Spring 6-25-2020

A Case for Delirium Risk Prediction Models to Aid in Triaging Resources to those Most at Risk an Integrative Literature Review

Tammy Perttula
tammylynn102018@gmail.com

A Case for Delirium Risk Prediction Models to Aid in Triaging Resources to those Most at Risk an Integrative Literature Review

A Scholarly Inquiry Paper
Submitted to the Faculty of the Department of Nursing
College of Nursing and Health Sciences
of Winona State University

by
Tammy L. Perttula

In Partial Fulfillment of the Requirements
For the Degree of
Master of Science: Acute Gerontology Acute Care Nurse Practitioner Program

June 25th, 2020

ACKNOWLEDGMENT

I would like to than Dr. Kim Langer and Diane Forsyth for the guidance, support, and advice throughout this writing process. Their expertise was an asset to this project, aiding in its completion. I would also like to thank Dr. Meller in the Fairview- University of Minnesota Hospital Psychiatry department for your interest, mentorship, and guidance on the subject of delirium. Dr. Meller introduced me to his colleagues whom have been instrumental in past delirium research, Dr. Steven Thurber and Dr. Yasuhiro Kishi. They are all so passionate in this subject area, giving me hope for the future.

I will be forever in debt to my mother Diane, aunt Eileen, and Mike Prohofsky for their support and many hours of reading and editing! My entire family has provided so much encouragement over the years, especially my children, Zach and Nick. Thank you for all your sacrifices! All the support I have been given is the reason I was able to complete this project and so many more.

Lastly, I thank all the researchers that poured their time into developing and validating delirium risk prediction models as a way to better the care received during hospitalization. I also want to mention my Grandmother Florence whom was the subject of my desperate search to slow the rates of iatrogenic delirium in an effort to prevent its negative impacts on an individual's quality of life.

Abstract

Delirium is a complex syndrome resulting from compounding effects of acute illness, comorbidities, and environment. It results in adverse outcomes: elevated mortality rates, length of stay, readmissions, institutionalization, long-term cognitive changes, and diminished quality of life. The rates of iatrogenic delirium are astounding, ranging from 10%-89%. There are no curative treatments; thus, primary prevention is the key. The purpose of this literature review is to identify and critique the research for accuracy of risk stratification and feasibility in practice. Support for interventions that prevent delirium is mounting; however, interventions are resource-intensive and often not implemented. Researchers have responded to this problem by developing risk stratification tools to triage interventions toward those of the highest risk. There is evidence that some of the models' implementation is successful; however, they are not yet widely operationalized. A compilation of seven published models of risk prediction were critiqued and compared using the Stetler Model of Evidence-Based Practice as a guiding model. The Newcastle-Ottawa Scale and the Critical Appraisal and the Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS checklist) are employed to aid in the critical appraisal, evaluation of the study's quality, and aid in data abstraction. The models show the ability to stratify risk, but their effectiveness in practice cannot be studied without directed interventions because they risk prediction models are created to aid healthcare staff in making clinical decisions. Therefore, a complete clinical pathway with evidence-based interventions should be employed with a delirium risk prediction model to triage the interventions to patients

at the highest risk. Recommendations are to implement an automated electronic model (automatic calculation using the EMR or a machine learning model) into clinical practice along with a delirium prevention care pathway. Electronic versions of risk scores allow for an opportunity to achievement clinical efficiency and show statistical superiority to the other models. Published evidence on the impact of the models is diminutive, their ability to triage patients and aid in clinical decision-making should be published in an impact study.
Keywords: Delirium, risk assessment, risk prediction, risk model, risk score, patient safety, patient-centered outcomes research.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

Figure Page

LIST OF TABLES

Introduction

According to the Diagnostic and Statistical Manual 5th Edition (DSM-5), delirium is an acute cognitive change resulting in an alteration in cerebral functionality of the brain, and its severity fluctuates over a short period (American Psychiatric Association [APA], 2013). Delirium is a medical emergency indicating the presence of a severe underlying illness resulting from a complex compilation of predisposing and precipitating factors that are at the root of its onset (Wass, Webster, & Nair, 2018). Its hallmark signs are disturbances in awareness, attention, and perception. The negative consequences of delirium are far-reaching, impacting the patients, families, healthcare staff, hospitals, and healthcare systems. Those with delirium are at a higher risk of morbidity, mortality, postoperative complications, prolonged length of stay, readmissions, institutionalization, and long-term cognitive changes resulting in a diminished quality of life (Douglas et al., 2013; Inouye et al., 1993; Rudolph et al., 2011).

