and timely delirium prediction model would formalize the highest impact risk factors into a powerful tool, facilitating early implementation of prevention measures.[11]

This systematic review expands on previous published reviews on delirium prediction models by integrating both medical and surgical populations while examining statistical aspects of each study including reporting metrics and includes recently published models. Our aim was to provide important recommendations on study design for future delirium prediction models while integrating knowledge gained from the study of both medical and surgical populations. We conducted a systematic review of the literature focusing on the identification and subsequent validity of existing prognostic delirium prediction models in the older adult (≥60 years old) acute hospital population.

METHODS

This systematic review followed the protocol developed from the PRISMA Statement and the

CHARMS checklist (Appendix A).[15 16] A delirium prediction model was defined as a statistical model that either stratified individuals for their level of delirium risk, or assigned a risk score to an individual based on the number and/or weighted value of predetermined modifiable and non-modifiable risk factors of delirium present. This review included studies focused on 1) older adult (≥ 60 years) population, (the U.S. CDC and UN define an older adult as 60 years of age and older)[17 18], 2) inpatient hospital setting, 3) publication dates of 1990/1/1 to 2016/12/31, and 4) developed and/or validated delirium prediction models. Studies were excluded if they 1) studied a different patient population (*i.e.*, emergency department, skilled nursing facilities, palliative care, and hospice) as these are not generalizable to an inpatient hospital setting, 2) related to alcohol withdrawal, or delirium tremens, as the presence of alcohol withdrawal complicates delirium assessment, and 3) had a sample size ≤ 50 for methodological reasons (*i.e.,* underpowered). All study designs were included. Studies were not limited by timeframe of delirium development (prevalent vs incident), however, only prognostic statistics were discussed. The search terms were as follows: ("Delirium" OR "postoperative delirium" OR "ICU delirium" OR "ICU psychosis" OR "ICU syndrome" OR "acute confusional state" OR "acute brain dysfunction") AND ("inpatient" OR "hospital*" OR "postoperative" OR surg* OR "critical care unit" OR "intensive care unit" OR CCU OR ICU) AND ("predict*" model OR risk*). Electronic databases of PubMed, CINAHL, PsychINFO, Cochrane Database of Systematic Reviews, SocINDEX, ISI, and EMBASE were searched. Studies using a language other than English were included if translation was available through the University of Wisconsin-Madison Health Sciences Librarian. Bibliographies of identified studies were hand-searched for additional references. Study quality was assessed through the Newcastle-Ottawa Scale (NOS)[19] for case-control and cohort studies. Two authors (HL, SP) independently performed data collection, data extraction,

and assessed study quality, with any disagreement resolved by RDS.

Data extracted included: 1) study characteristics (study design, population, sample size), 2) outcome measure (method of identification and diagnosis, frequency, and length of screening), 3) model performance information including the diagnostic accuracy of the delirium prediction models, calibration metrics, and events per variable

4) characteristics of the models (variables used in model, scoring/stratification system), 5) cognitive measures used in the study and 6) statistical methods applied for analysis. Five authors were contacted for missing or incomplete data. Four responses were received.

Statistics

Model performance was assessed through calibration and classification metrics.[15] The AUROC was the primary measure collected to evaluate the discriminatory ability of the delirium prediction models. We chose to designate delirium prediction models with an AUROC greater than 0·75, albeit arbitrary, as clinically relevant.[20] Sensitivity, specificity, positive predictive values and negative predictive values were also collected from each delirium prediction model. Goodness-of-fit statistics including Chi-square (X^2) and Hosmer-Lemeshow tests were collected to evaluate effective model calibration. Studies were also assessed for the inclusion of calibration plots and slopes. Model calibration refers to the agreement between observed outcomes and predictions.[21] Secondary pre-planned outcome measures included cognitive assessments, and predictive variable use per model.

Role of the Funding Source

The funding sources named have no role in this study. All authors had full access to all the data in the study and shared responsibility for the decision to submit the publication.

RESULTS

Twenty-seven studies were identified for inclusion.[22-46] The initial search resulted in 7,502 citations, with 192 studies chosen for full-text review as detailed in the PRISMA diagram (Figure 1). We did not identify any relevant, unpublished studies for this review. Two studies that included younger populations in the development cohort for the delirium prediction model were included due to the subsequent older external validation cohort thus meeting our inclusion criteria (age $\geq$ 60).[24 39] Twenty-three delirium prediction models were developed, thirteen were externally validated [22 26 28-30 32-34 40 42-45] and three were internally validated.[23 36 41] Prospective cohort design was used in 23 studies.[22 24-30 32-34 36-48] Retrospective design was used in four studies.[23 31 35 43] Eleven studies focused on the medical population [22 24 28-32 39 41 44 48], three included medical and surgical [23 42 43] and thirteen recruited a surgical population (seven-orthopaedic[25-27 33 37 40 47], one-cardiac [45], two-noncardiac [36 46], one general surgery[34], two-oncological[35 38]). Data collection occurred upon admission in seventeen studies [22 24 26 28-30 32-34 39-44 47 48]; participants were approached within forty- eight hours of admission. Seven studies collected data pre-operatively then followed participants post-operatively.[25 27 36-38 45 46] The average NOS quality ranking for included cohort studies was seven; five studies received the maximum of nine stars. Further characteristics of studies are listed in Table 1.

Delirium assessment

The outcome variable was measured using the Confusion Assessment Method in twenty-one studies. [22 24-30 32-39 42 45-48] The frequency of delirium assessment varied from two or more assessments daily (three studies)[25 34 40], to once daily (twelve studies)[24 27 29 31 33 35-37 43-45 47], every-

other day (eight studies)[22 26 28 30 32 41 42 48], once following surgery[46], and undefined (three studies).[23 38 39] Of the studies that assessed delirium twice or more daily, all of these studies relied on ward nurse observations or telephone interview with the nurse to identify delirium symptoms.[25 34 40] The principal investigator confirmed the presence of delirium following the nurse report of symptoms.[25 34] Twenty-one studies used trained research or clinical personnel to conduct the delirium assessments.[22 24-26 28-30 32-39 42-46 48] Three studies relied on delirium diagnosis, or keywords designated as representing delirium, to identify the outcome measure through retrospective chart review.[23 31 43] Three studies relied on clinical staff to recognize and chart delirium symptoms.[27 40 47] One of these studies retrospectively confirmed the diagnosis of delirium through consensus review of two authors, disagreement was resolved by a psychiatrist.[40] One study did not report details on personnel performing delirium assessments.[41]

Model design and statistical methods

Various statistical techniques were employed by the thirteen externally validated delirium prediction models in the selection of variables for model inclusion. Five used univariate or bivariate analyses and selected variables with a pre-determined statistical value (range for $p<0.05$ to $p<0.25$) for inclusion in the model.[22 24 39 42 45] One of these models paired bivariate analyses with a bootstrapping technique to address lower sample and event size.[45] Three models based their variable selection from a literature review of risk factors for delirium.[26 27 40 43 47] Two used proportional hazards regression modeling paired with bivariate analyses and included variables with either a *p*-value <0.25[31] or a relative risk of ≥ 1.5.[29] Five Studies published their power analysis.[24 32 34 39 45] To further refine and test the estimated models, the following

methods were employed: seven studies-stepwise logistic regression (LR),[22 24 29 34 39 42 45], four studies-multivariate LR[26 31 33 40], one study-continuation ratio model combined with log-binomial regression model[30], one study-multivariable binomial regression.[28]

Variables

Figure two demonstrates the frequency of variable use in the thirteen externally validated delirium prediction models. Baseline cognitive impairment was the most frequently used variable. Five models defined baseline cognitive impairment as a cognitive test score at or below the level of dementia.[26 29 33 42] This cognitive test was administered upon study enrollment. One study additionally evaluated chronic cognitive impairment through family or caregiver interview with the modified Blessed Dementia Rating Scale (mBDRS).[29 30] Four models combined the cognitive test score derived upon enrollment with a history of dementia to define baseline cognitive impairment.[30 32 40 43] History of dementia was defined as follows: Two studies-family or caregiver report supplemented with documented history in medical record [32 40], one study-medical record review and interview with mBDRS[30], and one study-dementia billing codes or prescription information.[43] One study defined baseline cognitive impairment as a pre-specified key term in the electronic health.[44]

Predictive ability

Reported AUROC in externally validated delirium prediction models ranged from 0.52-0.94 (Figure three). Five Five models attained an AUROC above 0.75 indicating potential clinical relevance and moderate predictive ability.[22 32 34] Of these five models, the highest performing model (AUROC 0.94, CI 0.91-0.97) was developed and validated in a surgical population.[34] Carrasco *et al.,* (2014) was developed and validated in a medical population (AUROC 0.78, CI

0.66-0.90).[22] The remaining three models were developed in separate medical cohort populations[24 31 39] but, were externally validated within the same cohort of medical patients and modified to share similar variable measures of cognition, functional status and illness severity (AUROC 0.78-0.83).[32] These five models share similarities with variable use, as seen in Figure two.

Model Calibration

Thirteen externally validated delirium prediction models reported calibration metrics.[28 29 33 44] The reported chi-square statistics were significant in three models[28 29 33] and did not reach significance in one model.[44] None of the included studies reported Hosmer-Lemeshow test statistics, calibration plots or slopes.

Risk of Overfitting

Events per variable (EPV) were examined in each of the thirteen externally validated models. Models estimating more parameters than events in a 1:10 ratio are at risk of statistical overfitting.[15 49 50] In 13 models with external validation, four had fewer than optimum events for the number of parameters estimated in the development stage of the models.[24 28 29 47] Five had fewer than optimum events in the external validation stage.[22 28-30 44] Two models did not reach optimum events for the number of parameters in either the development or the external validation studies.[28 29] Of the five models with an AUROC greater than 0.75, one of these models did not obtain sufficient EPV in the development stage[24] and another did not attain sufficient EPV in the external validation study, likely impacting the model's predictive ability (development – AUROC 0.86, CI-0.82-0.91, external validation – AUROC 0.78, CI-0.66-0.90).[22]

DISCUSSION

This review identified moderate predictive ability in five of the thirteen externally validated delirium prediction models, however three main limitations were identified. First, assessment of the outcome variable, delirium, was largely non-systematic, once daily, and avoided weekends. This is a major limitation for an acute condition that fluctuates and may occur suddenly. In the highest performing model, a major limitation was identified: data collection overlapped with the initial diagnosis of delirium, likely exaggerating model performance.[15 34] Second, model performance may be influenced by inadequate EPV leading to statistical overfitting and exaggerated model performance. Overall reporting of model performance measures was inconsistent with only four models reporting calibration statistics. Finally, variable definition was heterogeneous and often indistinct, making comparisons between models difficult and decreasing the ability to generalise models across populations. Further, broad variable definitions, particularly in functional and cognitive abilities, may have led to overlapping data capture. Pendlebury *et al.,* (2016) facilitated comparisons between three of the moderately performing models by externally validating these in the same cohort.[32] These models were re-developed to best fit the validation cohort. While this provides an opportunity to compare models, it is not known how these models will generalize to subsequent patient populations. Re-development is not equal to model validation.[15] Taken together, these findings suggest that current delirium prediction is limited by moderately performing, heterogeneous, non-generalizable models that may be improved with the application of frequent, systematic delirium assessments and the use of applicable statistical methods to evaluate and build clinical prediction models.

As delirium is a multifactorial syndrome representing an interrelationship between premorbid and precipitating factors,[28] the time course of data collection is important. Eight of the thirteen

externally validated delirium prediction models incorporate precipitating factors into their predictive model; two models [28 30] are intentionally constructed in this manner. The inclusion of a precipitating factor into a premorbid delirium prediction model may provide important predictive power if designed in the appropriate manner, as demonstrated by Inouye *et al.,* (1993).[29] However, if variables are collected after the onset of delirium this would exaggerate model performance (*e.g.,* ICU admission). As an example, one delirium prediction model has a robust AUROC of 0.94 (CI 0.91- 0.97).[34] This study excluded those with a MMSE <23 and prevalent delirium. Data collection occurred within the first 24-hours following surgery, however, delirium assessment began immediately after surgery, with a 50% delirium prevalence on the day of surgery. This overlap of data collection and delirium assessment likely exaggerated model performance for this outlier study. The remaining three models with AUROCs greater than 0.75 included data about the precipitating factor present upon admission and either excluded those with prevalent delirium or calculated separate AUROCs for prevalent delirium versus incident delirium.

Model underperformance may be explained through low powered studies leading to insufficient events per variable (EPV) resulting in statistical overfitting.[49 50] As overfitting of a model leads to an underestimation of event probability in low risk patients and overstates the probability in high-risk patients, it is an important consideration when evaluating the predictive performance of delirium prediction models.[51] This effect is highlighted in the Carrasco *et al.,* (2014) model as the AUROC decreased from the development study (0.82) to the external validation study (0.78). Future studies should attain adequate EPV to avoid overfitting. Further, past models validated with insufficient EPV should be interpreted with caution.

The identified studies largely used univariate or bivariate analysis then stepwise logistic

regression to develop the delirium prediction models. Although these are common methods to use for model development and may counter the effects of insufficient EPV, each approach has significant drawbacks.[51] Univariate analysis may reduce predictive ability by inclusion of variables that are not independent of each other, and stepwise regression disadvantages include conflation of *p*-values and a biased estimation of coefficients.[21 52] Statistical methods to counter low EPV could include penalized regression using either ridge or lasso regression and bootstrapping.[21 51]

Increasing age, pre-existing cognitive impairment, functional and sensory impairments were the most frequently used variables in the externally validated delirium prediction models. However, many studies employed different definition for these variables, making comparisons difficult between models and limiting generalizability across populations. Functional and physical impairments were broadly defined resulting in the inability to discern whether impairments resulted from truly physical origins or if the noted decrease in function was related to cognitive impairment leading to an overlap in data collection. Interestingly, these variables were also not consistently included in the five highest performing delirium prediction models, questioning their potential role in delirium prediction. Age may not be a relevant risk factor when considering an older cohort of patients; for example, a recent study found that global cognition may mediate the relationship between age and postoperative delirium[53] therefore the inclusion of age in a delirium prediction model may not add to the overall performance of the model if cognition is adequately captured or if only elderly patients are included in the study.

The highest performing delirium prediction model excluded those with pre-existing cognitive impairment, did not incorporate a cognitive variable and used hearing impairment as a predictive variable (note the methodological concerns of this study were discussed above).[34] Cognitive

impairment was the most frequently used variable and is a known risk factor for delirium development.[2] [53] Prior research demonstrates individuals with Mild Cognitive Impairment (MCI) are at a significantly higher risk of delirium development.[54] [55] All models used cut-off scores on cognitive tests that would indicate dementia, providing no evaluation of subtler cognitive decline such as MCI. Furthermore, Jones *et al.,* (2016) demonstrated a strong linear relationship between risk of delirium and all levels of cognitive function, even those considered unimpaired through formal testing.[53] In this study, a General Cognitive Performance score was developed using a complex battery of neuropsychological tests. Unfortunately, the neuropsychological battery is too complex to be practical for the clinical setting. Fong *et al.,* (2015) found associations between baseline executive functioning, complex attention and semantic networks to be associated with subsequent delirium development[56]. The inclusion of MCI, or simple cognitive tests as employed by Fong *et al.,* (2015), as a variable may increase the detection and prevalence of cognitive impairment as a variable thus increasing its predictive power. Further exploration into isolated cognitive tests that are feasible to administer in a clinical setting as well as sensitive to the spectrum of cognitive impairment may enhance delirium prediction.

Four of the best-performing models contained a measurement of functional or physical impairment.[22] [32] [34] This measurement may be representative of numerous underlying factors working to inhibit a biological compensatory mechanism and serve as a marker for a vulnerable individual.[57] [58] Carrasco *et al.,* (2014) used the Barthel Index, which evaluates basic functioning in ten different areas. A proxy completed this measure instead of self-report which has been shown to improve accuracy.[59] Kim *et al.,* (2016) did not report use of a standardized measurement tool, but defined impaired physical status as the inability to be self-sufficient.

Pendlebury *et al.,* (2016) defined functional impairment as an individual residing in a care home or receiving care at their home and applied this definition in two of the four models validated within that patient cohort. These broad definitions lead to the inability to discern whether the functional impairment was due to a physical or cognitive mechanism.

Strengths and weaknesses of this study

This systematic review benefitted from a prospectively developed protocol. A comprehensive literature search from multiple databases using broad search terms yielded twenty-seven studies with thirteen externally validated delirium prediction models. Our author team is interprofessional, providing the opportunity for different perspectives on model evaluation. Further, this review synthesizes evidence from both medical and surgical populations while providing statistical-based recommendations for study and model design for future delirium prediction model studies. The limitations of this systematic review may be that articles focused on a younger population were not included along with studies identifying predictive risk factors, not exclusively predictive models. This limitation could narrow the generalizability of the results of this systematic review to the broader population however delirium predominantly affects older adults.

Strengths and weaknesses in relation to other studies

Past systematic reviews concluded that the identified delirium prediction models were largely heterogeneous in variable inclusion and were not sufficiently developed for incorporation into practice.[60-62] Recommendations include further testing on existing delirium prediction models followed by integration in practice as well as further exploration into measurements that are feasible clinically. This review included eight models not previously identified in past systematic reviews of delirium prediction models. Further this review is the first to identify study and model design issues and discusses the paucity of measurements sensitive to the spectrum of cognitive

impairment.

Implications and future research

Future studies should focus on the development and validation of delirium prediction models using the following broad principles: (1) Delirium prediction models should be developed only using data available prior to the onset of delirium and likely should be focused in specific populations depending on whether the precipitating event has occurred or not; (2) should explore the use of further cognitive variables to enhance current model performance and should distinguish functional impairment due to physical conditions, cognitive impairment or both, (3) should include structured, twice daily assessment (regardless of weekends) using validated tools and trained research staff to identify delirium, (4) adhere to strict guidelines for both statistical methodology and metric reporting, (5) Delirium prediction model variables should have sufficient prevalence along with the number of events within the population studied to optimize model performance and (6) consider development of dynamic predictive models using AI methods and machine learning. In addition, rigorous statistical methods would improve the development and validation of models and avoid issues of under and overfitting of models. An example example of this would be to employ Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) in stepwise selection. This would avoid exclusion of variables that may not be statistically significant in standard hypothesis testing, yet may yield important variable prediction in model estimations.[21] Standardized metric reporting would augment model development and validation, facilitating the ability to compare model across populations and settings. Recommendations for future statistical reporting of delirium prediction models include sensitivity, specificity, positive predictive value, negative predictive value, Nagelkerke's R^2, area under the receiver operating curve (AUROC), and goodness-of-fit measures. Further,

calculating and reporting statistical metrics on the calibration and clinical usefulness of models would benefit delirium prediction.[21]

Two classes of delirium prediction models may be required, based on the acuity of the admission (elective or emergency). If precipitating factors are included in an elective admission delirium prediction model, where the patient is yet to incur the delirium provoking event, an individual's delirium risk may be overestimated. In the second option, inclusion of only premorbid factors may underestimate delirium risk given the emergency clinical scenario.

Conclusion

Twenty-three delirium prediction models were identified. Thirteen of these were externally validated and three were internally validated. Of the thirteen validated delirium prediction models, the overall predictive ability is moderate with only five models achieving an AUROC above 0.75.[22 32 34] Assessment of the outcome variable, delirium is often non-systematic and future studies would be improved with more standardized and frequent assessment. Overall, the variable inclusion and applied definitions in delirium prediction models are heterogeneous making comparisons difficult. To improve delirium prediction models, future models should consider using standard variables and definitions to work towards a prediction tool that is generalizable to several populations within the remit of understanding the relationship with the precipitating event.

Contributors

HL and SP with the mentorship of RDS formulated the aim, developed the study protocol, completed the search and extracted the data. HL and RDS synthesized the data. HL with the mentorship of RDS drafted the manuscript and designed the tables. RB designed the figures and assisted with statistical interpretation. LB provided expertise on content related to cognition and

reviewed the manuscript. DD and CMC assisted with synthesis and interpretation of results and discussion in relation to their expertise in geriatrics, cognition, and delirium. MC, MM, MTVC, and PP assisted with synthesis of results and discussion section, providing expertise in delirium in its respective settings.

Declaration of Interests

All authors have completed ICMJE disclosure forms and no conflicts of interest are declared.

Acknowledgements

HL and RDS acknowledge funding support from the Department of Anesthesiology at University of Wisconsin- Madison. PP acknowledges funding support from R01 NHLBI (HL111111) and research grant from Hospira Inc. in collaboration with NIH. We would like to express our sincere gratitude for the assistance of Mary Hitchcock in Ebling Health Sciences Library in the design of the search strategy and Dave Dwyer and Lily Turner for their assistance with proofreading the manuscript.

References

1. Sanders RD, Pandharipande PP, Davidson AJ, *et al.* Anticipating and managing postoperative delirium and cognitive decline in adults. *BMJ: British Medical Journal (Overseas & Retired Doctors Edition)* 2011;343(7816):197-200 4p. doi: 10.1136/bmj.d4331

2. Inouye SK, Westendorp RG, Saczynski JS. Delirium in elderly people. *Lancet* 2014;383(9920):911-22. doi: 10.1016/s0140-6736(13)60688 10.1016/S0140-6736(13)60688-1. Epub 2013 Aug 28.</p>

3. Ryan DJ, O'Regan NA, Caoimh RÓ, *et al.* Delirium in an adult acute hospital population: predictors, prevalence and detection. *BMJ Open* 2013;3(1) doi: 10.1136/bmjopen-2012-001772

4. Pandharipande PP, Girard TD, Jackson JC, *et al.* Long-Term Cognitive Impairment after Critical Illness. *The New England Journal of Medicine* 2013;369(14):1306-16. doi: 10.1056/NEJMoa1301372

5. Leslie DL, Inouye SK. The importance of delirium: economic and societal costs. *Journal of the American Geriatrics Society* 2011;59 Suppl 2:S241-3. doi: 10.1111/j.1532-5415.2011.03671.x [published Online First: 2011/12/07]

6. Han JH, Brummel NE, Chandrasekhar R, *et al.* Exploring Delirium's Heterogeneity: Association Between Arousal Subtypes at Initial Presentation and 6-Month Mortality in Older Emergency Department Patients. *The American journal of geriatric psychiatry: official journal of the American Association for Geriatric Psychiatry* 2016 doi: 10.1016/j.jagp.2016.05.016 [published Online First: 2016/09/14]

7. Pollard C, Fitzgerald M, Ford K. Delirium: The lived experience of older people who are delirious post-orthopaedic surgery. *International Journal of Mental Health Nursing*

2015;24(3):213-21. doi: 10.1111/inm. 12132

8. Gleason LJ, Schmitt EM, Kosar CM, *et al.* Effect of Delirium and Other Major Complications on Outcomes After Elective Surgery in Older Adults. *JAMA Surgery* 2015;150(12):1134-40. doi: 10.1001/jamasurg.2015.2606
9. Robinson TN, Raeburn CD, Tran ZV, *et al.* Postoperative delirium in the elderly: risk factors and outcomes. *Annals of surgery* 2009;249(1):173-8. doi: 10.1097/SLA.0b013e31818e4776 [published Online First: 2008/12/25]
10. Adams ST, Leveson SH. Clinical prediction rules. *BMJ (Clinical research ed)* 2012;344
11. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Annals of internal medicine* 2006:144(3):201-09.
12. Moons KGM, Kengne AP, Woodward M, *et al.* Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart* 2012;98(9):683-90. doi: 10.1136/heartjnl-2011- 301246
13. Debray TPA, Moons KGM, Ahmed I, *et al.* A framework for developing, implementing and evaluating clinical prediction models in an individual participant data meta-analysis. *Statistics in medicine* 2013;32(18):3158-80. doi: 10.1002/sim.5732
14. Hayhurst CJ, Pandharipande PP, Hughes CG. Intensive Care Unit DeliriumA Review of Diagnosis, Prevention, and Treatment. *Anesthesiology* 2016;125(6):1229-41. doi: 10.1097/ALN.0000000000001378
15. Moons KG, de Groot JA, Bouwmeester W, *et al.* Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS medicine* 2014;11(10):e1001744. doi: 10.1371/journal.pmed.1001744 [published Online First: 2014/10/15]

16. Moher De, LIberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg* 2010;8(5):336-41. doi: 10.1016/j.ijsu.2010.02.007 [published Online First: 2010/02/23]

17. Nations U U. Ageing 2016 [Available from: http://www.un.org/en/sections/issues-depth/ageing/index.html accessedMarch 7th 2017.

18. Centers for Disease C. Elder Abuse: Definitions 2016 [Available from: https://www.cdc.gov/violenceprevention/elderabuse/definitions.html accessed March 7th 2017.

19. Well GA, Shea, B., O'Connell, J.P., Welch, V., Losos, M., Tugwell, P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses Ottawa: Ottawa Hospital; 2014 [Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp accessed 05-31 2016.

20. Fan J, Upadhye S, Worster A. Understanding receiver operating characteristic (ROC) curves. *Canadian Journal of Emergency Medicine* 2015;8(1):19-20. doi: 10.1017/S1481803500013336

21. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating: New York: Springer, [2009] ©2009 2009.

22. Carrasco MP, Villarroel L, Andrade M, *et al.* Development and validation of a delirium predictive score in older people. *Age and ageing* 2014;43(3):346-51. doi: 10.1093/ageing/aft141 [published Online First: 2013/09/26]

23. de Wit HAJM, Winkens B, Mestres Gonzalvo C, *et al.* The development of an automated ward independent delirium risk prediction model. *Int J Clin Pharm* 2016 doi: 10.1007/s11096-016-0312-7

24. Douglas VC, Hessler CS, Dhaliwal G, *et al.* The AWOL tool: derivation and validation of a delirium prediction rule. *J Hosp Med* 2013;8(9):493-9. doi: 10.1002/jhm.2062 [published Online First: 2013/08/08]

25. Fisher BW, Flowerdew G. A simple model for predicting postoperative delirium in older patients undergoing elective orthopedic surgery. *Journal of the American Geriatrics Society* 1995;43(2):175-8. [published Online First: 1995/02/01]

26. Freter S, Dunbar M, Koller K, *et al.* Risk of Pre-and Post-Operative Delirium and the Delirium Elderly At Risk (DEAR) Tool in Hip Fracture Patients. *Can Geriatr J* 2015;18(4):212-16. doi: 10.5770/cgj.18.185

27. Freter SH, Dunbar MJ, MacLeod H, *et al.* Predicting post-operative delirium in elective orthopaedic patients: the Delirium Elderly At-Risk (DEAR) instrument. *Age and ageing* 2005;34(2):169-71.

28. Inouye SK, Charpentier PA. Precipitating factors for delirium in hospitalized elderly persons. Predictive model and interrelationship with baseline vulnerability. *Jama* 1996;275(11):852-7. [published Online First: 1996/03/20]

29. Inouye SK, Viscoli CM, Horwitz RI, *et al.* A predictive model for delirium in hospitalized elderly medical patients based on admission characteristics. *Annals of internal medicine* 1993;119(6):474-81. [published Online First: 1993/09/15]

30. Inouye Inouye SK, Zhang Y, Jones RN, *et al.* Risk factors for delirium at discharge: development and validation of a predictive model. *Archives of internal medicine* 2007;167(13):1406-13. doi: 10.1001/archinte.167.13.1406 [published Online First: 2007/07/11]

31. Isfandiaty R, Harimurti K, Setiati S, *et al.* Incidence and predictors for delirium in

hospitalized elderly patients: a retrospective cohort study. *Acta medica Indonesiana 2012;44(4):290-7. [published Online First: 2013/01/15]*

32. Pendlebury ST, Lovett N, Smith SC, *et al.* Delirium risk stratification in consecutive unselected admissions to acute medicine: validation of externally derived risk scores. *Age and ageing* 2016;45(1):60-65. doi: 10.1093/ageing/afv177
33. Kalisvaart KJ, Vreeswijk R, de Jonghe JF, *et al.* Risk factors and prediction of postoperative delirium in elderly hip-surgery patients: implementation and validation of a medical risk factor model. *Journal of* the American Geriatrics Society 2006;54(5):817-22. doi: 10.1111/j.1532-5415.2006.00704.x [published Online First: 2006/05/16]
34. Kim MY, Park UJ, Kim HT, *et al.* DELirium Prediction Based on Hospital Information (Delphi) in General Surgery Patients. *Medicine (Baltimore)* 2016;95(12):e3072. doi: 10.1097/MD.0000000000003072
35. Korc-Grodzicki B, Sun SW, Zhou Q, *et al.* Geriatric Assessment as a Predictor of Delirium and Other Outcomes in Elderly Patients With Cancer. *Annals of surgery* 2014 doi: 10.1097/sla.0000000000000742 [published Online First: 2014/06/03]
36. Leung JM, Sands LP, Lim E, *et al.* Does preoperative risk for delirium moderate the effects of postoperative pain and opiate use on postoperative delirium? *The American journal of geriatric psychiatry : official journal of the American Association for Geriatric Psychiatry* 2013;21(10):946-56. doi: 10.1016/j.jagp.2013.01.069 [published Online First: 2013/05/11]
37. Liang C-K, Chu C-L, Chou M-Y, *et al.* Developing a Prediction Model for Post-Operative Delirium and Long-Term Outcomes Among Older Patients Receiving Elective Orthopedic Surgery: A Prospective Cohort Study in Taiwan. *Rejuvenation Res*

2015;18(4):347-55. doi: 10.1089/rej.2014.1645

38. Maekaw Y, Sugimoto K, Yamasaki M, *et al.* Comprehensive Geriatric Assessment is a useful predictive tool for postoperative delirium after gastrointestinal surgery in old-old adults. *Geriatr Gerontol Int* 2015 doi: 10.1111/ggi.12587
39. Martinez JA Belastegui A, Basabe I, *et al.* Derivation and validation of a clinical prediction rule for delirium in patients admitted to a medical ward: an observational study. *BMJ Open* 2012;2(5) doi: 10.1136/bmjopen-2012- 001599 [published Online First: 2012/09/18]
40. Moerman S, Tuinebreijer WE, de Boo M, *et al.* Validation of the Risk Model for Delirium in hip fracture patients. *Gen Hosp Psychiatry* 2012;34(2):153-9. doi: 10.1016/j.genhosppsych.2011.11.011 [published Online First: 2012/01/10]
41. O'Keefe ST, Lavan JN. Predicting delirium in elderly patients: development and validation of a risk-stratification model. *Age and ageing* 1996;25(4):317-21. [published Online First: 1996/07/01]
42. Pompei P, Foreman M, Rudberg MA, *et al.* Delirium in hospitalized older persons: outcomes and predictors. *Journal of the American Geriatrics Society* 1994;42(8):809-15. [published Online First: 1994/08/01]
43. Rudolph JL, Doherty K, Kelly B, *et al.* Validation of a Delirium Risk Assessment Using Electronic Medical Record Information. *Journal of the American Medical Directors Association* 2016;17(3):244-48 5p. doi: 10.1016/j.jamda.2015.10.020
44. Rudolph JL, Harrington MB, Lucatorto MA, *et al.* Validation of a medical record-based delirium risk assessment. *Journal of the American Geriatrics Society* 2011;59 Suppl 2:S289-94. doi: 10.1111/j.1532-5415.2011.03677.x

45. Rudolph JL, Jones RN, Levkoff SE, *et al.* Derivation and validation of a preoperative prediction rule for delirium after cardiac surgery. *Circulation* 2009;119(2):229-36 8p. doi: 10.1161/CIRCULATIONAHA.108.795260
46. Dworkin A, Lee DS, An AR, *et al.* A Simple Tool to Predict Development of Delirium After Elective Surgery. *Journal of the American Geriatrics Society* 2016;64(11):e149-e53. doi: 10.1111/jgs.14428 [published Online First: 2016/09/22]
47. Freter SH, George J, Dunbar MJ, *et al.* Prediction of delirium in fractured neck of femur as part of routine preoperative nursing care. *Age and ageing* 2005;34(4):387-8. doi: 10.1093/ageing/afi099 [published Online First: 2005/06/16]
48. Pendlebury ST, Lovett NG, Smith SC, *et al.* Delirium risk stratification in consecutive unselected admissions to acute medicine: validation of a susceptibility score based on factors identified externally in pooled data for use at entry to the acute care pathway. *Age and ageing* 2016 doi: 10.1093/ageing/afw198 [published Online First: 2016/11/07]
49. Peduzzi P, Concato J, Kemper E, *et al.* A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49(12):1373-9. [published Online First: 1996/12/01]
50. Austin PC, Steyerberg EW. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Stat Methods Med Res*;0(0):0962280214558972. doi: doi:10.1177/0962280214558972
51. Pavlou M, Ambler G, Seaman SR, *et al.* How to develop a more accurate risk prediction model when there are few events. *BMJ (Clinical research ed)* 2015;351:h3868. doi: 10.1136/bmj.h3868 [published Online First: 2015/08/13]

52. Grobman WA, Stamilio DM. Methods of clinical prediction. *American journal of obstetrics and gynecology* 2006;194(3):888-94. doi: 10.1016/j.ajog.2005.09.002 [published Online First: 2006/03/09]

53. Jones RN, Marcantonio ER, Saczynski JS, *et al.* Preoperative Cognitive Performance Dominates Risk for Delirium Among Older Adults. *J Geriatr Psychiatry Neurol* 2016 doi: 10.1177/0891988716666380 [published Online First: 2016/09/21]

54. Kazmierski J, Banys A, Latek J, *et al.* Mild Cognitive Impairment with Associated Inflammatory and Cortisol Alterations as Independent Risk Factor for Postoperative Delirium. *Dementia & Geriatric Cognitive Disorders* 2014;38(1/2):65-78 14p. doi: 10.1159/000357454

55. Veliz- Reissmüller G, Agüero Torres H, van der Linden J, *et al.* Pre-operative mild cognitive dysfunction predicts risk for post-operative delirium after elective cardiac surgery. *Aging Clinical And Experimental Research* 2007;19(3):172-77.

56. Fong TG, Hshieh TT, Wong B, *et al.* Neuropsychological profiles of an elderly cohort undergoing elective surgery and the relationship between cognitive performance and delirium. *Journal of the American Geriatrics Society* 2015;63(5):977-82. doi: 10.1111/jgs.13383

57. Rozzini R, Sabatini T, Cassinadri A, *et al.* Relationship between functional loss before hospital admission and mortality in elderly persons with medical illness. *J Gerontol A Biol Sci Med Sci* 2005;60(9):1180-3. [published Online First: 2005/09/27]

58. Fimognari FL, Pierantozzi A, De Alfieri W, *et al.* The Severity of Acute Illness and Functional Trajectories in Hospitalized Older Medical Patients. *J Gerontol A Biol Sci Med Sci* 2016 doi: 10.1093/gerona/glw096 [published Online First: 2016/06/04]

59. Li M, Harris I, Lu ZK. Differences in proxy-reported and patient-reported outcomes: assessing health and functional status among medicare beneficiaries. *BMC Medical Research Methodology* 2015;15(1):62. doi: 10.1186/s12874- 015-0053-7

60. Van Meenen LC, van Meenen DM, de Rooij SE, *et al.* Risk prediction models for postoperative delirium: a systematic review and meta-analysis. *Journal of the American Geriatrics Society* 2014;62(12):2383-90. doi: 10.1111/jgs.1313810.1111/jgs.13138.</p>

61. Newman MW, O'Dwyer LC, Rosenthal L. Predicting delirium: A review of risk-stratification models. *General Hospital Psychiatry* 2015;37(5):408-13. doi: 10.1016/j.genhosppsych.2015.05.003

62. Campaner Campaner CP. Valoración del riesgo de delirium en pacientes mayores hospitalizados. = Delirium risk assessment in elderly hospitalized patients. *Revista Española de Geriatría y Gerontología* 2010;45(5):285-90. doi: 10.1016/j.regg.2010.03.011

Table 1					
Author	**Study Design Population Sample Size**	**Study Grade (NOS)**	**Outcome Variable & Rate (%)**	**Delirium measurement & frequency**	**DPM Design & (Name)**
Carrasco et al. (2014)[24]	P.Cohort Medical Dev: 374 Val: 104	S: **** C: - O: ** T: 6 Stars	Delirium Dev: 25 (.06) Val: 12 (12)	CAM Every 48 h	Predictive Risk Score
de Wit et al. (2016)[25]	Retro All hospital patients Dev: 1291	S: *** C: ** O: *** T: 8 stars	Delirium Dev: 225(17)	Chart abstraction EHR "diagnosis table"	Automated Delirium Prediction Model
Douglas et al. ** **(2013)[26]**	P.Cohort Medical Dev: 209 Val: 165	S: **** C: - O: *** T: 7 Stars	Delirium Dev: 25(12) Val: 14(8.5)	CAM-S & CAM Daily	Risk Stratification model (AWOL)
Dworkin et al. (2016)[48]	P.Cohort Elective noncardiac surg Dev: 76	S: C: O: T:	Delirium Dev: 10(13)	CAM or FAM-CAM 1xafter surgery	Mini-Cog Stratified into a five point score
Fisher and Flowerdew (1995)[27]	P.Cohort Elective Orthopedic Dev: 80	S: ** C: - O: ** T: 4 stars	Delirium Dev: 14(17.5)	CAM 2xDaily	Prediction Model using two variables.
Freter et al. (2005)[29]	P.Cohort Elective Hip surgery Dev: 132	S: ** C: ** O: ** T: 6 stars	Delirium Dev: 18(14)	CAM Daily	Risk Stratification Model (DEAR)
Freter et al. (2005)[49]	P.Cohort Hip Fx Dev: 100	S: ** C: ** O: ** T: 6 Stars	Delirium Dev: 24(24)	CAM Daily	Risk Stratification Model (DEAR)
Freter et al. (2015)[28]	P.Cohort Hip Fracture Val: 283	S: *** C: - O: ** T: 5 stars	Delirium Val: 119(42)	CAM POD1, 3 & 5	Risk stratification model (DEAR)
Inouye and Charpentier (1996)[30]	P.Cohort Medical Dev: 196 Val: 312	S: **** C: ** O: *** T: 9 stars	Delirium Dev: 35(18) Val: 47(15)	CAM Every other day	Risk stratification model based on precipitating factors

Inouye et al. (2007)[32]	P.Cohort Medical Dev: 491	S: **** C: ** O: ***	Delirium/ subsyndrome	CAM Every other day	Risk stratification model
	Val: 461	T: 9 stars	delirium at discharge Dev: 58(12)		
Inouye et la. (1993)[31]	P.Cohort Medical Dev: 107 Val: 174	S: **** C: ** O: *** T: 9 stars	Delirium Dev: 27(25) Val: 29(17)	CAM Daily	Risk stratification model
Isfandiaty et al. (2012)[33]	Retro Medical Dev: 457	S: ** C: - O: *** T: 5 Stars	Delirium Dev: 87(19)	Undefined Daily	Risk stratification model
Kalisvaart et al. (2006)[35]	P.Cohort Hip Surgery & Facture Val: 603	S: *** C: - O: *** T: 6 stars	Delirium Dev: 74(12)	CAM, DRS-98 Daily through POD5	Externally validated Inouye's '93 model.
Kim et al. (2016)[36]	P.Cohort Major General Surgery Dev: 561 Val: 533	S: *** C: ** O: *** T: 8 stars	Delirium Dev: 112(20) Val: 99(18)	Nu-Desc -every shift by RNs Confirmed with CAM.	Risk stratification model
Korc-Grodzicki et al. (2014)[37]	Retro Oncological Surgery Dev: 416	S: *** C: - O: *** T: 6 stars	Delirium Dev: 79(19)	CAM Daily	Comprehensive Geriatric Assessment (CGA) as model.
Leung et al. (2013)[38]	P.Cohort Noncardiac surgery Dev: 581	S: *** C: - O: ** T: 5 stars	Delirium Dev: 234(40)	CAM Daily	Risk stratification model
Liang et al. (2015)[39]	P.Cohort Elective Orthopedic Surgery Dev: 461	S: *** C: ** O: ** T: 7 stars	Delirium Dev: 37(8)	CAM Daily Confirmed by psychologist	Built 2 DPMs CGA Risk stratification models
Maekawa et al. (2015)[40]	P.Cohort Oncological; Gastrointestinal Surgery Dev: 517	S: ** C: * O: *** T: 6 stars	Delirium Dev: 124(24)	CAM Unknown frequency	Comprehensive Geriatric Assessment (CGA) as model.

Martinez et al.(2012)[41]**	P.Cohort Medical Dev: 397 Val: 302	S: *** C: - O: ** T: 5 stars	Delirium Dev: 52(13) Val: 76(25)	CAM Undefined	Clinical prediction rule

Moerman et al. (2012)[42]	P.Cohort Hip Fracture Val: 378	S: *** C: - O: *** T: 6	Delirium Val: 102(27)	Ward RN observation, 3xdaily Confirmed by chart	Risk stratification model (Risk Model for Delirium, RD)
O'Keeffe and Lavan (1996)[43]	P.Cohort Acute Geriatric Unit Dev:	S: **** C: - O: ** T: 6 star	Delirium Dev: 28(28) IVal: 25(30)	DAS Every 48 hours DSM III	Risk Stratification model
Pendlebury et al. (2016)[50]	P. Cohort Medical Dev: 308	S: **** C: * O: *** T: 8 star	Delirium Dev: 95(31)	CAM Every 48-hours Confirmed by DSM-IV interview	Susceptibility Score
Pendlebury et al. (2016)[34]	P.Cohort Medical Val: 308	S: **** C: - O: *** T: 7 star	Delirium Val: 95(31)	CAM Every 48-hours Confirmed by DSM-IV interview	Externally validated 4 DPMs
Pompei et al. (1994)[44]	P.Cohort Med/surg	S: **** C: ** O: *** T: 9	Delirium Dev: 64(14.8) Val: 86(26.3)	CAM 2xweekly. Confirmed with DSM	Risk stratification model
Rudolph et al. (2009)[47]	P.Cohort Cardiac Surger	S: *** C: * O: **	Delirium Dev: 63(52) Val: 48(44)	CAM, MDAS, DSI Daily	Risk stratification model
Rudolph et al. (2011)[46]	P.Cohort Medical V:	S: **** C: - O: ***	Delirium Dev: 23(23)	DSM-IV Daily clinical interview	Externally validated Inouye's '93 model.