In acute care settings delirium is the single most common acute disorder affecting aged (Carrasco et al., 2014; Health Research and Education Trust [HRET], 2018; & Inouye, 2018). It affects 10% to 20% of all hospitalized adults (over 18 years of age), 14% to 56% of all hospitalized patients aged 60 and over, 42% in the general medical settings (Carrasco, Villarroel, Andrade, Calderon, & Gonzalez, 2014), and up to 89% of patients admitted to an intensive care unit (ICU) (Hayhurst, Pandharipande & Hughes, 2016; HRET, 2018).

Preventative Interventions

Studies show that multifactorial and interdisciplinary interventions result in an overall reduction in delirium rates between 30% (Brown et al., 2017) and 53% (Halladay, Sillner, & Rudolph, 2018; Inouye et al., 1999; Wong et al., 2018). The interventions focus on early and

frequent mobilization, promotion of a healthy circadian rhythm, adequate hydration, urinary and fecal continence, reorientation to place, time, and situation, therapeutic activities (walking, watching tv, listening to music, playing cards, folding towels, or engaging in self-care), use of glasses and hearing aids or devices other sensory devices, anesthesia protocols (sedation medications to avoid and minimization of sedation levels), prompt removal of intravenous (IV) lines, drains, and restraints, nutritional optimization, and "non-pharmacologic" sleep promotion interventions, and pharmacist involvement in decreasing use of medications associated with delirium development (Douglas et al., 2013; HRET, 2018; Hospital Elder Life Program [HELP], 2013; Inouye et al., 1999; National Institute for Health and Care Excellence [NICE], 2010; Reston & Schoelles, 2013). The implementation of the interventions requires a clinician to recognize and prioritize the prevention of delirium; however, during acute hospitalizations, co-occurring illnesses often take precedence over delirium risk (Taft, Nelsen, Slager, & Weir, 2018).

Guideline and Societal Recommendations for Risk Determination

There is no known cure for delirium; thus, its management of a delirious patient includes: treating the potential causes (precipitating factors), supportive care during recovery, and, when necessary, the pharmacological treatment for behavioral symptoms (Michaud et al., 2007). NICE published a guideline for delirium recognition and management recommending that all adults at risk for delirium should receive tailored interventions to prevent the iatrogenic onset (NICE, 2014); they list risk factors for delirium but do not suggest a systematic method to determine risk. In response to the need for a systematic method, the HRET proposed that all healthcare facilities employ a risk prediction model to identify patients at high risk for the development of delirium (HRET, 2018). In an effort to remedy the gaps in delirium prevention innovative

researchers have been prompted to develop delirium risk prediction models. Individual healthcare systems may have also developed methods to systematically stratify the risk of delirium and provide guidance on interventions to implement based on the risk level. This author located two published models that are implemented into clinical practice within two healthcare systems (Douglas et al., 2013; Rudolph, Doherty, Kelly, Driver, & Archambault, 2016).

Introduction to Delirium Risk Prediction Models

Despite the high rates of hospital delirium and its adverse effects on a patient, family, and healthcare system, delirium risk prediction models rarely are adopted into practice. The reasons that delirium risk prediction models are not employed include a lack of consensus on the causes and pathophysiology, which model to use in which situation, the timing of the assessment, the accuracy of the current models in clinical settings, as well as the fact that the guidelines are not recommending one model over another (Newman et al., 2015; Wang et al., 2017).