Rudolph et al. (2016)[45]	Dev: Retro Val: P.Cohort Med/surg Dev: 27625	S: **** C: - O: ** T: 6 stars	Delirium Dev: 2343(8) Val: 64(26)	Dev: Chart audit Val: DSM-IV Daily	Risk stratification model

Key:
**=Models developed in population ≤60 years of age, but validated in population ≥60 years of age.
Study Design: P.Cohort=Prospective Cohort, Retro=Retrospective design, Dev=Development

Study Grade: NOS=Newcastle Ottawa Scale. S=Selection, C=Comparability, O=Ottawa
Outcome Variable: Dev=Development, Val=Validation
Delirium Measurement: CAM=Confusion Assessment Method, DSM=Diagnostic Statistical Manual, POD=Postoperative Day, MDAS=Memorial Delirium Assessment Scale, Nu-Desc=Nursing Delirium Screening Scale, DRS-98=Delirium Rating Scale, EHR=Electronic Health Record
Type of Model: How authors designed their delirium prediction model (DPM)
-Risk stratification model: Points (weighted or un-weighted) assigned per predictive risk factor present.
-CGA=Comprehensive Geriatric Assessment

Table 2

External Validated DPM Name	Citation	Delirium #(%)	Sens Spec PPV NPV (external)	AUROC (95%CI)	Model Components	Cog. Assess Tool & Cutoff
AWOL Tool	Pendlebury et al. (2016) [34]	1st Val: 14(9) 2nd Val: 95(31) (any delirium) 67-prevalent 28-incident	Mod. AWOL Cutoff - 3 Any Delirium Sens .7 Spec .66 PPV .55 NPV .79 Incident Del Sens .76 Spec .66 PPV .27 NPV .94	1st Val: 0.69 (0·54-0·83) Incident delirium 2nd Val: Cohort 1 (MMSE) 0·78 (0·68-0·88) Cohort 2 (AMTS) 0·73 (0·63-0·83)	Original AWOL Tool Age >80 1 pt Failure to spell WORLD backwards 1 pt Disorientation 1 pt Illness Severity 1 pt Modified AWOL Tool Age >80 1 pt Diag of Dementia 1 pt MMSE<24, AMTS<9 1 pt Illness severity 1 pt	MMSE < 24 AMTS < 9
Clinical Prediction Rule-Cardiac Surgery	Rudolph et al. (2009) [47]	Dev: 63(52) Val: 48(44) (incident delirium)	Not reported	Dev: 0·74 Val: 0·75 Did not report CI	Weighted Points-Regression MMSE ≤ 23 2 pt MMSE 24-27 1 pt Hx of Stroke/TIA 1 pt GDS >4 1 pt Abnormal Albumin 1 pt Stratified into point categories 0 pt 1 pt 2 pts ≥ 3 pts	MMSE -Stratified score
DEAR	Freter et al. (2015)[28]	Dev: (2005) 18(14) Val: (2015) Pre-Op= 163(58) Post-op= 118(42)	Sens .68 Spec .73 PPV .65 NPV .76 Optimal cut-off score: 3pts (Incident post-op delirium)	Dev: (2005) 0·77 (0·64-0·87) Val: (2015) AUROC Not published	MMSE ≤ 23 1 pt Functional dependence 1 pt Sensory impairment 1 pt Substance use 1 pt Age >80 1 pt Not weighted. 0-5 Score, cut-off of 2 or 3 indicating high risk.	MMSE Cut-off ≤ 23

Table 2

External Validated DPM Name	Citation	Delirium #(%)	Sens Spec PPV NPV (external)	AUROC (95%CI)	Model Components	Cog. Assess Tool & Cutoff
Delirium at Discharge Prediction Model	Inouye et al. (2007)[32]	Dev: 58(12) Val: 28(6) (incident delirium)	Not reported	Dev: 0.80 Val: 0.75 Did not report CI	Delirium at Discharge Prediction Dementia diagnosis or mBDRS≥4 — 1 pt Vision Impairment — 1 pt ADL Impairment — 1 pt Charlson Score — 1 pt Restraint use during delirium — 1 pt Not weighted. 0-1 pt = Low Risk 2-3 pt = Intermediate Risk 4-5 pt = High Risk	MMSE < 24 mBDR ≥ 4
Delirium Prediction Score (DPS)	Carrasco et al. (2014)[24]	Dev: 25(.06) Val: 12(12) (incident delirium)	Sens .88 Spec .74 PPV .22 NPV .99	Dev: 0.86 (0.82-0.91) Val: 0.78 (0.66-0.90)	DPS=[5xBUN/Cr ratio]-(3xBarthel Index). Cut off is: > -240 = High risk for Delirium In conventional units, cut-off is: > -160 = High Risk for Delirium	None. Pfeffer Functional Activities Questionnaire as a proxy for prior dementia
Delphi Score	Kim et al. (2016)[36]	Dev: 112(20) Val: 99(18) (incident delirium)	Sens .81 Spec .93 PPV .70 NPV .96 Optimal cut-off score: 6.5pts	Dev: 0.911 (0.88-0.94) Val: 0.938 (0.91-0.97)	Age (years) 60-69 — 0 70-79 — 1 ≥80 — 2 Low Physical Activity Self-sufficient — 0 Need assist. — 2 Heavy ETOH No — 0 Yes — 1 Open Surgery No — 0 Yes — 2	No measure of cognition. Excluded participants if

Table 2

External Validated DPM Name	Citation	Delirium #(%)	Sens Spec PPV NPV (external)	AUROC (95%CI)	Model Components	Cog. Assess Tool & Cutoff
					ICU Admission No 0 Yes 3 Pre-Op CRP (mg/dL) <10 0 ≥10 1 Max points: 15 Optimal cut-off: 6.5 High Risk: ≥7 points	
e-NICE Rule	Rudolph et al. (2016)[43]	Dev: 2343(8) Val:64(26) (incident delirium)	Cohort / AURO C / CI Dev 0.81 (0.80-0.82) Validation Original 0.69 (0.61-0.77) mRASS 0.72 (0.65-0.79) TMYB 0.73 (0.66-0.80) MoCA 0.74 (0.66-0.81) Any delirium Did not report sens, spec, PPV, NPV		Weighted Points/OR Cog impair -Medications diagnosis or both 4 pt Age ≥ 65 y 2 pt Age ≥ 80 y 3 pt Infection 2 pt Fracture 4 pt Vision 1 pt Severe Illness 2 pt 0-2 pts = Low Risk 2-5 pts = Intermediate Risk 6-8 pts = High Risk ≥ 9 pts = Very High Risk	e-NICE Tool Diagnosis of dementia, medications for dementia or both qualified as "cognitive impairment" in model. Prospective cohort, additional: -mRASS -TMYB -MoCA ≤ 18
Inouye Prediction Rule (IPR)	Inouye et al. (1993)[31]	Dev: 27(25) Val: 29(17) (incident delirium)	Did not report	Dev: 0.74 (0.63-0.85) Val: 0.66 (0.55-0.77)	Baseline cognitive impairment 1 pt High BUN/Cr ratio 1 pt Severe illness (Composite score: APACHE II >16 + RN rating) 1 pt Vision impairment 1 pt Not weighted. 0 pts = Low risk 1-2 pts = Intermediate risk 3-4 pts = High risk	MMSE Cut-off < 24 Family/caregiver bDRS Excluded those w/history of severe dementia

Table 2

External Validated DPM Name	Citation	Delirium #(%)	Sens Spec PPV NPV (external)	AUROC (95%CI)	Model Components	Cog. Assess Tool & Cutoff
IPR	Kalisvaart et al. (2006)[35]	Val: 74(12)	Did not report	Val: 0.73 (0.65-0.78)	Externally validated IPR in surgical hip fracture population.	MMSE Cut-off < 24
IPR	Rudolph et al. (2011)[46]	Val: 23(23) Any delirium 10-Prevalent 13-Incident	Did not report	Val: 0.56 (0.42-0.74) Incident delirium	Externally validated IPR in medical VA population.	MMSE Cut-off < 24
IPR	Pendlebury et al. (2016)[34]	Val: 95(31) Any delirium 67-prevalent 28-incident	Cutoff 2pts All Delirium Sens .57 Spec .80 PPV .64 NPV .76 Incident D Sens .52 Spec .80 PPV .31 NPV .91	Val: Incident delirium Cohort 1 (MMSE) 0.73 (0.62-0.84) Cohort 2- (AMTS) 0.70 (0.60-0.81)	Baseline cognitive impairment 1 pt High BUN/Cr ratio 1 pt Severe illness (SIRS ≥ 2) 1 pt Vision impairment 1 pt 4pts=Incident delirium	Original Model: MMSE <24 Modified Model: MMSE < 24 AMTS < 9
Isfandiaty model	Pendlebury et al. (2016)[34]	Dev: 87(19) Val: 95 (31) Any delirium 67-prevalent 28-incident	Cutoff 4pts Any Delirium Sens .74 Spec .71 PPV .60 NPV .82 Incident Del Sens .81 Spec .71 PPV .31 NPV .96	Dev: 0.82 (0.77-0.88) Val: Incident delirium Cohort 1 (MMSE) 0.83 (0.74-0.91) Cohort 2 (AMTS) 0.77 (0.67-0.86)	Baseline cognitive impairment 3 pt Functional dependency 2 pt Infection w/sepsis 2 pt Infection w/out sepsis 1 pt Weighted score Score = 7 for incident delirium Cohort 1: MMSE Cohort 2: AMTS	Original Model: Chart review Modified Model: MMSE < 24 AMTS < 9
Martinez et al. 2012 model	Pendlebury et al. (2016)[34]	1st Val: 76(25) 2nd Val: 95(31) Any delirium 67-prevalent 28-incident	Modified Model Cutoff 2pts Any Delirium Sens .62 Spec .68 PPV .54 NPV .75 Incident Del	1st Val: 0.85 (0.80-0.88) Incident delirium 2nd Val: Cohort 1 (MMSE) 0.78 (0.68-0.88)	Martinez et al. 2012 Original Model Age >85 1 pt Dependent in ≥5 ADLs 1 pt Drugs on admit: 1pt/drug -Antidepressants 2pt/ -Antidementia antipsych -anticonvulsants -antipsychotics Score 0-3 Score >1 = high risk for delirium	Original Model: -No cognitive measure Modified Model: MMSE < 24 AMTS

Table 2

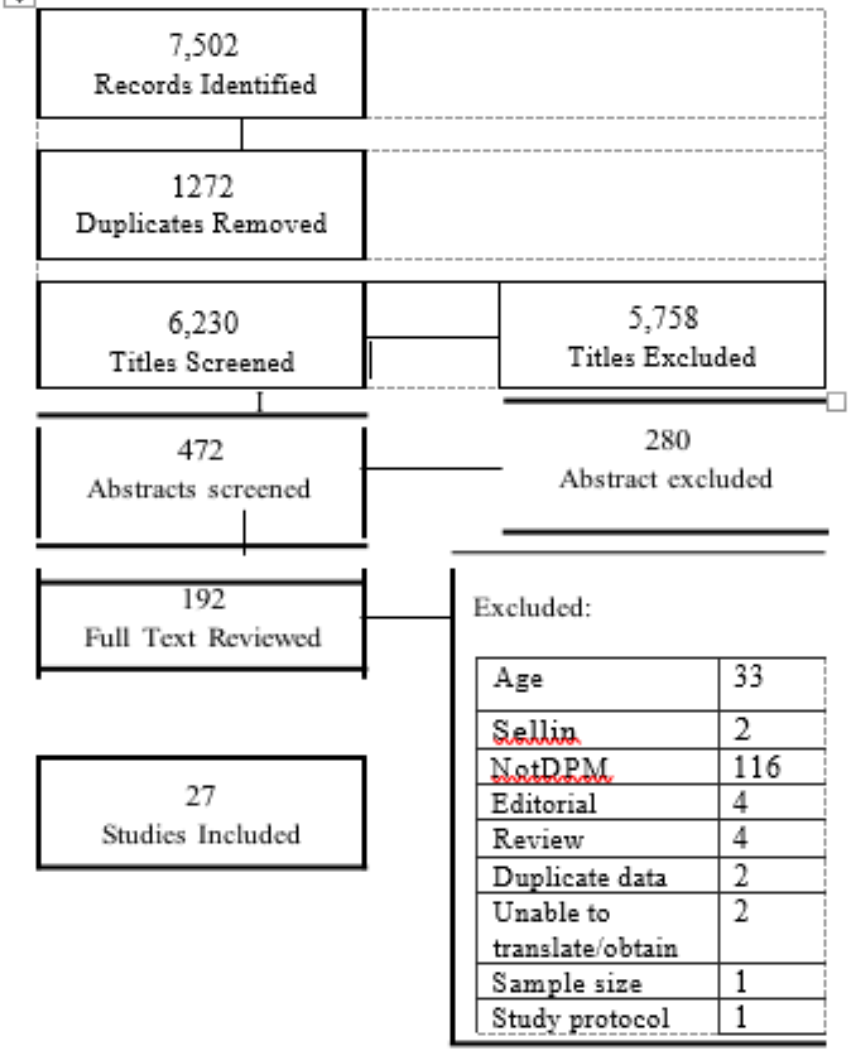
External Validated DPM Name	Citation	Delirium #(%)	Sens Spec PPV NPV (external)	AUROC (95%CI)	Model Components	Cog. Assess Tool & Cutoff
			Sens .81 Spec .68 PPV .29 NPV .96	Cohort 2 (AMTS) 0.75 (0.65-0.84)	Modified Model Age >85 — 1 pt Dependency in ≥ 5 ADLs — 1 pt Diag of Dementia MMSE<24 AMTS<9 — 1 pt	< 9
Pompei et al. 1994 model	Pompei et al. (1994)[44]	Dev: 64(15) Val: 86(26) (21=prevalent delirium)	Sens Spec PPV NPV *Pts stratifie low or moderat high-ris .83 .50 .38 .89	Dev: 0.74 +/- 0.05 Val: 0.64 +/- 0.05	Weighted Points Baseline cognitive impairment — 2 pt Depression — 2 pt Alcoholism — 3 pt > 4 comorbidities — 3 pt 0-3 pts = Low risk 4-7 pts = Moderate risk 8-10 pts = High risk	MMSE Less than HS <21 High school <23 College edu < 24
Precipitating Risk Factors	Inouye and Charpentier (1996)[30]	Dev: 35(18) Val: 47(15) (incident delirium)	Not reported	No AUROC reported	Physical restraint use — 1 pt Malnutrition — 1 pt ≥3 medications added — 1 pt Bladder catherization — 1 pt Any iatrogenic event — 1 pt Not weighted. 0 pt = Low Risk 1-2 pt = Intermediate ≥ 3 pt = High Risk	None used in model
Risk Model for Delirium (RD)	Moerman et al. (2012)[42]	Val: 102(27) (incident delirium)	Sens .81 Spec .56 PPV .41 NPV .89 Optimal cut-off score: 4 pts	Val: 0.73 (0.68-0.77)	Weighted Points Delirium-previous hospitalization — 5 pt Dementia — 5 pt Clock Drawing -Sm mistake — 1 pt -big mistake — 2 pt Age -70 to 85 years old — 1 pt - >85 years — 2 pt Impaired hearing — 1 pt Impaired vision — 1 pt Problems w/ADL -Help w/meal prep — .5p -help w/physical — .5p Use of heroin, methadone, morphine — 2 pt Daily >4 alcohol — 2 pt ≥ 5 pts = High risk	CDT -11:10 -Two Categories 1 Small mistakes 2 Big mistakes

Table 2

External Validated DPM Name	Citation	Delirium #(%)	Sens Spec PPV NPV (external)	AUROC (95%CI)	Model Components	Cog. Assess Tool & Cutoff

Key:
Dev=Development, Val=Validation
Sens=Sensitivity, Spec=Specificity, PPV=Positive Predictive Value, NPV=Negative Predictive Value
Area Under the Receiver Operating Curve Statistic, Dev=Development, Val=Validation, mRASS=Modified Richmond Agitation-Sedation Scale, TMTYB=The Months of the Year Backwards
ADL=Activities of Daily Living
MMSE=Mini Mental Status Exam, AMTS=Abbreviated Mental Test Score, CDT=Clock Drawing Test, mBDR=Modified Blessed Dementia Rating, MoCA=Montreal Cognitive Assessment.

Figure 1-PRISMA Diagram-Study Selection



Age	33
Sellin	2
NotDPM	116
Editorial	4
Review	4
Duplicate data	2
Unable to translate/obtain	2
Sample size	1
Study protocol	1

Figure 2: Frequency of Variable use in the 13 externally validated Delirium Prediction Models

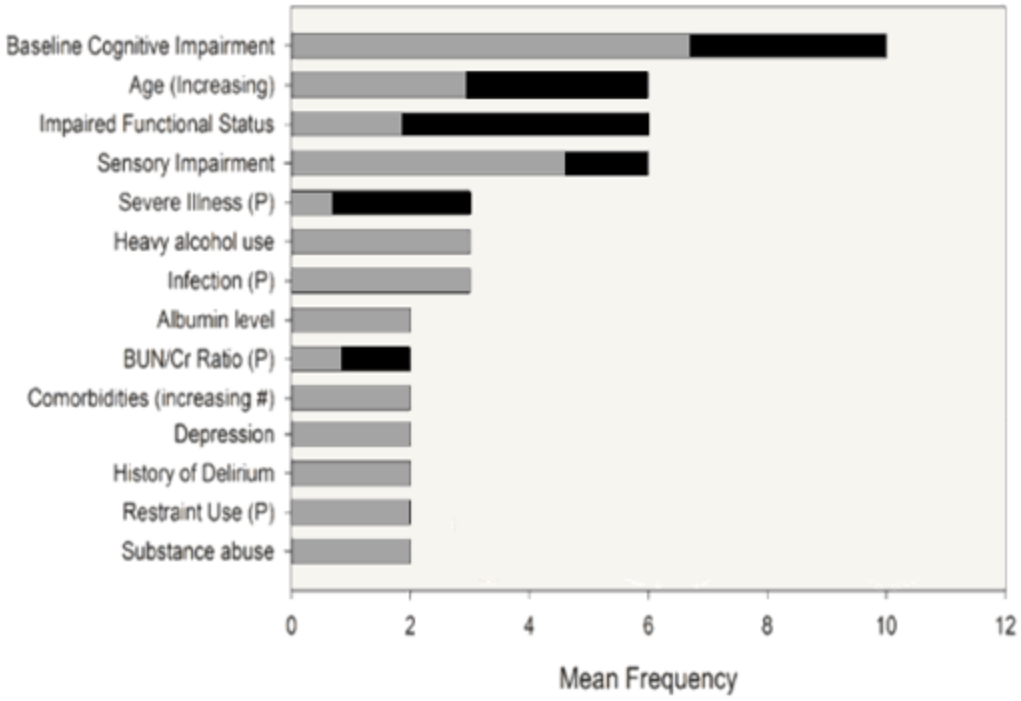


Figure 2 displays the mean frequency of variable use in the thirteen externally validated delirium prediction models. The black back represents the frequency of variable use in the top five moderately performing delirium predictive models (AUROC >0.75).
(P) indicates a precipitating risk factor used in a delirium prediction model
The following variables were used once and are not represented in this figure: addition of >3 medications, bladder catheter use, C-Reactive Protein, emergency surgery, presence of fracture upon admission, history of cerebrovascular accident, iatrogenic event, intensive care unit admission, low physical activity, malnutrition (using a validated scale) and open surgery.

Figure 3: AUROC Statistics of the 13 Externally Validated Delirium Prediction Models

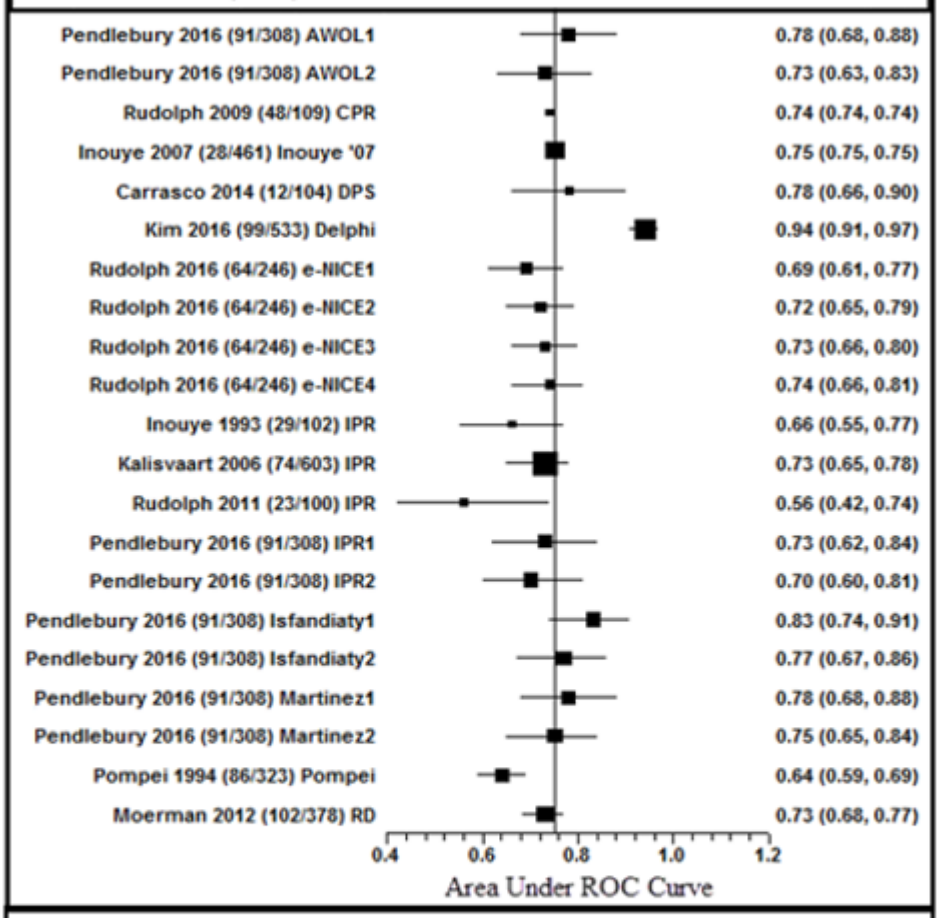


Figure 3 shows the published AUROC Statistic for the 13 externally validated Delirium Prediction Models

#D/N: Number of confirmed delirium in study/overall sample size of study
DPM: Delirium prediction model name. The corresponding number references the different AUROCs calculated based on different cognitive tests applied to the model by the authors
Squares w/error bars: Size of square corresponds to sample size of study
Vertical Line: Indicating, albeit arbitrary, a potential clinical relevance AUROC of 0.75. Those models with an AUROC greater than 0.75 are considered the highest performing models with moderate to high predictive ability.

<table>
<tr><td colspan="5">Appendix A – Review Protocol</td></tr>
<tr><td>Working Title of Review</td><td colspan="3">Systematic Review of Delirium Prediction Models</td><td>Support</td><td>Modifications</td></tr>
<tr><td>Authors</td><td colspan="3"><table>
<tr><td>1st & Corresponding</td><td>Heidi Lindroth</td><td>Literature search, data extraction, data synthesis and manuscript preparation.</td></tr>
<tr><td>Data Extraction</td><td>Heidi Lindroth
Suzanne Purvis</td><td>Literature search, data extraction, data synthesis.</td></tr>
<tr><td>Content Experts</td><td>Lisa Bratzke</td><td>Assisted with content related to cognition. Results review.</td></tr>
<tr><td></td><td>Roger Brown</td><td>Statistical content expert</td></tr>
<tr><td></td><td>Mark Coburn</td><td>Results review, Manuscript preparation</td></tr>
<tr><td></td><td>Marko Mrkobrada</td><td>Results review, Manuscript preparation</td></tr>
<tr><td></td><td>Matthew TV Chan</td><td>Results review, Manuscript preparation</td></tr>
<tr><td></td><td>Daniel Davis</td><td>Geriatrician expertise, reviewed results, manuscript preparation.</td></tr>
<tr><td></td><td>Pratik Pandharipande</td><td>Results review, Manuscript preparation</td></tr>
<tr><td></td><td>Cynthia M. Carlsson</td><td>Geriatrician expertise, reviewed results, manuscript preparation.</td></tr>
<tr><td>Mentoring</td><td>Robert D. Sanders</td><td>Mentoring author, resolved content/data disagreements b/w authors, manuscript preparation.</td></tr>
</table></td><td></td></tr>
<tr><td>Aim</td><td colspan="3">To identify existing prognostic delirium prediction models and evaluate their validity and statistical methodology in the older adult (≥60yo) acute hospital population.</td><td></td><td></td></tr>
<tr><td>Search Terms</td><td colspan="3">(“Delirium” OR “postoperative delirium” OR “ICU delirium” OR “ICU psychosis” OR “ICU syndrome” OR “acute confusional state” OR “acute brain dysfunction”) AND (“inpatient” OR “hospital*” OR “postoperative” OR surg* OR “critical care unit” OR “intensive care unit” OR CCU OR</td><td>UW-Madison Health Sciences librarian. Three meetings to refine search</td><td></td></tr>
</table>

	ICU) AND ("predict*" model OR risk*)	terms.	
Databases searched	PubMed, CINAHL, PsychINFO, Cochrane, SocINDEX and Medline	Health Sciences librarian.	Expanded to include SocINDEX
Timelines established	01/01/1990-12/31/2016		Originally was 12/31/15. Expanded to include all of 2016.
Inclusion criteria	• Age $\geq$ 60 • Inpatient population • Developing and/or validating a delirium prediction model		Age expanded from $\geq$ 70 years of age due to the literature
Exclusion criteria	• Emergency department • Hospice/palliative care • Pediatric population • Related to alcohol withdrawal • $\leq$50 sample size	Mentoring author	Sample size criteria added to build rigor in the studies that were included in the sys review
Selection process	Studies will be selected based on the inclusion/exclusion criteria. The data extraction authors (HL and SP) will conduct the literature search independently and meet monthly to discuss findings. Any disagreements will be resolved by the mentoring author (RDS)		
Data Management	A shared folder on the UW-Madison Box account will be created to share documents, data and meeting information.		
Data collection process	Data will be collected independently by HL and SP then data points will be shared at monthly meetings. Data collection tables will be created using Microsoft Excel then uploaded to the shared Box account. Any disagreement between authors will be resolved by the mentoring author (RDS).		
Data points collected	• Characteristics of studies (design, population, sample size) • Outcome measure including how it was identified, measured, defined. Prevalence. • Statistical methods applied • Statistical information about the delirium prediction models		

	(sensitivity, specificity, positive predictive value, negative predictive value, AUROC) • Characteristics of DPMs (variables used, scoring, development) • Cognitive measures used in studies. • Criteria to fulfill the Newcastle Ottawa Scale.		
Outcomes	• AUROC will be the primary outcome measure • Characteristics of DPMs (variables, statistics) • Cognitive tests used		
Data synthesis	The first/corresponding author (HL) will synthesize the data into the manuscript. The co-authors will verify this. RB will complete the meta-analysis.		
Manuscript preparation	HL will complete manuscript preparation. All co-authors are responsible for reviewing content and data to assure correctness and complete synthesis of data gathered.		

Manuscript 2

Title: Derivation of a simple postoperative delirium incidence and severity model

Target Journal: The British Journal of Anaesthesia (BJA), https://academic.oup.com/bja

- 250 word limit for abstract
- 3000 word limit total

Abstract

Background: Delirium is an important postoperative complication, yet a simple and effective delirium prediction model remains elusive. We investigated whether the combination of the National Surgical Quality Improvement Program (NSQIP) risk calculator for serious complications (NSQIP-SC) or risk of death (NSQIP-D), with cognitive tests of executive function (Trail Making Test A and B [TMTA, TMTB]), could provide a parsimonious model to predict postoperative delirium incidence or severity.

Methods: Data were collected from 100 adults (≥65yo) undergoing major non-cardiac surgery. In addition to NSQIP-SC, NSQIP-D, TMTA and TMTB, we collected age, sex, ASA score, smoking, type of surgery, depression, Framingham risk score, and preoperative blood pressure. Delirium was diagnosed with the Confusion Assessment Method (CAM) and the Delirium Rating Scale-R-98 (DRS) was used to assess symptom severity. LASSO and Best Subsets logistic and linear regression were employed in line with TRIPOD guidelines.

Results: Three subjects were excluded due to intraoperative deaths (2) and alcohol withdrawal (1). Ninety-seven participants with a mean age of 71.68±4.55, 55% male (31/97 CAM+, 32%) and a mean Peak DRS of 21.5±6.40 were analyzed. Of the variables included, only NSQIP-SC and TMTB were identified to be predictors of postoperative delirium incidence ($p<0.001$, AUROC 0.81, 95% CI: 0.72, 0.90) or severity ($p<0.001$, Adj. R2: 0.30).

Conclusions: In this cohort, preoperative NSQIP-SC and TMTB were identified as predictors of postoperative delirium incidence and severity. Future studies should verify whether this two-factor model can be used for accurate delirium prediction.

Key words: aging, delirium, perioperative, prediction, surgical risk Clinical Trials:

NCT03124303, NCT01980511

Introduction

Delirium, an acute brain failure, is a common surgical complication experienced by approximately 50% of patients, incurring an estimated annual U.S. cost of $152 billion[1-4]. As such, delirium is a crucial public health concern as it is significantly associated with increased mortality[5] and morbidity[6], subsequent cognitive decline[7-9], and loss of independence.[10 11] One in three cases of delirium may be preventable when multicomponent delirium prevention measures are implemented[12 13]. Yet, clinicians are encumbered by an arduous list of potential patient and perioperative risk factors to identify at risk individuals. Prediction models facilitate the identification of high-risk individuals and are evaluated on their statistical ability to accurately distinguish between health and disease. The area under the receiver operator curve (AUROC) statistic evaluates predictive ability with a value of 0.5 being no better than chance at 1.0 being perfect. Moderate predictive ability was identified in our recent systematic review with four postoperative delirium prediction models reporting area under the receiver operator curves (AUROC) of 0.73-0.94. While an AUROC of 0.94 is excellent, this model is multifactorial and data collection included postoperative data, inflating model performance, and invalidating its use as a preoperative prediction model. Further, we did not identify a prediction model for delirium severity in the literature[14]. A severity model may have most utility as it may identify the subjects that really must have a delirium prevention plan in place prior to surgery. An ideal model that is clinically applicable would be both predictive of delirium incidence and its severity and be parsimonious, such that it would use a limited number of variables that can be cheaply and easily obtained.

The ability to identify high-risk individuals prior to their surgery is critical to delirium prevention. To identify potential candidate predictors, we considered the pathogenesis of

delirium and sought to identify both predisposing (age, depression, and medical comorbidities) and precipitating factors (surgery). The online surgical risk calculator built and copyrighted by the American College of Surgeons, National Surgical Quality Improvement Program (ACS NSQIP)[15 16] offers an avenue to input several predisposing risk factors and include the estimated magnitude of the precipitating event, the surgery. Built using data from 1,414,006 patients including 1,557 distinct surgical procedures, the NSQIP risk score has been widely validated and applied to predict outcomes in various surgical populations,[15-20] but has not been applied in delirium prediction. The estimated risk of serious complications (NSQIP-SC) and death (NSQIP-

D) are applicable to delirium, and we hypothesized that NSQIP-SC would be a stronger predictor of delirium over NSQIP-D as the causal relationship between delirium and complications is likely stronger than the association between the risk of death, calculated preoperatively, and postoperative delirium[21] Further, our recent systematic review identified that current delirium prediction models do not evaluate specific cognitive domains, such as executive function.[14] Executive function facilitates attention and problem-solving[22] Significant associations between preoperative executive function and postoperative delirium incidence have been reported.[23-26] As NSQIP does not include a cognitive factor, we sought to enhance delirium prediction through the combination of medical factors and a measure of executive function. Our aim was to examine the predictive ability of the NSQIP risk scores with a measure of executive function among other potential delirium risk factors and to develop a parsimonious model to predict postoperative delirium incidence and severity for future use in delirium prediction.

Methods

This analysis is a sub-study drawn from an ongoing prospective perioperative cohort study that is

approved by the University of Wisconsin-Madison, Health Sciences Institutional Review Board (#2015-074) and registered with ClinicalTrials.gov (ref: NCT03124303, NCT01980511).

Between August 2015 and May 2018, 1,049 potential participants were screened from vascular, urology, general and spine surgical clinics. As shown in Figure 1, 100 subjects were recruited and 97 were included in the final analysis.

Preoperative Predictors and Assessment

Preoperatively, participants underwent an interview and completed measurements of executive function, functional status, depression, and a formal delirium assessment using the Confusion Assessment Method (CAM)[27]. Executive function was assessed through the use of two well-validated and widely used measures, Trail Making Test A (TMTA) and Trail Making Test B (TMTB)[28]. These tests are completed by connecting a series of circles in ascending order. TMTA is composed of 25 encircled numbers; TMTB alternates between 23 encircled numbers and letters. Scoring is based on time to completion, a longer completion time indicates worse executive functioning. Raw scores are used in the analysis. Functional ability was assessed using the Instrumental Activities of Daily Living (iADL)[29]. Depression was assessed using the Geriatric Depression Scale (GDS) composed of 15, yes/no questions that are widely validated in the older adult population[30]. Demographic information included age, sex, years of education, and tobacco use history. Data extracted from the electronic health record included preoperative blood pressure values, height, weight, past medical history including comorbidities, current outpatient medication use, and the American Society of Anesthesiologists Physical Status (ASA) classification for the planned surgical procedure. ASA classification ranges from 1-6 and are determined by the attending anesthesiologists prior to surgical intervention [31-33]. Each level has a specific definition with higher levels indicating increasing levels of disease, comorbidities and

risk. Vascular surgery (yes/no) was selected a priori to be included as a covariate to examine whether a planned surgery was sufficient to predict delirium or if the additional information provided by a surgical risk score was necessary for a prediction model. Perioperative and cardiovascular risk were assessed through three different metrics; ASA31, Framingham Cardiovascular Disease 10-year Risk Calculator (Framingham CVD)[34 35], and the ACS NSQIP online risk calculator15 16 20 36. Framingham CVD was calculated through the online, interactive risk score calculator including BMI using data extracted from electronic health records and participant interview. This calculator includes sex, age, systolic blood pressure (SBP), treatment for hypertension, current smoker, diabetes and BMI. The most recent SBP value prior to surgery was used and BMI was calculated from collected height/weight values. The ACS NSQIP online surgical risk calculator, (http://riskcalculator.facs.org) was used to obtain the risk scores for serious complications (NSQIP-SC) and death (NSQIP-D). This calculator employs twenty patient predictors and pairs these with the Current Procedural Terminology (CPT) code using a linear risk approach, providing a risk score specific to that specified procedure.15 The inputted variables are as follows: age, sex, functional status (independent, partially dependent, dependent), emergency case, ASA classification, steroid use for chronic condition, ascites within 30-day prior to surgery, systemic sepsis within 48-hours prior to surgery, ventilator dependency, disseminated cancer, diabetes, hypertension with medications, congestive heart failure (within 30-days prior to surgery), dyspnea, current smoker (within 1-year), history of severe COPD, dialysis, acute renal failure, height and weight as well as surgical procedure. There are 1,557 distinct CPT codes, ranging from minor surgeries such as a cholecystectomy to major surgeries such as thoracoabdominal aortic aneurysm. This calculator is completed online from information contained in the electronic health record and supplemented

with patient report. Currently, ACS NSQIP does not allow the data extraction to be automated. Nonetheless, it encapsulates several medical comorbidities and provides an estimate of the risk based on the planned surgery. Supplementary figure 1 provides further information on the ACS NSQIP calculator.

Delirium Assessment

Pre- and postoperatively, participants were formally assessed for delirium and symptoms using the widely-validated Confusion Assessment Method (CAM)[27], 3D-CAM[37], and Delirium Rating Scale-R-98 (DRS)[38] twice daily, between the hours of 0500-1000 and 1600-2200 on postoperative days 1-4. The CAM and 3D-CAM were administered concurrently to provide both a global and comprehensive view of delirium symptoms while providing a structured interview format. If the participant was CAM positive at the postoperative day 4, afternoon assessment, the participant was followed until delirium resolved to collect data on delirium duration. If participants were ventilated in the intensive care unit (ICU), the CAM-ICU[39] [40] was administered. The research team met at least weekly to discuss delirium assessment findings and DRS ratings. The DRS-R-98 (DRS) is a 16-item assessment tool that measures delirium symptoms and severity. The max score is 44-points, an increasing score indicates worse delirium.[38]

Research Team

Each research team member underwent intensive training on delirium interview completion including CAM, 3D-CAM, CAM-ICU, and DRS. CAM training was completed as part of the NeuroVISION cohort study[41] and the first author was officially trained on the CAM at the 2016 CEDARTREE Delirium Bootcamp and the CAM-ICU at the 2016 American Delirium Society pre-conference. Team members viewed 6 videos on CAM and 3D-CAM administration from the

Hospital Elder Life Program[42] website followed by an interactive team discussion on CAM/3D-CAM completion and observations. Team members shadowed the first author for six in-person CAM/3D-CAM assessments followed by in-depth discussion on observations, symptom rating (DRS), and bedside manner. The first author shadowed the team members for six CAM/3D-CAM/CAM-ICU/DRS assessments to ensure competency.

Sample Size

The sample size was determined based on the need for a parsimonious delirium prediction model. We estimated three to four risk factors would form a reasonable clinical model if an AUROC >0.80 was obtained. Sample size was based on logistic regression and determined using the rule of 8-10 outcome events (delirium) per variable[43 44] with a known delirium incidence rate of 32%. The decision to analyze was made after 100 participants were recruited.

Statistical Analysis

Patient characteristics were described using means $\pm$ standard deviations for continuous variables and frequency counts with percentages for categorical variables. Dependent on the distribution of the data confirmed by statistical plots, continuous variables were compared using Student's t-test or Mann-Whitney U-test. Categorical variables were compared using *X2*. The outcome variables were Delirium Yes/No (DELYN) for logistic regression and delirium severity using the Peak DRS Total Score (DRS) for linear regression. Missing data was identified in the following variables (#missing): TMTB (1), GDS15 (1), TMTA (6), and Tobacco Pack Years (10). Little's test of missing completely at random (MCAR) was not significant indicating that the missing data are missing completely at random and do not influence the analysis. Therefore, as outlined by Little (1988) the listwise deletion of participants with missing data is appropriate.[45] Significance was notated with a p-value $\leq$0.05. NCSS v12.0, Stata/IC v15.0 and R v1.1453 were

used for statistical analysis. These statistical packages were necessary to verify results across packages.

First, to evaluate the predictive ability of NSQIP-SC over NSQIP-D, logistic (DELYN) regression model were completed. The predictive ability of both ASA classification and the Framingham Cardiovascular Risk score were also evaluated. Second, a delirium prediction model was developed. We did not employ univariate statistics to select candidate predictors as this may lead to poor performing predictors and overfitting.[46] To counter the effects of small sample sizes and reduce bias within data, we employed a statistical shrinkage regression technique, using Least Absolute Shrinkage and Selection Operator (LASSO).[47-49] This technique reduces the noise within the data, allowing true signals to be detected and avoids common problems such as model overfitting. Candidate variables demonstrating the smallest Mallow's C*p* value[50], indicating precise predictors, were then applied in Best Subsets regression. Best Subsets regression is an automated regression approach that evaluates all possible combinations of candidate predictors.[51] [52] The output provides a set of models with model fit statistics. Model selection was based on assessment of model fit using Akaike information criteria (AIC), Bayesian information criteria (BIC), and McKelvey and Zavoina's R2.[53] The area under the receiver operating characteristic curve (AUROC) with 95% CI was calculated. Calibration was assessed through goodness-of-fit tests calculated by the Hosmer-Lemeshow statistic. Sensitivity, specificity, positive predictive and negative predictive values were calculated and reported.

The peak delirium severity (DRS) score was transformed using the Box-Cox Method[54] with the optimal Lambda value due to the positive skew, please refer to supplementary figure 2 for raw and transformed plots. The regression modeling procedures outlined in the paragraph above for logistic regression were repeated for the linear regression model. Model selection was based on

assessment of model fit using Akaike information criteria (AIC), Bayesian information criteria (BIC), and adjusted R2.[53]

Results

Thirty-one participants (32%) experienced postoperative delirium with a mean peak DRS severity total score of 21.48 (±SD 6.40). The median delirium duration was one day (24-hours). Participant characteristic are summarized in Table 1.

Significant associations with postoperative delirium incidence were demonstrated with the preoperative NSQIP risk scores of serious complications (NSQIP-SC) and death (NSQIP-D), the executive function tests, TMTA and TMTB, vascular surgery, and ASA status (univariate, $p<0.05$). Significant pairwise correlations were demonstrated between the DRS and NSQIP-SC, NSQIP-D, TMTA and TMTB (univariate, $p<0.05$). No significant differences were identified between the non-delirious and delirious groups in terms of age, sex, education level, past/present tobacco use, blood pressure metrics, functional status, and Framingham Risk Score.

Derivation of a Delirium Incidence Prediction Model

First, we confirmed preoperative NSQIP-SC as a robust predictor of postoperative delirium incidence using single factor logistic regression models for NSQIP-SC and NSQIP-D. Both demonstrated moderate to fair predictive ability with an AUROC of 0.76 (95% CI: 0.66, 0.87) and 0.73 (95% CI: 0.62, 0.84), respectively. Although similar in their projected ability to predict individuals at a higher risk for delirium development, strong support for the NSQIP-SC model, over the NSQIP-D model, is provided by optimal AIC (1.086) and the BIC (-333.254) metrics. We further evaluated NSQIP-SC as a robust predictor of postoperative delirium incidence against the ASA classification and Framingham risk score. The NSQIP-SC AUROC of 0.76

(95% CI: 0.66, 0.87) is robust compared to ASA classification (AUROC 0.63, 95% CI: 0.53, 0.73) and Framingham risk score (0.53, 95% CI: 0.40, 0.65). The reported odds ratios, coefficients, AUROC, sensitivity and specificity, and model fit statistics including AIC and BIC are in supplementary table 2.

Second, we applied the statistical shrinkage technique followed by Best Subsets regression using the variables age, sex, NSQIP-SC, NSQIP-D, tobacco pack years, vascular surgery, ASA classification, Framingham risk score, GDS15, TMTA, TMTB and blood pressure metrics. To predict postoperative delirium incidence, a two-factor logistic regression model containing preoperative NSQIP-SC and TMTB was identified from best model fit statistics by both LASSO and Best Subsets regression. Hosmer-Lemeshow goodness-of-fit test was not significant (p=0.37), indicating accurate model calibration. The model maintained significance and predictive ability (AUROC 0.81, 95% CI: 0.72-0.90). The logistic regression models including classification metrics are displayed in Table 2.

Derivation of a Delirium Symptom Prediction Model

Preoperative NSQIP-SC was confirmed as a robust predictor of postoperative peak DRS using simple linear regression models for NSQIP-SC ($p<0.0001$, AdjR2: 0.184). Similar to the logistic regression results above NSQIP-D was also significantly associated with DRS (p=0.04, AdjR2: 0.03). However, the NSQIP-SC model demonstrated higher adjusted R^2, and lower AIC (1.525) and BIC (-290.718) metrics, providing strong support for that predictor. Similar to logistic regression for postoperative delirium incidence, ASA (p=0.04, AdjR2: 0.03) and Framingham risk score (p=0.69, AdjR2: -0.01) did not demonstrate a strong predictive relationship with DRS.

The model demonstrating optimal fit statistics for postoperative peak delirium severity was a

two-factor linear regression model containing preoperative NSQIP-SC and TMTB using LASSO and Best Subsets regression. This two-factor model reports an adjusted R2 of 0.30 ($p<0.001$), demonstrating the ability to explain 30% of the variability in observed delirium symptoms. For every 1-percentage increase in the NSQIP-SC score, the DRS will increase by 0.42 points.