The first delirium risk prediction model (DRPM) was developed in 1993 by Dr. Inouye and colleagues, using a dual prospective cohort study approach, in tandem, to develop and validate a risk prediction model (Inouye et al., 1993). This model was never widely adopted into practice due to the required acute physiologic and chronic health evaluation scoring that created a necessity to draw additional laboratory sample (arterial blood gas). Some models rely on questionnaires administered by health care professionals making integration into clinical practice impractical. Usability in clinical practice is a requirement and includes accuracy of risk stratification, ease of use, and timeliness. In an external validation study by Pendlebury (2016a) found that DRPMs must be adapted and simplified to allow for use of routinely collected clinical assessments (Wong et al., 2018).

Delirium risk prediction models are developed to allow healthcare providers to target high-risk individuals to increase delirium screening assessments and implement targeted delirium prevention interventions. Assessing risk is not a new concept to healthcare; however, it is becoming more commonly used by healthcare providers to aid in the decision-making process when resources are limited, or risk of illness decreases by the implementation of interventions appropriate to the clinical question. Delirium is a complex, iatrogenic syndrome caused by numerous predisposing and precipitating factors. The use of clinical and technologically advanced prediction models may allow for triaging of resource-intensive interventions that make delirium preventable. Risk prediction models are most useful in situations such as delirium when the outcome is difficult to ascertain due to the immensity of causative factors.

Background

Despite early management and treatment of incident delirium it may result in considerable consequences for the patients and health care systems. Increased mortality rates during and post hospitalization, an average of eight days prolonged hospital stay, increased risk of complications, poor physical recovery and cognitive recovery, increased risk of development of dementia, and higher chance of placement in a residential care facility after discharge. Frailty prior to delirium and delays in diagnosis and treatment increase the odds of the occurrence of these negative outcomes.

Delirium carries an in hospital mortality risk of up to 75% whilst in the hospital and after discharge 40% in the first year (Wass et al., 2018). Delirium has devastating consequences that have domino effects for the patient, family, healthcare system, and the population in general. According to LaHue et al. (2019), iatrogenic delirium is significantly associated with hospital readmissions within 30 days of initial discharge with an odds ratio (OR) of 2.60 (95% CI: 1.96-

3.44), as well as post-discharge emergency room visits within 30 days of discharge OR: 2.18 (95% CI: 1.77-2.69). Medicare and Medicaid impose penalties on hospitals with elevated 30-day readmission rates; thus, there is a national effort to decrease readmissions rates (LaHue et al., 2019) thus preventing delirium as part of this effort seems necessary.

The cost of hospitalization for patients with delirium increase by $16,303 to $64,421 per patient (Leslie, Marcantonio, Zhang, Leo-Summers, & Inouye, 2008) resulting in $38 billion to $352 billion annually (Douglas et al., 2013). Less than half of patients fully recover before discharge, which incurs additional costs associated with residential care, rehabilitation, and home services (Wass et al., 2018). The 2050 projections on aging note that 88.5 million people will be over the age of 65, which is more than double that in the year 2010 (U S Census Bureau, 2010). Individuals over 65 are at a higher risk for developing delirium (Douglas et al., 2013) since most have multiple predisposing risk factors (male gender, history of cognitive impairment, renal disease, liver disease, cancer) and in general, are increasingly vulnerable to insults when multiple risk factors are present, or illness is severe (Wass, et al., 2018). Without actions to curb the rates of delirium, incidence and consequences will continue to impact patients and the financial burden will increase as the elderly population explodes.

Manifestations of Delirium

Delirium can affect an individual's ability to rest, wake, converse, and their awareness of surroundings. It alters a patient's ability to reasoning resulting in agitation, hallucinations, or delusions (Ford, 2016). The most frequently observed symptom is moderate to severe inattention most often detected during a physical exam; elicitation of mild inattention could require a formal cognitive test (e.g., digit span, serial sevens, or naming the months in reverse order) (Cerejeira, & Mukaetova-Ladinska, 2011). This syndrome can manifest clinically on the

three domains: cognitive, executive, and circadian rhythm disturbances (Thurber et al., 2015). An individual's disorientation, memory impairment, and inattention demonstrate cognitive effects. Executive functioning deficits show in a patient's inability to complete a task or thought, difficulty with self-regulation of behavior, impaired ability to interact socially, changes in speaking ability and speech patterns, and inability to problem solve. Circadian rhythm disturbances result in disturbed sleep-wake cycles, often reversed into insomnia at night with fatigue or exhaustion during the daylight hours.

Three subtypes can categorize delirium; hyperactive, hypoactive, or mixed with both hyperactive and hypoactive features. Hyperactive types are the most recognized by healthcare providers, as it presents with agitation, emotionally lability, restlessness, sleeplessness, and are potentially combative behaviors (Vasilevskis, Pandharipande, Girard, & Ely, 2010). This subtype often requires increasing nursing interventions with frequent calls to physicians for medical or pharmacological interventions; to manage aggressive, unsafe patient behaviors, and providing safety for the staff caring for these patients. Hyperactive delirium is much more involved concerning behavioral issues, and these patients are likely to need restraints or chemical sedation with medications such as Haloperidol or Lorazepam.

Hypoactive delirium is the most serious of all subtypes; patients characterized by apathy, decreased responsiveness defined as lethargy, unresponsiveness, or coma. These patients are typically older than 75 years of age, have many co-morbidities, and present with greater severity of illness (Cerejeira & Mukaetova-Ladinska, 2011). It is the most under-recognized, under-treated subtype (Vasilevskis, Ely et al., 2010), and is likely to be overlooked and misdiagnosed as either depression or fatigue (Cerejeira & Mukaetova-Ladinska, 2011). Evidence indicates a poorer prognosis in patients with the hypoactive type of delirium; this is perhaps due to the poor

recognition and treatment or the immobility associated with this subtype (Vasilevskis et al., 2010).

Mixed hyperactive/hypoactive delirium is the subtype in which patients alternate between lethargy and hyperactivity. This combination type is unlikely to be the most commonly diagnosed because a patient may be misdiagnosed with "sundowners" during periods of hyperactivity. In a study of the prognostic effects of these motor types, Avelino-Silva, Campora, Curiati, and Jacob-Filho (2018) state, "hypoactive delirium was the predominant motor subtype (53%), followed by mixed delirium (30%) and hyperactive delirium (17%). Hospital mortality rates were respectively 33%, 34%, and 15%" (2018, p. 1). This study also noted that hypoactive delirium had an independent hazard ratio for in-hospital mortality of 2.43 (95%CI =1.64–3.59) and mixed delirium resulted in a hazard ratio for in-hospital mortality of 2.31 (95%CI = 1.53–3.50) (Avelino-Silva et al., 2018, p. 2).

Understanding the manifestations, subtypes, and the implications of them can aid healthcare providers in recognizing its subtle or not so subtle onset. Recognition of delirium improves with the implementation of assessment tools for the onset of delirium in practice (Inouye et al., 1993). These may include tools such as the confusion assessment method (CAM), the nursing delirium screening tool (NuDeSC), the delirium observation screening scale (DOSS).

Pathophysiology

Due to complexity of physiologic responses to illness and injury, the pathophysiology of delirium is poorly understood and rarely researched (MacLullich, Ferguson, Miller, de Rooij, & Cunningham, 2008). MacLullich et al., (2008) proposed two very distinctive classifications of etiologies of delirium. The first are direct brain insults such as trauma, intracranial hemorrhages, cerebral infarcts, hypoxemia, hypercapnia, and hypoglycemia. The second are "aberrant stress

responses" which result from normal protective functions of the human body in response to infections, surgeries, anxiety, and pain (MacLullich et al., 2008). This stress-diathesis model is dominant in literature stating the interaction of predisposing factors and the adequate or inadequate stress response (Newman et al., 2015). Elevated levels of dopamine, impaired acetylcholine synthesis and cholinergic synapses, low levels of norepinephrine, 5-hydroxytryptamine, and y-aminobutyric acid result (Wang, Lyu, Tan, Wang, & Liu, 2017). Wang et al. (2017) also report that the insults driven by the external factors (surgery, trauma, and infection) active the vascular endothelial cells causing destruction of the blood-brain barrier which allows the inflammatory factors to cross into the cerebral tissues, stimulating further release of proinflammatory factors and ultimately resulting in neurotoxicity and delirium. Thus, delirium should not be regarded as a psychological issue, instead it should be prevented and managed comparably to other diseases (MacLullich et al., 2008; Wang et al., 2017).