Further model details are displayed in Table 2. Age, sex, NSQIP-D, ASA, tobacco pack years, vascular surgery, Framingham risk score, GDS-15 score, TMTA, and blood pressure metrics were not identified as significant predictors.

Discussion

Our analysis of a prospective perioperative cohort found the National Surgical Quality Improvement Program online risk calculator score for serious complications (NSQIP-SC) to be a robust predictor of postoperative delirium incidence or severity. The preoperative NSQIP-SC score combines data on both predisposing risk factors and the estimated magnitude of the surgery, *i.e.,* the precipitating event and is well positioned to contribute information on the risk of delirium. While NSQIP-D was found to be significantly associated with postoperative delirium incidence and severity, it was not selected in the prediction modeling procedures. The addition of the executive function measure, Trail Making Test B, improved performance of predictions models for postoperative delirium incidence and symptom severity. Previous research has identified executive function to be significantly associated with delirium incidence.[23 25 26 55]

However, as identified by our recent systematic review of delirium prediction models in older adults, an executive function measure has not been applied in model development[14]. Given the breakdown in executive function during delirium and prior data on the predisposition to delirium

by impaired cognition, incorporating a cognitive variable is both biologically and statistically important. This study expands current knowledge by examining both the utility of NSQIP risk scores and executive function in predicting delirium incidence and severity. A recent systematic review questioned the strength of the association between postoperative delirium and mortality. Furthermore, delirium often results from a complicated perioperative course; particularly following major surgery.[21] Hence *a priori* we hypothesized that NSQIP-SC would perform better than NSQIP-D for predicting postoperative delirium incidence and severity.

Strengths

The strengths of this study include its prospective perioperative design, statistical methods chosen, and rigorous delirium assessments including outcomes based on incidence and severity of delirium. This novel application of a readily available, simple tool has potential for broad application in delirium-focused clinical care. The NSQIP-SC score combines several potential preoperative risk factors for delirium (age, functional status, current tobacco use, vascular burden) with the precipitating event, the planned surgery and provides a single risk score that is easy to interpret, *i.e.,* a 5 point increase in NSQIP-SC results in a 10% increase in the probability of that patient experiencing delirium. Given that the subject becomes becomes delirious postoperatively, quantifying the potential impact of this precipitating event is clearly a key feature of a delirium prediction model.

Delirium represents a severe breakdown in an individual's executive function. The *Cognitive Disintegration Model* posits that there is a critical threshold (*i.e.,* the "Delirium Threshold") in cognition (or network connectivity) that must be crossed for delirium to result. Individuals with more vulnerability (predisposing factors) going into surgery, *i.e.,* older age, multiple comorbidities and worsening executive function, are closer to that "Delirium Threshold" and

may need a smaller precipitating event to push them into delirium. In this context, inclusion of executive function that is not captured in NSQIP-SC, improved model performance.

Furthermore, this two-factor model captures information on both the predisposing and precipitating factors for delirium. Its parsimonious nature is a clinical strength.

Limitations

This study has several limitations to consider. While these models were built using statistical methods optimized for modeling a limited number of events, the small sample size may still have had an effect on the results. The recommendations by the TRIPOD guidelines and the CHARMS checklist were followed to use statistical shrinkage procedures were undertaken to minimize model overfitting. Nonetheless, it is worth emphasizing the study sample is small with only 31 delirious participants. The population is largely homogenous in terms of years of education and ethnicity. In larger and more diverse populations, additional factors may enhance model performance. Future studies should focus on the broad external validation of these models following the statistical methodology outlines by the TRIPOD guidelines. This requires a new perioperative cohort study that will recruit patients from diverse populations and will be the target of future grant applications.

Conclusion

This analysis of a prospective perioperative cohort study identified NSQIP-SC and TMT-B as predictors for delirium incidence and severity. A preliminary delirium prediction model was created for both delirium incidence and severity and this should be further tested in future studies.

References

1. Inouye SK, Westendorp RG, Saczynski JS. Delirium in elderly people. *Lancet* 2014;383(9920):911-22. doi: 10.1016/s0140-6736(13)60688-1 [published Online First: 2013/09/03]
2. Leslie DL, Inouye SK. The importance of delirium: economic and societal costs. *Journal of the American Geriatrics Society* 2011;59 Suppl 2:S241-3. doi: 10.1111/j.1532-5415.2011.03671.x [published Online First: 2011/12/07]
3. Leslie DL, Marcantonio ER, Zhang Y, *et al.* One-year health care costs associated with delirium in the elderly population. *Archives of internal medicine* 2008;168(1):27-32. doi: 10.1001/archinternmed.2007.4 [published Online First: 2008/01/16]
4. Diagnostic and Statistical Manual of Mental Disorders-5. Diagnostic and Statistical Manual of Mental Disorders-5: American Psychiatric Association 2013.
5. Ha A, Krasnow RE, Mossanen M, *et al.* A contemporary population-based analysis of the incidence, cost, and outcomes of postoperative delirium following major urologic cancer surgeries. *Urologic oncology* 2018 doi: 10.1016/j.urolonc.2018.04.012 [published Online First: 2018/05/29]
6. Dasgupta M, Brymer C. Poor functional recovery after delirium is associated with other geriatric syndromes and additional illnesses. *International Psychogeriatrics* 2015;27(5):793-802. doi: 10.1017/S1041610214002658
7. Inouye SK, Marcantonio ER, Kosar CM, *et al.* The short-term and long-term relationship between delirium and cognitive trajectory in older surgical patients. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 2016 doi: 10.1016/j.jalz.2016.03.005
8. Hshieh TT, Inouye SK, Oh ES. Delirium in the Elderly. *The Psychiatric clinics of North*

America 2018;41(1):1-17. doi: 10.1016/j.psc.2017.10.001 [published Online First: 2018/02/08]

9. Oh ES, Fong TG, Hshieh TT, *et al.* Delirium in older persons: Advances in diagnosis and treatment. *Jama* 2017;318(12):1161-74. doi: 10.1001/jama.2017.12067
10. Drews T, Franck M, Radtke FM, *et al.* Postoperative delirium is an independent risk factor for posttraumatic stress disorder in the elderly patient: a prospective observational study. *European journal of anaesthesiology* 2015;32(3):147-51. doi: 10.1097/EJA.0000000000000107
11. Sanders RD, Pandharipande PP, Davidson AJ, *et al.* Anticipating and managing postoperative delirium and cognitive decline in adults. *BMJ (Clinical research ed)* 2011;343:d4331.
12. Siddiqi N, Harrison JK, Clegg A, *et al.* Interventions for preventing delirium in hospitalised non-ICU patients. *Cochrane Database Syst Rev* 2016;3:CD005563. doi: 10.1002/14651858.CD005563.pub3
13. Hshieh TT, Yue J, Oh E, *et al.* Effectiveness of Multicomponent Nonpharmacological Delirium Interventions: A Meta-analysis. JAMA internal medicine2015.
14. Lindroth H, Bratzke L, Purvis S, *et al.* Systematic review of prediction models for delirium in the older adult inpatient. *BMJ Open* 2018;8(4):e019223. doi: 10.1136/bmjopen-2017-019223 [published Online First: 2018/05/01]
15. Bilimoria KY, Liu Y, Paruch JL, *et al.* Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *Journal of the American College of Surgeons* 2013;217(5):833-42.e1-3. doi: 10.1016/j.jamcollsurg.2013.07.385

16. Shiloach M, Frencher SK, Jr., Steeger JE, *et al.* Toward robust information: data quality and inter- rater reliability in the American College of Surgeons National Surgical Quality Improvement Program. *Journal of the American College of Surgeons* 2010;210(1):6-16. doi: 10.1016/j.jamcollsurg.2009.09.031
17. Kubasiak JC, Landin M, Schimpke S, *et al.* The effect of tobacco use on outcomes of laparoscopic and open ventral hernia repairs: a review of the NSQIP dataset. *Surg Endosc* 2017;31(6):2661-66. doi: 10.1007/s00464-016-5280-4 [published Online First: 2016/10/19]
18. Cohen ME, Liu Y, Huffman KM, *et al.* On-demand Reporting of Risk-adjusted and Smoothed Rates for Quality Profiling in ACS NSQIP. *Annals of surgery* 2016 doi: 10.1097/SLA.0000000000001551
19. Guillamondegui OD, Gunter OL, Hines L, *et al.* Using the National Surgical Quality Improvement Program and the Tennessee Surgical Quality Collaborative to improve surgical outcomes. *Journal of the American College of Surgeons* 2012;214(4):709-14; discussion 14-16. doi: 10.1016/j.jamcollsurg.2011.12.012
20. Hamilton GM, Wheeler K, Di Michele J, *et al.* A Systematic Review and Meta-analysis Examining the Impact of Incident Postoperative Delirium on Mortality. *Anesthesiology* 2017;127(1):78-88. doi: 10.1097/aln.0000000000001660 [published Online First: 2017/05/02]
21. Diamond A. Executive functions. *Annu Rev Psychol* 2013;64:135-68. doi: 10.1146/annurev-psych- 113011-143750 [published Online First: 2012/10/02]
22. Fong TG, Hshieh TT, Wong B, *et al.* Neuropsychological Profiles of an Elderly Cohort Undergoing Elective Surgery and the Relationship Between Cognitive Performance and

Delirium. *Journal of the American Geriatrics Society* 2015;63(5):977-82 6p. doi: 10.1111/jgs.13383

23. Greene NH, Attix DK, Weldon BC, *et al.* Measures of executive function and depression identify patients at risk for postoperative delirium. *Anesthesiology* 2009;110(4):788-95 8p.
24. Smith PJ, Attix DK, Weldon BC, *et al.* Executive function and depression as independent risk factors for postoperative delirium. *Anesthesiology* 2009;110(4):781-87 7p.
25. Rudolph JL, Jones RN, Grande LJ, *et al.* Impaired executive function is associated with delirium after coronary artery bypass graft surgery. *Journal of the American Geriatrics Society* 2006;54(6):937- 41. doi: 10.1111/j.1532-5415.2006.00735.x
26. Inouye SK, van Dyck CH, Alessi CA, *et al.* Clarifying confusion: the confusion assessment method. A new method for detection of delirium. *Annals of internal medicine* 1990;113(12):941-8. [published Online First: 1990/12/15]
27. Strauss E. A compendium of neuropsychological tests : administration, norms, and commentary: 3rd ed. Oxford ; New York : Oxford University Press, 2006. 2006.
28. Graf C. The Lawton Instrumental Activities of Daily Living Scale. *American Journal of Nursing* 2008;108(4):52-63.
29. Dennis M, Kadri A, Coffey J. Depression in older people in the general hospital: a systematic review of screening instruments. Age and ageing. England2012:148-54.
30. Delegates AHo. ASA Physical Status Classification System https://www.asahq.org/resources/clinical- information/asa-physical-status-classification-system: American Society of Anesthesiologists 2014 [updated 2014. Available from: https://www.asahq.org/resources/clinical- information/asa-physical-status-classification-

system accessed June 3, 2018 2018.

31. Parenti N, Reggiani ML, Percudani D, *et al.* Reliability of American Society of Anesthesiologists physical status classification. *Indian J Anaesth* 2016;60(3):208-14. doi: 10.4103/0019- 5049.177875 [published Online First: 2016/04/08]
32. Dripps RD, Lamont A, Eckenhoff JE. The role of anesthesia in surgical mortality. *Jama* 1961;178:261-6. [published Online First: 1961/10/21]
33. D'Agostino RB, Sr., Vasan RS, Pencina MJ, *et al.* General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008;117(6):743-53. doi: 10.1161/circulationaha.107.699579 [published Online First: 2008/01/24]
34. D'Agostino V, Pencina, Wolf, Cobain, Massaro, Kannel. Framingham Heart Study; Cardiovascular Disease (10-year risk) 2008 [Available from: https://www.framinghamheartstudy.org/fhs-risk- functions/cardiovascular-disease-10-year-risk/# accessed June 3 2018 2018.
35. American College of Surgeons NSQIP. Surgical Risk Calculator 2018 [Available from: https://riskcalculator.facs.org/RiskCalculator/ accessed December 2017-May 15th, 2018 2018.
36. Marcantonio ER, Ngo LH, O'Connor M, *et al.* 3D-CAM: derivation and validation of a 3-minute diagnostic interview for. *Annals of internal medicine* 2014;161(8):554-61. doi: 10.7326/M14- 0865
37. Trzepacz PT, Mittal D, Torres R, *et al.* Validation of the Delirium Rating Scale-revised-98: comparison with the delirium rating scale and the cognitive test for delirium. *The Journal of neuropsychiatry and clinical neurosciences* 2001;13(2):229-42.
38. Ely EW, Inouye SK, Bernard GR, *et al.* Caring for the critically ill patient. Delirium in

mechanically ventilated patients: validity and reliability of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). *JAMA: Journal of the American Medical Association* 2001;286(21):2703.

39. Ely EW, Margolin R, Francis J, *et al.* Evaluation of delirium in critically ill patients: validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). *Critical Care Medicine* 2001;29(7):1370-79.
40. Devereaux PJ, Mrkobrada M. NeuroVISION. Detection and neurological impact of cerebrovascular events in noncardiac surgical patients: a cohort evaluation. In: 3.0 PV, ed. Hamilton, Ontario, Canada: Population Health Research Institute, 2014:1-51.
41. Program HEL. Hospital Elder Life Program (HELP) for Prevention of Delirium: The Hospital Elder Life Program, 2018.
42. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol* 2007;165(6):710-8. doi: 10.1093/aje/kwk052 [published Online First: 2006/12/22]
43. Green SB. HOW MANY SUBJECTS DOES IT TAKE TO DO A REGRESSION-ANALYSIS. *Multivariate Behav Res* 1991;26(3):499-510. doi: 10.1207/s15327906mbr2603_7
44. Roderick JAL. A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association* 1988;83(404):1198-202. doi: 10.2307/2290157
45. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. *J Clin Epidemiol* 2016;76:175-82. doi: 10.1016/j.jclinepi.2016.02.031 [published Online First: 2016/03/12]

46. Lockhart R, Taylor J, Tibshirani RJ, *et al.* A SIGNIFICANCE TEST FOR THE LASSO. *Annals of statistics* 2014;42(2):413-68. doi: 10.1214/13-aos1175 [published Online First: 2015/01/13]

47. Tibshirani R, Bien J, Friedman J, *et al.* Strong rules for discarding predictors in lasso-type problems. *Journal of the Royal Statistical Society Series B, Statistical methodology* 2012;74(2):245-66. doi: 10.1111/j.1467-9868.2011.01004.x [published Online First: 2012/03/01]

48. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med* 1997;16(4):385-95. [published Online First: 1997/02/28]

49. Mallows CL. Some Comments on C_P. *Technometrics* 1973;15(4):661-75. doi: 10.2307/1267380

50. King JE. Running a Best-Subsets Logistic Regression: An Alternative to Stepwise Methods. *Educational and Psychological Measurement* 2003;63(3):392-403. doi: 10.1177/0013164403063003003

51. Hosmer DW, Jovanovic B, Lemeshow S. Best Subsets Logistic Regression. *Biometrics* 1989;45(4):1265-70. doi: 10.2307/2531779

52. Steyerberg EW, Vickers AJ, Cook NR, *et al.* Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology (Cambridge, Mass)* 2010;21(1):128-38. doi: 10.1097/EDE.0b013e3181c30fb2 [published Online First: 2009/12/17]

53. Arzideh F, Wosniok W, Haeckel R. Indirect reference intervals of plasma and serum thyrotropin (TSH) concentrations from intra-laboratory data bases from several German and Italian medical centres. *Clinical chemistry and laboratory medicine* 2011;49(4):659-

64. doi: 10.1515/cclm.2011.114 [published Online First: 2011/02/24]

54. Price CC, Garvan C, Hizel LP, *et al.* Delayed Recall and Working Memory MMSE Domains Predict Delirium follwing Cardiac Surgery. *J Alzheimers Dis* 2017;59(3):1027-35. doi: 10.3233/jad- 170380 [published Online First: 2017/07/13]

Figure 1 displays the study screening, recruitment, consent, and attrition numbers.

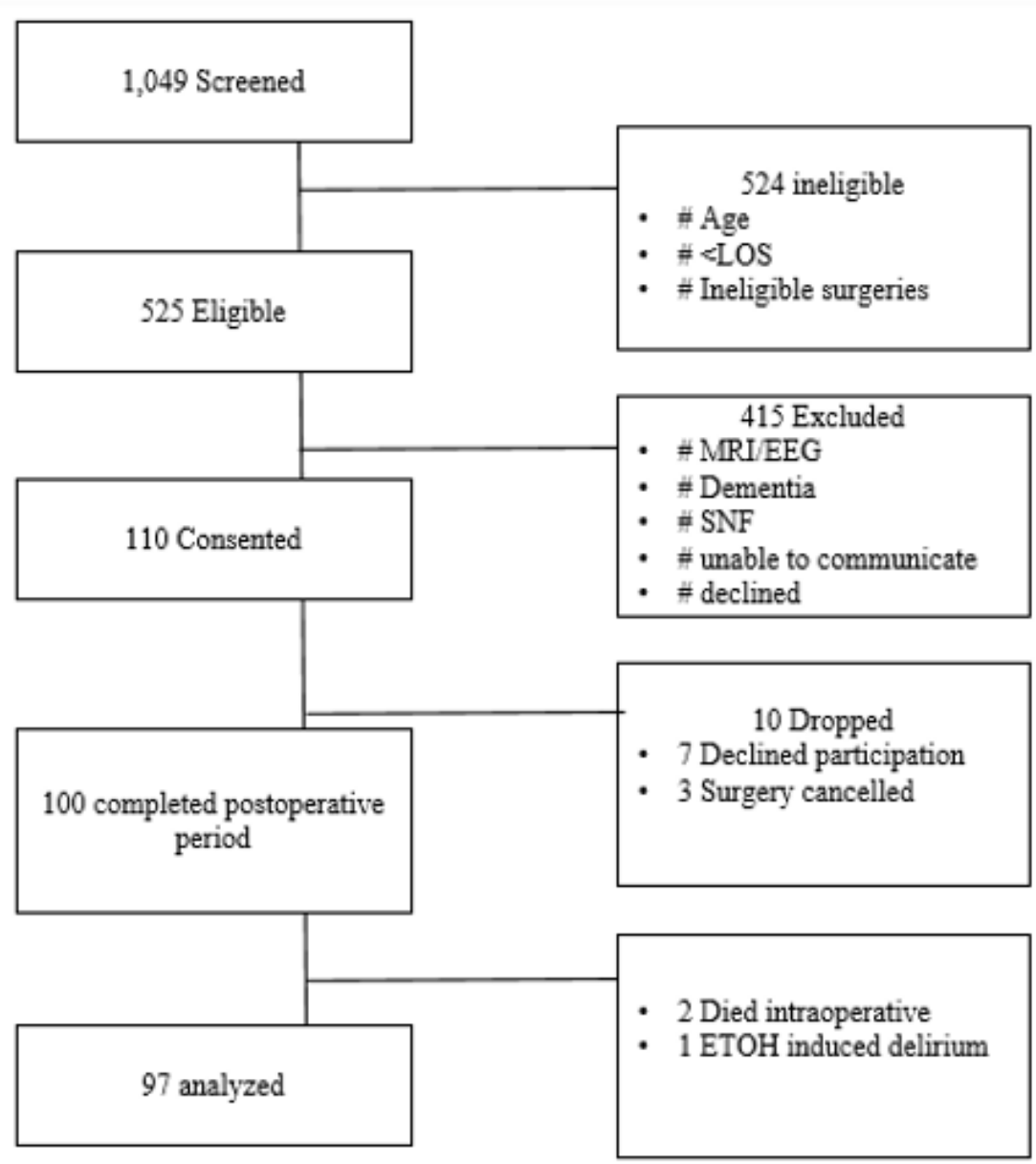


Table 1: Description of sample and significant differences between no delirium and			
Variable	**Mean (SD)**	**No Delirium**	**Delirium**
Age	71.68 (4.55)	71.71 (4.79)	71.61 (4.09)
NSQIP-SC	17.56 (11.66)	13.93 (9.34)	25.27 (12.49)***
NSQIP-D	2.52 (3.87)	1.85 (3.45)	3.95 (4.38)**
Framingham Risk	35.86 (19.6)	35.99 (20.74)	26.52 (16.98)
ASA	2.64 (3.87)	2.56 (0.61)	2.84 (0.52)*
Preoperative SBP	133 (17)	133 (17)	135 (17)
Preoperative DBP	74 (10)	74 (11)	74
Preoperative PP	59 (16)	59 (16)	61
Preoperative MAP	94 (11)	94 (12)	94
Tobacco Pack Years	18 (24)	16 (22)	23
GDS-15 (93)	2.44 (2.51)	2.38 (2.52)	2.59 (2.53)
Preoperative TMTA	41.39 (16.47)	38.47 (12.79)	47.66 (21.35)*
Preoperative TMTB	98.5 (52.1)	89.54 (46.71)	117.48 (60.01)*
Peak DRS Total Score	11.5 (8.3)	6.82 (3.74)	21.48 (6.40)***
Delirium duration			1 day***
	Frequency (type,		
Sex	55 (male, 55%)	39 (59%)	16 (52%)
Type of Surgery			
Vascular	38 (vascular, 39%)	22 (33%)	16 (51%)*
Other	12 (general, 12%) 36 (Spine, 37%) 11 (Urology, 11%)	44 (66%)	15 (48%)
Years of Education	3 (<12yrs, 3%)	1 (1%)	2 (6%)
	29 (12yrs, 30%)	19 (29%)	10 (32%)
	65 (>12yrs, 67%)	46 (70%)	19 (61%)
Current tobacco user	19 (yes, 16%)	10 (15%)	6 (19%)
Past tobacco user	65 (yes, 67%)	43 (65%)	22 (71%)
Significance levels: * = p<0.05, **p<0.001, ***p<0.0001			
Abbreviations: ASA=American Society of Anesthesiologists classification score, DBP=Diastolic Blood Pressure, DRS=Delirium Rating Scale-98-R, Framingham Risk Score=Framingham Cardiovascular Risk Score, GDS15=Geriatric Depression Scale-15 questions, MAP=Mean Arterial Pressure, NSQIP_SC=National Surgical Quality Improvement Program Risk for Serious Complications, NSQIP-D=National Surgical Quality Improvement			

Table 2: Developed Delirium Prediction Models for Delirium Incidence and Peak Severity

Table 2. Results of LASSO and Best Subsets Regression. Model Statistics shown below.						
Delirium Incidence Prediction Model						
LR chi2 p-	Variable	Odds Ratio	Std. Error (95%CI)	p-value	AUROC (95%C	McKelvey & Zavoina's
34.75 <0.0001	NSQIP-SC TMT	1.10 1.01	0.03 (1.05, 1.15) 0.01 (1.00, 1.02)	<0.001 <0.05	0.81 (0.72,0.90)	0.33
Model Classification: Sensitivity: 0.52, Specificity: 0.88, PPV: 0.67, NPV: 0.79. AUROC						
Delirium Symptom Prediction Model						
Transformed DRS						
F-Statistic p-	Variable	Coefficient	Std. Error (95%CI)	p-value	Std. Betas	Adj R-Square
21.35 <0.000	NSQIP-SC	0.02 0.003	0.004 (0.01, 0.03) 0.001 (0.002	<0.001 <0.001	0.41 0.35	0.30
Raw DRS						
34.73 <0.000	NSQIP-SC	0.29 0.05	0.06 (0.17, 0.42) 0.01 (0.02, 0.08)	<0.001 <0.001	0.41 0.33	0.29
Abbreviations: Adj.=adjusted, AUROC=Area Under the Receiver Curve Operator, CI=Confidence Interval, DRS=Delirium Rating Scale-R-98, LR=Likelihood ratio, NPV: Negative Predictive Value, P						

Figure 2: Area under the ROC Curve for Delirium Incidence Prediction Model

Figure 2 displays the Area Under the Receiver Operator Curve (AUROC) for the prognostic delirium incidence prediction model.

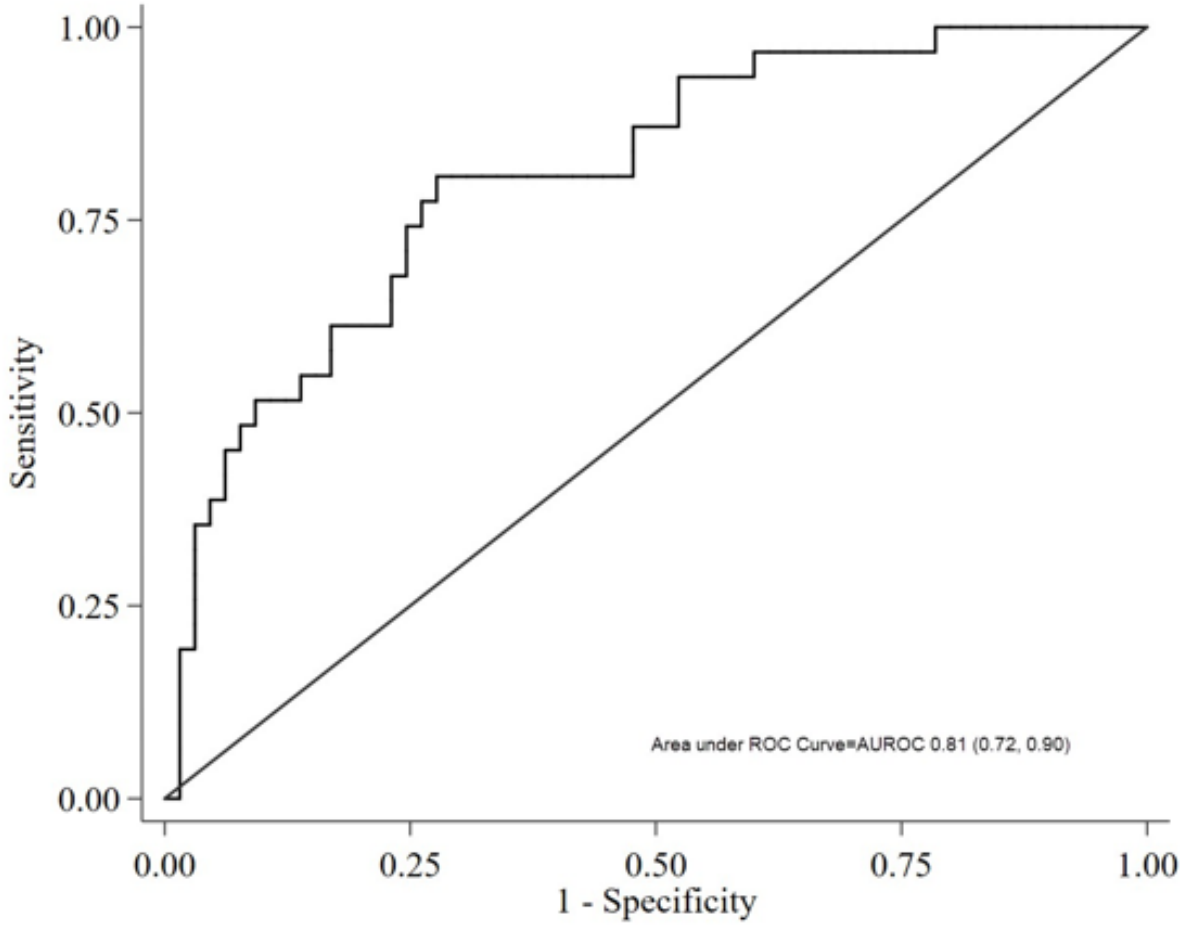
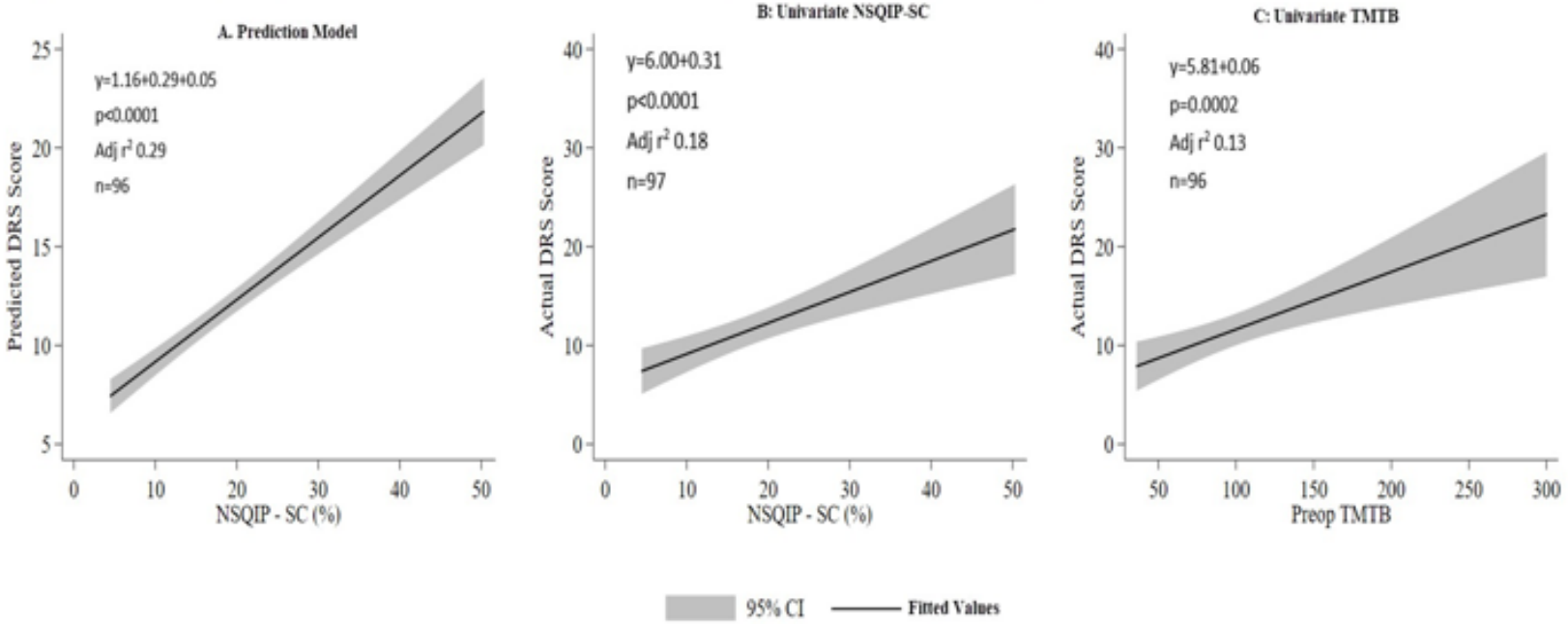
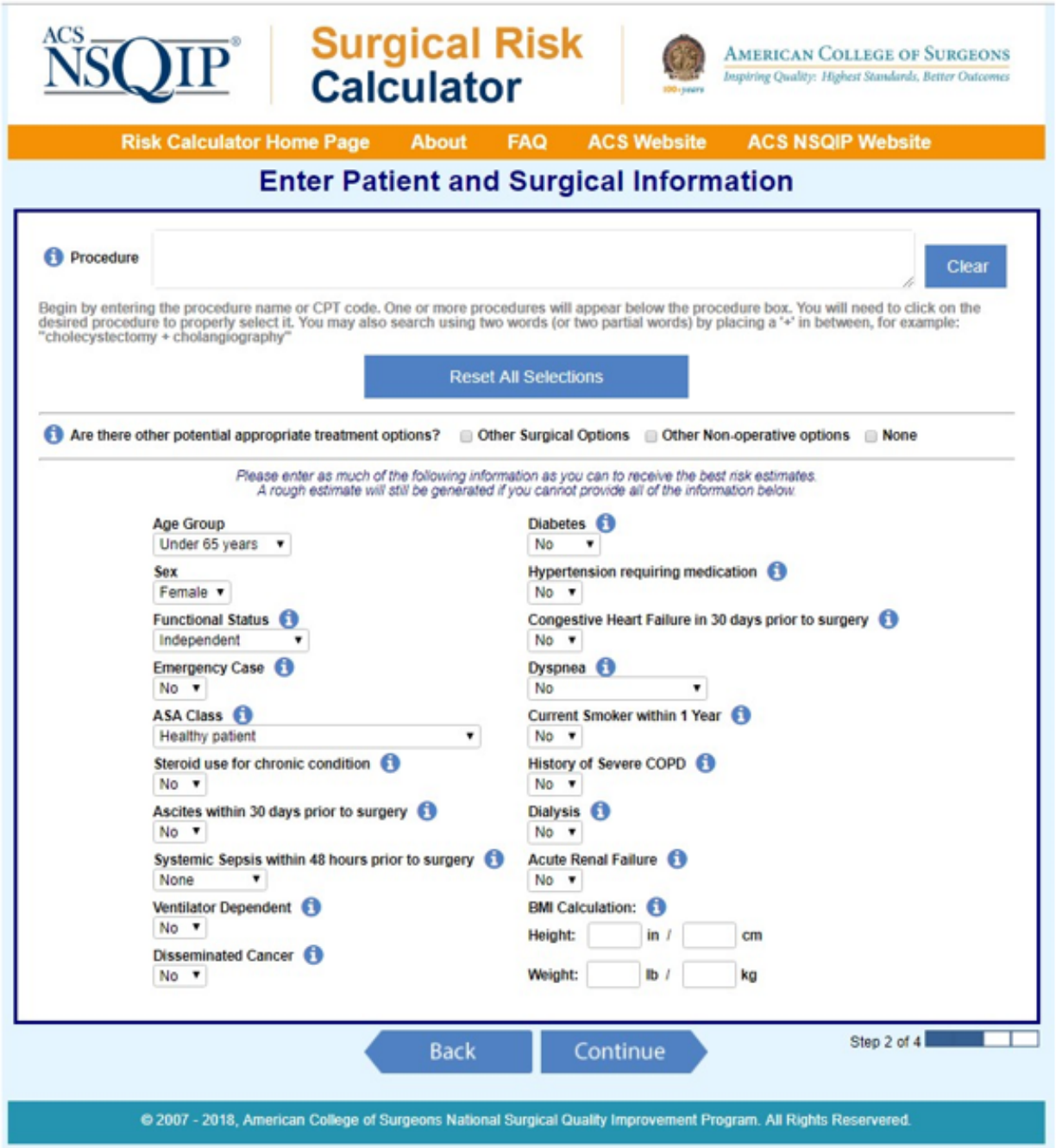


Figure 3: Postoperative Delirium Symptom Model.
Prediction Model: This is the predicted burden of delirium symptoms (DRS score) by NSQIP-SC and TMTB prediction model (Plot A).
Plot B and C display univariate models predicting the burden of delirium symptoms based on NSQIP-SC (Plot A, NSQIP) versus TMTB (Plot B, TMTB). The statistics from the regression models are in the upper left corner of each plot. The univariate NSQIP-SC sample is 97. However, due to one missing assessment of TMTB, plots A and C are analyzed with a n of 96.



Supplementary Material

Supplementary Figure 1

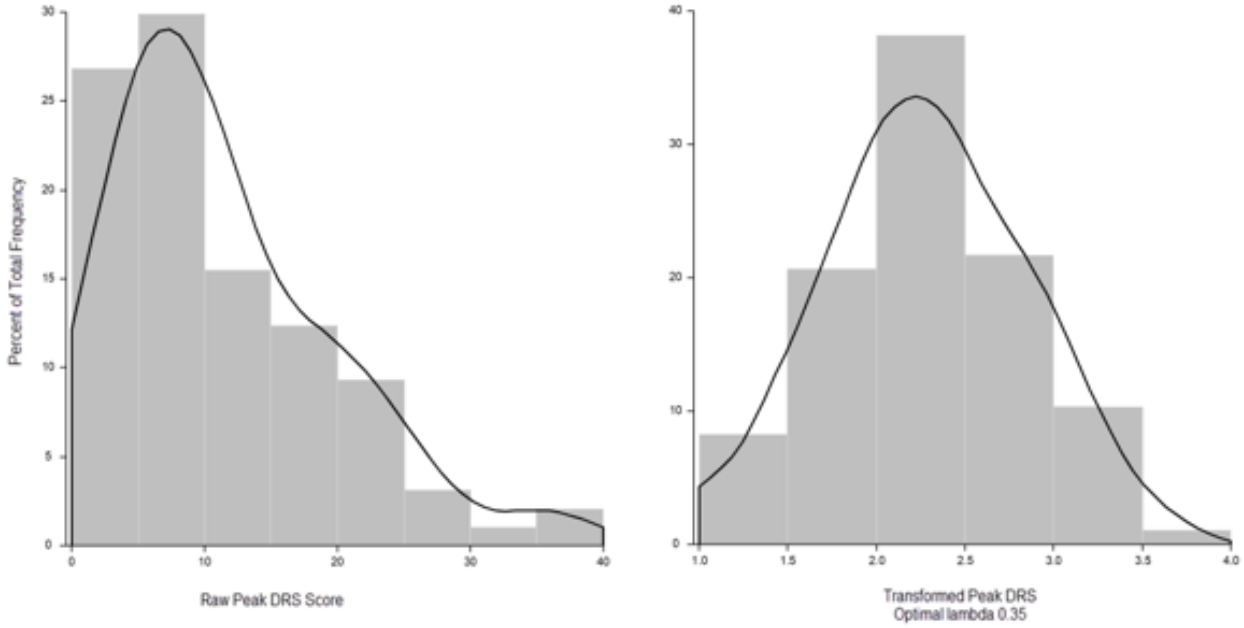


Definitions for each variable are available by clicking the information button to the right of the variable.
https://riskcalculator.facs.org/RiskCalculator/PatientInfo.jsp
Risk prediction is projected for 30-days following the surgery

Supplementary Table 2 – Regression Model Statistics for Surgical Risk Scores							
Logistic Regression with Delirium Incidence							
LR chi2 p-	Variable	Odds Rati	Std. Error (95%CI)	p-value	AUROC (95%CI)	McKelvey & Zavoina's	AIC BI
-50.67 <0.000	NSQIP-SC	1.09	0.03 (1.05-1.15)	<0.001	0.76 (0.66, 0.87)	0.25	1.086 -15.63
-57.80 <0.05	NSQIP-D	1.15	0.07 (1.02, 1.29)	<0.05	0.73 (0.62, 0.84)	0.08	1.233 -1.38
-58.44 <0.05	ASA	2.24	0.86 (1.06, 4.76)	<0.05	0.63 (0.53, 0.73)	0.07	1.246 -0.09
-60.74 P=0.82	Framing.	1.00	0.11 (0.98, 1.02)	P=0.82	0.53 (0.40, 0.65)	-0.03	1.294 4.52
Linear Regression with Transformed DRS							
F-Stat p-value	Variable	Coef.	Std. Error (95%CI)	p-value	Std. Betas	Adj R-Square	AIC BI
22.68 <0.000	NSQIP-SC	0.02	0.004 (0.01, 0.03)	<0.001	0.44	0.18	1.525 -16.19
4.25 <0.05	NSQIP-D	0.03	0.01 (0.001, 0.06)	<0.05	0.21	0.03	1.695 0.33
4.18 <0.05	ASA	0.20	0.096 (0.01, 0.39)	<0.05	0.21	0.03	1.696 0.40
0.16 P=0.69	Framing.	0.001	0.003 (-0.004, 0.007)	P=0.69	0.04	0.002	1.737 4.14
Abbreviations: Adj.=adjusted, AIC= Akaike information criteria, ASA=American Society of Anesthesiologists, AUROC=Area Under the Receiver Curve Operator, BIC=Bayesian Information Criteria, CI=Confidence Interval, DRS=Delirium Rating Scale-R-98, Framing.=Framingham Cardiovascular Risk Score, LR=Likelihood ratio, NSQIP							

Supplementary Figure 2

Supplementary Figure 2 displays the distribution of the raw peak Delirium Rating Scale (DRS) score, left, and the transformed distribution on the right. This was done using the Box-Cox transformation with an optimal lambda of 0.35.



Manuscript 3

Title: Examining differences between T1 and T1+T2-FLAIR cortical thickness measurements using latent profile analysis

https://www.frontiersin.org/about/author-guidelines

Can have 15 figures. Times new roman, 12point font. Section headings are laid out per Frontiers.

12,000 words for overall manuscript. 350 words for abstract

Abstract

Cortical thickness measurements are traditionally derived from T1-weighted MRI imaging and applied to investigate the association between two or more variables. Multimodal imaging, involving the combination of T1+T2-FLAIR anatomical scans to obtain cortical thickness measurements are not widely used even though it has been suggested to improve accuracy in segmentation and signal intensity. This study endeavored to investigate the statistical differences between T1-only and T1+T2-FLAIR derived cortical thickness measures and if statistical findings from T1-only cortical thickness measurements are replicated when multimodal imaging is applied. Standard t-tests, correlations, and regression were applied as well as latent profile analysis. Latent profile analysis (LPA) investigates patterns across data, identifying sub-groups or classes or like-individuals. This approach is not widely used in cortical thickness research yet interest is growing because it facilitates the discovery of unknown patterns and characteristics within a dataset. T1-weighted and T2-FLAIR anatomical images on 54 older adults (mean age 71 years, 65-81 years, # females) from 1.5T and 3T scanners were examined. Cortical parcellation statistics were extracted using the Desikan-Killiany (DKT) Atlas, 34 regions per hemisphere. LPA models were built using unadjusted, scanner adjusted, and scanner and intracranial volume (ICV) adjusted cortical thickness measures. T1+T2-FLAIR measurements demonstrated significantly higher means in all DKT regions (68). Significant correlations in 65/68 DKT regions between T1-only and T1+T2-FLAIR images were identified. Age was associated with T1+T2-FLAIR measurements following false discovery rate correction. T1+T2-FLAIR LPA models did not replicate the findings of the T1-only LPA models. Classification metrics and clinical characteristics of the identified sub-groups were different. In summary, multimodal imaging using T1+T2-FLAIR did not replicate the results of T1-only images. T1+T2-FLAIR

cortical thickness cortical thickness measurements demonstrate higher precision with segmentation and signal intensity. It was more sensitive to age-associated atrophy and contributed more information to LPA model classification procedures. Future studies should consider employing multimodal imaging to obtain precise cortical thickness measurements.

Keywords: multimodal segmentation, cortical thickness, latent profile analysis, aging

Introduction

Cortical thickness is a widely-used neuroimaging measurement that quantifies the amount of cortical atrophy occurring in the brain. Cortical atrophy occurs as a natural process of aging as well as in different types of disease processes leading to focal or regional atrophy.[1] This atrophy equals a loss of neuron containing tissue and the connections between those neurons leading to functional and cognitive deficits.[1 2] Cortical thickness measurements may be used to identify relationships between patient-level clinical variables and regions of brain atrophy as well as predict those most likely to convert from mild cognitive impairment (MCI) to Alzheimer's Disease (AD)[1 3-7]. Cortical thickness measures are typically derived from T1-weighted anatomical magnetic resonance imaging (MRI) scans. While this source has been documented to be reliable, segmentation errors are often identified when the dura layer is included in the gray matter measurements.[8] Traditionally, these are manually corrected, or scans are discarded if there are significant segmentation errors. FreeSurfer[9-16], a widely-used, open source, processing stream for cortical thickness measures, introduced a multi-modal processing option using both T1-only and T2-FLAIR anatomical scans to improve pial surface segmentation. The derivation of cortical thickness measures depends on the correct segmentation of gray matter from the surrounding white matter and dura surfaces leading to inaccuracies in analysis and results if the segmentation is inaccurate. There are few studies to date that evaluate the difference between T1-only and T1+T2-FLAIR processed scans. These studies identified that multimodal imaging, using both T1-only and T1+T2 anatomical scans decreased segmentation error and misidentification of tissue.[8 17-19] Further, T1+T2-FLAIR data demonstrated higher correlation with age than T1-only data.[18] These studies did not examine if T1-only results are replicated when T1+T2-FLAIR data are used. Therefore, the purpose of the present study was to investigate

the statistical differences between T1-only and T1+T2-FLAIR data and examine whether T1+T2-FLAIR data replicate the findings of T1-only cortical thickness measurements.

Cortical thickness measures are generally examined for associations between biologic and cognitive variables. While this type of investigation is important, data is evaluated at the aggregate level and does not reveal unknown relationships.[20] An alternate approach that facilitates the identification of patterns, or sub-groups, within data is termed person-oriented analysis because it views the data as a whole.[21] Recently, person-oriented approaches have been applied with increasing frequency in neuroimaging analysis because it's ability to identify unknown patterns, or sub-groupings, of individuals.[22 23] The demographic, biologic and cognitive characteristics of each sub-group are then examined, allowing for the discovery of unidentified associations. Latent profile analysis is one type of person-oriented approach that results in parsimonious profiles, *i.e.*, subgroups, of variables.[24 25] Therefore, we sought to evaluate the use of latent profile analysis to identify latent sub-groups of individuals with similar patterns of cortical atrophy then evaluate the clinical characteristics of each sub-group using biologic and cognitive variables. Previous research has reported significant associations between vascular comorbidities such as blood pressure, smoking, and diabetes with cortical atrophy in the frontal, temporal, and parietal regions[5 7 26-30]. Executive function, the cognitive domain that facilitates day-to-day functioning through problem-solving, attention, and reasoning, is reported to correlate with cortical atrophy in the frontal, parietal, and temporal lobes.[6 31-35] Atrophy in the medial temporal cortex, inferior temporal gyrus, temporal pole, angular gyrus, superior frontal gyrus, superior parietal lobule, supramarginal gyrus, precuneus, and inferior frontal sulcus represent the AD Signature[36]. This signature has demonstrated a significant association with

postoperative delirium severity, an acute brain failure, as well as within MCI, identifying those most likely to convert to AD.[36 37] Prior research is based on traditional variable-centered methodology and while informative, we sought to expand on this past research by using a person-oriented approach to identify sub-groups of cortical atrophy in an older adult perioperative cohort and describe the vascular and cognitive clinical characteristics within each sub-group. Based on prior research, we hypothesized that latent profile analysis would classify sub-groups based on atrophy. Further, we hypothesized that atrophy would be associated with higher age, vascular burden and lower cognition.

The present manuscript is outlined as follows. The Materials and Methods section provides information on the sample size, descriptive variables collected, MRI sequences and processing procedures, and statistical analysis. The Results section details the following: 1) T1-only and T1+T2-FLAIR cortical thickness measures are examined and analyzed for statistical differences between mean thickness, ICV, and segmentation. Correlations are completed between scan types. A general linear model is used to regress age. Explanatory figures are employed to demonstrate the difference in cortical segmentation between T1-only and T1+T2-FLAIR processed images; 2) Latent profile analysis (LPA) models are compared between T1-only and T1+T2-FLAIR cortical thickness data. Each LPA model is represented with a detailed figure showing classification and significance statistics along with regions of identified cortical atrophy; and 3) Univariate entropy is reported between different scan types.

Materials and Methods

This is a cross-sectional, descriptive analysis of sixty-seven participants whom underwent a preoperative brain MRI as part of an ongoing perioperative cohort study registered ClinicalTrials.gov (ref: NCT03124303, NCT01980511) and approved by the University of

Wisconsin-Madison Institutional Review Board (#2015-0374). Seven MRI scans were discarded due to motion artifact or anatomical anomalies, and an additional six MRI scans were discarded for an incomplete anatomical MRI sequence, for a final sample size of fifty-four. Figure 1 describes study flow.

All participants completed full written informed consent prior to study participation. Older adults, sixty-five years of age and older, scheduled to undergo a non-cardiac surgery with an estimated length of stay in the hospital of two days or greater following their surgery were eligible to participate in this study. Eligible participants were excluded if they had contraindications to an MRI, unable to communicate with research staff due to language or sensory barriers, had a documented history of dementia, and resided in a nursing home or assisted care facility. Prior to surgery, consented participants completed a structured interview including questions on past medical history, demographics, functional status, depression, All participants completed full written informed consent prior to study participation. Older adults, sixty-five years of age and older, scheduled to undergo a non-cardiac surgery with an estimated length of stay in the hospital of two days or greater following their surgery were eligible to participate in this study. Eligible participants were excluded if they had contraindications to an MRI, unable to communicate with research staff due to language or sensory barriers, had a documented history of dementia, and resided in a nursing home or assisted care facility. Prior to surgery, consented participants completed a structured interview including questions on past medical history, demographics, functional status, depression, burden largely through the correlation of single risk factors to regions of cortical atrophy. This study sought to use two different types of risk assessment scores that combine several vascular risk factors into one score. FCDR calculates a 10-year risk of a cardiovascular event based on sex, age, systolic blood

pressure (SBP), treatment for hypertension, current smoker, diabetes and body mass index (BMI). ASA classification is assigned by the attending anesthesiologist prior to surgery and takes into account physical status along with comorbidities such as diabetes, hypertension, COPD and obesity. The patient is assigned a ranking from 1-6 with 1 being a "normal healthy patient" independent of type of surgery planned. While the FCDR is a well-established marker of vascular disease burden, ASA considers physical status, function, and comorbidities. Preoperative blood pressure metrics (Systolic Blood Pressure, SBP, Diastolic Blood Pressure, DBP, Pulse Pressure, PP, and Mean Arterial Pressure, MAP) are included in this analysis along with age, sex, FCDR, ASA, current tobacco use, and past tobacco use to provide a comprehensive picture of vascular disease burden and comorbidities.

MRI Images

T1-weighted MPRAGE (TR: 2,530ms, TE: 3.09ms, Flip Angle: 10˚, 256 x 256 matrix, 208 coronal slices, 1 mm isotropic resolution) and T2-fluid attenuated inversion recovery (FLAIR, T2*-weighted gradient-echo echo planar imaging pulse sequence sensitive to BOLD contrast: field of view 224mm, matrix 64 x 64, TR: 2600ms, TE: 22ms, Flip Angle 60˚, 40 axial plane slices of 3.5mm thickness with 3.5mm spacing between slices) images were obtained on 1.5-telsa and 3-telsa General Electric scanners using an 8-channel head coil at the University of Wisconsin Hospital and the Wisconsin Institute for Medical Research center. Cortical thickness (mm) measures were obtained using a publicly available software package, FreeSurfer v6.0 (https://surfer.nmr.mgh.harvard.edu/). First T1-weighted MPRAGE images were processed using the well-documented recon-all processing stream which includes motion correction, skull-stripping, registration, segmentation, smoothing, and parcellation mapping Coronal slices were manually inspected and ranked on quality level by two independent observers. The quality

rankings were as follows: 0) 0-1 segmentation error; 1) ≤2 segmentation errors; 2) ≤5 segmentation errors; 3) moderate segmentation error with dura overlap; 4) poor quality, discard. Seven scans were discarded due to motion artifact, poor segmentation, or anatomical abnormalities. All scans were then re-processed using T1+T2-FLAIR multimodal recon-all processing stream. Cortical parcellation statistics were extracted using the Desikan-Killiany Atlas (DKT), which contains 68 regions, 34 per hemisphere. Means, standard deviations, and data distribution were then evaluated. Following re-processing, manual segmentation was completed by two independent observers and scans were re-ranked on their level of quality. Statistical analysis was performed to evaluate the differences between the T1-only recon-all images versus the T1-only/T2-FLAIR recon-all images (n=54).

Statistical Analysis

Descriptive patient characteristics were evaluated using means ± standard deviations for continuous variables, frequency counts with percentages for categorical variables and data distributions were visually inspected. Cortical thickness measures were adjusted for scanner-type and intracranial volume (ICV). The ComBat harmonization correction tool was applied to account for inter-site scanner variation38. This tool performs well in small sample sizes and is able to maintain biologic variability while correcting for inter-site scanner variation. T1-only statistics and T1+T2-FLAIR statistics were compared using Pearson correlations and Student's t-test. If data was non-normally distributed, non-parametric Mann-Whitney U-test was performed.

Latent profile analysis (LPA) were performed using the mixture-modeling package in M*plus* version 8.1. LPA was chosen because it offers a "model-based clustering" approach that derives clusters/classes using a probabilistic model, describing the distribution of the data. So instead of findings clusters/classes with some arbitrary chosen distance measure, the LPA describes the

distribution of the data and probabilities on class assignment are provided. This is considered a top-down approach. Statistical threshold was $p \leq 0.05$. Each hemisphere (DKT atlas, 34 regions per hemisphere) was loaded into M*plus* as classification variables. Twelve LPA models, 6 per FreeSurfer processing stream (T1, T1+T2-FLAIR, right and left hemisphere), were ran and included: 1) T1-only Unadjusted CT; 2) T1-only Scanner-adjusted CT; 3) T1-only ICV-Scanner adjusted CT; 4) T1+T2-FLAIR Unadjusted CT, 5) T1-T2-FLAIR Scanner-adjusted CT; and 6) T1+T2-FLAIR ICV-Scanner adjusted CT. LPA models were bootstrapped with 1000 repetitions to confirm model reproducibility. The selection of the optimal number of classes was guided by a priori criteria designed to choose a LPA model that would be clinically relevant and statistically justified. The LPA model that provides an optimal balance of statistical fit and clinical applicability is selected based on the following criteria:: 1) interpretability; 2) theoretical justification; 3) parsimony; 4) relative entropy >0.75; 5) lowest adjusted Bayesian Information Criteria (BICa) score; 7) average posterior probability in each class >0.75 and no more than 10% overlap between non-contiguous classes; 8) at least 2.5% of total count in each group; and 9) no significant improvement as assessed by likelihood ratio test (Lo-Mendall-Rubin). Each criteria contributes to the selection, however, theoretical justification, parsimony, and interpretability were prioritized followed by relative entropy and the remaining statistical metrics. Theoretical justification is used to demonstrate that a system has importance and is typically rhetorical. It includes reasoning on whether the model is feasible and functional. Interpretability indicates the degree to which one can consistently predict the model's results. Univariate entropy signifying the quality of a latent class qualifier will be reported for the selected classes39. Auxiliary analysis was performed to evaluate statistically significant differences between groups using the BCH method40. False Discovery Rate (FDR) and effect sizes with 95% confidence intervals

were calculated for the auxiliary variables. To describe the regions of cortical atrophy associated with class assignment, FreeSurfer v6.0 gui QDEC was used. Posterior probabilities generated by MPlus for LPA class assignment per participant were loaded into QDEC. Both unadjusted and scanner-adjusted (scanner type as a nuisance factor) models were described.

Results

Fifty-four participants with a mean age of 71 (4.84) and 48% female were included in this cortical thickness analysis. Table 1 summarizes participant characteristics including a breakdown on the distribution of participants between scanners and statistical significance identified perLPA model. Cortical thickness data extracted from FreeSurfer using the 68-region, 34-per hemisphere, DKT atlas were adjusted for scanner-type with the ComBat harmonization tool.[38] As a sensitivity analysis, scanner type was regressed out of each mean using a general linear model and compared to the Combat adjusted data. Four regions were identified to be significantly different ($p<0.05$) from the ComBat harmonization data.

T1-only versus T1+T2-FLAIR Processed Images

Manual inspection of segmentations revealed overall improvement, with 82% of the T1+T2-FLAIR images ranking as "0=perfect" or "1=minimal error". This is contrasted to 55% of T1-only images categorized accordingly. Figure 3 demonstrates the segmentation improvement with the T1+T2-FLAIR image segmentation not overlapping into the dural layer.

Cortical parcellation statistics for T1-only and T1+T2-FLAIR processed images were extracted from FreeSurfer using the 34-per hemisphere, DKT atlas. T1-only processed images demonstrated a normal distribution with minimal outliers. The distribution of T1+T2-FLAIR cortical thickness data shifted to the right. The cortical thickness means per DKT region were significantly thicker when compared to the T1-only processed images using Mann-Whitney U-

Tests ($p<0.05$), Figure 2. Total intracranial volume was not significantly different ($p=0.25$).

Sixty-five of 68 DKT T1+T2-only regions were significantly correlated ($p<0.05$) with T1-only DKT regions. The three exceptions were the right hemisphere postcentral gyrus, rostral anterior cingulate, and rostral middle frontal gyrus. The highest r2 demonstrated was 0.77 ($p<0.05$), and the mean r2 was 0.55 ± 0.13 ($p<0.05$).

Age was regressed onto T1-only and T1+T2-FLAIR cortical thickness data using FreeSurfer's gui QDEC, using a general linear model. No vertices survived FDR correction at 0.05 with unadjusted and scanner-adjusted data in the T1-only images. The T1+T2-FLAIR identified 20034 vertices, 43 clusters, to survive FDR correction in the left hemisphere and 32216 vertices, 38 clusters, to survive FDR correction in the right hemisphere. These results remained after adjustment for scanner-type. These results are not shown.

LPA Model Selection

Two-class, three-class and four-class solutions were evaluated using automated separation of 34 cortical thickness regions per hemisphere as classifiers for the LPA modeling procedure. Two-class models were selected for the unadjusted, scanner adjusted, and ICV-scanner adjusted T1-only and T1+T2-FLAIR cortical thickness data based on the parsimony principle and model fit statistics displayed in Table 2. For all six models, Class 1 is named the cortical atrophy Class (CAC) and Class 2 is named the Non-atrophy Class (NC).

T1-only Cortical Thickness LPA Models

T1-Unadjusted CT LPA

The 2-class unadjusted cortical thickness measures demonstrated unadjusted significant differences between classes (unadjusted, $p<0.05$). In the left hemisphere, the CAC demonstrated significantly higher means for age, FCDR risk score, current tobacco use, and a lower RBANS

sub-test story recall score. In bilateral hemispheres, the subtest RBANS picture naming score means were higher in the CAC. The FCDR risk score survived FDR correction ($P<0.001$).

Statistical details are shown in Table 3. Classification estimated means are displayed in Figure 4. Mean cortical thickness was significantly different between classes, CAC vs NC, differences are displayed in Figure 4 (unadjusted, $p<0.05$). Overall atrophy is observed. There were significant differences in cortical atrophy in the 2 groups ($p<0.05$, FDR corrected). These regiosn include: 2 regions in the right hemisphere (precentral, superior frontal gyrus) and 4 regions in the left hemisphere (rostral middle frontal gyrus, supramarginal gyrus, medial orbitofrontal gyrus, superior frontal gyrus). Vertices, size and talairach coordinates are listed in supplementary Table 1.

In summary, the CAC sub-group had a significantly higher mean FCDR score ($p<0.05$, FDR corrected), indicating vascular burden in the left hemisphere. Supplementary Figure 1 displays the standardized effects sizes showing the data trend; the CAC contained older adults with more vascular burden and worse cognitive scores when compared to the NC.

T1-Scanner-Adjusted CT LPA

Following scanner adjustment using the ComBat harmonization, class separation was comparable to the unadjusted models with similar class counts. In the left hemisphere, age and the RBANS sub-test picture naming were significantly different (unadjusted, $p<0.05$). These did not survive FDR correction. Significant differences between classes are displayed in Table 4. Standardized effect sizes are available in supplementary Figure 2.

Mean cortical thickness was significantly different between classes, CAC vs NC, differences are displayed in Figure 5 (unadjusted, $p<0.05$). Scanner adjustment resulted in less overall atrophy, but increased the number of regions that survived FDR correction ($p<0.05$) in both hemispheres.

Vertices, size and talairach coordinates are listed in supplementary Table 2.

In summary, scanner adjustment resulted in the identification of a 2-class LPA model, but the clinical differences between those two classes were diminished.

T1-ICV/Scanner Adjusted CT LPA

Cortical thickness means were further adjusted for ICV and the 2-class model closely resembled the T1-Scanner adjusted 2-class LPA. Mean cortical thickness between classes, CAC vs NC, did not survive FDR correction. These data are not displayed in tables or figures.

Synopsis: T1-only LPA Model

LPA classification ability was unaffected by scanner and scanner+ICV adjustment. However, the significant clinical differences between classes were impacted. In the unadjusted LPA model, the Framingham Cardiovascular Risk score survived FDR correction in the left hemisphere (FDR corrected, $p<0.001$). However, the significance of this variable dropped when scanner adjusted data was used. Scanner+ICV adjustment further reduced significant findings between classes indicating its possible importance as a covariate.

T1+T2-FLAIR Cortical Thickness LPA Models

Unadjusted T1+T2-FLAIR CT LPA

The classification of the unadjusted T1+T2-FLAIR data differed from the T1-only data classification; class counts in the bilateral hemispheres demonstrated <10 in class 1 with the majority of data points being classified into class 2. In the left hemisphere, age, SBP, and years of education means were significantly higher in the CAC (unadjusted, $p<0.05$). ASA classification and the RBANS coding sub-test means were significantly lower (unadjusted, $p<0.05$) in the CAC. Age, ASA classification, SBP, and education survived FDR ($p<0.05$). Further statistics are shown in Table 5. Standardized effect sizes are shown in supplementary

Figure 3.

Bilateral Cortical thickness means between classes were significant (unadjusted, $p<0.05$). There were significant differences in cortical atrophy in the 2 groups ($p<0.05$, FDR corrected) and are displayed in Figure 6. These regions include: 1 region in the left hemisphere (precuneus) and ten regions in the right hemisphere (5 largest: superior temporal gyrus, inferior temporal gyrus, isthmus cingulate, posterior cingulate, and the medal orbitofrontal gyrus). Vertices, size and talairach coordinates are listed in supplementary Table 3.

In summary, the T2-unadjusted LPA model did identify a 2-class solution, however, the class membership in the CAC is much smaller when compared to the T1-unadjusted LPA model (4-LH/9-RH vs 27-LH/30-RH). Within this class, age, ASA ranking, SBP, and education survived FDR correction.

T1+T2-FLAIR Scanner Adjusted CT LPA

The two-class model using cortical thickness measurements adjusted with the ComBat harmonization tool[38] demonstrated higher CAC classification than the T1+T2-FLAIR unadjusted cortical thickness measures (LH-15/39, RH-13/41). Class separation is visualized in Figure 7. As shown in Table 6, statistical differences in the left hemisphere are markedly different than the T1+T2-FLAIR unadjusted model. Age, ASA, SBP, and education are not statistically significant while GDS15 and RBANS Coding subtest demonstrate significantly lower means (unadjusted, $p<0.05$). Contrary to previous models, the right hemisphere identified age and depression to be significantly different between classes (unadjusted, $p<0.05$). No regions survived FDR correction. Standardized effect sizes are shown in supplementary Figure 3.

Mean cortical thickness was significant between the two classes, CAC vs NC, in both hemispheres (unadjusted, $p<0.05$). As in the T1-only scanner adjusted LPA model, scanner

adjustment resulted in additional areas of atrophy that survived FDR correction in the two groups ($p<0.05$). Vertices, size and talairach coordinates are listed in supplementary Table 4.

In summary, harmonization of the data using the ComBat tool38 resulted in an increase in membership between class 1 and 2. GDS15 and RBANS Coding Total Subtest demonstrated significance between classes in the left hemisphere but did not survive FDR correction.

T1+T2-FLAIR ICV & Scanner Adjusted CT LPA

Similar to the T1-only ICV & Scanner adjusted CT 2-class LPA, age, depression and the RBANS coding subtest were identified to be significant between classes but did not survive FDR correction. Mean cortical thickness was significant between the two classes, CAC vs NC, in both hemispheres ($p<0.05$, unadjusted). No regions survived FDR correction ($p<0.05$) in the left hemisphere. 2 regions in the right hemisphere survived FDR correction ($p<0.05$) (inferior parietal gyrus, postcentral region). No tables or figures are displayed for this analysis.

Synopsis: *T1+T2-FLAIR LPA Models*

Multimodal imaging changed how LPA classified on cortical atrophy. In the unadjusted model, class membership in the CAC was greatly reduced with the majority of participants placed into the NC group. This is in contrast to T1-only unadjusted model class membership of an approximate even division. The few individuals placed into the CAC demonstrated significance following FDR correction in age, SBP, ASA, and years of education. These findings were no longer significant in the T1+T2-FLAIR scanner-adjusted or ICV+Scanner adjusted models.

Discussion

This analysis demonstrated significant differences between T1-only and T1+T2-FLAIR cortical thickness measures and statistical analysis. Although significantly correlated, T1+T2-FLAIR processed images had significantly thicker DKT cortical thickness means. This difference in measurement appears to lead to difficulties replicating statistical findings between T1-only and

T1+T2-FLAIR generated data as shown in this analysis. Further, different results were obtained between LPA classes analyzed with T1-only versus T1+T2-FLAIR processed cortical thickness data. In the T1-only unadjusted 2-Class LPA Model, the Framingham Cardiovascular Disease Risk score survived false discovery rate in the left hemisphere of the Cortical Atrophy class. In the T1+T2-FLAIR Unadjusted 2-class LPA model, age, ASA status, SBP, and years of education survived false discovery rate correction. These findings did not sustain when the cortical thickness measures were harmonized using the ComBat harmonization tool to control for inter-scanner variability. This is likely due to confounding introduced by scanner type. This is important for studies that use different scanner strengths and manufacturers – harmonization appears critical to obtaining accurate measurement and results. Further, it accounts for variation in the distribution of participants between scanner types especially as this may be non-random. The combination of T1+T2-FLAIR anatomical images yielded enhanced segmentation of gray matter from the dura layer, resulting in significantly higher cortical thickness means per DKT Region. This finding confirms reports from previous studies that examined the difference between T1-only and T1+T2-FLAIR cortical thickness measures using different processing programs.[8 17- 19] T1-only and T2-FLAIR are anatomical scans that use different pulse sequences to identify different brain tissue and cerebrospinal fluid. T1-only images typically have a shorter pulse sequences in contrast to T2-FLAIR images, which have a much longer pulse sequence leading to an increased sensitivity to pathology and a differentiation from the dura layer. [8 17] Past studies have identified that multimodal imaging increases accuracy through the differentiation of the dura layer and vessels from gray matter to improve segmentation, as well as an improved signal intensity.[8 18 41] Salat *et al.* (2009) reported that signal intensity between gray and white matter increases with age leading to a decreased contrast between tissue classes. This

is a regional phenomenon, mostly affecting the superior frontal, precentral, postcentral, occipital, medial frontal, and superior temporal regions.[42] The ability of T2-FLAIR to enhance signal intensity from gray matter likely contributed to the overall, higher cortical thickness means particularly in the three regions (rostral middle gyrus, postcentral, rostral anterior cingulate) that did not significantly correlate with T1-only data. The rostral middle gyrus was identified by Salat *et al.* (2009) as a region with significantly reduced gray and white matter signal intensity. Therefore, T1+T2-FLAIR may be better at detecting gray matter in this region. This study builds on previous research by examining the difference in statistical findings between T1-only versus T1+T2-FLAIR cortical thickness data. Further, this study explores a person-oriented approach to identify sub-groups of atrophy within an older adult cohort, scheduled to undergo a non-cardiac surgical procedure. The finding that statistical results are different not only between T1-weighed and T1+T2-FLAIR data but also between unadjusted and scanner-harmonized data is important for future research as it questions the generalizability of findings between cortical thickness studies using different types of anatomical scans as well as data from different scanners. This study should be replicated with a larger sample size as it is a preliminary analysis.

Latent profile analysis identified a sub-group of cortical atrophy. This is an important finding for future studies as identification of sub-groups of cortical atrophy may elucidate regions of the brain that are vulnerable to surgery or inflammation. The investigation of biologic and cognitive variables within each sub-group may elucidate unknown symptoms or clinical characteristics that may aid in understanding risk prior to surgery. In this exploratory analysis of latent profile analysis, only the T1-only and T1+T2-FLAIR unadjusted models demonstrated significant differences between the classes ($p<0.05$, FDR corrected). Vascular burden assessed by the FCDR risk score was significant in the T1-only unadjusted left-hemisphere of the CAC. In this model,

the rostral middle frontal region in the left hemisphere demonstrated the most atrophy and has been identified in previous studies to be significantly associated with vascular burden.[7 26 27] This is similar to the T1+T2-FLAIR unadjusted model. Systolic blood pressure was significantly higher in the CAC and the superior and inferior temporal gyri were significantly atrophied along with the isthmus and posterior cingulate (FDR corrected, $p<0.05$), this association has been found in previous research.[27 43] These findings did not remain significant when cortical thickness measures were adjusted for scanner-type emphasizing the importance of controlling for scanner variance.

Previous research has reported the use of a cortical thickness signature (AD signature) to be an accurate neuroimaging biomarker to identify those that will transition from mild cognitive impairment to Alzheimer's disease (AD).[4 36 44 45] The AD signature represents atrophy in the following regions: medial temporal cortex, inferior temporal gyrus, temporal pole, angular gyrus, superior frontal gyrus, superior parietal lobule, supramarginal gyrus, precuneus, and inferior frontal sulcus[36]. Further, the AD signature is reported to significantly correlate with postoperative delirium severity.[37] Delirium is an acute brain failure that may signify a vulnerable brain, demonstrating a heightened risk for future dementia.[46 47] The present study identified significant cortical atrophy within AD signature regions in both the T1-unweighted and T1+T2-FLAIR scanner adjusted LPA models in a sample of older adults ($\geq$65yo) scheduled to undergo a non-cardiac surgery. This finding may represent areas of vulnerability in the brain and should be explored further in future research studies. This is a preliminary finding from an exploratory analysis that was part of a larger analysis examining the statistical differences and replicability of results between T1-only and T1+T2-FLAIR derived cortical thickness measures. Nonetheless, this finding highlights the potential benefit of using latent profile analysis to identify sub-groups

of individuals with distinct vulnerabilities within the brain.[22] [23] Examining the clinical characteristics of the identified sub-group may elucidate unknown risk factors that signify pre-delirium vulnerability. The pattern of cortical atrophy within that sub-group could become a biomarker for pre-clinical delirium and the associated clinical characteristic could inform future delirium prediction, facilitating the ability to identify vulnerable individuals before delirium occurs.

Strengths and Limitations

The present study contributes to current literature by examining statistical differences between T1-only and T1+T2-FLAIR cortical thickness derived measures and evaluates the ability to replicate findings. Further, this study explores a person-oriented approach, a novel method in neuroimaging data, to identify subgroups of cortical atrophy. The present study is limited by sample size. The sample size is small for latent profile analysis leading to more parameters being estimated than data points, however, these models were bootstrapped with 1000 repetitions and the 2-class solutions were replicated for all 12 models indicating stability. This study is preliminary and generated effect sizes will be used to power future studies to further explore LPA modeling using cortical thickness data. The criteria used to select the best class is iterative, however, this process does require judgment decisions during the analysis on theoretical justifications and interpretability and can be subjective. However, the incorporation of multiple fit statistics with clinical reasoning aid in the selection of the model representing an optimal balance between clinical application and statistical fit.

Recommendations for Future Research

Use of T1 versus T1+T2-FLAIR

Multimodal imaging, T1+T2-FLAIR, improved cortical thickness measurements by reducing

segmentation error and increasing the signal intensity of the gray matter. Future studies should consider the addition of T2-FLAIR scans as it provides a more accurate picture of cortical thickness. The addition would increase scan time as well as processing time, however, the advantages surpass the impediment of time as a precise measurement of cortical thickness is captured

Adjustment for scanner

Following harmonization using the ComBat Harmonization tool[38] , statistical findings did change between LPA models, which is important for future studies to consider as it appears that inter-scanner variance as well as between-participant variation impacted cortical thickness measurements. The traditional method of treating scanner-type as a confounder in regression, and "regressing out" this variable, may not sufficiently account for the variance observed between scanner strengths and manufacturer due to omnipresent inconsistencies. Instead, ComBat accounts for omnipresent variance by considering not only the effect of the scanner strength and manufacturer, but also the site effects. Additionally, the application of a Bayesian Framework enhances the stability of the estimated parameters for smaller sample sizes. Future studies should consider the use of this tool as it was feasible to implement and free script is available for R, Matlab, and Python.

Adjustment for Intracranial Volume (ICV)

ICV adjustment is a common practice in prior cortical thickness studies. This analysis demonstrated that it is an important factor to consider as it did change the statistical findings including the ability to predict atrophy based on LPA class assignment. Future studies should continue to include this as a covariate.

Latent Profile Analysis

The application of latent profile analysis facilitated the identification of sub-groups, or latent classes, of the cortical thickness measures. Future neuroimaging research should consider the use of person-oriented research because it allows for the investigation of unknown variables or symptoms within sub-groups that may lead to important findings regarding vulnerability to delirium and future cognitive decline.

Conclusion

This cross-sectional, descriptive neuroimaging study identified that multimodal imaging, T1+T2-FLAIR, improved the accuracy of cortical thickness measures be reducing segmentation errors and increasing signal intensity. Further, an exploratory analysis using latent profile analysis revealed sub-groups that share patterns of cortical atrophy and these were described with clinical characteristics. Future studies should consider the use of multimodal imaging to improve cortical thickness measurement.

References:

1. Shaw ME, Sachdev PS, Anstey KJ, *et al.* Age-related cortical thinning in cognitively healthy individuals in their 60s: the PATH Through Life study. *Neurobiol Aging* 2016;39:202-9. doi: 10.1016/j.neurobiolaging.2015.12.009 [published Online First: 2016/03/01]
2. Stroke NIoNDa. Cerebral Atrophy Information Page 2018 [Available from: https://www.ncbi.nlm.nih.gov/pubmed/ accessed July 4th 2018.
3. Doherty BM, Schultz SA, Oh JM, *et al*. Amyloid burden, cortical thickness, and cognitive function in the Wisconsin Registry for Alzheimer's Prevention. *Alzheimer's & dementia (Amsterdam, Netherlands)* 2015;1(2):160-69. doi: 10.1016/j.dadm.2015.01.003 [published Online First: 2015/07/15]
4. Peters F, Villeneuve S, Belleville S. Predicting progression to dementia in elderly subjects with mild cognitive impairment using both cognitive and neuroimaging predictors. *J Alzheimers Dis* 2014;38(2):307-18. doi: 10.3233/jad-130842 [published Online First: 2013/08/22]
5. Lee J, Seo SW, Yang JJ, *et al.* Longitudinal cortical thinning and cognitive decline in patients with early- versus late-stage subcortical vascular mild cognitive impairment. *Eur J Neurol* 2018;25(2):326- 33. doi: 10.1111/ene.13500 [published Online First: 2017/12/21]
6. Yuan P, Raz N. Prefrontal cortex and executive functions in healthy adults: a meta-analysis of structural neuroimaging studies. *Neuroscience and biobehavioral reviews* 2014;42:180-92. doi: 10.1016/j.neubiorev.2014.02.005 [published Online First: 2014/02/27]
7. Seo SW, Lee JM, Im K, *et al.* Cardiovascular risk factors cause cortical thinning in

cognitively impaired patients: relationships among cardiovascular risk factors, white matter hyperintensities, and cortical atrophy. *Alzheimer Dis Assoc Disord* 2012;26(2):106-12. doi: 10.1097/WAD.0b013e31822e0831 [published Online First: 2011/09/29]

8. Viviani R, Pracht ED, Brenner D, *et al.* Multimodal MEMPRAGE, FLAIR, and [Formula: see text] Segmentation to Resolve Dura and Vessels from Cortical Gray Matter. *Front Neurosci* 2017;11:258. doi: 10.3389/fnins.2017.00258 [published Online First: 2017/05/09]
9. Desikan RS, Segonne F, Fischl B, *et al*. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 2006;31(3):968-80. doi: 10.1016/j.neuroimage.2006.01.021 [published Online First: 2006/03/15]
10. Makris N, Schlerf JE, Hodge SM, *et al.* MRI-based surface-assisted parcellation of human cerebellar cortex: an anatomically specified method with estimate of reliability. *NeuroImage* 2005;25(4):1146-60. doi: 10.1016/j.neuroimage.2004.12.056 [published Online First: 2005/04/27]
11. Fischl B, Salat DH, van der Kouwe AJ, *et al.* Sequence-independent segmentation of magnetic resonance images. *NeuroImage* 2004;23 Suppl 1:S69-84. doi: 10.1016/j.neuroimage.2004.07.016 [published Online First: 2004/10/27]
12. Fischl B, van der Kouwe A, Destrieux C, *et al.* Automatically parcellating the human cerebral cortex. *Cerebral cortex (New York, NY : 1991)* 2004;14(1):11-22. [published Online First: 2003/12/05]
13. Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic

resonance images. *Proc Natl Acad Sci U S A* 2000;97(20):11050-5. doi: 10.1073/pnas.200033797 [published Online First: 2000/09/14]

14. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage* 1999;9(2):179-94. doi: 10.1006/nimg.1998.0395 [published Online First: 1999/02/05]
15. Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis. II: Inflation, flattening, and a surface- based coordinate system. *NeuroImage* 1999;9(2):195-207. doi: 10.1006/nimg.1998.0396 [published Online First: 1999/02/05]
16. Fischl B, Sereno MI, Tootell RB, *et al.* High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human brain mapping* 1999;8(4):272-84. [published Online First: 2000/01/05]
17. Helms G, Kallenberg K, Dechent P. Contrast-driven approach to intracranial segmentation using a combination of T2- and T1-weighted 3D MRI data sets. *J Magn Reson Imaging* 2006;24(4):790-5. doi: 10.1002/jmri.20692
18. Lindig T, Kotikalapudi R, Schweikardt D, *et al.* Evaluation of multimodal segmentation based on 3D T1-, T2- and FLAIR-weighted images - the difficulty of choosing. *NeuroImage* 2018;170:210-21. doi: 10.1016/j.neuroimage.2017.02.016 [published Online First: 2017/02/07]
19. van der Kouwe AJW, Benner T, Salat DH, *et al.* Brain morphometry with multiecho MPRAGE. *NeuroImage* 2008;40(2):559-69. doi: 10.1016/j.neuroimage.2007.12.025 [published Online First: 2008/02/01]
20. Malpas CB. Structural neuroimaging correlates of cognitive status in older adults: A person-oriented approach. *Journal of clinical neuroscience : official journal of the*

Neurosurgical Society of Australasia 2016;30:77-82. doi: 10.1016/j.jocn.2016.01.025 [published Online First: 2016/04/09]

21. Bergman LR, Magnusson D, El Khouri BM. Studying individual Development in An interindividual Context A Person-oriented Approach. Paths Through Life Series v Vol 4. Hoboken: Taylor and Francis,, 2002:1 online resource (211 p.
22. Poulakis K, Pereira JB, Mecocci P, *et al.* Heterogeneous patterns of brain atrophy in Alzheimer's disease. *Neurobiol Aging* 2018;65:98-108. doi: 10.1016/j.neurobiolaging.2018.01.009 [published Online First: 2018/01/31]
23. Schwedt TJ, Si B, Li J*, et al.* Migraine Subclassification via a Data-Driven Automated Approach Using Multimodality Factor Mixture Modeling of Brain Structure Measurements. *Headache* 2017;57(7):1051-64. doi: 10.1111/head.13121 [published Online First: 2017/06/19]
24. Stanley L, Kellermanns FW, Zellweger TM. Latent Profile Analysis: Understanding Family Firm Profiles. *Family Business Review* 2016;30(1):84-102. doi: 10.1177/0894486516677426
25. Collins LM, Lanza ST. Latent class and latent transition analysis with applications in the social, behavioral, and health sciences. Wiley Series in Probability and Statistics v 718. Hoboken, NJ: Wiley,, 2010:1 online resource (331 p.
26. Tchistiakova E, MacIntosh BJ, Initiative AsDN. Summative effects of vascular risk factors on cortical thickness in mild cognitive impairment. *Neurobiol Aging* 2016;45:98-106. doi: 10.1016/j.neurobiolaging.2016.05.011 [published Online First: 2016/05/20]
27. Villeneuve S, Reed BR, Madison CM, *et al.* Vascular risk and Aβ interact to reduce cortical thickness in AD vulnerable brain regions. *Neurology* 2014;83(1):40-7. doi:

10.1212/WNL.0000000000000550 [published Online First: 2014/06/06]

28. Seo SW, Ahn J, Yoon U, *et al.* Cortical thinning in vascular mild cognitive impairment and vascular dementia of subcortical type. *Journal of neuroimaging : official journal of the American Society of Neuroimaging* 2010;20(1):37-45. doi: 10.1111/j.1552-6569.2008.00293.x [published Online First: 2009/02/18]
29. Zi W, Duan D, Zheng J. Cognitive impairments associated with periventricular white matter hyperintensities are mediated by cortical atrophy. *Acta neurologica Scandinavica* 2014;130(3):178-87. doi: 10.1111/ane.12262 [published Online First: 2014/05/20]
30. Cardenas VA, Reed B, Chao LL, *et al.* Associations among vascular risk factors, carotid atherosclerosis, and cortical volume and thickness in older adults. *Stroke; a journal of cerebral circulation* 2012;43(11):2865-70. doi: 10.1161/STROKEAHA.112.659722 [published Online First: 2012/09/13]
31. Burzynska AZ, Nagel IE, Preuschhof C, *et al.* Cortical thickness is linked to executive functioning in adulthood and aging. *Human brain mapping* 2012;33(7):1607-20. doi: 10.1002/hbm.21311 [published Online First: 2011/07/09]
32. Grambaite R, Selnes P, Reinvang I, *et al.* Executive dysfunction in mild cognitive impairment is associated with changes in frontal and cingulate white matter tracts. *J Alzheimers Dis* 2011;27(2):453-62. doi: 10.3233/jad-2011-110290 [published Online First: 2011/08/16]
33. Manard M, Bahri MA, Salmon E, *et al.* Relationship between grey matter integrity and executive abilities in aging. *Brain research* 2016;1642:562-80. doi: 10.1016/j.brainres.2016.04.045 [published Online First: 2016/04/25]
34. Schmidt EL, Burge W, Visscher KM, *et al.* Cortical thickness in frontoparietal and cingulo-

opercular networks predicts executive function performance in older adults. *Neuropsychology* 2016;30(3):322-31. doi: 10.1037/neu0000242 [published Online First: 2015/10/16]

35. Vasconcelos Lde G, Jackowski AP, Oliveira MO, *et al.* The thickness of posterior cortical areas is related to executive dysfunction in Alzheimer's disease. *Clinics (Sao Paulo, Brazil)* 2014;69(1):28- 37. doi: 10.6061/clinics/2014(01)05 [published Online First: 2014/01/30]
36. Bakkour A, Morris JC, Dickerson BC. The cortical signature of prodromal AD: regional thinning predicts mild AD dementia. *Neurology* 2009;72(12):1048-55. doi: 10.1212/01.wnl.0000340981.97664.2f [published Online First: 2008/12/26]
37. Racine AM, Fong TG, Travison TG, *et al.* Alzheimer's-related cortical atrophy is associated with postoperative delirium severity in persons without dementia. *Neurobiol Aging* 2017;59:55-63. doi: 10.1016/j.neurobiolaging.2017.07.010 [published Online First: 2017/08/29]
38. Fortin JP, Cullen N, Sheline YI, *et al.* Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage* 2018;167:104-20. doi: 10.1016/j.neuroimage.2017.11.024 [published Online First: 2017/11/21]
39. Muthen TAaB. Variable-Specific Entropy Contribution. In: MPlus, ed. MPlus, 2014.
40. Bolck A, Croon M, Hagenaars J. Estimating Latent Structure Models with Categorical Variables: One- Step Versus Three-Step Estimators. *Political Analysis* 2004;12(1):3-27.
41. Viviani R, Stocker T, Stingl JC. Multimodal FLAIR/MPRAGE segmentation of cerebral cortex and cortical myelin. *NeuroImage* 2017;152:130-41. doi: 10.1016/j.neuroimage.2017.02.054 [published Online First: 2017/03/04]

42. Salat DH, Lee SY, van der Kouwe AJ, *et al*. Age-associated alterations in cortical gray and white matter signal intensity and gray to white matter contrast. *NeuroImage* 2009;48(1):21-8. doi: 10.1016/j.neuroimage.2009.06.074 [published Online First: 2009/07/04]

43. Leritz EC, Salat DH, Williams VJ, *et al.* Thickness of the human cerebral cortex is associated with metrics of cerebrovascular health in a normative sample of community dwelling older adults. *Neuroimage* 2011;54(4):2659-71. doi: 10.1016/j.neuroimage.2010.10.050 [published Online First: 2010/11/03]

44. Risacher SL, Saykin AJ, West JD, *et al.* Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. *Curr Alzheimer Res* 2009;6(4):347-61. [published Online First: 2009/08/20]

45. Lee JS, Kim C, Shin JH, *et al*. Machine Learning-based Individual Assessment of Cortical Atrophy Pattern in Alzheimer's Disease Spectrum: Development of the Classifier and Longitudinal Evaluation. *Sci Rep* 2018;8(1):4161. doi: 10.1038/s41598-018-22277-x [published Online First: 2018/03/09]

46. Jones RN, Manly J, Glymour MM, *et al.* Conceptual and measurement challenges in research on cognitive reserve. *Journal of the International Neuropsychological Society : JINS* 2011;17(4):593- 601. doi: 10.1017/s1355617710001748 [published Online First: 2011/03/18]

47. Fong TG, Davis D, Growdon ME, *et al.* The interface between delirium and dementia in elderly adults. *The Lancet Neurology* 2015;14(8):823-32. doi: 10.1016/s1474-4422(15)00101-5 [published Online First: 2015/07/04]

Table 1: Means and Standard Deviations (SD) for descriptive variables. Significance reported per LPA Class.