Predisposing and Precipitating Factors

Recognizing the risk of delirium currently hinges on a clinician's or nurse's ability to effectively assess patients for predisposing and precipitating factors and using clinical judgement decide if a patient is at risk. Based on past experience and education of a clinician or nurse preventative measures may be implemented. Predisposing risk factors are those that describe the vulnerabilities of the individual, which are non-modifiable. Precipitating factors are events that occur with illness or within the healthcare facility, that trigger delirium, which may be modifiable.

According to Mehta et al. (2015), the predisposing factors of significant importance are age, history of cognitive impairment (dementia or a history of delirium), and sensory impairments. The precipitating factors are severity of illness, administration of opioids,

benzodiazepines, antipsychotics, oral analgesics, anticholinergic, and sedatives (Mehta et al., 2015). According to the NICE guideline, the precipitating risk factors are acute infection, fracture, emotional or physical stress, surgery or other medical procedures that include anesthesia, pain, sedatives, hypnotic medications, anticholinergics, anemia or blood loss, dehydration, malnutrition, and electrolyte disturbances (NICE, 2010). Precipitating factors include hospital environments, particularly in the ICU, where frequent disruptions disrupt the circadian rhythm. Over one hundred various triggering events are associated with delirium in the ICU alone (Vasilevskis, Pandharipande et al., 2010). Each source and organization vary in their lists of these risk factors and each risk prediction model vary in the risk factors used and how they are weighted in relative importance. Tables 1-7 show the models with their risk factors including the modified tools by Pendlebury et al. in their model update study (2016a).

Groves and Huskin (2011) believe that baseline risk is a predictor of delirium likelihood, and when a patient has a low baseline risk, despite triggering events, their likelihood of becoming delirious is low. However, if vulnerabilities are high and baseline risk is High (many or complex comorbidities or severity of illness), then delirium may occur even with the most innocuous of insults.

The consensus is still out as researchers continue to evaluate the relationships between the predisposing and precipitating factors and other newer findings such as biochemical, environmental, and genetic factors. In all accounts, delirium is a result of the complex interactions of both predisposing and precipitating factors (Inouye, 2018; Rudolph et al., 2011) as well as the individual's status of baseline protective mechanisms (Groves & Huskin, 2011).

Risk Stratification or Prediction Models

Risk stratification is a technique for systematically categorizing patient risk levels based on their health status, predisposing, and precipitating factors. Clinical risk prediction models (also known as risk prediction rules, stratification tools or models, risk assessment models, risk decision rules, risk scores, risk assessment tools, and indexes) are tools that can determine the probability of an event occurring. Healthcare risk prediction or stratification models are essential for optimizing of healthcare research, quality improvement, and clinical decision-making (Bernard, 2017).

Assessing risk is not a new concept to healthcare. It is becoming more common as healthcare resources are increasingly limited. Implementation of risk assessment tools are successful in the stratification of other complex medical situations such as the risk of falls (Hendrich II), pressure sores (Braden scale), stroke risk in patients with atrial fibrillation (CHA_2DS_2-VASc), osteoporosis fracture risk assessment tool (FRAX), and the 10-year heart disease risk score (atherosclerotic cardiovascular disease (ASCVD)). The National Institute of Health (NIH) and Patient-Centered Outcomes Research initiatives (PCORI) support risk prediction tools used to personalize an individuals' healthcare needs in a personalized dynamic manner (Columbia University Mailman School of Public Health, 2019). Risk stratification allows for customized medicine with an emphasis on specific treatments (Agyeman & Ofori-Asenso, 2015) before any symptoms occur to diminish the risk of illness. Delirium biomarker research is a growing and expanding area of study. Presently there is no biomarker test to determine the presence or the risk for delirium. Some of the previously mentioned risk prediction models incorporate biomarkers or medical technology in their scores. The lack of this technology

for the diagnosis of delirium makes its prediction more complicated than the prediction of other conditions.