Variable	Mean (SD) N=54	T1-Unadj LPA 2-Class		T1-Scanner LPA 2-Class		T1-ICVScan LPA 2-Class		T2-Unadj LPA 2-Class		T2-Scanner LPA-2-Class		T2-ICVScan LPA 2-Class	
		LH	RH	LH	RH	LH	RH	LH	RH	LH	RH	LH	RH
Age	71.74±4.84	*>C1		*>C1		*>C1		*F>C1	*>C1		*>C1	*>C1	*>C1
ASA	2.61±0.63							*F>C1					
FCDR	35.31±20.32	*F>C1											
SBP	134.3±17.69							*F>C1					
DBP	74.11±11.41												
PP	60.19±15.37												
MAP	94.17±11.78												
GDS15	2.24 ±2.49									*<C1	*<C1	*<C1	*<C1
EDUC	1.74 ± 0.56							*F>C1					
TMTA	36.18±10.27												
TMTB	93.69±52.42												
BNT	28.43±1.81												
COWAT-adj	-0.34±1.1												
RBANS Tot	91.43±14.26												
RBANS LL	22.74±5.53												
RBANS SM	16.06±4.27												
RBANS FC	15.54±2.83												
RBANS LO	16.22±3.06							*<C1					
RBANS PN	9.93±0.26	*>C1	*>C1	*>C1	*>C1	*>C1	*>C1						
RBANS SF	18.61±4.90												
RBANS DS	11.44±2.41						*>C1						
RBANS Cod	40.57±11.26								*<C1	*<C1		*<C1	
RBANS LRec	3.5±2.95												
RBANS List	17.5±2.39												
RBANS SR	7.89±2.46			*<C1									
RBANS FCR	11.17±3.54												
Sex	48% (female)												
Past Smoke	67%												
Cur Smoke	22%	*>C1											
Scanner Type	3T #1: 34 3T #2: 14 3T #3: 5 1.5T #4: 1	*	*										

Significance levels: * = p<0.05, F=Survived False Discovery Rate correction. C1=CAC class. > or < = indicates direction of mean

Abbreviations: ASA=American Society of Anesthesiologists Classification Score, BNT=Boston Naming Test, C1=Cortical Atrophy Class, COWAT=Controlled Oral Word Association Test adjusted for age and education, EDUC=Years of Education, DBP=Diastolic Blood Pressure, GDS15=Depression symptom score, MAP=Mean Arterial Pressure, PP=Pulse Pressure, RBANS=Repeatable Battery for the Assessment of Neuropsychological Status (Tot=Total Score, LL=List Learning, SM=Story Memory, FC=Figure Copy, LO=Line Orientation, PN=Picture Naming, SF=Semantic Fluency, DS=Digit Symbol, Cod=Coding, LRec=List recall, List=List recognition, SR=Story Recall, FCR=Figure Copy Recall).

Table 2: LPA Model Fit Statistics. 4 Classes run for each model, the null model is not shown.

Data	Class	Hemi	Relative Entrop	BIC adjusted	Relative BIC	Average Posterior Probilities per cl	Class Count	Lo-Mendall-Rubin p-va
T1unadj	**2**	**LH**	**0.988**	**-**	**0.999**	**0.996/0.999**	**27/27**	**0.0307**
		RH	**0.973**	**-2020.57**	**0.999**	**0.999/0.988**	**30/24**	**0.1194**
T1unadj.	3	LH	0.998	-	0.999	1.00/0.999/0.999	19/28/7	0.3543
		RH	0.993	-2192.28	0.999	0.998/0.998/0.999	27/10/17	0.4112
T1unadj.	4	LH	0.987	-	0.999	0.990/0.997/0.997/	17/16/14/	0.8863
		RH	0.992	-2271.80	0.999	1.00/0.997/1.00/0.995	5/17/17/15	0.7193
T1scana	**2**	**LH**	**0.989**	**-**	**0.999**	**0.997/0.999**	**30/24**	**0.0397**
		RH	**0.980**	**-2124.49**	**0.999**	**0.998/0.996**	**27/27**	**0.0777**
T1scana	3	LH	0.993	-	0.999	0.999/0.998/0.997	16/29/9	.4352
		RH	0.997	-2353.49	0.999	1/0.999/1.0	12/27/15	0.3134
T1scana	4	LH	0.981	-	0.999	0.993/0.979/1.0/1.	19/11/15/	0.8988
		RH	0.988	-2418.06	0.999	0.991/0.999/0.997/1.0	16/5/21/12	0.8412
T1ICVS	**2**	**LH**	**0.994**	**-**	**0.999**	**1/0.999**	**30/24**	**0.0252**
		RH	**0.969**	**-2163.98**	**0.999**	**0.981/1.000**	**26/28**	**0.40**
T1ICVS	3	LH	0.991	-	0.999	0.995/0.999/0.999	14/29/11	0.4863
		RH	0.995	-2378.32	0.999	0.999/0.999/0.999	27/13/14	0.2594
T1ICVS	4	LH	0.725	-	0.999	0.995/0.449/0.550/	14/13/16/	1.000
		RH	0.995	-2454.99	0.999	1.000/0.997/1.00/0.99	10/17/13/14	0.8084
T2unadj	**2**	**LH**	**1.000**	**-838.553**	**0.999**	**1/1**	**4/50**	**0.2871**
		RH	**1.000**	**-852.441**	**0.999**	**1/1**	**9/45**	**0.2491**
T2unadj	3	LH	0.996	-	0.999	0.994/1.00/1.00	11/39/4	0.7628
		RH	0.988	-1103.98	0.999	1/1/.981	8/31/15	0.7193
T2unadj	4	LH	0.994	-	0.999	1/1/0.990/1	4/6/14/30	0.4521
		RH	0.993	-1304.71	0.999	1/0.997/1.0/0.996	5/31/6/12	0.6390
T2scan	**2**	**LH**	**1.000**	**-961.168**	**0.999**	**1/1**	**15/39**	**0.1907**
		RH	**1.000**	**-978.141**	**0.999**	**1/1**	**13/41**	**0.1054**
T2scan	3	LH	1.000	-	0.999	1/1/1	4/12/38	0.6430
		RH	0.998	-1217.18	0.999	0.999/1/1	19/7/28	0.5076
T2scan	4	LH	0.995	-	0.999	1/0.994/1/1	6/14/1/33	0.7201
		RH	0.996	-1412.40	0.999	1/1/0.998/1	9/4/27/14	0.7629
T2ICVS	**2**	**LH**	**1.000**	**-**	**0.999**	**1/1**	**12/42**	**0.1732**
		RH	**1.000**	**-992.215**	**0.999**	**1/1**	**41/13**	**0.1151**
T2ICVS	3	LH	0.995	-	0.999	1/1/0.998	5/14/35	0.6709
		RH	0.999	-1240.84	0.999	1/0.999/1	9/14/31	0.5034
T2ICVS	4	LH	0.998	-	0.999	1/1/1/1	1/14/4/35	0.2708
		RH	1.000	-1457.73	0.999	1/1/1/1	9/14/27/4	0.6541

Bold= Selected models

Abbreviations: T1-unadjust=T1-weighted unadjusted cortical thickness measures, T1scanad=T1-weighted ComBat adjusted cortical thickness measures, T1ICVS=T1-weighted Combat and Intracranial Volume adjusted cortical thickness measures. T2unadj=T1+T2-

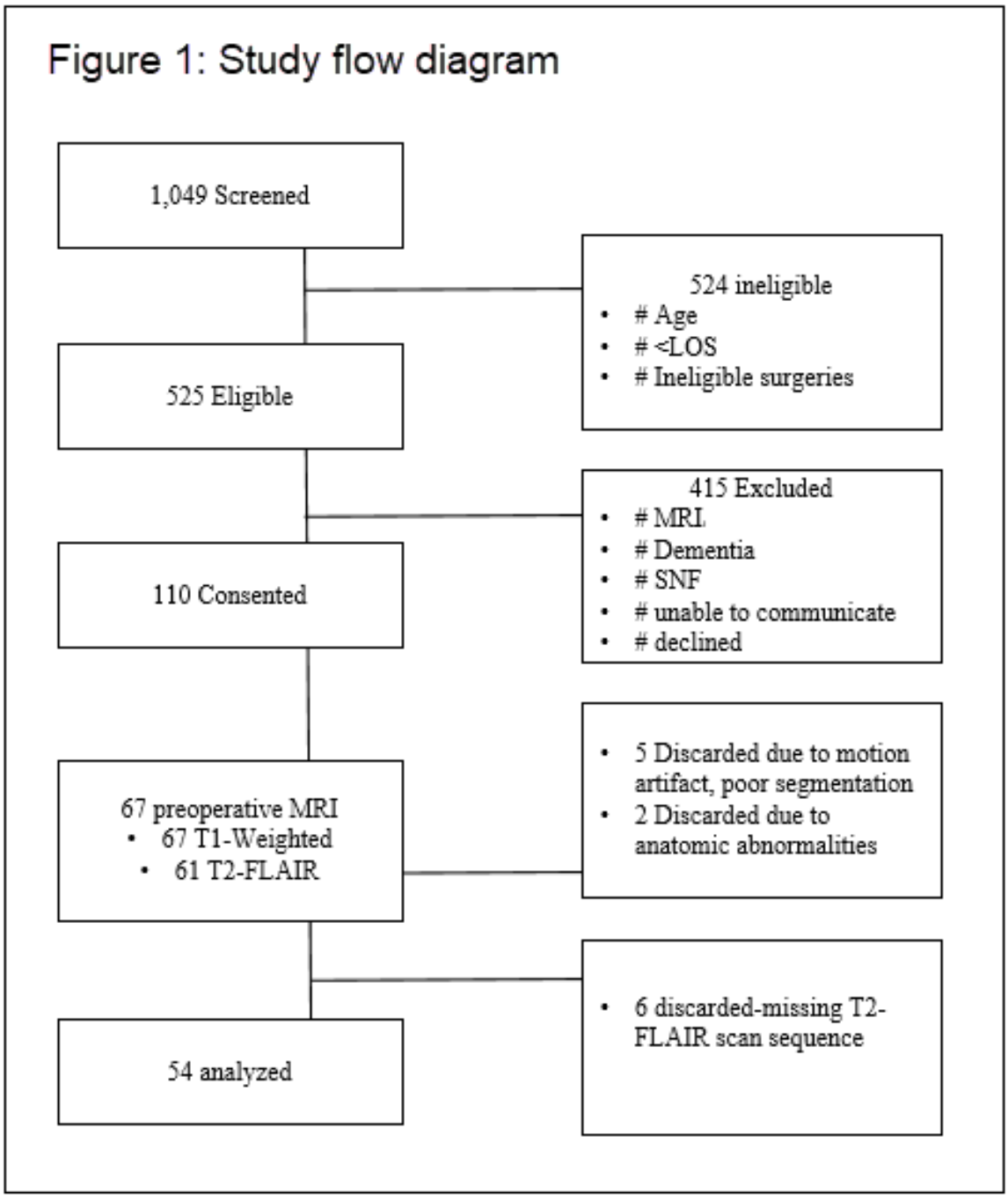


Figure 1: Study flow diagram
1,049 Screened
524 ineligible
Age
<LOS
Ineligible surgeries
525 Eligible
415 Excluded
MRI
Dementia
SNF
unable to communicate
declined
110 Consented
67 preoperative MRI
67 T1-Weighted
61 T2-FLAIR
5 Discarded due to motion artifact, poor segmentation
2 Discarded due to anatomic abnormalities
6 discarded-missing T2-FLAIR scan sequence
54 analyzed

Figure 2: This boxplot displays the statistical differences between the T1-only and T1+T2-FLAIR cortical thickness means. T1+T2-FLAIR data is significantly thicker, representing more gray matter, when compared to T1 only.

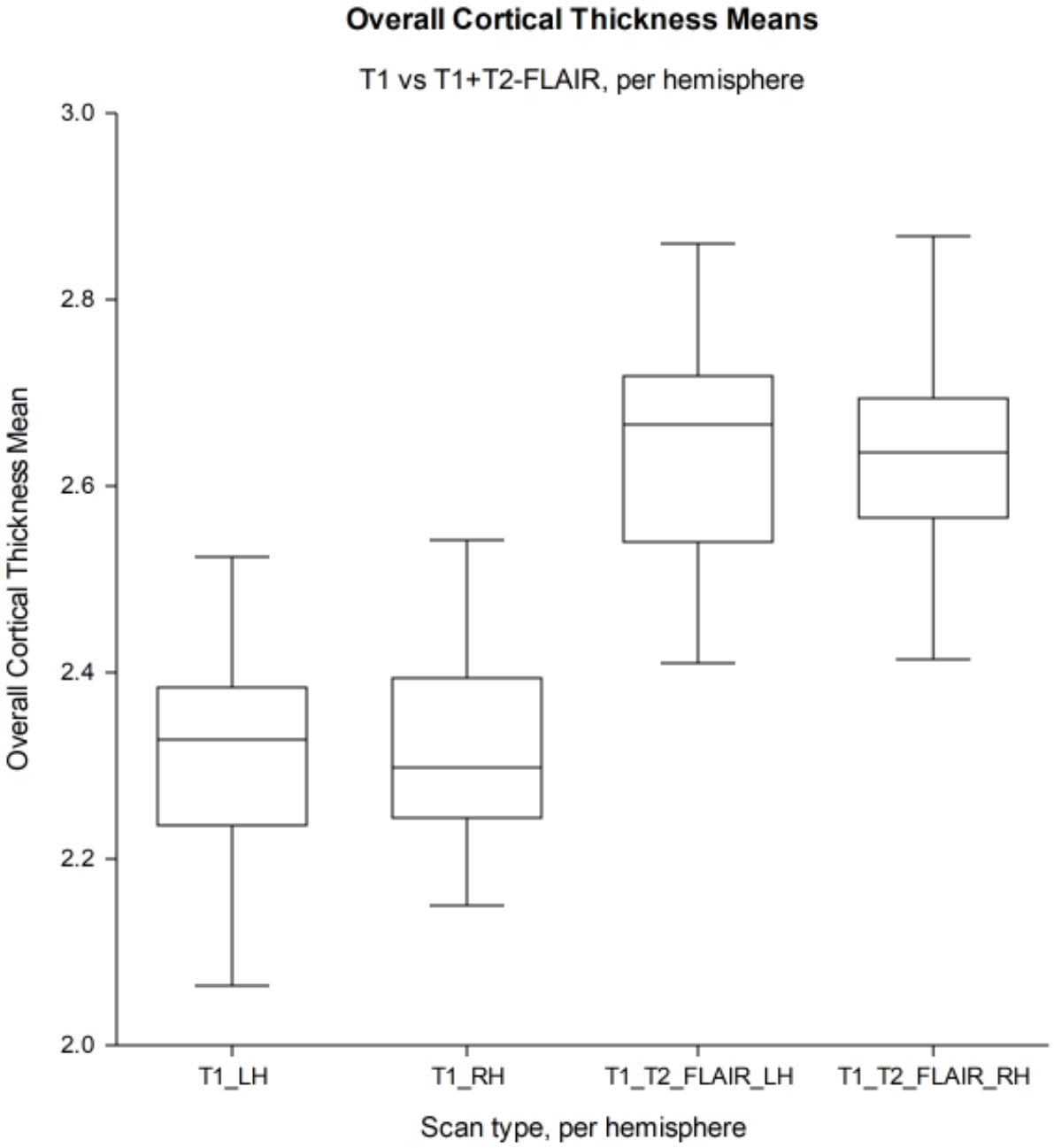
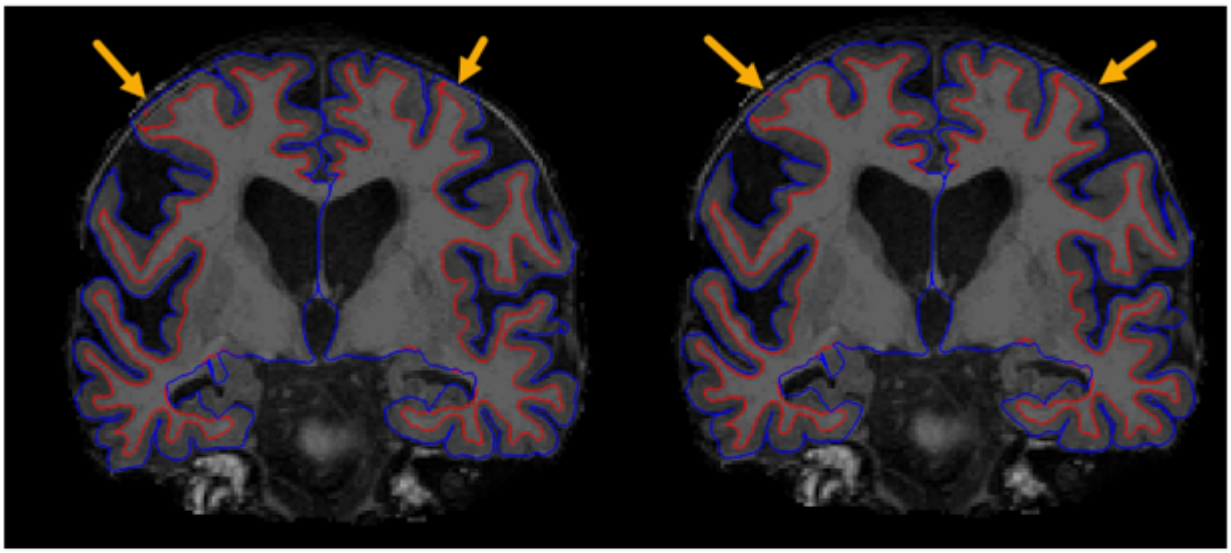


Figure 3: Represents the segmentation changes from T1-weighted to T1+T2-FLAIR processing. The red line represents the segmented white matter (lighter gray color) from the gray matter (darker gray color). the blue line represents the gray matter segmentation from dural and outlying brain material. Yellow arrows on the left point to segmentation overlap into the dura. The segmentation error is corrected in the image on the left, indicated by the yellow arrows. The bottom figures compare alignment.

T1 Segmentation

T1+T2-FLAIR Alignment



T1 Alignment

T1+T2-FLAIR Alignment

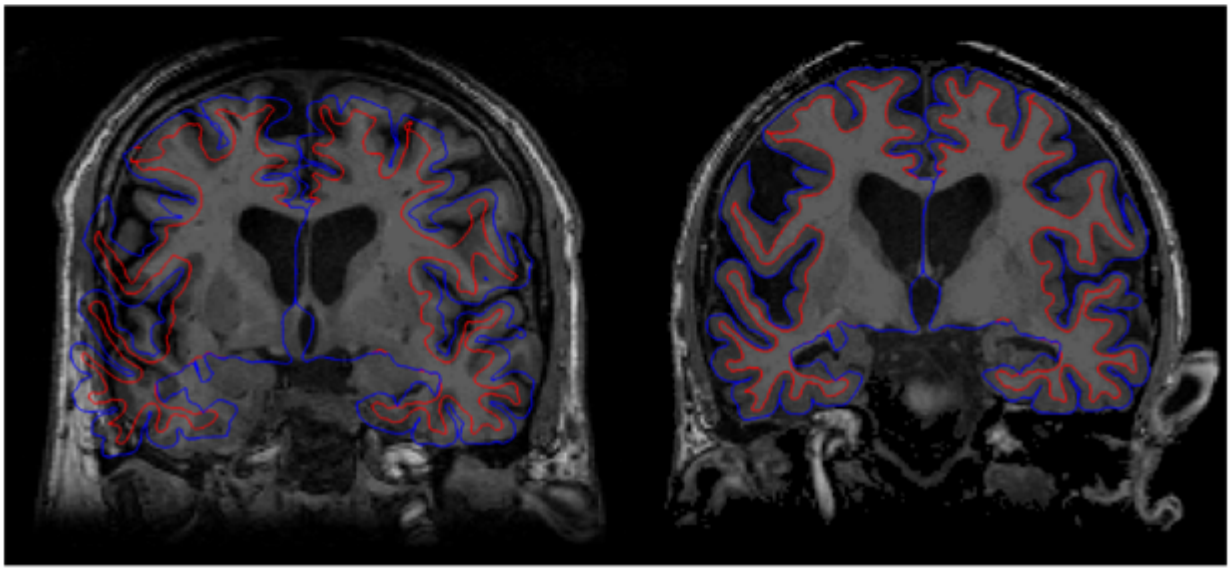


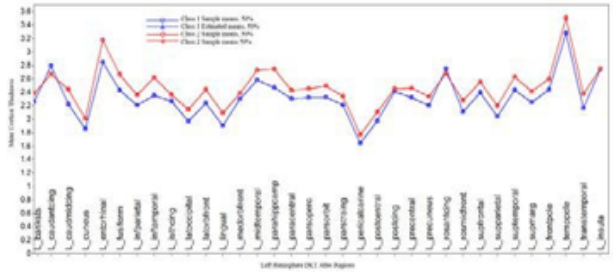
Table 3: Statistical differences between T1-only Unadjusted LPA Model				
	Class 1=Cortical		Class 2= Non-atrophy	
Hemisphere	L H	R H	L H	RH 24
Class Membership	27	30	27	Mean±S
Cortical	2.32 ± 0.31	2.31 ±	2.47 ± 0.32	2.43 ± 0.34
Age	73.1 ± 0.94*		70.4 ± 0.83	
FCDR Score Current	44.0 ± 3.91***		26.63 ± 3.04	
tobacco	0.33 ± 0.09*	10.0±0.00*	0.11 ± 0.06	9.831±0.07
RBANS PN	10.0 ± 0.00*		9.85 ± 0.07	
FDR Adj Framingham	44.0 ±3.91***			
Significance p-levels: *p<0.05, **p<0.01, ***p<0.001				
FCDR: Framingham Cardiovascular Disease BMI Risk Score RBANS: Repeatable Battery for the Assessment of Neuropsychological Status RBANS PN: Picture Naming subtest RBANS SR: Story Recall subtest				

Table 4: Statistical differences between T1-only scanner adjusted LPA				
	Class 1=Cortical		Class 2= Non-atrophy	
Hemisphere Class Membership Count Significant	L H 24	R H 27	L H 30	RH 27 Mean+S
Cortical	2.32±0.32**	2.31 ±	2.46 ± 0.32	2.46 ± 0.34
Age	73.2 ± 0.95*		70.4 ± 0.82	
RBANS PN	10.0 ± 0.00*	10.0+0.00*	9.87 ± 0.06	9.831+0.07
FDR	None			
Significance p-levels: *p<0.05, **p<0.01, ***p<0.001				
RBANS: Repeatable Battery for the Assessment of Neuropsychological Status RBANS PN: Picture Naming subtest				

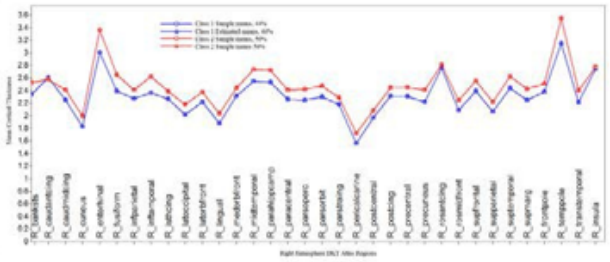
Figure 4:

Figure 4: T1-only unadjusted 2-Class LPA Models, left and right hemispheres are displayed. Plots A and B display the sample and estimated cortical thickness means per class. This shows the separation between the classes based on mean cortical thickness. Figures C and D are regions of atrophy predicted by class membership. Atrophy is indicated by the blue color, with increasing values showing as lighter blue. These lighter blue regions survived False Discovery Rate correction at 0.05. Supplementary table 1 lists the largest 5 identified clusters with the talairach coordinates.

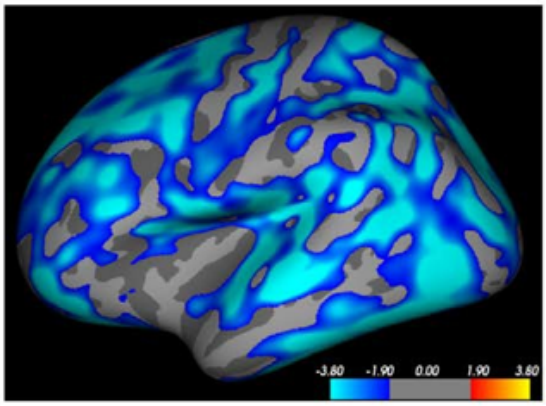
Plot A. Left Hemisphere T1-only unadjusted 2-Class LPA Model



Plot B. Right Hemisphere T1-only unadjusted 2-Class LPA Model



Plot C: Left hemisphere cortical atrophy predicted by class membership



Plot D: Right hemisphere cortical atrophy predicted by class membership

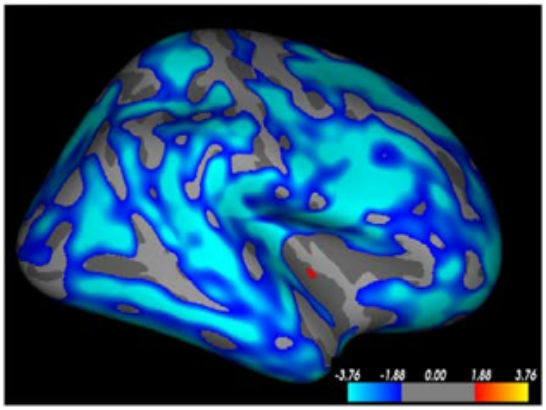
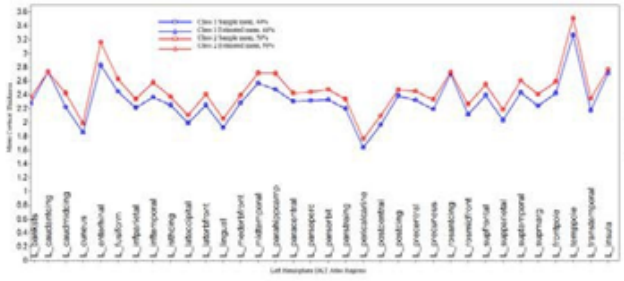
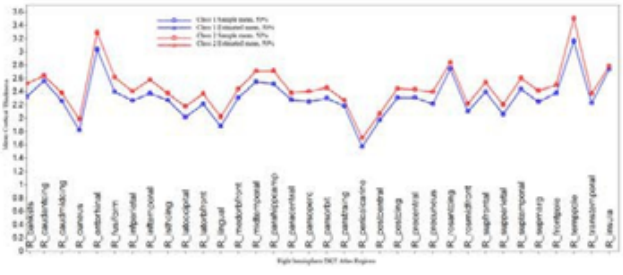


Figure 5: T1-only Scanner adjusted 2-Class LPA Models, left and right hemispheres are displayed. Plots A and B display the sample and estimated cortical thickness means per class. This shows the separation between the classes based on mean cortical thickness. Figures C and D are regions of atrophy predicted by class membership. Atrophy is indicated by the blue color, with incresing values showing as lighter blue. These lighter blue regions survived False Discovery Rate correction at 0.05. Supplementary table 2 lists the largest 5 identified clusters with the talairach coordinates.

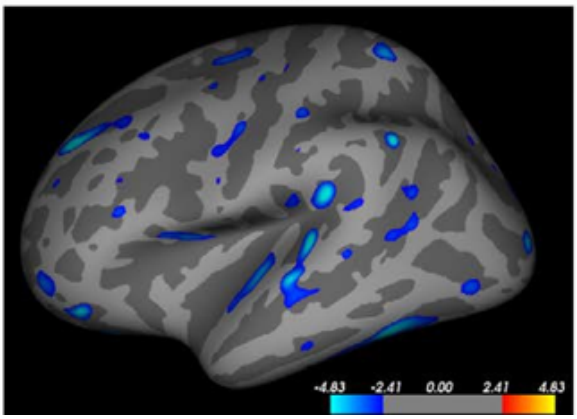
Plot A. Left Hemisphere T1-only Scanner adjusted 2-Class LPA Model



Plot B. Right Hemisphere T1-only Scanner adjusted 2-Class LPA Model



Plot C: Left hemisphere cortical atrophy predicted by class membership



Plot D: Right hemisphere cortical atrophy predicted by class membership

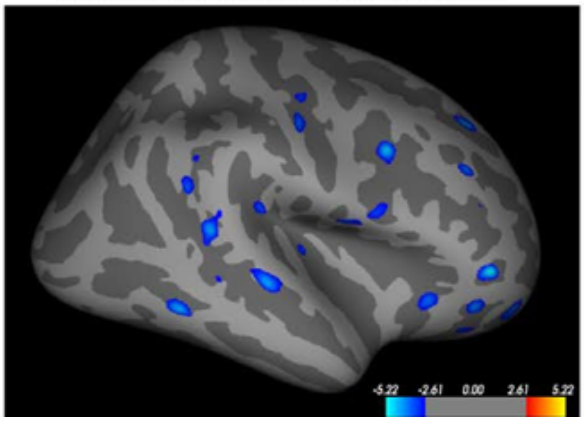
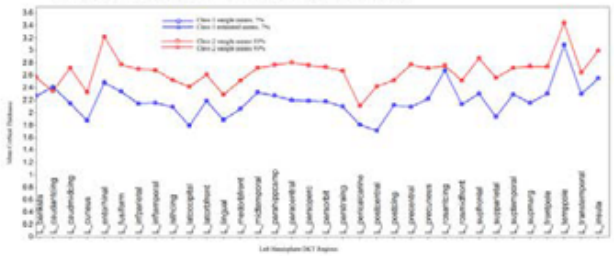


Table 5: Statistical comparison between T1+T2-FLAIR Unadjusted LPA					
		Class 1=Cortical		Class 2= Stable	
Hemisphere		L H	R H	L H	R H
Class Membership		4	9	50	45
Cortical		2.20 ±	2.33 ±	2.67 ± 0.25	2.70 ± 0.28
	Age	78.5 ±	76.11 ±	71.2 ± 4.41	70.87 ± 4.27
	AS	4.15***	4.86**	2.67 ± 0.57	
	A	2.00 ±		133.32 ±	
	SB	0.00***		17.71	
	P	146.5 ±		1.7 ± 0.57	
	Education	8.29***		16.40 ± 3.03	2.44 ± 2.63
		2.0 ± 0.00***	1.22 ± 0.92*		
FDR	Ag	78.5 ± 2.07**			
	e	2.00 ± 0.00***			
	AS	146.5 ± 4.1***			
	A	2.0 ± 0.00***			
Significance p-levels: *p<0.05, **p<0.01, ***p<0.001					
RBANS: Repeatable Battery for the Assessment of Neuropsychological Status RBANS LO: Line Orientation subtest					

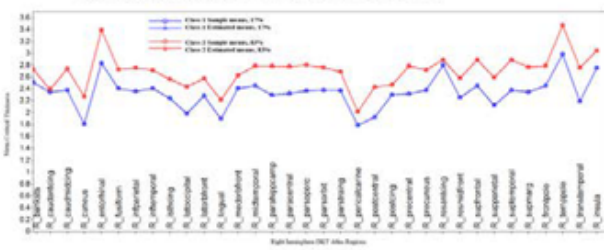
Table 6: Statistical differences between T1+T2-FLAIR Scanner adjusted LPA				
	Class 1=Cortical		Class 2= Stable	
Hemisphere	L H	R H	L H	RH 41
Class Membership	15	13	39	Mean±S
Cortical	2.42 ±	2.42 ±	2.71 ± 0.26	2.71 ± 0.28
Age		74.69 ± 5.50*		70.81 ± 4.13
Depression,	1.13 ±	1.23 ± 0.97*	2.67 ± 2.73	
GDS15 RBANS	0.96**		42.08 ±	2.56 ± 2.71
FDR Adj	None			
Significance p-levels: *p<0.05, **p<0.01, ***p<0.001				
RBANS: Repeatable Battery for the Assessment of Neuropsychological Status				

Figure 6: T1+T2-FLAIR Unadjusted 2-Class LPA Models, left and right hemispheres are displayed. Plots A and B display the sample and estimated cortical thickness means per class. This shows the separation between the classes based on mean cortical thickness. Figures C and D are regions of atrophy predicted by class membership. Atrophy is indicated by the blue color, with increasing values showing as lighter blue. These lighter blue regions survived False Discovery Rate correction at 0.05. Supplementary table 3 lists the largest 5 identified clusters with the talairach coordinates.

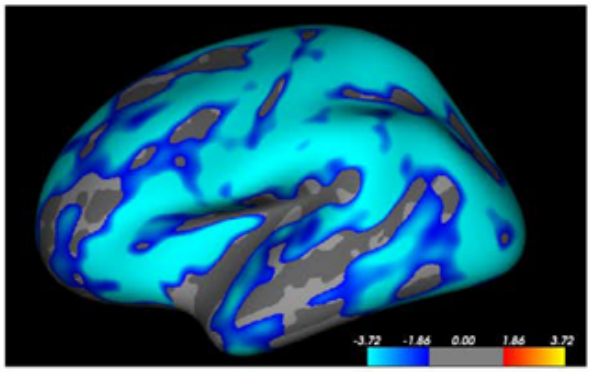
Plot A. Left Hemisphere T1+T2-FLAIR 2-Class LPA Model



Plot B. Right Hemisphere T1+T2-FLAIR 2-Class LPA Model



Plot C: Left hemisphere cortical atrophy predicted by class membership



Plot D: Right hemisphere cortical atrophy predicted by class membership

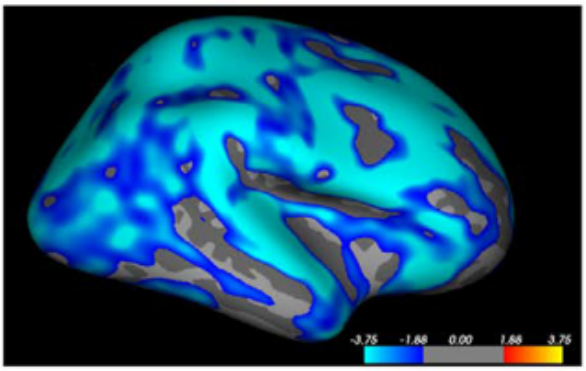
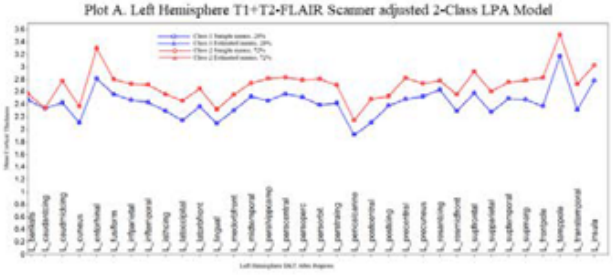
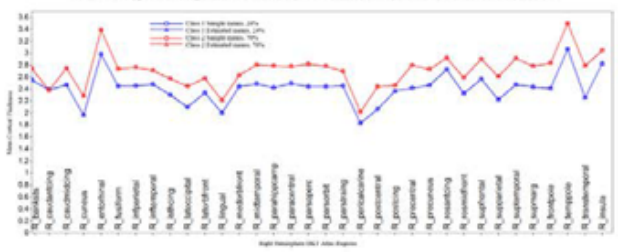


Figure 7: T1+T2-FLAIR Scanner adjusted 2-Class LPA Models. left and right hemispheres are displayed. Plots A and B display the sample and estimated cortical thickness means per class. This shows the separation between the classes based on mean cortical thickness. Figures C and D are regions of atrophy predicted by class membership. Atrophy is indicated by the blue color, with increasing values showing as lighter blue. These lighter blue regions survived False Discovery Rate correction at 0.05. Supplementary table 4 lists the largest 5 identified clusters with the talairach coordinates.

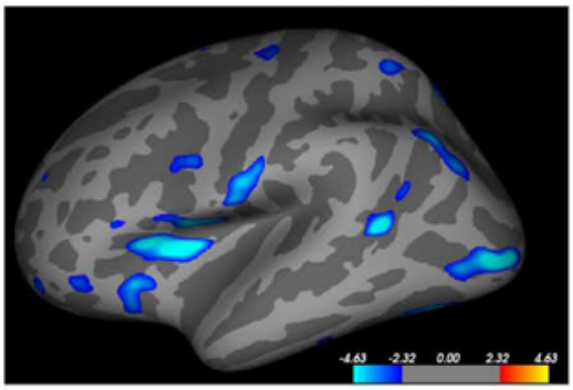
Plot A. Left Hemisphere T1+T2-FLAIR Scanner adjusted 2-Class LPA Model



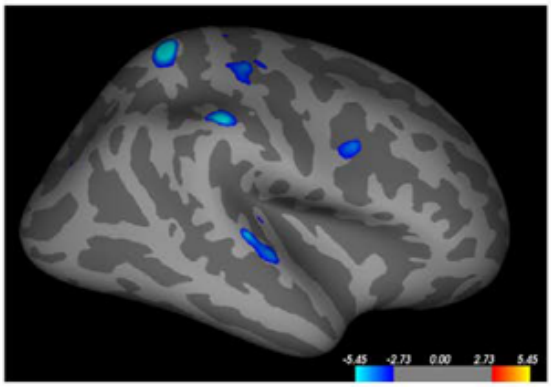
Plot B. Right Hemisphere T1+T2-FLAIR Scanner adjusted 2-Class LPA Model



Plot C: Left hemisphere cortical atrophy predicted by class membership

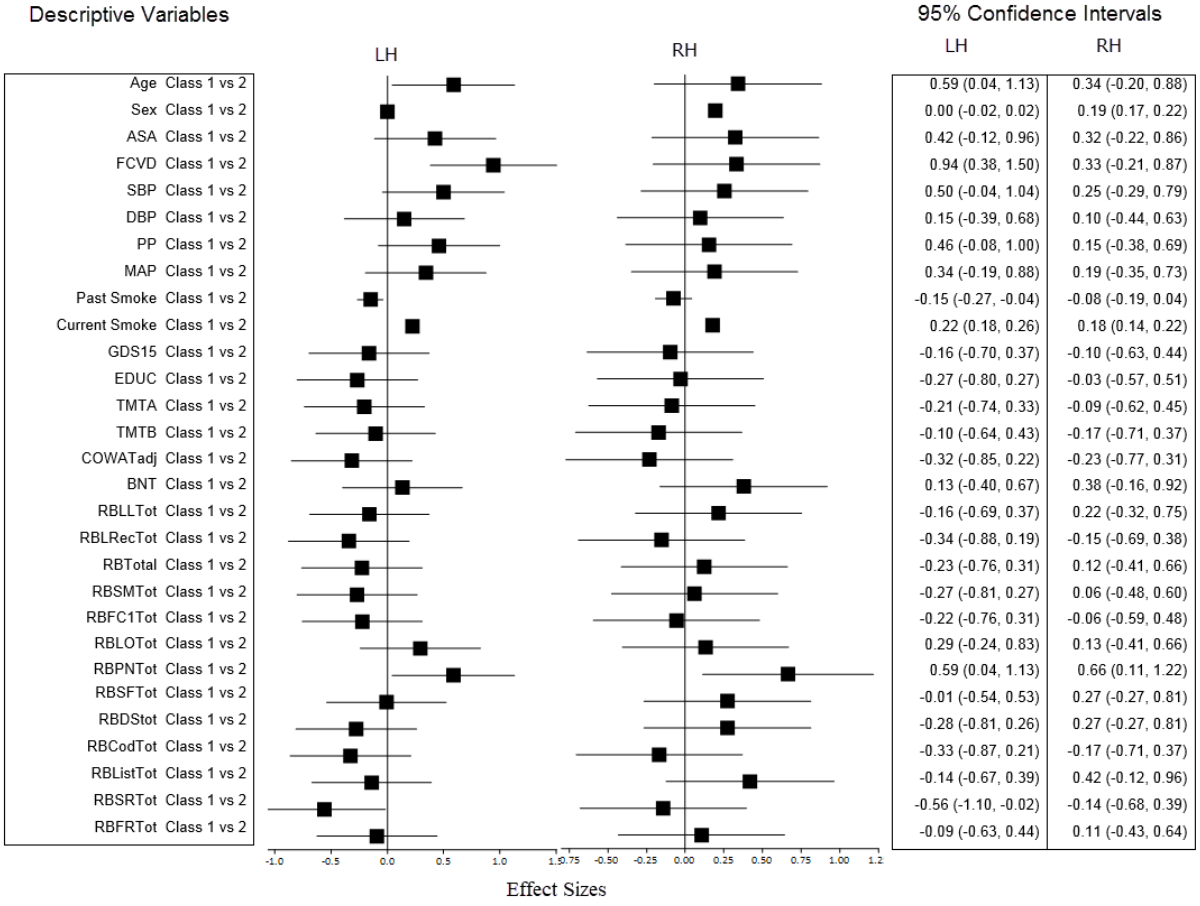


Plot D: Right hemisphere cortical atrophy predicted by class membership



Supplementary Materials

Supplementary Figure 1, T1 Unadjusted 2-Class LPA model displays the calculated standardized effect sizes for the left and right hemisphere, 2-class models. Error bars and 95% confidence intervals are included. These show descriptive data trends between the 2-class models.

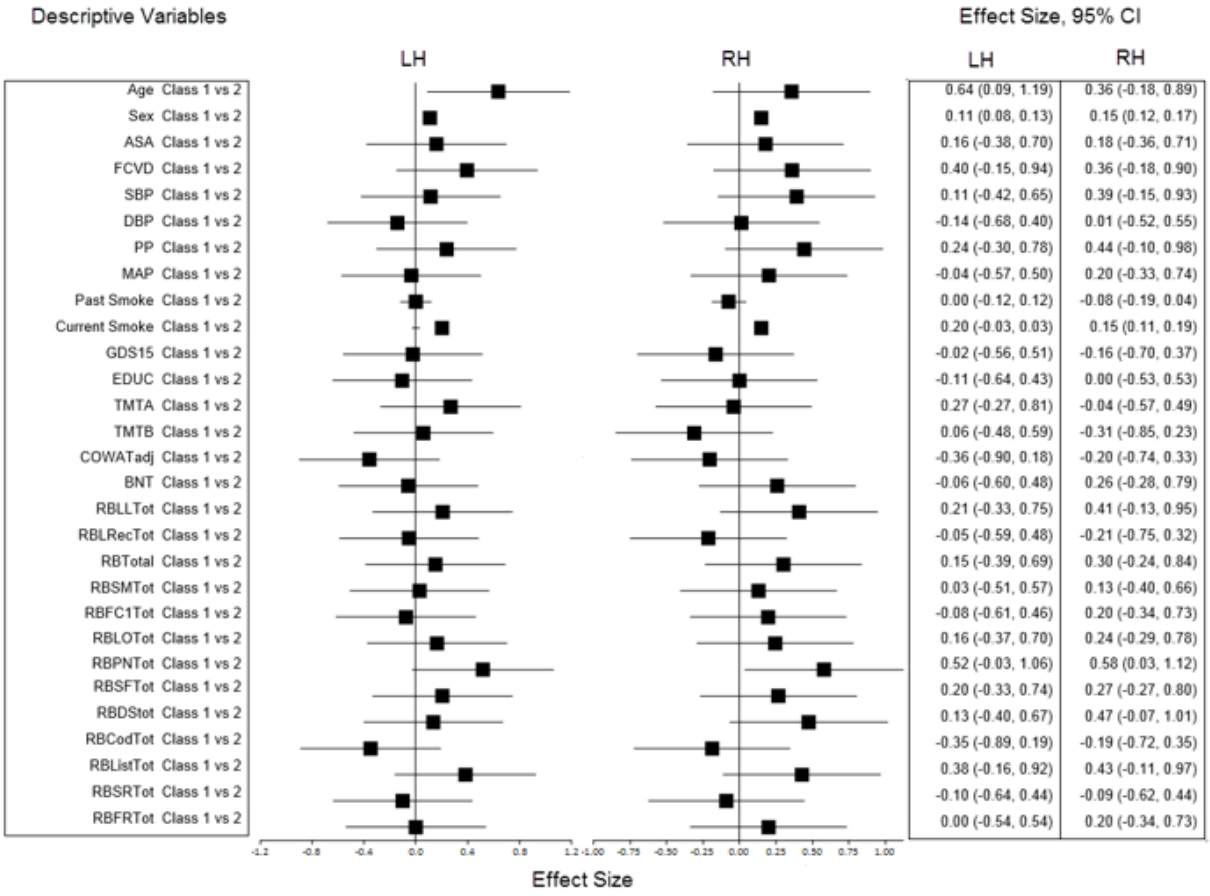


Abbreviations: ASA=American Society of Anesthesiologists classification, BNT=Boston Naming Test, CI=Confidence Intervals, COWATadj=Controlled Oral Word Association Test adjusted for age and education, DBP=Diastolic Blood Pressure, EDUC=Years of education, FCVD=Framingham Cardiovascular Disease Risk Score, GDS15=Geriatric Depression score, LPA=Latent Profile Analysis, MAP=Mean Arterial Pressure, PP=Pulse Pressure, RBANS=Repeatable Batttery Assessment for Neurpsychological Status, RBCodTot=RBANS Coding sub-test score, RBDSTot=RBANS Digit Symbol sub-test score, RBFC1Tot=RBANS Figure Copy 1 subtest score, RBListTot=RBANS List recognition subtest score, RBLLTot=RBANS list learning subtest score, RBLRecTot=RBANS List Recall sub-test total score, RBLOTot=RBANS Line Orientation subtest score, RBPNTot: RBANS Picture Naming subtest score, RBSFTot=RBANS Semantic Fluency subtest score, RBSMTot=RBANS Story Memory subtest score, RBSRTot=RBANS Story Recall subtest score, RBTotal=Total RBANS Score,

Supplementary Table 1: **T1 unadjusted 2-Class LPA. Class Membership Predicting**					
Right					
Cortex Region	**Cluste**	**Size**	**Talx**	**Taly**	**Talz**
RH-Precentral	-8.23	52637.98	49.5	5.6	10.7
RH-Superior Frontal	-2.25	38.96	15.2	37.4	15.7
RH-Superior Frontal	-1.75	18.71	12.1	2.4	39.9
RH-Insula	1.67	16.03	35.6	-13.5	-5.2
RH-101677 of 163842 Vertices survived FDR at 0.05, identifying					
Left					
Cortex Region	**Cluste**	**Size**	**Talx**	**Taly**	**Talz**
LH-	-11.05	47873.09	-19.8	41.4	33.2
LH-Supramarginal	-3.44	121.91	.-54.9	.27.5	32.1
LH-Medialorbitofrontal	-2.45	79.73	-14.1	45.2	-2.5
LH-Superior frontal	-2.07	46.64	-13.2	44.6	14.6
LH-	1.77	16.08	-5.8	8.3	29.5
LH-92292 of 163842 vertices survived FDR at 0.05, identifying 8 clusters. The five largest regions are listed					
Abbreviations/Definitions: Cluster: Maximum –log10(p-value), LH=Left Hemisphere, Size: Surface area of cluster, Talx,y,z:					

Supplementary Table 2: **T1 Scanner Adjusted 2-class LPA. Class membership predicting**					
Cortex Region	**Cluste**	**Size**	**Talx**	**Taly**	**Talz**
RH-Parstriangularis	-5.13	89.50	40.4	39.7	-2.9
RH-	-4.95	86.89	11.8	23.1	-15.3
RH-	-4.75	73.31	34.2	5.6	33.5
RH-Fusiform	-4.55	142.80	40.3	-6.8	-32.7
RH-Inferior temporal	-4.46	82.70	54.3	-50.3	-11.2
RH: 3499 of 163842 Vertices survived FDR at 0.05. Five largest					
Left					
Cortex Region	**Cluste**	**Size**	**Talx**	**Taly**	**Talz**
LH-	-5.74	271.92	-20.2	38.9	33.0
LH-Supramarginal	-5.47	67.86	-48.2	-52.3	34.4
LH-	-5.40	467.69	-16.8	31.3	-22.3
LH-Supramarginal	-5.28	208.54	-45.3	-34.4	24.0
LH-Inferior Temporal	-5.28	463.18	-45.2	-47.6	-12.9
LH: 8714 of 163842 Vertices survived FDR at 0.05. Five largest					
Abbreviations/Definitions: Cluster: Maximum –log10(p-value), LH=Left Hemisphere, Size: Surface area of cluster, Talx,y,z: Talairach (MNI305) coordinates					

Supplementary Figure 2, T1 Scanner adj. 2-Class LPA model displays the calculated standardized effect sizes for the left and right hemisphere, 2-class models. Error bars and 95% confidence intervals are included. These show descriptive data trends between the 2-class models.



Abbreviations: ASA=American Society of Anesthesiologists classification, BNT=Boston Naming Test, CI=Confidence Intervals, COWATadj=Controlled Oral Word Association Test adjusted for age and education, DBP=Diastolic Blood Pressure, EDUC=Years of education, FCVD=Framingham Cardiovascular Disease Risk Score, GDS15=Geriatric Depression score, LPA=Latent Profile Analysis, MAP=Mean Arterial Pressure, PP=Pulse Pressure, RBANS=Repeatable Batttery Assessment for Neurpsychological Status, RBCodTot=RBANS Coding sub-test score, RBDSTot=RBANS Digit Symbol sub-test score, RBFC1Tot=RBANS Figure Copy 1 subtest score, RBListTot=RBANS List recognition subtest score, RBLLTot=RBANS list learning subtest score, RBLRecTot=RBANS List Recall sub-test total score, RBLOTot=RBANS Line Orientation subtest score, RBPNTot: RBANS Picture Naming subtest score, RBSFTot=RBANS Semantic Fluency subtest score, RBSMTot=RBANS Story Memory subtest score, RBSRTot=RBANS Story Recall subtest score, RBTotal=Total RBANS Score,

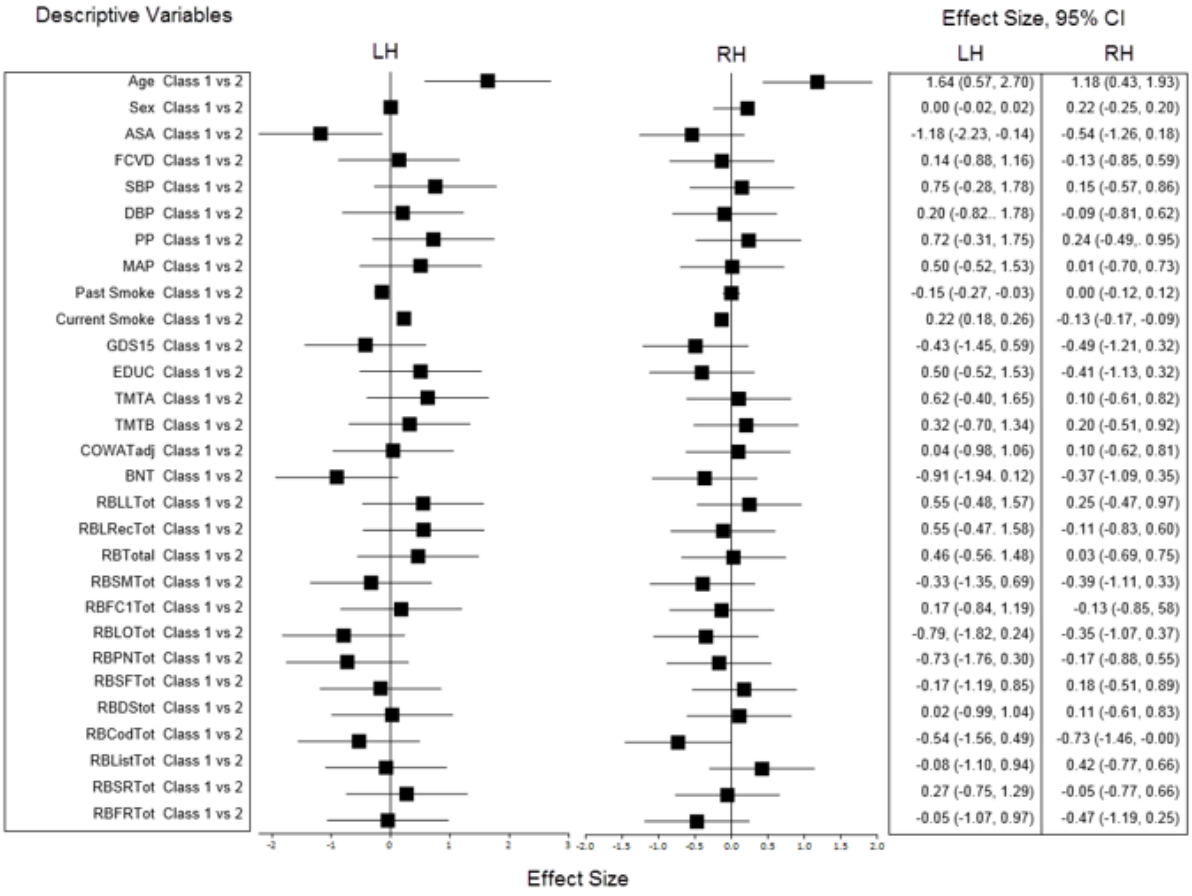
Supplementary Table 3:
T1+T2-FLAIR unadjusted LPA. Class Membership Predicting

Right					
Cortex Region	**Cluster**	**Size**	**Talx**	**Taly**	**Talz**
RH-Superior	-13.42	21313.80	64.3	-20.9	4.8
RH-Inferior temporal	-3.49	711.78	52.2	-43.6	-21.0
RH-Isthmus	-3.29	154.44	6.3	-50.2	17.7
RH-Isthmus	-2.79	26.51	4.4	-37.2	30.9
RH-Posterior	-2.46	11.25	4.9	-30.6	33.3
RH: 103088 of 163842 Vertices survived FDR at 0.05, identifying 10 clusters. Largest five clusters are listed.					
Abbreviations/Definitions: * Crosshair reference, Cluster: Maximum –log10(p- value), RH=Right Hemisphere, Size: Surface area of cluster, Talx y z: Talairaich (MNI305)					
Left					
Cortex Region	**Cluster**	**Size**	**Talx**	**Taly**	**Talz**
LH-Precuneus*	-14.35	58035.57	-4.6	-64.9	28.0
LH: 111331 of 163842 Vertices survived FDR at 0.05, identifying 1					
Abbreviations/Definitions: Cluster: Maximum –log10(p-value), LH=Left Hemisphere, Size: Surface area of cluster, Talx y z: Talairach (MNI305) coordinates					

Supplementary Table 4:
T1+T2-FLAIR Scanner adjusted LPA. Class membership

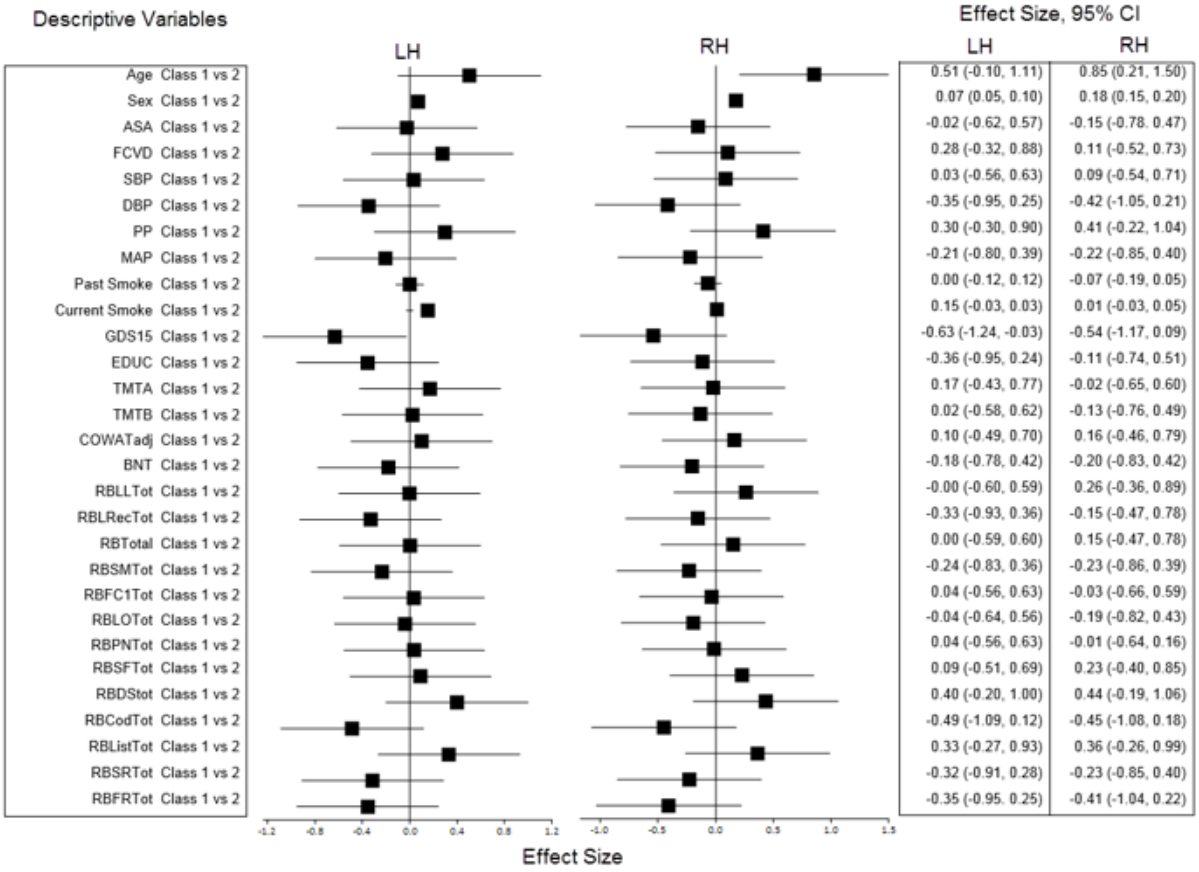
Right					
Cortex Region	**Cluste**	**Size**	**Talx**	**Taly**	**Talz**
RH-Superior Parietal	-7.57	245.73	27.6	-47.3	63.2
RH-Supramarginal	-5.94	182.54	53.2	-31.2	42.1
RH-Superior Temporal	-5.06	164.90	64.6	-19.9	4.5
RH-Postcentral	-4.41	212.81	35.2	-28.2	50.8
RH-Precentral	-4.36	96.85	39.6	2.1	30.1
RH: 2065 of 163842 Vertices survived FDR at 0.05, identifying 9 clusters. The top 5 are listed.					
Left					
Cortex Region	**Cluste**	**Size**	**Talx**	**Taly**	**Talz**
LH Bankssts	-7.53	153.99	-46.6	-48.7	7.3
LH-Insula	-6.81	393.12	-30.7	9.5	9.4
LH-Lateral occipital	-6.25	766.74	-35.6	-87.6	-5.0
LH-Fusiform	-6.00	634.36	-41.4	-60.7	-19.5
LH-Precentral	-5.77	329.94	-45.4	4.7	7.2
LH: 13576 of 163842 Vertices survived FDR at 0.05, identifying 39 clusters. The top five are listed.					
Abbreviations/Definitions: BankSSTS=Bank of the superior temporal sulcus, Cluster: Maximum –log10(p-value), LH=Left Hemisphere, Size: Surface area of cluster, Talx y z: Talairach					

Supplementary Figure 3, T1/T1+T2-FLAIR Unadjusted 2-Class LPA model displays the calculated standardized effect sizes for the left and right hemisphere, 2-class models. Error bars and 95% confidence intervals are included. These show descriptive



Abbreviations: ASA=American Society of Anesthesiologists classification, BNT=Boston Naming Test, CI=Confidence Intervals, COWATadj=Controlled Oral Word Association Test adjusted for age and education, DBP=Diastolic Blood Pressure, EDUC=Years of education, FCVD=Framingham Cardiovascular Disease Risk Score, GDS15=Geriatric Depression score, LPA=Latent Profile Analysis, MAP=Mean Arterial Pressure, PP=Pulse Pressure, RBANS=Repeatable Batttery Assessment for Neurpsychological Status, RBCodTot=RBANS Coding sub-test score, RBDSTot=RBANS Digit Symbol sub-test score, RBFC1Tot=RBANS Figure Copy 1 subtest score, RBListTot=RBANS List recognition subtest score, RBLLTot=RBANS list learning subtest score, RBLRecTot=RBANS List Recall sub-test total score, RBLOTot=RBANS Line Orientation subtest score, RBPNTot: RBANS Picture Naming subtest score, RBSFTot=RBANS Semantic Fluency subtest score, RBSMTot=RBANS Story Memory subtest score, RBSRTot=RBANS Story Recall subtest score, RBTotal=Total RBANS Score,

Supplementary Figure 4, T1/T1+T2-FLAIR Scanner adj 2-Class LPA model displays the calculated standardized effect sizes for the left and right hemisphere, 2-class models. Error bars and 95% confidence intervals are included. These show descriptive data trends between the 2-class models.



Abbreviations: ASA=American Society of Anesthesiologists classification, BNT=Boston Naming Test, CI=Confidence Intervals, COWATadj=Controlled Oral Word Association Test adjusted for age and education, DBP=Diastolic Blood Pressure, EDUC=Years of education, FCVD=Framingham Cardiovascular Disease Risk Score, GDS15=Geriatric Depression score, LPA=Latent Profile Analysis, MAP=Mean Arterial Pressure, PP=Pulse Pressure, RBANS=Repeatable Batttery Assessment for Neurpsychological Status, RBCodTot=RBANS Coding sub-test score, RBDSTot=RBANS Digit Symbol sub-test score, RBFC1Tot=RBANS Figure Copy 1 subtest score, RBListTot=RBANS List recognition subtest score, RBLLTot=RBANS list learning subtest score, RBLRecTot=RBANS List Recall sub-test total score, RBLOTot=RBANS Line Orientation subtest score, RBPNTot: RBANS Picture Naming subtest score, RBSFTot=RBANS Semantic Fluency subtest score, RBSMTot=RBANS Story Memory subtest score, RBSRTot=RBANS Story Recall subtest score, RBTotal=Total RBANS Score,

Discussion

Summary of findings

Delirium, an acute brain failure, is a crucial public health concern as it is significantly associated with increased mortality and morbidity (Hshieh *et al.,* 2018; S. K. Inouye *et al.,* 2014). Despite these poor outcomes, delirium remains bereft of accurate markers of vulnerability that are able to sufficiently predict delirium incidence and severity. The present dissertation investigated novel preoperative vulnerabilities to postoperative delirium incidence, severity, and duration in older adults, sixty-five years of age and older, undergoing a non-cardiac surgery. The outcomes of the present dissertation are the following: 1) a systematic review of delirium prediction models in older adults (≥60yo) medical and surgical inpatients; 2) development of two parsimonious delirium prediction models using the National Surgical Quality Improvement Program, serious complications risk score, and a measure of preoperative executive function, Trail Making Test B, for postoperative delirium incidence and severity; 3) comparison of statistical outcomes of two anatomical MRI scans, T1-weighted and T2-FLAIR; and 4) identification of a preliminary Latent Profile Analysis model, the Cortical Atrophy Class, that will serve as a foundation for future studies where I will validate the class using a larger sample size.

Several gaps in the literature were addressed by this present dissertation. The National Surgical Quality Improvement Program (Bilimoria *et al.,* 2013; Shiloach *et al.,* 2010), risk score for serious complications (NSQIP-SC), was identified as a significant preoperative predictor of postoperative delirium incidence, severity, and duration. Preoperative executive function measured with Trail Making Test B improved the ability of NSQIP-SC to predict postoperative delirium incidence and postoperative delirium severity; two parsimonious prediction models

were developed as a result. This is the first prediction model to use the preoperative NSQIP-SC score and TMTB to predict postoperative delirium incidence and severity. NSQIP-SC is valuable to delirium prediction because it incorporates several baseline risk factors including age, sex, functional status and multiple vascular comorbidities with the planned surgery, which provides information on both the premorbid risk factors and the precipitating event, prior to its occurrence. NSQIP-SC does not contain a cognitive variable and the addition of TMTB, an executive function measure, improved its predictive ability. Age, gender, blood pressure measures, tobacco pack years, vascular surgery, depression, ASA classification, and the Framingham Cardiovascular Risk Score were not identified as important candidate predictors by LASSO regression. Several of these variables are included in the NSQIP score, indicating that they are stronger predictors when combined with other factors than when they are used independently. Model building procedures and regression diagnostics are included in appendix 1 and 2.

Both models have clinical utility and could be integrated into the electronic health record (EHR) to provide an immediate risk score of delirium incidence and severity, *i.e.,* anticipated symptom burden. While the NSQIP surgical risk calculator does require 20 variables including the planned surgery, all 20 variables are readily accessible in the EHR as well as routinely collected during a preoperative clinical visit and can be supplemented by patient report. The executive function test, TMTB, takes less than 5 minutes to complete and is easily interpretable; longer time to completion indicates worse performance. An iPad application is built for TMTB further increasing the practicality (NeuRA Trail making test from Neuroscience Research Australia $4.99). Ideally, NSQIP-SC and TMTB scores would be obtained at the preoperative clinical visit, scores inputted into the EHR, and the anticipated risk of delirium communicated to the

perioperative interprofessional clinical care team. Implications for the perioperative interprofessional clinical care team include the following: 1) avoidance of deliriogenic medications such as benzodiazepines and anticholinergic medication (Hshieh *et al.,* 2018); 2) preoperative education of delirium including expected symptoms and prevention to the patient and the family (Eghbali-Babadi, Shokrollahi, & Mehrabi, 2017; J. Lee, Jung, Noh, Yoo, & Hong, 2013); 3) postoperative initiation of multicomponent non-pharmacologic delirium prevention bundles which include early mobility, sleep hygiene, cognitive stimulation, sensory stimulation, and maintaining adequate hydration and nutrition (Siddiqi *et al.,* 2016). Nursing is at the forefront for postoperative delirium prevention as they are involved in patient care throughout the perioperative spectrum and can use the delirium severity prediction model to anticipate symptom burden, implementing strategies to minimize and manage delirium symptoms. Past literature has not examined the prediction or prevention of delirium severity in terms of symptom management. The present dissertation is the first to develop a postoperative delirium severity prediction model. My future research will focus on the following: 1) partner with NSQIP to build an application embedded within the EHR to automatically calculate the NSQIP score; 2) broad external validation in a multisite perioperative cohort study; 3) implementation of prediction models in the perioperative clinical setting; and 4) customization of current delirium prevention measures to address and manage the anticipated delirium symptom burden.

As hypothesized, the preoperative NSQIP serious complications risk score outperformed the NSQIP risk of death score. Postoperative delirium has been associated with an increased risk of mortality (Ha *et al.,* 2018; Mosk *et al.,* 2017; K. Radinovic *et al.,* 2015). However, a recent literature review reported that this association should be considered with caution due to their finding that as study bias decreased, so did the association between postoperative delirium and

subsequent mortality (Hamilton, Wheeler, Di Michele, Lalu, & McIsaac, 2017). Past studies do not evaluate the spectrum of mortality risk. It is not known whether the individual that died following postoperative delirium was at a high-risk for mortality before delirium occurred. The present dissertation identified that the NSQIP risk model. Future perioperative delirium research should investigate this relationship further in prospective longitudinal cohorts.

A prediction model for postoperative delirium duration was investigated during the present dissertation. This model is not included in the second manuscript due to word count limitations and clinical utility. NSQIP-SC was identified as the single predictor for the postoperative delirium duration prediction model and details of the model building procedure are included in appendix 3.

The third manuscript of this dissertation aimed to compare anatomical MRI sequencing and analytic differences in the measurement of cortical thickness. The resulting methods paper focused on evaluating the statistical differences between the anatomical T1 and T1+T2-FLAIR MRI images and their impact on cortical thickness measures in terms of statistical findings. This addressed a gap in the literature surrounding the use of multimodal imaging (T1+T2-FLAIR) and how its application effects statistical results, *i.e.,* the replication of results from T1 derived cortical thickness measures. Prior research has evaluated segmentation improvement, age correlations, and the overall quality of T1 versus multimodal imaging (Lindig *et al.,* 2018; Viviani, Pracht, *et al.,* 2017). The present dissertation builds on previous work by identifying that multimodal imaging did not replicate the results of T1 derived cortical thickness measures and were found to be significantly thicker when compared to T1 only. This is likely due to T1-FLAIR images' longer pulse sequence leading to improved differentiation between anatomical borders and increased signal intensity from gray and white matter (Salat *et al.,* 2009; Viviani,

Pracht, *et al.,* 2017; Viviani, Stocker, & Stingl, 2017). Future research should incorporate multimodal imaging into their MRI scan sequence to increase precision in studies using cortical thickness measurements. The present dissertation employed a novel statistical technique, latent profile analysis, to cortical thickness measurements to build pilot data for future analysis. Latent profile analysis is considered a person-oriented approach, meaning that the whole individual is considered as the unit of analysis (Bergman *et al.,* 2002). This allows for the investigation of patterns, or sub- groups, of individual characteristics. This is in contrast to the traditional variable-oriented research that investigates associations between a set of variables (L. M. Collins & Lanza, 2010). This type of analysis has not been widely employed in cortical thickness studies (Malpas, 2016). The benefit of a person-oriented approach is it embodies a holistic perspective of the individual and accepts that the individual lives within a complex, dynamic, and adaptive system, meaning that they are in constant interaction with their environment (Bergman et al., 2002). It is through the combination of both approaches, variable and person-oriented, that we advance understanding and improve clinical outcomes. A prior study by Racine *et al*., (2017) found regions of the Alzheimer's disease (AD) signature to correlate with postoperative delirium severity. This is an important indicator of vulnerability in the brain to delirium, yet it does not describe the risk factors related to, or vulnerabilities associated with, those regions. Whether or not delirium ultimately manifests is based on the presence of risk factors and individual vulnerability. Therefore, delirium may be similar to AD in that amyloid burden within the brain is not a consistent indicator for the risk of future cognitive decline and manifestation of AD (Lyons *et al.,* 2018). Instead, cognitive decline may occur when sufficient conditions are met, *i.e.,* risk factors coupled with vulnerability. The recent findings by Rabin *et al.*, (2018) exemplify this potential relationship as faster cognitive decline was associated with a higher amyloid

burden and a higher vascular burden, measured by the Framingham Cardiovascular risk score. Further, the interaction between these two variables was significant indicating a combined effect (Rabin *et al.,* 2018). Delirium may result from a similar interaction between pre-existing risk (*i.e.,* amyloid burden) and a vulnerability (*i.e.,* vascular burden), however, past research has not examined how these risk factors and vulnerabilities are represented in the brain, specifically in cortical thickness measurements. The present dissertation identified a sub-group of cortical atrophy in a cross-sectional study of older adults, scheduled to undergo surgery and evaluated the clinical characteristics associated with each sub-group. No distinct conclusions can be drawn from this analysis as no significant findings survived false discovery rate correction.

Nonetheless, the calculated standard effect sizes will be used to power a future perioperative cohort study. Further, this preliminary analysis informs future study design and variable collection. In the current dissertation study, standard demographic and cognitive variables were used and observed data trends are well-supported by previous literature, *i.e.*, the cortical atrophy sub-group tends to have an older mean age. Future research will consider the use of behavioral, emotional, and social variables such as resilience, self-reported belief of health, and emotional cognitive scales. These have not been examined in delirium research to-date and may yield important insight into individual variations in delirium development. The overall aim of the future studies involves the development of a biomarker of vulnerability, using patterns of cortical atrophy. It is not feasible for each individual to undergo an MRI prior to surgery. However, my goal is to establish a cortical atrophy pattern that signifies vulnerability, *i.e.,* higher risk of delirium prior to surgery, and then incorporate the clinical characteristics associated with that cortical atrophy pattern into a delirium prediction model.

The systematic review of delirium prediction models completed for the present dissertation

revealed that current delirium prediction models focus on individuals with dementia- level cognitive impairment. Overlapping data capture may result as these also employ increasing age and impaired functional status as predictors. These models hold value for individuals with those factors as it is well established that dementia significantly increases the risk for delirium. However, these models are largely missing the individuals that still develop delirium, and have more subtle changes in executive function. For example, Jones *et al.,* (2016) demonstrated that with every 1-point decrease in the preoperative General Cognitive Performance score, the risk for postoperative delirium increased (Jones *et al.,* 2016). Further, a recent study found that high-functioning individuals experienced significantly higher mortality rates following delirium than those that had low function (Dani *et al.,* 2017). My future research will focus on elucidating risk factors in these high-functioning individuals, so prevention efforts can benefit the spectrum of the patient population.

Implications of New Science

In summary, the present dissertation addressed several gaps in the literature and provides broad directions for future research. These are organized by manuscript.

1. Future studies aimed at improving current delirium prediction in non-demented older adults should incorporate advanced statistical techniques to facilitate the identification of unknown or unused predictive variables that are sensitive to, and represent, vulnerability. Further, these studies should follow the Transparent Reporting of multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines (G. S. Collins, Reitsma, Altman, & Moons, 2015) developed to guide studies on proper statistical application and reporting. As identified by the first manuscript of this dissertation, identified delirium prediction models were limited by statistical methods and insufficient statistical reporting making model evaluation and

comparison difficult. This will aid in increasing the rigor and generalizability of delirium prediction models.

2. The developed delirium prediction model for postoperative delirium incidence and severity has several potential implications for researchers and clinicians. For future research studies, confirmation of predictive ability is needed in a large, multicenter perioperative cohort study. Further, engagement of the ACS NSQIP association is needed to initiate conversations surrounding the creation of an integrated EHR tool to automatically pull needed information. This model is preliminary so caution must be taken with interpretation, however, the findings could inform policy work focused on improving surgical outcomes. Delirium is a common surgical complication associated with a significant burden of morbidity. Improving our ability to identify high-risk, or vulnerable individuals, prior to surgery with the NSQIP risk scores has potential to decrease delirium incidence and severity. For nurses and the healthcare team, the NSQIP risk score for serious complications could be incorporated into daily clinic and inpatient care to inform perioperative prevention efforts. Quality improvement projects could use this information to generate evidence for best practices involving identifying delirium risk and tailoring prevention measures.

3. Future research studies should consider obtaining multimodal imaging using T1+T2-FLAIR for cortical thickness measurement as multimodal imaging resulted in a more accurate measurement of gray matter. This is an important finding as it addresses a limitation present in prior cortical thickness studies, the use of T1-only may have underestimated the amount of gray matter present in the brain.

Strengths and Limitations

The present dissertation investigated the novel application of a preoperative surgical risk

score and executive function to predict postoperative delirium incidence and severity. Further, the present dissertation employed a novel statistical method and person-oriented approach to analyze cortical thickness data. Nonetheless, the present dissertation is limited by sample size. Statistical methods such as Least Absolute Shrinkage and Selection Operator (LASSO) and Best Subsets Regression along with Bootstrapping were employed to counter the effects of the small sample size. Further, the study population is largely homogenous with no minority groups represented.

Next Steps

My overriding long-term goal of my developing program of research is delirium prevention in the older adult hospital population. I believe that accurate delirium prediction will facilitate the identification of high-risk individuals prior to delirium development and allow for the implementation of delirium prevention measures. My scientific perspective and process has been greatly informed by my education and experiences as a nurse and it has evolved throughout the formation, analysis and writing of the present dissertation. As a result, I view delirium through a holistic, interactional paradigm that lends towards the use of a holistic model that strives to understand the functioning of an individual, or a system, from the integration of biological, behavioral, and mental characteristics (Bergman *et al.,* 2002). My next steps are the following: 1) design a multi-site perioperative study to externally validate the NSQIP-SC and TMTB prediction model for delirium incidence and severity; and 2) conduct a retrospective analysis using a publicly available neuroimaging database, employing a person-oriented approach to identify sub-groups of cortical atrophy that will inform future prospective perioperative studies on vulnerabilities prior to surgery and postoperative delirium.

References

Abelha, F. J., Luís, C., Veiga, D., Parente, D., Fernandes, V., Santos, P., . . . Santos, C. (2013). Outcome and quality of life in patients with postoperative delirium during an ICU stay following major surgery. *Critical Care (London, England), 17*(5), R257-R257. doi:10.1186/cc13084

Adams, S. T., & Leveson, S. H. (2012). Clinical prediction rules. *Bmj, 344*.

American Psychiatric Association, A. (2013). Neurocognitive Disorders : Diagnostic and Statistical Manual of Mental Disorders-5.

Arzideh, F., Wosniok, W., & Haeckel, R. (2011). Indirect reference intervals of plasma and serum thyrotropin (TSH) concentrations from intra-laboratory data bases from several German and Italian medical centres. *Clin Chem Lab Med, 49*(4), 659-664. doi:10.1515/cclm.2011.114

Bakkour, A., Morris, J. C., & Dickerson, B. C. (2009). The cortical signature of prodromal AD: regional thinning predicts mild AD dementia. *Neurology, 72*(12), 1048-1055. doi:10.1212/01.wnl.0000340981.97664.2f

Bergman, L. R., Magnusson, D., & El Khouri, B. M. (2002). *Studying individual Development in An interindividual Context A Person-oriented Approach*. In *Paths Through Life Series v Vol 4* (pp. 1 online resource (211 p).

Bilimoria, K. Y., Liu, Y., Paruch, J. L., Zhou, L., Kmiecik, T. E., Ko, C. Y., & Cohen, M. E. Development and Evaluation of the Universal ACS NSQIP Surgical Risk Calculator: A Decision Aid and Informed Consent Tool for Patients and Surgeons. *Journal of the American College of Surgeons, 217*(5), 833-842.e833. doi:10.1016/j.jamcollsurg.2013.07.385

Bilimoria, K. Y., Liu, Y., Paruch, J. L., Zhou, L., Kmiecik, T. E., Ko, C. Y., & Cohen, M. E. (2013). Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *Journal of the American College of Surgeons, 217*(5), 833-842.e831-833. doi:10.1016/j.jamcollsurg.2013.07.385

Bohnen, J. D., Mavros, M. N., Ramly, E. P., Chang, Y., Yeh, D. D., Lee, J., . . . Kaafarani, H. M. A. (2017). Intraoperative Adverse Events in Abdominal Surgery: What Happens in the Operating Room Does Not Stay in the Operating Room. *Ann Surg, 265*(6), 1119-1125. doi:10.1097/sla.0000000000001906

Brandt, J., Aretouli, E., Neijstrom, E., Samek, J., Manning, K., Albert, M. S., & Bandeen-Roche, K. (2009). Selectivity of executive function deficits in mild cognitive impairment. *Neuropsychology, 23*(5), 607-618. doi:10.1037/a0015851

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Bmj, 350*, g7594. doi:10.1136/bmj.g7594

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis with applications in the social, behavioral, and health sciences*. In *Wiley Series in Probability and Statistics v 718* (pp. 1 online resource (331 p).

Dani, M., Owen, L. H., Jackson, T. A., Rockwood, K., Sampson, E. L., & Davis, D. (2017). Delirium, frailty and mortality: interactions in a prospective study of hospitalized older people. *J Gerontol A Biol Sci Med Sci*. doi:10.1093/gerona/glx214

Davis, D. H. J., Skelly, D. T., Murray, C., Hennessy, E., Bowen, J., Norton, S., . . . Cunningham, C. (2015). Worsening cognitive impairment and neurodegenerative pathology

progressively increase risk for delirium. *Am J Geriatr Psychiatry, 23*(4), 403-415. doi:10.1016/j.jagp.2014.08.005

Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., . . . Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage, 31*(3), 968-980. doi:10.1016/j.neuroimage.2006.01.021

Diamond, A. (2013). Executive functions. *Annu Rev Psychol, 64*, 135-168. doi:10.1146/annurev-psych-113011-143750

Drews, T., Franck, M., Radtke, F. M., Weiss, B., Krampe, H., Brockhaus, W. R., . . . Spies, C. D. (2015). Postoperative delirium is an independent risk factor for posttraumatic stress disorder in the elderly patient: a prospective observational study. *Eur J Anaesthesiol, 32*(3), 147-151. doi:10.1097/EJA.0000000000000107

Eghbali-Babadi, M., Shokrollahi, N., & Mehrabi, T. (2017). Effect of Family-Patient Communication on the Incidence of Delirium in Hospitalized Patients in Cardiovascular Surgery ICU. *Iran J Nurs Midwifery Res, 22*(4), 327-331. doi:10.4103/1735-9066.212985

Ely, E. W., Inouye, S. K., Bernard, G. R., Gordon, S., Francis, J., May, L., . . . Dittus, R. (2001). Caring for the critically ill patient. Delirium in mechanically ventilated patients: validity and reliability of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). *JAMA: Journal of the American Medical Association, 286*(21), 2703.

Ely, E. W., Margolin, R., Francis, J., May, L., Truman, B., Dittus, R., . . . Inouye, S. K. (2001). Evaluation of delirium in critically ill patients: validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). *Critical Care Medicine, 29*(7), 1370-

1379.

Fimognari, F. L., Pierantozzi, A., De Alfieri, W., Salani, B., Zuccaro, S. M., Arone, A., . . . Palleschi, L. (2016). The Severity of Acute Illness and Functional Trajectories in Hospitalized Older Medical Patients. *J Gerontol A Biol Sci Med Sci*. doi:10.1093/gerona/glw096

Fortin, J. P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., . . . Shinohara, R. T. (2018). Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage, 167*, 104-120. doi:10.1016/j.neuroimage.2017.11.024

Franco, J. G., Valencia, C., Bernal, C., Ocampo, M. V., Trzepacz, P. T., Pablo, J. d., & Mejia, M. A. (2010). Relationship between cognitive status at admission and incident delirium in older medical inpatients. *J Neuropsychiatry Clin Neurosci, 22*(3), 329-337. doi:10.1176/jnp.2010.22.3.329

Galyfos, G. C., Geropapas, G. E., Sianou, A., Sigala, F., & Filis, K. (2017). Risk factors for postoperative delirium in patients undergoing vascular surgery. *J Vasc Surg, 66*(3), 937-946. doi:10.1016/j.jvs.2017.03.439

Gleason, L. J., Schmitt, E. M., Kosar, C. M., Tabloski, P., Saczynski, J. S., Robinson, T., . Rogers, S. O., Jr. (2015). Effect of Delirium and Other Major Complications on Outcomes After Elective Surgery in Older Adults. *JAMA Surgery, 150*(12), 1134-1140 1137p. doi:10.1001/jamasurg.2015.2606

Gleason, L. J., Schmitt, E. M., Kosar, C. M., Tabloski, P., Saczynski, J. S., Robinson, T., . . Inouye, S. K. (2015). Effect of Delirium and Other Major Complications on OutcomesAfter Elective Surgery in Older Adults. *JAMA surgery, 150*(12), 1134-1140. doi:10.1001/jamasurg.2015.2606

Glisky, E. L. (2007). Changes in Cognitive Function in Human Aging. In D. R. Riddle (Ed.),

Brain Aging: Models, Methods, and Mechanisms. Boca Raton (FL): CRC Press.

Greene, N. H., Attix, D. K., Weldon, B. C., Smith, P. J., McDonagh, D. L., & Monk, T. G. (2009). Measures of executive function and depression identify patients at risk for postoperative delirium. *Anesthesiology, 110*(4), 788-795 788p.

Guillamondegui, O. D., Gunter, O. L., Hines, L., Martin, B. J., Gibson, W., Clarke, P. C., . . . Cofer, J. B. (2012). Using the National Surgical Quality Improvement Program and the Tennessee Surgical Quality Collaborative to improve surgical outcomes. *Journal of the American College of Surgeons, 214*(4), 709-714; discussion 714-716. doi:10.1016/j.jamcollsurg.2011.12.012

Gunther, M. L., Morandi, A., Krauskopf, E., Pandharipande, P., Girard, T. D., Jackson, J. C., . Ely, E. W. (2012). The association between brain volumes, delirium duration, and cognitive outcomes in intensive care unit survivors: the VISIONS cohort magnetic resonance imaging study*. *Critical Care Medicine, 40*(7), 2022-2032. doi:10.1097/CCM.0b013e318250acc0

Ha, A., Krasnow, R. E., Mossanen, M., Nagle, R., Hshieh, T. T., Rudolph, J. L., & Chang, S. L. (2018). A contemporary population-based analysis of the incidence, cost, and outcomes of postoperative delirium following major urologic cancer surgeries. *Urol Oncol*. doi:10.1016/j.urolonc.2018.04.012

Hamilton, G. M., Wheeler, K., Di Michele, J., Lalu, M. M., & McIsaac, D. I. (2017). A Systematic Review and Meta-analysis Examining the Impact of Incident Postoperative Delirium on Mortality. *Anesthesiology, 127*(1), 78-88. doi:10.1097/aln.0000000000001660

Helman, S. N., Brant, J. A., Moubayed, S. P., Newman, J. G., Cannady, S. B., & Chai, R. L.

(2017). Predictors of length of stay, reoperation, and readmission following total laryngectomy. *Laryngoscope, 127*(6), 1339-1344. doi:10.1002/lary.26454

Hessler, J. B., Brönner, M., Etgen, T., Gotzler, O., Förstl, H., Poppert, H., . . . Bickel, H. (2015). Smoking increases the risk of delirium for older inpatients: a prospective population-based study. *General Hospital Psychiatry, 37*(4), 360-364. doi:10.1016/j.genhosppsych.2015.03.009

Hosmer, D. W., Jovanovic, B., & Lemeshow, S. (1989). Best Subsets Logistic Regression. *Biometrics, 45*(4), 1265-1270. doi:10.2307/2531779

Hshieh, T. T., Inouye, S. K., & Oh, E. S. (2018). Delirium in the Elderly. *Psychiatr Clin North Am, 41*(1), 1-17. doi:10.1016/j.psc.2017.10.001

Hshieh, T. T., Yue, J., Oh, E., Puelle, M., Dowal, S., Travison, T., & Inouye, S. K. (2015). Effectiveness of Multicomponent Nonpharmacological Delirium Interventions: A Meta-analysis. In *JAMA Intern Med*.

Hsieh, S. J., Shum, M., Lee, A. N., Hasselmark, F., & Gong, M. N. (2013). Cigarette smoking as a risk factor for delirium in hospitalized and intensive care unit patients. A systematic review. *Annals Of The American Thoracic Society, 10*(5), 496-503. doi:10.1513/AnnalsATS.201301-001OC

Inouye, S. K., Marcantonio, E. R., Kosar, C. M., Tommet, D., Schmitt, E. M., Travison, T. G., . .. Jones, R. N. (2016). The short-term and long-term relationship between delirium and cognitive trajectory in older surgical patients. *Alzheimers Dement*. doi:10.1016/j.jalz.2016.03.005

Inouye, S. K., Viscoli, C. M., Horwitz, R. I., Hurst, L. D., & Tinetti, M. E. (1993). A predictive model for delirium in hospitalized elderly medical patients based on admission

characteristics. *Ann Intern Med, 119*(6), 474-481.