Organizations proposing ways to identify people at risk for delirium include the American Nurses Association (acronym) (American Nurses Association [ANA], 2016), the HELP (lists of risk factors) (2013), HRET (2018), and the NICE Guidelines (2010) for the prevention of delirium (lists of risk factors). These organizations have websites that are integral for clinical practice for updates on delirium, including its prevention and management (Table 8).

Due to the high mortality, morbidity, cost of care, and impact on patients that develop delirium and the health systems that care for them, evaluation of existing DRPMs must be done. Researchers continue to create models, but what is lacking are studies on impact, implementation, and quality improvement. Nurses are a key to the assessment and appraisal of these models because they are at the bedside caring for the delirious patients. They aid in the implementation of risk assessments and the interventions aimed to prevent this syndrome. Nurses additionally are involved in quality improvement projects. Guidelines state healthcare providers need to assess for risk but do not currently recommend a DRPM or any other systematic method to stratify risk.

Purpose

The purpose of this integrative literature review is to identify and critique the research on delirium risk prediction models for adults admitted to general medical units. Recommendations for assessing delirium risk in clinical practice are outlined.

Clinical Nursing Questions

The following clinical nursing questions were developed to guide the literature review. Do the validated primary delirium risk prediction model studies: (a) support their claim of feasibly in practice, (b) show accuracy and (c) ability to stratify the risk of delirium development in the general medical hospital population?

Method of Inquiry

A systematic literature search was completed using the methods outlined by the Winona State University Library (n.d.). It is a five-step approach to provide the seeker with relevant articles that will provide information to aid with clinical practice or research. A five-step method provides a simplistic approach for structuring the inquiry and gathering of the literature. The five steps are:

- Define the project by refining a question and brainstorm related ideas to develop a list of key terms
- Complete a preliminary search
- Refine the focus and look for quality articles
- Arrange the ideas in groupings and synthesize the literature
- Write chapters or sections, placing the literature within each, noting the gaps (Winona State University Library, n.d.).

Literature Review

The following sections are a comprehensive review of the literature including a synthesis and analysis of the research findings. Current practices in delirium risk prediction, risk factors for each model, statistical significance of each of the eight delirium risk prediction models, their capacity to stratify those at high risk accurately, and the ease of use of each model will be

discussed. The studies were included based on their level of evidence, as defined by Ackley, Swan, Ladwig, and Tucker (2008). The studies were appraised by use of the Newcastle Ottawa Scale for Quality Assessment which aided in determining the quality of study (Table 9). The levels of evidence description are located in table 10 and a table 11 was created to show the level of evidence for each DRPM included in this review. The Stetler model of research utilization (2001) is explained as a model for completion of this literature review. To highlight and clarify the contents of each DRPM and how they were scored (tables 1-7).

Search Strategy

To identify the primary studies of DRPMs using clinical data to predict iatrogenic delirium an extensive search was completed. Guided by the Methodological recommendations described by the Winona State University Library, the clinical questions guided the search. Cumulative index to nursing and allied health literature (CIHNAL), PubMed, PsychINFO, and the Cochrane library databases were searched using variations in keywords “delirium” or “acute confusional state” with Boolean connections to “rates”, “treatments”, “differential diagnosis”, “workup”, “predictors”, “pathophysiology”, “causation”, “cost”, “impact on families”, and/or “impact on patients”. These articles added background knowledge and support to this project and are referenced throughout the paper.

A second exploration was then completed within the same databases to narrow the discovery to stratification models which included multiple combinations of the following terms; “delirium” or “acute confusional state” or “acute confusion” or “acute brain dysfunction” and “risk assessment” or “tools” or “models” or “predict*” or “risk screen*” or “risk stratification” or “risk prediction” and “acute care”, “general medicine” or “medical” or “hospital admission” or

"hospital*". Exclusions included "alcohol" or "drugs" or "withdrawal" or "ICU" or "fracture" or "surgery".