Inouye, S. K., Westendorp, R. G., & Saczynski, J. S. (2014). Delirium in elderly people. *Lancet, 383*(9920), 911-922. doi:10.1016/s0140-6736(13)60688-1

Jackson, J. C., Pandharipande, P. P., Girard, T. D., Brummel, N. E., Thompson, J. L., Hughes, C. G., . . . Ely, E. W. (2014). Depression, post-traumatic stress disorder, and functional disability in survivors of critical illness in the BRAIN-ICU study: a longitudinal cohort study. *Lancet Respir Med, 2*(5), 369-379. doi:10.1016/s2213-2600(14)70051-7

Jones, R. N., Marcantonio, E. R., Saczynski, J. S., Tommet, D., Gross, A. L., Travison, T. G., . . . Inouye, S. K. (2016). Preoperative Cognitive Performance Dominates Risk for Delirium Among Older Adults. *J Geriatr Psychiatry Neurol*. doi:10.1177/0891988716666380

Kalisvaart, K. J., Vreeswijk, R., de Jonghe, J. F., van der Ploeg, T., van Gool, W. A., & Eikelenboom, P. (2006). Risk factors and prediction of postoperative delirium in elderly hip-surgery patients: implementation and validation of a medical risk factor model. *J Am Geriatr Soc, 54*(5), 817-822. doi:10.1111/j.1532-5415.2006.00704.x

Kazmierski, J., Banys, A., Latek, J., Bourke, J., Jaszewski, R., Sobow, T., & Kloszewska, I. (2014). Mild cognitive impairment with associated inflammatory and cortisol alterations as independent risk factor for postoperative delirium. *Dementia and Geriatric Cognitive Disorders, 38*(1-2), 65-78. doi:10.1159/000357454

King, J. E. (2003). Running a Best-Subsets Logistic Regression: An Alternative to Stepwise Methods. *Educational and Psychological Measurement, 63*(3), 392-403. doi:10.1177/0013164403063003003

Koskderelioglu, A., Onder, O., Gucuyener, M., Altay, T., Kayali, C., & Gedizlioglu, M. (2017). Screening for postoperative delirium in patients with acute hip fracture: Assessment of

predictive factors. *Geriatrics & gerontology international, 17*(6), 919-924. doi:10.1111/ggi.12806

Large, M. C., Reichard, C., Williams, J. T. B., Chang, C., Prasad, S., Leung, Y., . . . Steinberg, G. D. (2013). Incidence, risk factors, and complications of postoperative delirium in elderly patients undergoing radical cystectomy. *Urology, 81*(1), 123-128. doi:10.1016/j.urology.2012.07.086

Lee, J., Jung, J., Noh, J. S., Yoo, S., & Hong, Y. S. (2013). Perioperative psycho-educational intervention can reduce postoperative delirium in patients after cardiac surgery: a pilot study. *Int J Psychiatry Med, 45*(2), 143-158. doi:10.2190/PM.45.2.d

Lee, K.-H., Ha, Y.-C., Lee, Y.-K., Kang, H., & Koo, K.-H. (2011). Frequency, risk factors, and prognosis of prolonged delirium in elderly patients after hip fracture surgery. *Clinical Orthopaedics And Related Research, 469*(9), 2612-2620. doi:10.1007/s11999-011-1806- 1

Leslie, D. L., & Inouye, S. K. (2011). The importance of delirium: economic and societal costs. *J Am Geriatr Soc, 59 Suppl 2*, S241-243. doi:10.1111/j.1532-5415.2011.03671.x

Leslie, D. L., Marcantonio, E. R., Zhang, Y., Leo-Summers, L., & Inouye, S. K. (2008). One-year health care costs associated with delirium in the elderly population. *Arch Intern Med, 168*(1), 27-32. doi:10.1001/archinternmed.2007.4

Lindig, T., Kotikalapudi, R., Schweikardt, D., Martin, P., Bender, F., Klose, U., . . . Bender, B. (2018). Evaluation of multimodal segmentation based on 3D T1-, T2- and FLAIR-weighted images - the difficulty of choosing. *Neuroimage, 170*, 210-221. doi:10.1016/j.neuroimage.2017.02.016

Lindroth, H., Bratzke, L., Purvis, S., Brown, R., Coburn, M., Mrkobrada, M., . . . Sanders, R. D.

(2018). Systematic review of prediction models for delirium in the older adult inpatient. *BMJ open, 8*(4), e019223. doi:10.1136/bmjopen-2017-019223

Lockhart, R., Taylor, J., Tibshirani, R. J., & Tibshirani, R. (2014). A SIGNIFICANCE TEST FOR THE LASSO. *Ann Stat, 42*(2), 413-468. doi:10.1214/13-aos1175

Lyons, C. E., Tudorascu, D., Snitz, B. E., Price, J., Aizenstein, H., Lopez, O., . . . Cohen, A. D. (2018). The Relationship of Current Cognitive Activity to Brain Amyloid Burden and Glucose Metabolism. *The American Journal of Geriatric Psychiatry*. doi:https://doi.org/10.1016/j.jagp.2018.03.018

MacPherson, S. E., Cox, S. R., Dickie, D. A., Karama, S., Starr, J. M., Evans, A. C., . . . Deary, I. J. (2017). Processing speed and the relationship between Trail Making Test-B performance, cortical thinning and white matter microstructure in older adults. *Cortex, 95,* 92-103. doi:10.1016/j.cortex.2017.07.021

Mallows, C. L. (1973). Some Comments on C_P. *Technometrics, 15*(4), 661-675. doi:10.2307/1267380

Malpas, C. B. (2016). Structural neuroimaging correlates of cognitive status in older adults: A person-oriented approach. *J Clin Neurosci, 30*, 77-82. doi:10.1016/j.jocn.2016.01.025 Mangusan, R. F., Hooper, V., Denslow, S. A., & Travis, L. (2015). Outcomes associated with postoperative delirium after cardiac surgery. *Am J Crit Care, 24*(2), 156-163. doi:10.4037/ajcc2015137

Mansouri, F. A., Tanaka, K., & Buckley, M. J. (2009). Conflict-induced behavioural adjustment: a clue to the executive functions of the prefrontal cortex. *Nat Rev Neurosci, 10*(2), 141-152. doi:http://www.nature.com/nrn/journal/v10/n2/suppinfo/nrn2538_S1.html

Mogal,H., Vermilion, S. A., Dodson, R., Hsu, F. C., Howerton, R., Shen, P., & Clark, C. J. (2017). Modified Frailty Index Predicts Morbidity and Mortality After Pancreaticoduodenectomy. *Ann Surg Oncol, 24*(6), 1714-172doi:10.1245/s10434-016-5715-0

Morandi, A., Rogers, B. P., Gunther, M. L., Merkle, K., Pandharipande, P., Girard, T. D., . . . Hopkins, R. O. (2012). The relationship between delirium duration, white matter integrity, and cognitive impairment in intensive care unit survivors as determined by diffusion tensor imaging: the VISIONS prospective cohort magnetic resonance imaging study*. *Critical Care Medicine, 40*(7), 2182-2189. doi:10.1097/CCM.0b013e318250acdc

Mosk, C. A., Mus, M., Vroemen, J. P., van der Ploeg, T., Vos, D. I., Elmans, L. H., & van der Laan, L. (2017). Dementia and delirium, the outcomes in elderly hip fracture patients. *Clin Interv Aging, 12*, 421-430. doi:10.2147/cia.s115945

Moskowitz, E. E., Overbey, D. M., Jones, T. S., Jones, E. L., Arcomano, T. R., Moore, J. T., & Robinson, T. N. (2017). Post-operative delirium is associated with increased 5-year mortality. *Am J Surg, 214*(6), 1036-1038. doi:10.1016/j.amjsurg.2017.08.034

Muthen, T. A. a. B. (2014). *Variable-Specific Entropy Contribution*. Retrieved from MPlus: https://www.statmodel.com/recentpapers.shtml

Nightingale, E. O., & Fischhoff, B. (2002). Adolescent risk and vulnerability: overview. *J Adolesc Health, 31*(1 Suppl), 3-9.

Nowrangi, M. A., Lyketsos, C., Rao, V., & Munro, C. A. (2014). Systematic review of neuroimaging correlates of executive functioning: converging evidence from different clinical populations. *J Neuropsychiatry Clin Neurosci, 26*(2), 114-125.

doi:10.1176/appi.neuropsych.12070176

O'Malley, G., Leonard, M., Meagher, D., & O'Keeffe, S. T. (2008). The delirium experience: a review. *Journal Of Psychosomatic Research, 65*(3), 223-228. doi:10.1016/j.jpsychores.2008.05.017

O'Sullivan, R., Inouye, S. K., & Meagher, D. (2014). Delirium and depression: inter-relationship and clinical overlap in elderly people. *Lancet Psychiatry, 1*(4), 303-311. doi:10.1016/s2215-0366(14)70281-0

Ogundimu, E. O., Altman, D. G., & Collins, G. S. (2016). Adequate sample size for developing prediction models is not simply related to events per variable. *J Clin Epidemiol, 76*, 175- 182. doi:10.1016/j.jclinepi.2016.02.031

Oh, E. S., Fong, T. G., Hshieh, T. T., & Inouye, S. K. (2017). Delirium in older persons: Advances in diagnosis and treatment. *Jama, 318*(12), 1161-1174. doi:10.1001/jama.2017.12067

Oldroyd, C., Scholz, A. F. M., Hinchliffe, R. J., McCarthy, K., Hewitt, J., & Quinn, T. J. (2017). A systematic review and meta-analysis of factors for delirium in vascular surgical patients. *J Vasc Surg, 66*(4), 1269-1279.e1269. doi:10.1016/j.jvs.2017.04.077

Pandharipande, P. P., Girard, T. D., Jackson, J. C., Morandi, A., Thompson, J. L., Pun, B. T., . . . the, B.-I. C. U. S. I. (2013). Long-Term Cognitive Impairment after Critical Illness. *The New England Journal Of Medicine, 369*(14), 1306-1316. doi:10.1056/NEJMoa1301372

Partridge, J. S. L., Martin, F. C., Harari, D., & Dhesi, J. K. (2013). The delirium experience: what is the effect on patients, relatives and staff and what can be done to modify this? *Int J Geriatr Psychiatry, 28*(8), 804-812. doi:10.1002/gps.3900

Pendlebury, S. T., Lovett, N., Smith, S. C., Cornish, E., Mehta, Z., & Rothwell, P. M. (2016).

Delirium risk stratification in consecutive unselected admissions to acute medicine: validation of externally derived risk scores. *Age Ageing, 45*(1), 60-65. doi:10.1093/ageing/afv177

Pendlebury, S. T., Lovett, N. G., Smith, S. C., Wharton, R., & Rothwell, P. M. (2016). Delirium risk stratification in consecutive unselected admissions to acute medicine: validation of a susceptibility score based on factors identified externally in pooled data for use at entry to the acute care pathway. Age Ageing. doi:10.1093/ageing/afw198

Raats, J. W., Steunenberg, S. L., Crolla, R. M. P. H., Wijsman, J. H. H., te Slaa, A., & van der Laan, L. (2015). Postoperative delirium in elderly after elective and acute colorectal surgery: A prospective cohort study. *International Journal Of Surgery (London, England), 18*, 216-219. doi:10.1016/j.ijsu.2015.04.080

Rabin, J. S., Schultz, A. P., Hedden, T., Viswanathan, A., Marshall, G. A., Kilpatrick, E., . . . Chhatwal, J. P. (2018). Interactive Associations of Vascular Risk and beta-Amyloid Burden With Cognitive Decline in Clinically Normal Elderly Individuals: Findings From the Harvard Aging Brain Study. *JAMA Neurol*. doi:10.1001/jamaneurol.2018.1123

Racine, A. M., Fong, T. G., Gou, Y., Travison, T. G., Tommet, D., Erickson, K., . . . Inouye, S. J. (2017). Clinical outcomes in older surgical patients with mild cognitive impairment. *Alzheimers Dement*. doi:10.1016/j.jalz.2017.10.010

Racine, A. M., Fong, T. G., Travison, T. G., Jones, R. N., Gou, Y., Vasunilashorn, S. M., . . . Dickerson, B. C. (2017). Alzheimer's-related cortical atrophy is associated with postoperative delirium severity in persons without dementia. *Neurobiology of aging, 59*, 55-63. doi:10.1016/j.neurobiolaging.2017.07.010

Radinovic, K., Markovic-Denic, L., Dubljanin-Raspopovic, E., Marinkovic, J., Milan, Z., & Bumbasirevic, V. (2015). Estimating the effect of incident delirium on short-term outcomes in aged hip fracture patients through propensity score analysis. *Geriatr Gerontol Int, 15*(7), 848-855. doi:10.1111/ggi.12358

Radinovic, K. S., Markovic-Denic, L., Dubljanin-Raspopovic, E., Marinkovic, J., Jovanovic, L. B., & Bumbasirevic, V. (2014). Effect of the overlap syndrome of depressive symptoms and delirium on outcomes in elderly adults with hip fracture: a prospective cohort study. J Am Geriatr Soc, 62(9), 1640-1648. doi:10.1111/jgs.12992

Randolph, C., Tierney, M. C., Mohr, E., & Chase, T. N. (1998). The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): preliminary clinical validity. *J Clin Exp Neuropsychol, 20*(3), 310-319. doi:10.1076/jcen.20.3.310.823

Reilly, B. M., & Evans, A. T. (2006). Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med, 144*(3), 201-209.

Robinson, T. N., Raeburn, C. D., Tran, Z. V., Angles, E. M., Brenner, L. A., & Moss, M. (2009). Postoperative delirium in the elderly: risk factors and outcomes. *Ann Surg, 249*(1), 173-178. doi:10.1097/SLA.0b013e31818e4776

Rudolph, J. L., Jones, R. N., Grande, L. J., Milberg, W. P., King, E. G., Lipsitz, L. A., . . . Marcantonio, E. R. (2006). Impaired executive function is associated with delirium after coronary artery bypass graft surgery. *J Am Geriatr Soc, 54*(6), 937-941. doi:10.1111/j.1532-5415.2006.00735.x

Rudolph, J. L., Jones, R. N., Levkoff, S. E., Rockett, C., Inouye, S. K., Sellke, F. W., . . . Marcantonio, E. R. (2009). Derivation and validation of a preoperative prediction rule for delirium after cardiac surgery. *Circulation, 119*(2), 229-236 228p.

doi:10.1161/CIRCULATIONAHA.108.795260

Rudolph, J. L., Jones, R. N., Rasmussen, L. S., Silverstein, J. H., Inouye, S. K., & Marcantonio, E. R. (2007). Independent vascular and cognitive risk factors for postoperative delirium. *The American Journal Of Medicine, 120*(9), 807-813.

Salat, D. H., Lee, S. Y., van der Kouwe, A. J., Greve, D. N., Fischl, B., & Rosas, H. D. (2009). Age-associated alterations in cortical gray and white matter signal intensity and gray to white matter contrast. Neuroimage, 48(1), 21-28. doi:10.1016/j.neuroimage.2009.06.074

Sanders, R. D. (2011). Hypothesis for the pathophysiology of delirium: role of baseline brain network connectivity and changes in inhibitory tone. *Medical Hypotheses, 77*(1), 140-143. doi:10.1016/j.mehy.2011.03.048

Sanders, R. D., Pandharipande, P. P., Davidson, A. J., Ma, D., & Maze, M. (2011). Anticipating and managing postoperative delirium and cognitive decline in adults. *Bmj, 343*, d4331.

Scholz, A. F., Oldroyd, C., McCarthy, K., Quinn, T. J., & Hewitt, J. (2016). Systematic review and meta-analysis of risk factors for postoperative delirium among older patients undergoing gastrointestinal surgery. *Br J Surg, 103*(2), e21-28. doi:10.1002/bjs.10062

Sheline, Y. I., Barch, D. M., Garcia, K., Gersing, K., Pieper, C., Welsh-Bohmer, K., . . . Doraiswamy, P. M. (2006). Cognitive function in late life depression: relationships to depression severity, cerebrovascular risk factors and processing speed. *Biol Psychiatry, 60*(1), 58-65. doi:10.1016/j.biopsych.2005.09.019

Shiloach, M., Frencher, S. K., Jr., Steeger, J. E., Rowell, K. S., Bartzokis, K., Tomeh, M. G., . . . Hall, B. L. (2010). Toward robust information: data quality and inter-rater reliability in the American College of Surgeons National Surgical Quality Improvement Program.

Journal of the American College of Surgeons, 210(1), 6-16. doi:10.1016/j.jamcollsurg.2009.09.031

Siddiqi, N., Harrison, J. K., Clegg, A., Teale, E. A., Young, J., Taylor, J., & Simpkins, S. A. (2016). Interventions for preventing delirium in hospitalised non-ICU patients. *The Cochrane database of systematic reviews, 3*, CD005563. doi:10.1002/14651858.CD005563.pub3

Smith, P. J., Attix, D. K., Weldon, B. C., Greene, N. H., & Monk, T. G. (2009). Executive function and depression as independent risk factors for postoperative delirium. *Anesthesiology, 110*(4), 781-787 787p.

Smith, P. J., Attix, D. K., Weldon, B. C., & Monk, T. G. (2016). Depressive Symptoms and Risk of Postoperative Delirium. *Am J Geriatr Psychiatry, 24*(3), 232-238. doi:10.1016/j.jagp.2015.12.004

Steffens, D. C., & Potter, G. G. (2008). Geriatric depression and cognitive impairment. *Psychol Med, 38*(2), 163-175. doi:10.1017/s003329170700102x

Steyerberg, E. W. (2009). *Clinical prediction models : a practical approach to development, validation, and updating*: New York : Springer, [2009] ©2009.

Steyerberg, E. W., & Vergouwe, Y. (2014). Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J, 35*(29), 1925-1931. doi:10.1093/eurheartj/ehu207

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., . . . Kattan, M. W. (2010). Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology, 21*(1), 128-138. doi:10.1097/EDE.0b013e3181c30fb2

Strauss, E. (2006). *A compendium of neuropsychological tests : administration, norms, and commentary*: 3rd ed. Oxford ; New York : Oxford University Press, 2006.

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Stat Med, 16*(4), 385-395.

Tibshirani, R., Bien, J., Friedman, J., Hastie, T., Simon, N., Taylor, J., & Tibshirani, R. J. (2012). Strong rules for discarding predictors in lasso-type problems. *J R Stat Soc Series B Stat Methodol, 74*(2), 245-266. doi:10.1111/j.1467-9868.2011.01004.x

Trzepacz, P. T., Mittal, D., Torres, R., Kanary, K., Norton, J., & Jimerson, N. (2001). Validation of the Delirium Rating Scale-revised-98: comparison with the delirium rating scale and the cognitive test for delirium. *J Neuropsychiatry Clin Neurosci, 13*(2), 229-242.

Vasunilashorn, S. M., Marcantonio, E. R., Gou, Y., Pisani, M. A., Travison, T. G., Schmitt, E. M., . . . Inouye, S. K. (2016). Quantifying the Severity of a Delirium Episode Throughout Hospitalization: the Combined Importance of Intensity and Duration. *J Gen Intern Med, 31*(10), 1164-1171. doi:10.1007/s11606-016-3671-9

Veiga, D., LuÃ-s, C., Parente, D., & Abelha, F. (2013). Outcome after hepatectomy-delirium as an independent predictor for mortality. *BMC Anesthesiology, 13*(1), 4-12 19p. doi:10.1186/1471-2253-13-4

Veliz-Reissmuller, G., Aguero Torres, H., van der Linden, J., Lindblom, D., & Eriksdotter Jonhagen, M. (2007). Pre-operative mild cognitive dysfunction predicts risk for post-operative delirium after elective cardiac surgery. *Aging Clinical And Experimental Research, 19*(3), 172-177.

Vickers, A. J., & Cronin, A. M. (2010). Traditional statistical methods for evaluating prediction models are uninformative as to clinical value: towards a decision analytic framework.

Semin Oncol, 37(1), 31-38. doi:10.1053/j.seminoncol.2009.12.004

Viviani, R., Pracht, E. D., Brenner, D., Beschoner, P., Stingl, J. C., & Stöcker, T. (2017). Multimodal MEMPRAGE, FLAIR, and [Formula: see text] Segmentation to Resolve Dura and Vessels from Cortical Gray Matter. *Front Neurosci, 11*, 258. doi:10.3389/fnins.2017.00258

Viviani, R., Stocker, T., & Stingl, J. C. (2017). Multimodal FLAIR/MPRAGE segmentation of cerebral cortex and cortical myelin. *Neuroimage, 152*, 130-141. doi:10.1016/j.neuroimage.2017.02.054

Williams-Russo, P., Sharrock, N. E., Mattis, S., Szatrowski, T. P., & Charlson, M. E. (1995). Cognitive effects after epidural vs general anesthesia in older adults. A randomized trial. *Jama, 274*(1), 44-50.

Yevchak, A., Fick, D. M., Kolanowski, A. M., McDowell, J., Monroe, T., LeViere, A., & Mion, L. (2017). Implementing Nurse-Facilitated Person-Centered Care Approaches for Patients With Delirium Superimposed on Dementia in the Acute Care Setting. *J Gerontol Nurs, 43*(12), 21-28. doi:10.3928/00989134-20170623-01

Yevchak, A. M., Fick, D. M., McDowell, J., Monroe, T., May, K., Grove, L., . . . Inouye, S. K. (2014). Barriers and facilitators to implementing delirium rounds in a clinical trial across three diverse hospital settings. *Clin Nurs Res, 23*(2), 201-215. doi:10.1177/1054773813505321

Yoon, S., Shin, C., & Han, C. (2017). Depression and Cognitive Function in Mild Cognitive Impairment: A 1-Year Follow-Up Study. *J Geriatr Psychiatry Neurol, 30*(5), 280-288. doi:10.1177/0891988717723741

Yuan, P., & Raz, N. (2014). Prefrontal cortex and executive functions in healthy adults: a meta-

analysis of structural neuroimaging studies. *Neurosci Biobehav Rev, 42*, 180-192. doi:10.1016/j.neubiorev.2014.02.005

Zaubler, T. S., Murphy, K., Rizzuto, L., Santos, R., Skotzko, C., Giordano, J., . . . Inouye, S. K. (2013). Quality improvement and cost savings with multicomponent delirium interventions: replication of the Hospital Elder Life Program in a community hospital. *Psychosomatics, 54*(3), 219-226. doi:10.1016/j.psym.2013.01.010

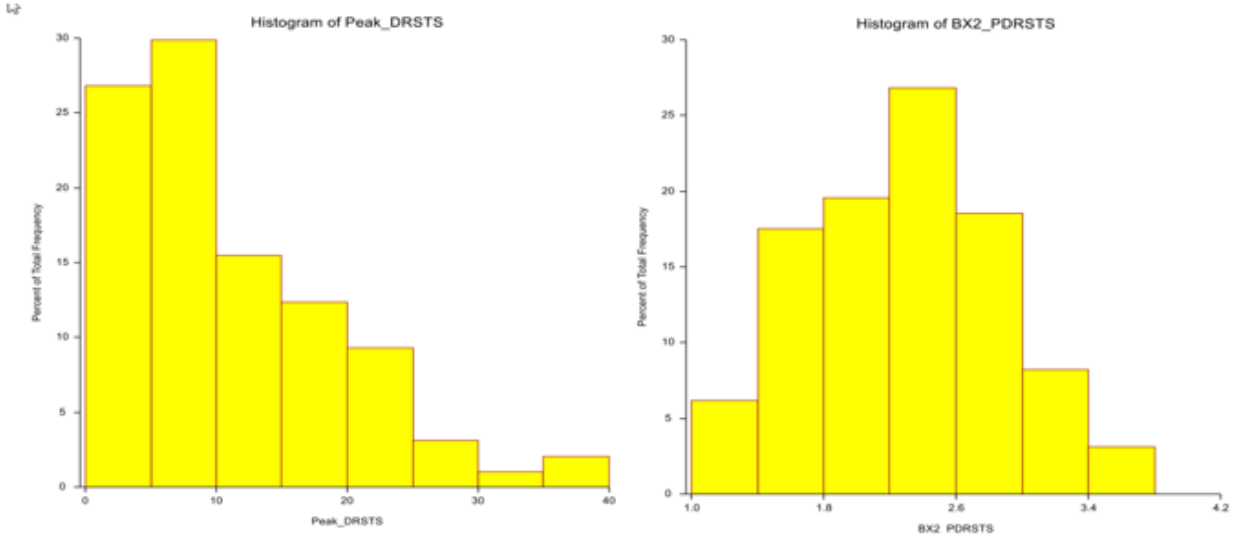
Zhou, M., Zhang, F., Zhao, L., Qian, J., & Dong, C. (2016). Entorhinal cortex: a good biomarker of mild cognitive impairment and mild Alzheimer's disease. *Rev Neurosci, 27*(2), 185-195. doi:10.1515/revneuro-2015-0019

Zhou, Q., Goryawala, M., Cabrerizo, M., Barker, W., Loewenstein, D., Duara, R., & Adjouadi, L. (2014). Multivariate analysis of structural MRI and PET (FDG and 18F-AV-45) for Alzheimer's disease and its prodromal stages. *Conf Proc IEEE Eng Med Biol Soc, 2014*, 1051-1054. doi:10.1109/embc.2014.6943774

Zhou, Y., & Lui, Y. W. (2013). Small-World Properties in Mild Cognitive Impairment and Early Alzheimer's Disease: A Cortical Thickness MRI Study. *ISRN Geriatr, 2013*. doi:10.1155/2013/542080

Appendix 1: Regression Analysis & Diagnostics of Peak DRS-98-R Score N=97

NV26.1 was dropped from the analysis due to experienced alcohol-induced delirium. This is a uniquely different phenomenon that postoperative delirium.



Peak DRS Total Score was skewed to the left, so was transformed using the Box-Cox Method with the optimal lambda value of 0.3466, calculated using NCSS's transformation report. The histograms from before (left) and after (right) are displayed as figure 1. This is included as supplementary material to the 2nd manuscript.

- Box-Cox equation: (Peak_DRSTS+1)^0.3466

1. Best Subsets regression ran using vselect command in Stata v15.1.
    a. 1st Analysis is with raw TMTA & B scores
        i. vselect BX2_PDRSTS nsqip_sc framingrisk Pre_SBP Pre_PP PreopTMTA PreopTMTB Sex nsqip_d ASA Pre_DBP Pre_MAP Pre_PP TobPackYears GDS15 centered_age, best
            1. Model #1-2 has the lowest BIC, 111.7359
                a. Nsqip_sc and preoptmtb are predictors in model
            2. Model #1-6 has the lowest AIC (100.1722), AICC (102.2641)

a. Nsqip_sc, preoptmtb, framingrisk, Pre_PP, Pre_MAP, and Pre_SBP are predictors

b. 2nd Analysis is with adjusted TMTA & B scores along with TMTB-TMTA calculation (also an adjusted score)

i. vselect BX2_PDRSTS nsqip_sc framingrisk Pre_SBP Pre_PP TMTA_adj TMTB_adj Sex nsqip_d ASA Pre_DBP Pre_MAP TobPackYears GDS15 TMTBTMTA PreopTMTA PreopTMTB centered_age, best

1. Model #2-4 has lowest BIC, 120.273

a. Nsqip_sc, PreopTMTA, Pre_MAP, TMTB_adj are predictors in model

2. Model #2-8 has lowest AIC (100.9645), AICC (104.2481)

a. Nsqip_sc, Pre_MAP, framingrisk, TMTA_adj, PreopTMTA, TMTBTMTA, Pre_SBP, Pre_DBP

2. Lasso Regression is ran with the above models

a. 1st Analysis with raw TMTA & TMTB scores

i. Model #1-2: lars BX2_PDRSTS nsqip_sc PreopTMTB, a(lasso) g

1. Cp of 3.000 with R-square of 0.3147, both variables included

ii. Model #1-6 lars BX2_PDRSTS nsqip_sc PreopTMTB framingrisk Pre_PP Pre_MAP Pre_SBP, a(lasso) g

1. Nothing was identified as a predictor – the Cp of -93.9961 was the lowest with no variable

b. 2nd Analysis with adjusted TMT scores

i. #2-4 Model lars BX2_PDRSTS nsqip_sc PreopTMTA Pre_MAP TMTB_adj, a(lasso) g

1. Cp of 5.000 with nsqip_sc, PreopTMTA, Pre_MAP, TMTB_adj

ii. #2-8 Model lars BX2_PDRSTS nsqip_sc framingrisk PreopTMTA Pre_MAP TMTA_adj TMTBTMTA Pre_SBP Pre_DBP, a(lasso) g

1. Nothing was identified as a predictor – the Cp of -88.9995 was the lowest with no variable.

3. This process was done in reverse to confirm results. LASSO regression was completed

first with all variables loaded into the model (BX2_PDRSTS nsqip_sc framingrisk Pre_SBP Pre_PP PreopTMTA PreopTMTB Sex nsqip_d ASA Pre_DBP Pre_MAP Pre_PP TobPackYears GDS15 centered_age).

a. This identified nsqip_sc and PreopTMTB model to have the lowest Cp.

4. Standard multiple linear regression was completed with models that survived Lasso to run diagnostics

```
. regress BX2_PDRSTS nsqip_sc PreopTMTB, beta

      Source |       SS           df       MS      Number of obs   =        96
-------------+----------------------------------   F(2, 93)        =     21.35
       Model |  9.75137276         2  4.87568638   Prob > F        =    0.0000
    Residual |  21.2381319        93  .228367009   R-squared       =    0.3147
-------------+----------------------------------   Adj R-squared   =    0.2999
       Total |  30.9895046        95  .326205312   Root MSE        =    .47788

------------------------------------------------------------------------------
  BX2_PDRSTS |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
    nsqip_sc |   .0200414   .0041958     4.78   0.000                 .411387
  PreopTMTB |   .0037863   .0009329     4.06   0.000                .3495423
       _cons |   1.551683   .1230968    12.61   0.000                       .
------------------------------------------------------------------------------

. fitstat

Measures of Fit for regress of BX2_PDRSTS

Log-Lik Intercept Only:      -81.945     Log-Lik Full Model:          -63.808
D(93):                       127.615     LR(2):                        36.274
                                         Prob > LR:                    0.000
R2:                            0.315     Adjusted R2:                  0.300
AIC:                           1.392     AIC*n:                      133.615
BIC:                        -296.869     BIC':                       -27.145
```

a. Rstudent

residuals

created. i.

0 >2.5

ii. 3>2.00 (NV28: 2.088, NV55.1: 2.001, NV11: -2.02)

b. Leverage values were created

c. Leverage was plotted against Rstudent residuals

i. NV003 and NV004 may be influential

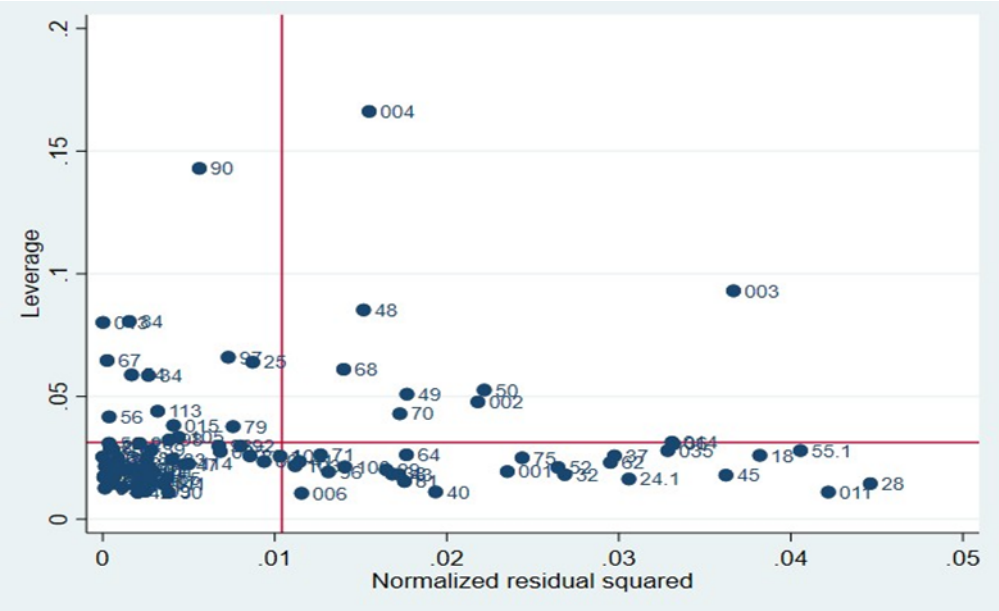
d. Cooks D, Dfit and DFbeta were examined

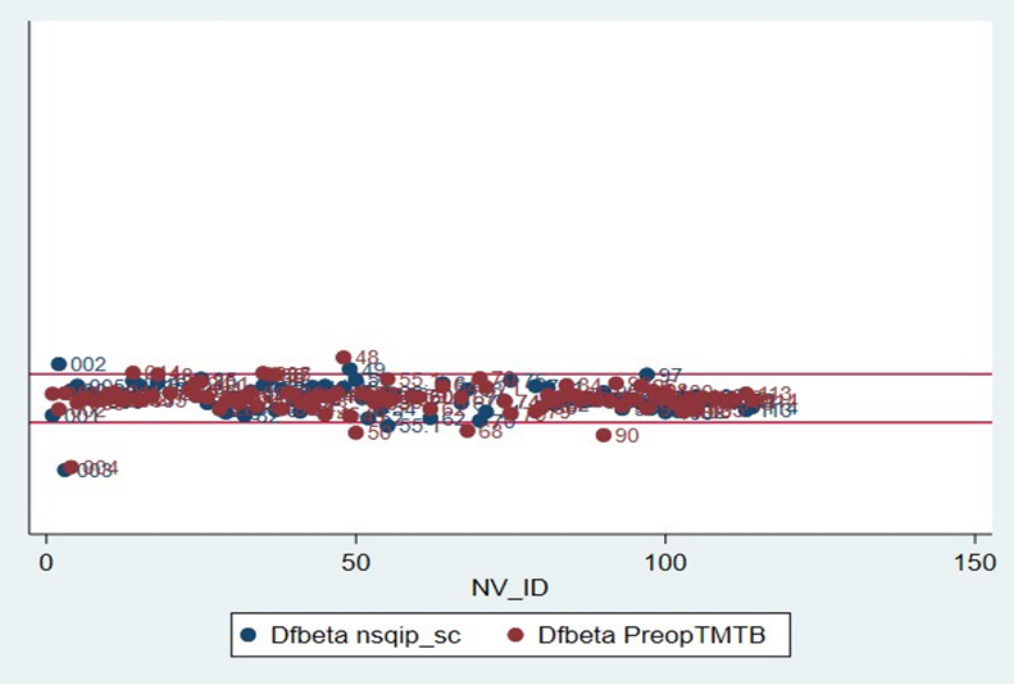
   i. NV003, 004, and 048 have cooksd >4/96

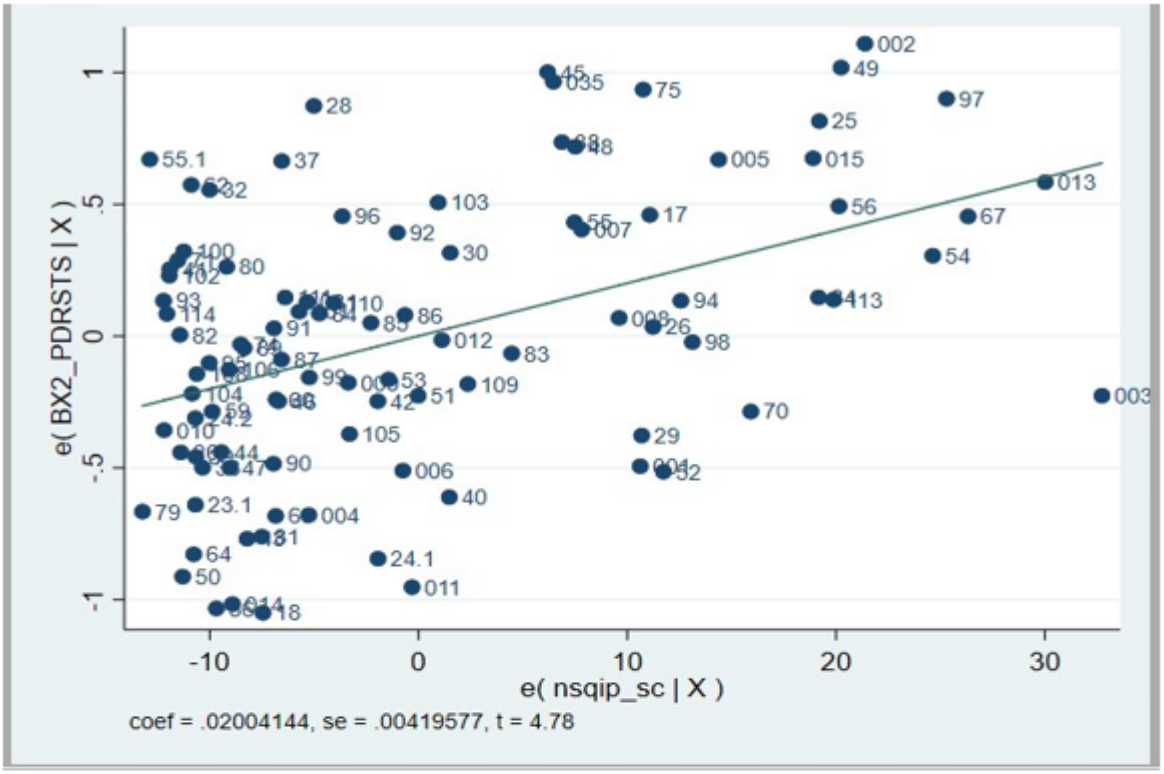
   ii. Dfbetas were calculated, figure below shows that NV003 and NV004 are clearly below the cutoff line of 0.20 (calculated by 2/sqrt(n))

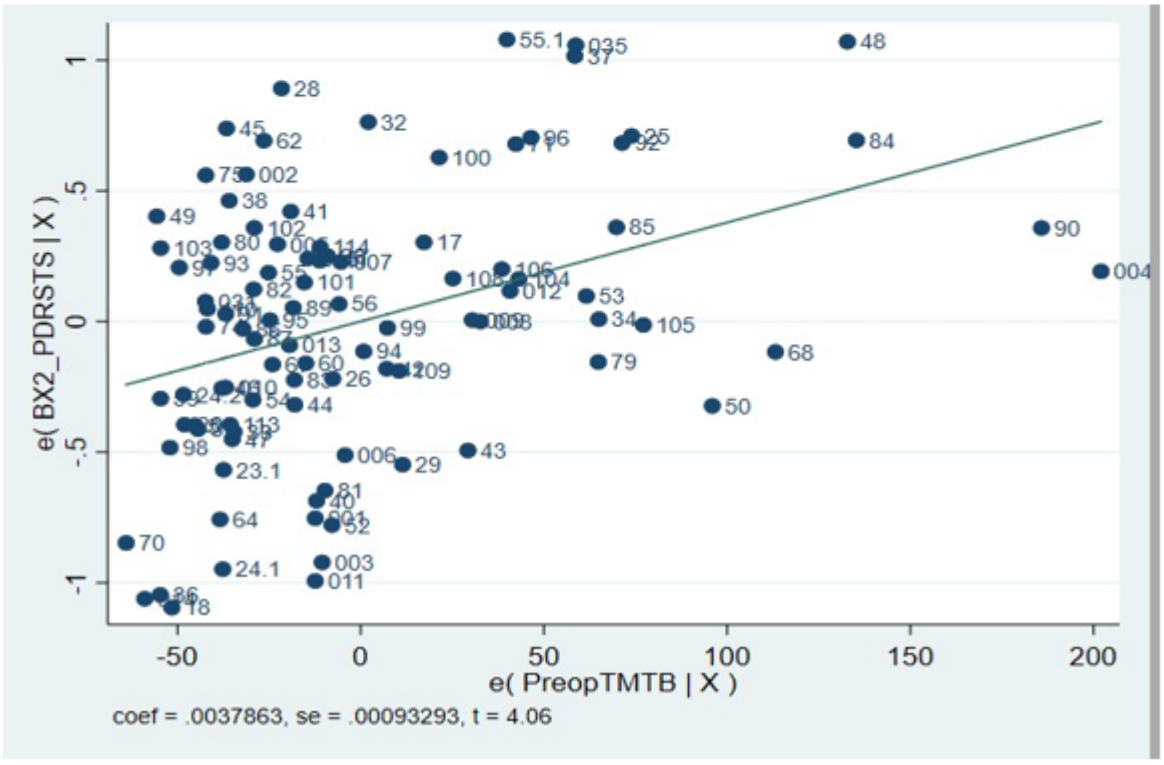
   iii. The AVplots below show that NV003 is an outlier for NSQIP_SC (serious complications)

   iv. NV004 is an outlier for PreopTMTB



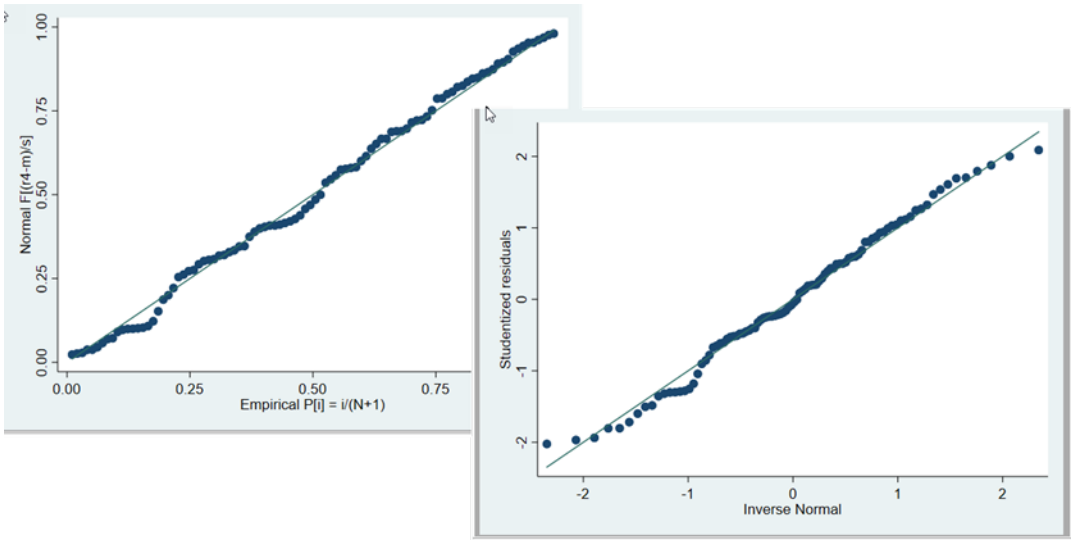

002
48
97
90
0
50
100
150
NV_ID
Dfbeta nsqip_sc
Dfbeta PreopTMTB


e(BX2_PDRSTS | X)
e(nsqip_sc | X)
coef = .02004144, se = .00419577, t = 4.78

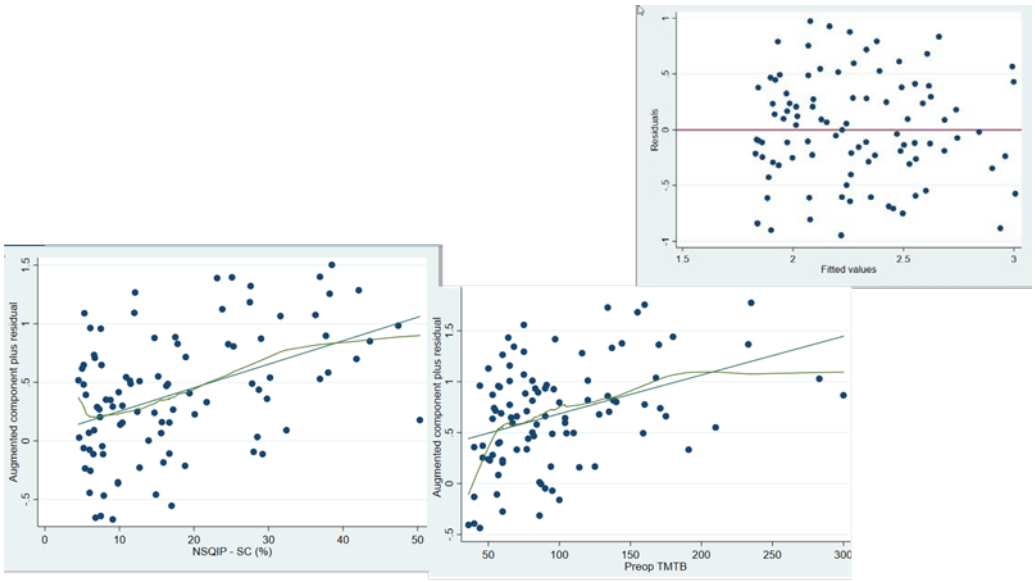

e(BX2_PDRSTS | X)
e(PreopTMTB | X)
coef = .0037863, se = .00093293, t = 4.06

As a test, I removed NV003 and NV004 from the analysis and re-ran the regression. The BIC dropped approximately 8 points from -296 to -292.10. Since the model did not change drastically, these two data points remained in the analysis. No other reason to exclude them (biological, etc.)

Next, Kernal density plot show normal distributions, as does pnorm and qnorm



e. Shapiro-Wilk Test is not significant indicating normality in the data
f. The White's test and Breush-pagan test are not significant indicating that the residuals are homogenous and that there is not heteroscedasticity in the data
g. VIF: No data points >10.00,
   both nsqip_sc and
   PreopTMTB have values of
   1.01
h. OVtest not significant



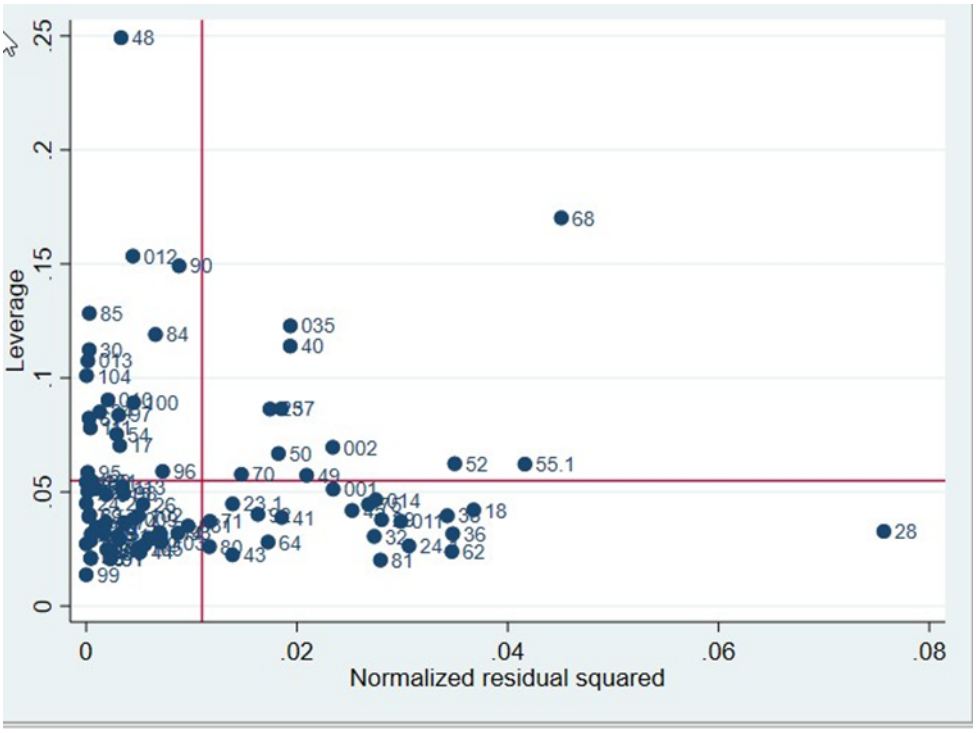

```
. regress BX2_PDRSTS nsqip_sc PreopTMTA Pre_MAP TMTB_adj, beta
```

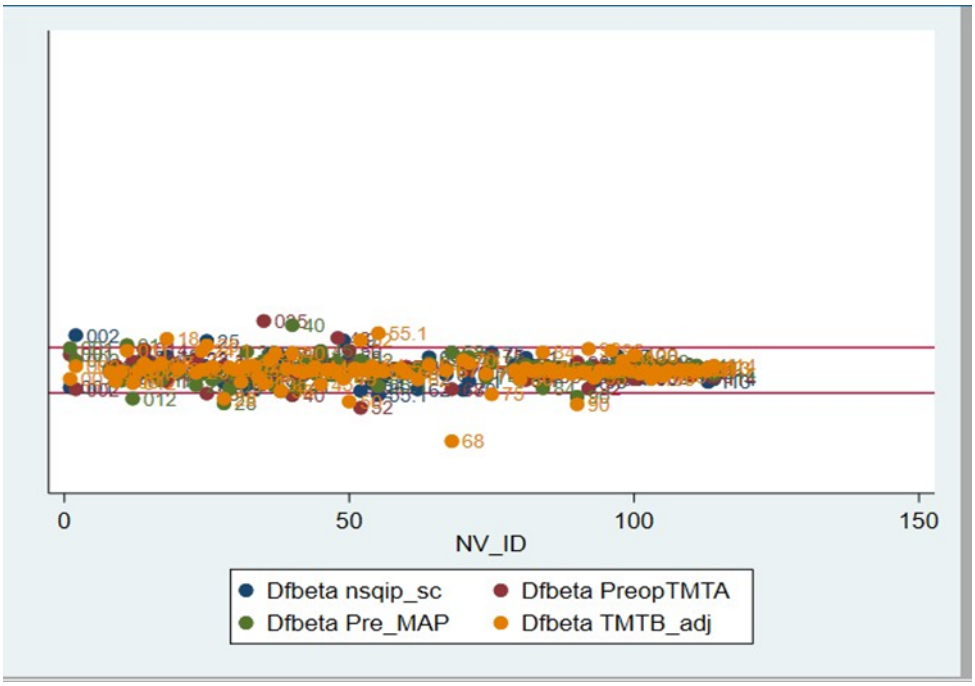
Source	SS	df	MS		
Model	12.1450846	4	3.03627114	Number of obs	= 91
Residual	17.978929	86	.209057314	F(4, 86)	= 14.52
				Prob > F	= 0.0000
				R-squared	= 0.4032
				Adj R-squared	= 0.3754
Total	30.1240135	90	.334711262	Root MSE	= .45723

BX2_PDRSTS	Coef.	Std. Err.	t	P>|t|	Beta
nsqip_sc	.0220017	.0042471	5.18	0.000	.4334642
PreopTMTA	.0050516	.0034845	1.45	0.151	.1438041
Pre_MAP	.0095223	.0046354	2.05	0.043	.1748412
TMTB_adj	.2044621	.0612352	3.34	0.001	.3362754
_cons	.8158156	.4594557	1.78	0.079	.

```
. fitstat

Measures of Fit for regress of BX2_PDRSTS

Log-Lik Intercept Only:      -78.821     Log-Lik Full Model:          -55.338
D(86):                       110.676     LR(4):                        46.967
                                         Prob > LR:                     0.000
R2:                            0.403     Adjusted R2:                   0.375
AIC:                           1.326     AIC*n:                       120.676
BIC:                        -277.258     BIC':                        -28.924

. listcoef

regress (N=91): Unstandardized and Standardized Estimates

 Observed SD: .57854236
 SD of Error: .45722786
```

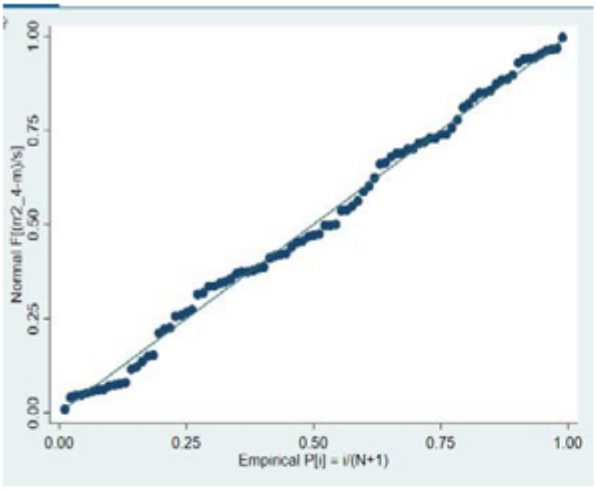
5. Model 2-4 Regression Diagnostics
    a. Student residuals
        calculated (r2_4) i.
        1 > 2.5 NV028: 2.68
            ii. 1>2.0 NV068: -2.20
    b. Leverage calculated
            i. NV068 may be an influential variable
    c. Cooks' D, Dfit, Dfbetas calculated
            i. NV068 has high Cooks D and is displayed in the Dfbetas plot below

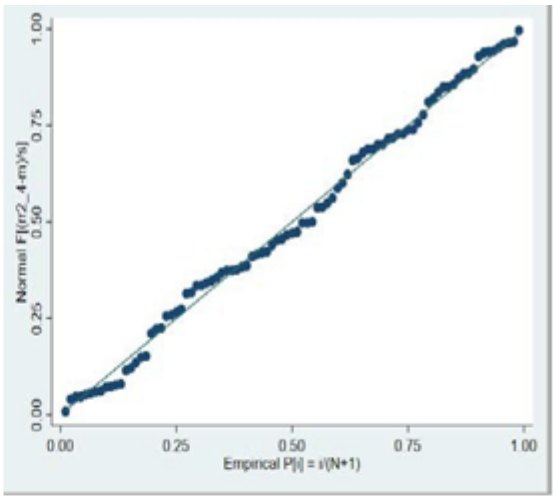


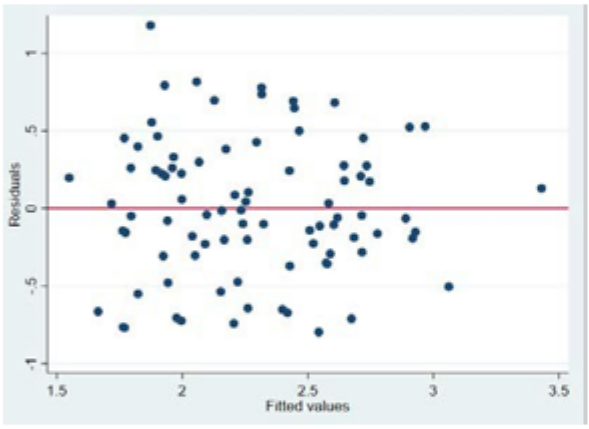


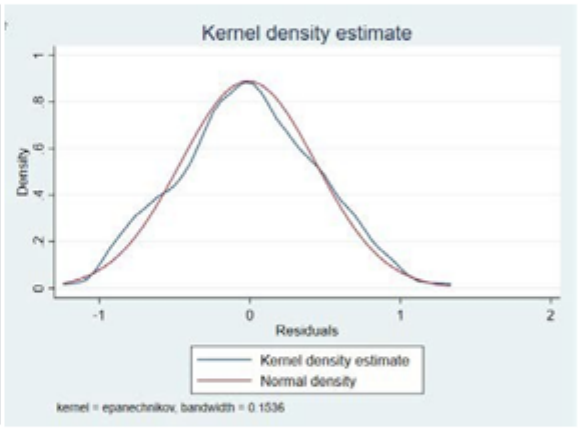
Regression ran without NV068 did not result in an improvement of the model, the BIC did not change. Therefore, NV068 will remain in the model.

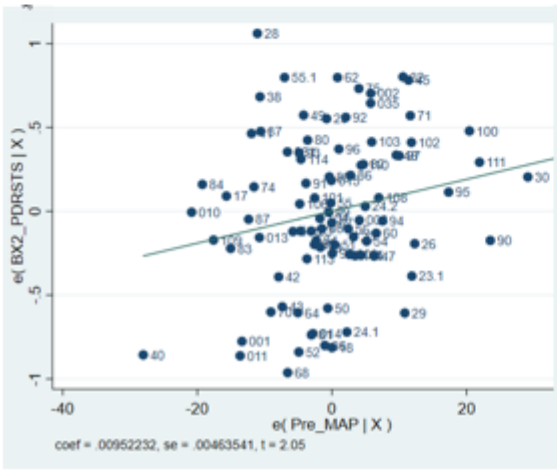
Shapiro-Wilk Test for normality was not significant. White and Breush-pagan test were not significant for heteroskedasity. VIFs were <10. (1.3 or 1.0 values). OVtest was not significant indicating there are not any specification errors.

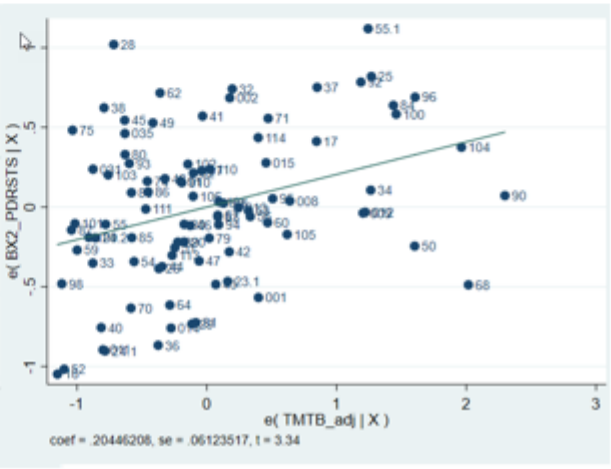

Normal F[(r2_4-m)/s]
Empirical P[i] = i/(N+1)
0.00
0.25
0.50
0.75
1.00

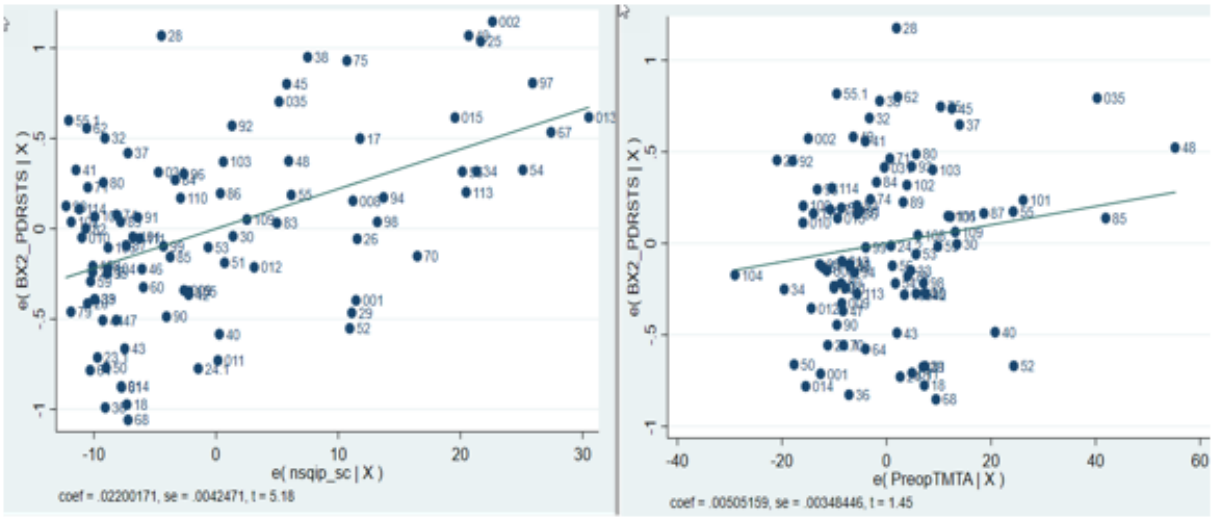

Normal F[(r2_4-m)/s]
Empirical P[i] = i/(N+1)
0.00
0.25
0.50
0.75
1.00

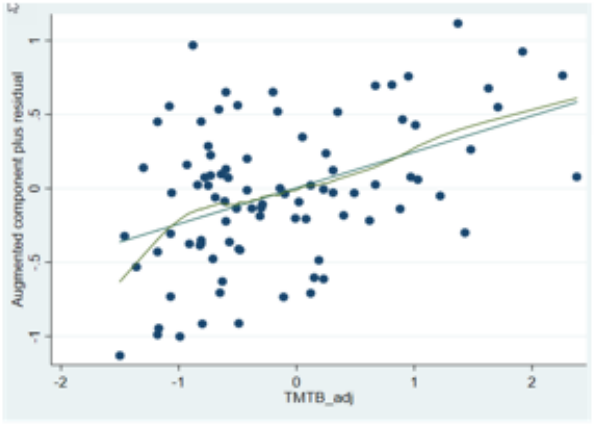

Residuals
Fitted values
1.5
2
2.5
3
3.5

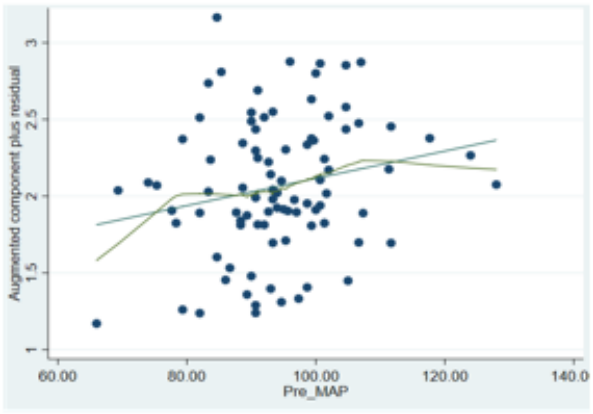

Kernel density estimate
Density
Residuals
Kernel density estimate
Normal density
kernel = epanechnikov, bandwidth = 0.1536

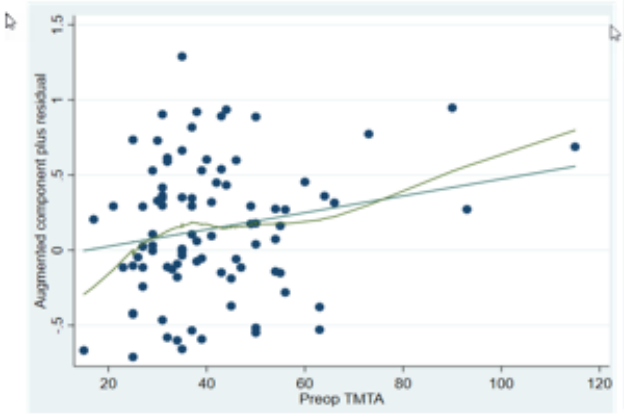

e(BX2_PDRSTS | X)
e(Pre_MAP | X)
coef = .00952232, se = .00463541, t = 2.05

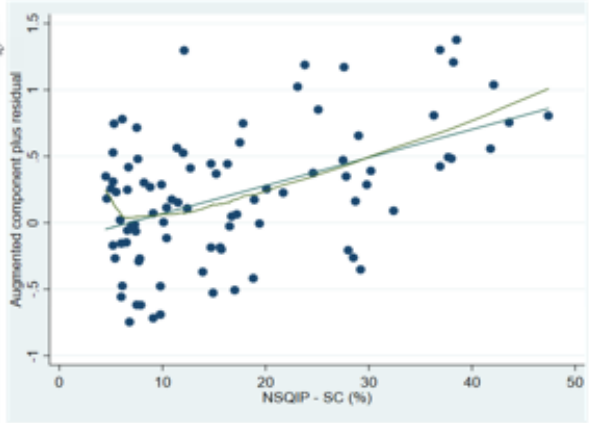

e(BX2_PDRSTS | X)
e(TMTB_adj | X)
coef = .20446208, se = .06123517, t = 3.34


e(BX2_PDRSTS | X)
e(nsqip_sc | X)
coef = .02200171, se = .0042471, t = 5.18
e(BX2_PDRSTS | X)
e(PreopTMTA | X)
coef = .00505159, se = .00348446, t = 1.45


Augmented component plus residual
TMTB_adj


Augmented component plus residual
Pre_MAP


Augmented component plus residual
Preop TMTA


Augmented component plus residual
NSQIP - SC (%)

Appendix 2: Regression Analysis & Diagnostics of Postoperative Delirium

Incidence N=97

1 missing TMTB, final n=96

This analysis is focused on building a prognostic delirium prediction model for postoperative delirium incidence. No delirium = 66, Delirium = 31. R v3.4.4 was used for LASSO and Best Subsets Regression. Stata does not have an updated version of this type of binomial analysis using LASSO/Best Subsets. The R packages glmnet() for LASSO and glmbest() for Best Subsets were used.

LASSO identified nsqip_sc with the best value of Cp, with the coefficient 0.02.

Best Subsets identified the best model to be nsqip_sc with a BIC of 81.63140 (relative to nested models). The second best model was nsqip_sc with PreopTMTB with a BIC of 81.63443. Since these values were within 0.003 of each other, AUROC and McKelvey and Zavoina's R2 values were examined with nsiqp_sc only versus nsqip_sc and PreopTMTB models. The addition of PreopTMTB improved the AUROC by 5 points, from 0.76 to 0.81. The Adjusted R2 value improved from 0.25 to 0.33. Therefore, the model with both nsqip_sc and PreopTMTB was explored further using regression diagnostics.

Model Diagnostics: logistic DELYN nsqip_sc PreopTMTB. Done with Stata v15.1.

https://stats.idre.ucla.edu/stata/webbooks/logistic/chapter3/lesson-3-logistic-regression-diagnostics/

Pearson residuals: standardized difference b/w the observed frequency and the predicted frequency. Relative deviations between the observed and fitted values.

Deviance residuals: measures disagreement b/w the maxima of the observed and the fitted log likelihood functions. Goal is to minimize the sum of the deviance residuals.

Hat diagonal/Pregibon leverage: measures leverage of the observation. These three statistics are the building blocks for logistic regression.

Linktest was negative, indicating no important variables are missing from the model. Goodness-of-fit with 50 iterations was performed with negative results, the model is fit well.

VIF: 1.01 for both.

Estimating robust standard errors, by using logit DELYN nsqip_sc PreopTMTB, vce(robust) results in the same exact model with same SE.

```
. logistic DELYN nsqip_sc PreopTMTB

Logistic regression                             Number of obs     =         96
                                                LR chi2(2)        =      25.82
                                                Prob > chi2       =     0.0000
Log likelihood = -47.476587                     Pseudo R2         =     0.2138

------------------------------------------------------------------------------
       DELYN | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nsqip_sc |   1.096865   .0253308     4.00   0.000     1.048324    1.147653
   PreopTMTB |   1.010676   .0047278     2.27   0.023     1.001452    1.019984
       _cons |   .0269296   .0208363    -4.67   0.000     .0059105    .1226966
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.

. fitstat

Measures of Fit for logistic of DELYN

Log-Lik Intercept Only:      -60.389     Log-Lik Full Model:          -47.477
D(93):                        94.953     LR(2):                        25.824
                                         Prob > LR:                     0.000
McFadden's R2:                 0.214     McFadden's Adj R2:             0.164
Maximum Likelihood R2:         0.236     Cragg & Uhler's R2:            0.329
McKelvey and Zavoina's R2:     0.325     Efron's R2:                    0.257
Variance of y*:                4.876     Variance of error:             3.290
Count R2:                      0.760     Adj Count R2:                  0.258
AIC:                           1.052     AIC*n:                       100.953
BIC:                        -329.531     BIC':                        -16.695

. estat ic

Akaike's information criterion and Bayesian information criterion

------------------------------------------------------------------------------
       Model |  Obs  ll(null)  ll(model)  df        AIC        BIC
-------------+----------------------------------------------------------------
           . |   96 -60.38865 -47.47659   3   100.9532   108.6462
------------------------------------------------------------------------------
Note: N=Obs used in calculating BIC; see [R] BIC note.

Logistic model for DELYN, goodness-of-fit test

(Table collapsed on quantiles of estimated probabilities)

      number of observations =        96
            number of groups =        50
  Hosmer-Lemeshow chi2(48) =        52.04
                Prob > chi2 =         0.3195
```

Logistic model for DELYN

Classified	True D	~D	Total
+	16	8	24
-	15	57	72
Total	31	65	96

Classified + if predicted Pr(D) >= .5
True D defined as DELYN != 0

Sensitivity	Pr(+| D)	51.61%
Specificity	Pr(-|~D)	87.69%
Positive predictive value	Pr(D| +)	66.67%
Negative predictive value	Pr(~D| -)	79.17%
False + rate for true ~D	Pr(+|~D)	12.31%
False - rate for true D	Pr(-| D)	48.39%
False + rate for classified +	Pr(~D| +)	33.33%
False - rate for classified -	Pr(D| -)	20.83%
Correctly classified		76.04%

. linktest, nolog

Logistic regression

Number of obs	=	96
LR chi2(2)	=	26.26
Prob > chi2	=	0.0000
Pseudo R2	=	0.2174

Log likelihood = -47.258521

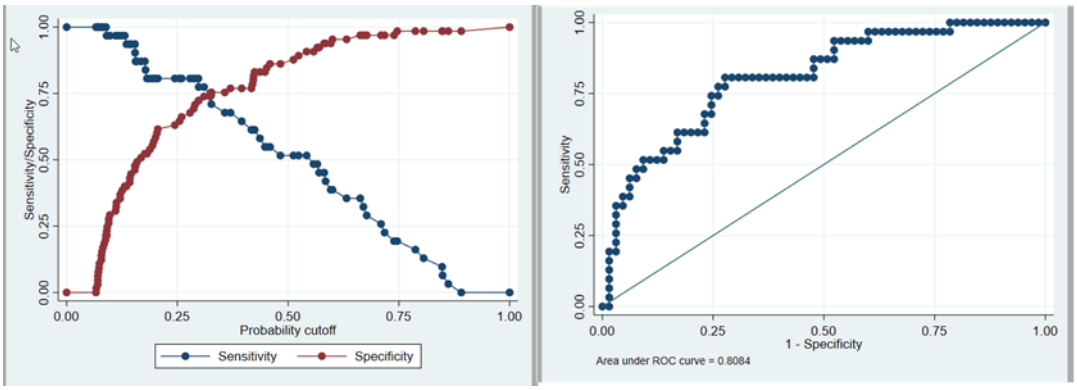
DELYN	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_hat	.8817193	.2783665	3.17	0.002	.336131	1.427308
_hatsq	-.1176025	.1767963	-0.67	0.506	-.4641168	.2289118
_cons	.1152218	.3337392	0.35	0.730	-.5388949	.7693386

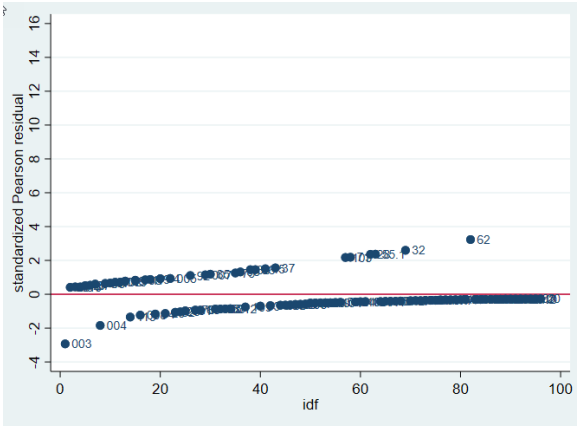
. listcoef

logistic (N=96): Factor Change in Odds

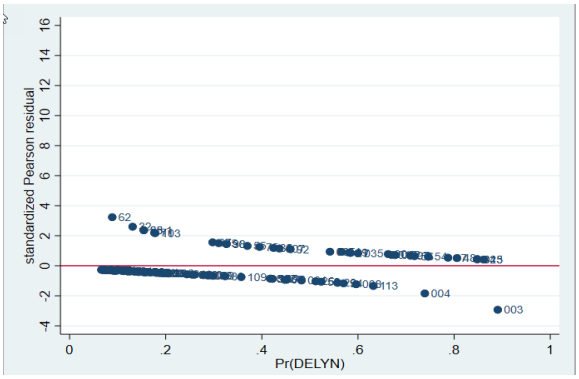
Odds of: 1 vs 0

DELYN	b	z	P>|z|	e^b	e^bStdX	SDofX
nsqip_sc	0.09246	4.004	0.000	1.0969	2.9563	11.7238
PreopTMTB	0.01062	2.270	0.023	1.0107	1.7505	52.7267

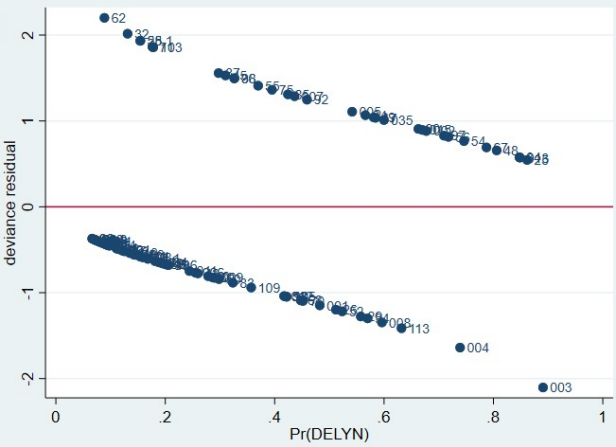




This plot is examining the standardized pearson residuals against the probabilities. We want to examine for influential participants, *i.e.*, 62. However, this distribution is tight and does not warrant much concern.



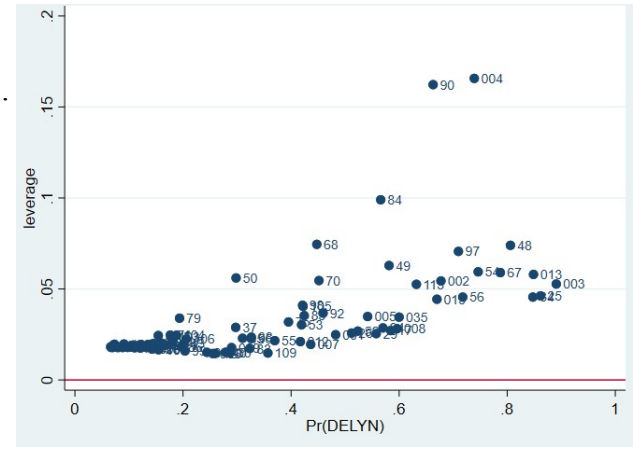
This plot also displays the standardized pearson residuals, however, the x access is the index id (id=_n). This numbers the covariate patterns. Observations with the same covariate have the same number. Again, NV062 and now NV003/NV004 are away from the general distribution indicating they have higher standardized residuals and may influence the model. This is a tighter data distribution and does not cause much concern.



This plot is displaying the deviance residual against the DELYN probabilities. Again, NV062, NV004 and NV003 are observations that may have influence on the model. NV004 is an outlier for TMTB (TMTB value of 300).

NV003 is an outlier for NSQIP-SC (NSQIP-SC value of 50.3. Neither became delirious. NV062 did become delirious, but has a low nsqip_sc & TMTB score.

This plot displays the leverage statistic (hat) against the predicted probabilities of DELYN



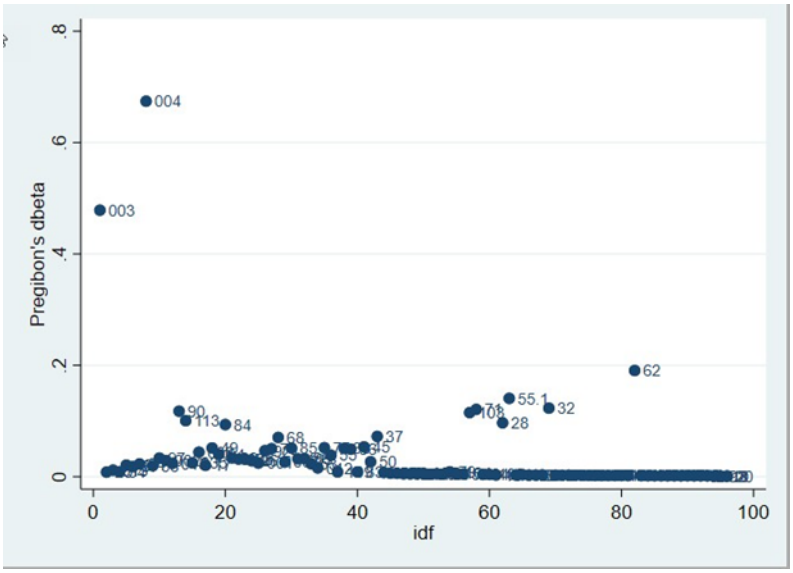
Dx2 measures the decrease in the Pearson X2 goodness-of-fit statistic that would be caused by deleting an observation and any others that may be sharing that covariate pattern.

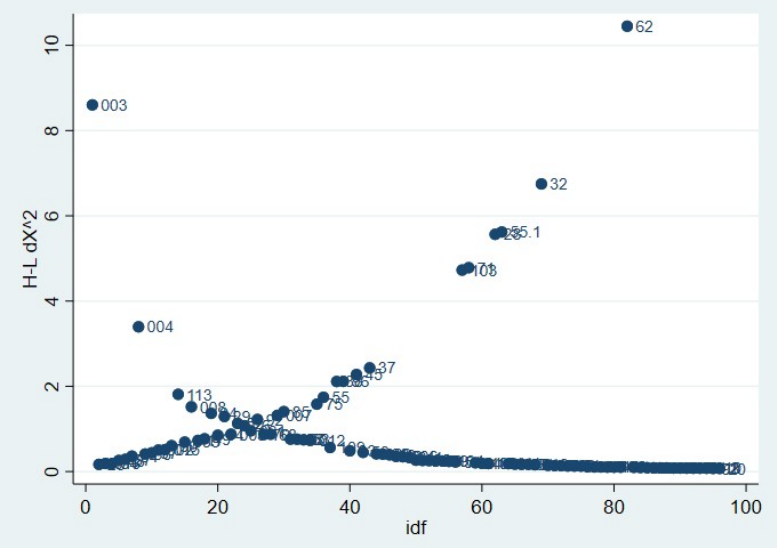
From Hosmer, Lemeshow and Sturdivant (2013):

- Top left to bottom right points are consistent with the covariate patterns, the number of positive outcomes equal to the number in the group.
- Top right to bottom left points are consistent with positive outcomes.

- Since the data points are unique, the dots lie along these two curves.
- We are looking for larger values of Dx2 (NV062, NV003)

Pregibon (1981) dfbeta statistic measures how the coefficient vector would change if that observation/participant was deleted. NV004, NV003 have largest values (see outlier discussion above).



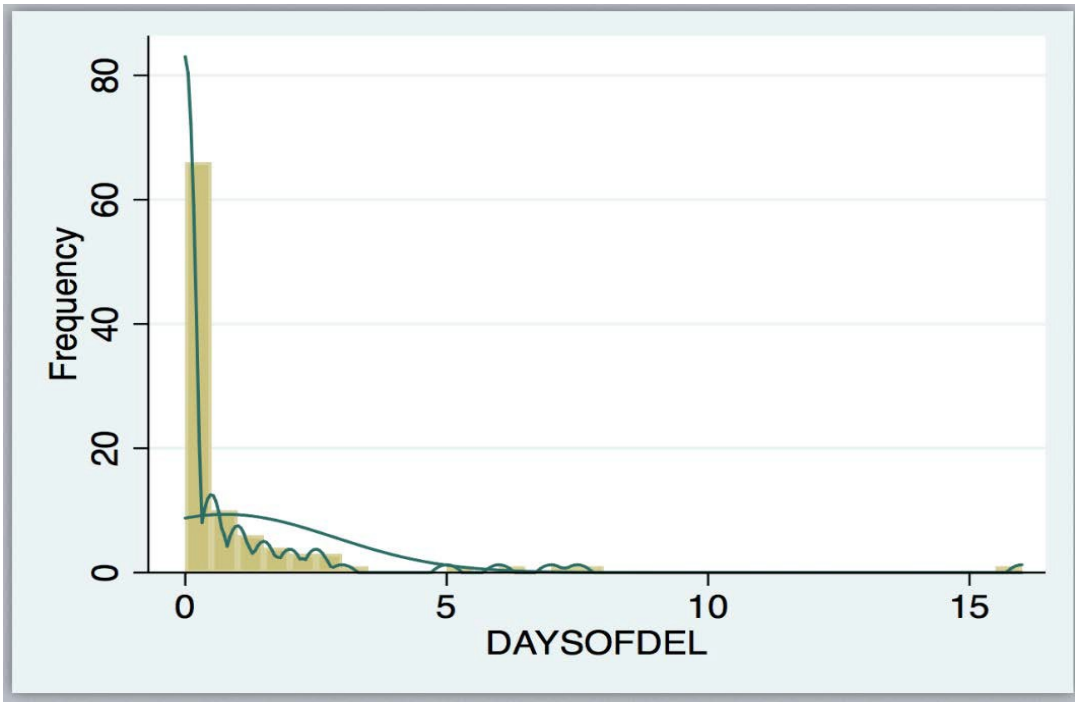


Appendix 3: Regression Analysis & Diagnostics of Delirium

Duration N=97

This analysis is focused on building a prediction model for the duration of delirium. Each CAM+ measurement is counted as 0.5 days of delirium. The data distribution is as follows, with a mean delirium duration of 0.75 days. Due to the distribution of the data as well as being count data, a zero-inflated negative binomial regression model was used. This type of regression model is applied in situations when there is excess zeros and an overdispersed distribution. The theory assumes that a separate process generates the extra zeros from the count values and that the excess zeros should be modeled independently (cite https://stats.idre.ucla.edu/stata/dae/zero-inflated-negative-binomial-regression/)

Distribution of Days of Delirium:



Variable Selection:

Lasso on DaysofDelirium, count data using STATA v15.1:

- lars DAYSOFDEL (nsqip_sc nsqip_d ASA Sex framingrisk Pre_MAP Pre_PP

TobPackYears PreopTMTA PreopTMTB centered_age GDS15 vascsurg), a(lasso) g

- nsqip_sc (0.02), Pre_PP (0.01), vascsurg (0.30) had lowest Cp of 5.23.
  - The outlier of 16 days duration of delirium had vascular surgery. Since this outlier was shown to highly influence the data in total delirium rating scale score, vascsurg was removed as it is likely biased.

- lars DAYSOFDEL (nsqip_sc nsqip_d Sex ASA framingrisk Pre_MAP Pre_PP TobPackYears PreopTMTA PreopTMTB centered_age GDS15), a(lasso) g
  - nsqip_sc, has lowest Cp of 3.27, coefficients 0.02.

To perform LASSO on the zero distribution, R v3.4.4 was used as this is a binomial distribution (no delirium=1, delirium=0). The package "glmnet" was used.

- covariates <- dd[,c("Age","Sex","GDS15","nsqip_sc","nsqip_d","PreopTMTA","PreopTMTB", "framingrisk", "Pre_PP", "Pre_MAP", "ASA", "TobPackYears", "DELN")]
- `%ni%`<-Negate(`%in%`)
- x<-model.matrix(DELN~age + sex + GDS15 + nsqip_sc + nsqip_d + TMTA + TMTB +framingrisk + Pre_PP + Pre_MAP+ASA+pack,data=dd)
- x=x[,-1]
- glmnet1<- cv.glmnet(x,y=na.omit(covariates)$DELN,alpha=1,family="binomial",type.measure = "mse")
- lambda_min <- glmnet1$lambda.min
- lambda_1se <- glmnet1$lambda.1se

NSQIP-SC was identified with the best lambda value of -0.01. This will be used in the inflated portion of the zero-inflated binomial model.

Zero-inflated negative binomial regression model, Stata:

- zinb DAYSOFDEL nsqip_sc, inflate(nsqip_sc) zip
- fitstat
  - The zip option performs a likelihood ratio test of alpha=0, testing whether the zero-inflated negative binomial model is a better fit of the data then a zero- inflated Poisson regression model.

- The zip test was $p<0.001$, indication the zero-inflated negative binomial model was a good fit.

- Model output is displayed below.

```
Zero-inflated negative binomial regression      Number of obs     =         97
                                                Nonzero obs       =         31
                                                Zero obs          =         66

Inflation model = logit                         LR chi2(1)        =       6.68
Log likelihood  = -95.74973                     Prob > chi2       =     0.0098
```

DAYSOFDEL	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
DAYSOFDEL						
nsqip_sc	.0693555	.0292769	2.37	0.018	.011974	.1267371
_cons	-1.766342	.828431	-2.13	0.033	-3.390037	-.1426474
inflate						
nsqip_sc	-.2249213	.2801854	-0.80	0.422	-.7740746	.324232
_cons	.8461639	2.724668	0.31	0.756	-4.494087	6.186415
/lnalpha	.8295742	.3483256	2.38	0.017	.1468687	1.51228
alpha	2.292342	.7984814			1.158202	4.537062

```
Likelihood-ratio test of alpha=0: chibar2(01) =    51.21 Pr>=chibar2 =  0.0000

.
end of do-file

. fitstat

Measures of Fit for zinb of DAYSOFDEL

Log-Lik Intercept Only:     -104.634     Log-Lik Full Model:          -95.750
D(92):                       191.499     LR(3):                        17.768
                                         Prob > LR:                     0.000
McFadden's R2:                 0.085     McFadden's Adj R2:             0.037
Maximum Likelihood R2:         0.167     Cragg & Uhler's R2:            0.189
AIC:                           2.077     AIC*n:                       201.499
BIC:                        -229.374     BIC':                         -4.044
```

```
. predict pnd, pr

. table nsqip_sc, con(mean pnd)
```

NSQIP - SC (%)	mean(pnd)
4.5	.4585995
4.6	.4530205
5	.4308357
5.2	.4198409
5.3	.4143724
5.4	.4089251
5.5	.4034999
5.9	.3820463
6	.3767505
6.1	.371484
6.5	.3507317
6.6	.3456272
6.7	.340558
6.8	.335525
7	.3255712
7.3	.3109329
7.4	.3061345
7.5	.3013778
7.6	.2966633
7.7	.2919918
7.8	.2873638
7.9	.2827798
8.2	.2692973
8.8	.2435815
9.1	.231365
9.8	.204555
9.9	.2009196

The predicted probability of having a certain nsqip_sc score and being a zero (no delirium) was calculated. As shown below, the lower the nsqip_sc score, the higher the probability of nothaving delirium.

As a summary, nsqip_sc was identified as a predictor for the duration of delirium (positive relationship, higher nsqip_sc score equals longer delirium duration). Nsqip sc was also identified to predict the excess zeros, *i.e.,* no delirium. This was a negative relationship, *i.e.*, a lower or decreasing nsqip_sc score indicates a higher probability of not becoming delirious.