Further addition of the subject matter was obtained by a third search focused to answer the questions in this review by using 'One Search' from the Winona State University Library. This search strategy allowed for an expansive search of many databases in less time since it pulls together all the library's resources into one single search. This search provided most of the articles included in this review. The search was limited to peer-reviewed literature; articles, dissertations, Journals; the English language, publish dates of 2008-2018 (to ensure the most up to date evidence was gathered), adults, older people, geriatrics, aged-medicine, hospitals, and medicine with the main subject of delirium. Titles of articles were reviewed for the keywords delirium or acute confusional state (required) and risk assessment, risk tools, risk scores, risk models, or risk stratification. Next the abstracts for the articles of interest were read, further narrowing the studies to fit the clinical questions. After reviewing the abstracts, studies of interest were saved for a full review. When a full-text article was not available online it was requested through the Winona State University library.

A review of full articles revealed further narrowing of the subject was needed to focus on general medical admissions because there were many models created for specific areas of the hospital or clinical condition such as post-cardiac surgery or orthopedic surgery, or ischemic/hemorrhagic stroke focus. Tools that predicted only prevalent delirium were disregarded, keeping the focus on obtaining studies on DRPMs created to determine the risk of incident or iatrogenic delirium (occurring after hospitalization).

Search for Recommendations of Statistical Cut-off Points

Information was sought, to aid in determining ideal AUC cut-off points for the accuracy of the risk stratification models. The literature showed that risk stratification models differ from diagnostic models in that the accuracy in diagnosing is not the goal of prediction modeling. Instead, they are created to rule out low risk people and limit expended resources by triaging people by the level of risk. An accurate prediction model may limit the resources to 50% of the population or less without false negatives. The most popular statistics of a predictive model is the receiver operating characteristic (ROC) curve which is a plot of sensitivity (Se) and specificity (Sp) (Cook, 2008). The ROC or area under the curve (AUC) can also be called a *c*-statistic or *c* index. According to Lee, Bang, and Kim (2016), a model with an AUC of 0.50-0.59 has poor discrimination power, an AUC of 0.60-0.69 has better than average power to discriminate, an AUC of 0.70-0.80 has adequate power of discrimination, and 0.80-0.90 is excellent (Lee, Bang, & Kim, 2016).

The following statistical levels are employed based on the previous recommendations and the evidence in the statistical models: a poor performance rating occurs when the AUC, ROC, or *c*-statistic are between 0.50 and 0.59 (as this indicates the model is stratifying risk lightly better than a coin toss), moderate performance is between 0.60-0.75, and excellent performance is between 0.76-1.0. Studies of poor performance (≤0.59) were not included in this review as they lack statistical significance. The range of AUC in this review is 0.69-0.85, thus all are proven to have moderate or better performance.

A total of fourteen models were discovered and are presented in the literature review table (table 12). During the process of the literature review seven of the studies were not validated internally or externally. According to the literature on risk prediction models

validation of a model equates to proof of its applicability in a clinical setting. Validation of a model can be internal (use of the same data set for development and validation of the model), temporal (model validation on subsequent patients from the same facility), or external validation (model validation in a different facility with a similar population). Validation of a study proves or disproves its clinical credibility, accuracy, generalizability, and preferably shows clinical effectiveness (Altman et al., 2009). The unvalidated studies are relevant to the future of delirium risk prediction, thus were retained within the literature review tables (Table 12). The seven models critiqued are marked by an asterisks before the first authors name in the literature review tables.

Appraisal and Synthesis of Evidence

The highest levels of evidence attainable are randomized controlled trials (RCT) or meta-analysis to answer clinical questions related to the primary prevention of illness. Next in order of the hierarchy are prospective studies, cohort studies, case-control studies, and case series (Ackley, et al. 2008). The level V studies include the highest level of evidence for systematic reviews of qualitative studies such as a meta-synthesis or an RCT (Ackley et al., 2008). Three studies, that are rated as level V evidence, were found and included in this review which are: *Models for predicting incident delirium in hospitalized older adults: A systematic review* by Kalimisetty, S., Wajih, Fay, & Khan, (2017); *Systematic review of prediction models for delirium in the older adult inpatient* by Lindroth, et al., (2018); and *Predicting delirium: a review of risk-stratification models* by Newman, O'Dwyer, & Rosenthal, (2015). One systematic review of delirium risk prediction models focused on understanding the barriers to implementation (Newman, O'Dwyer, & Rosenthal, 2015). The aim of the systematic review by Kalimisetty et al., was to develop a risk prediction model based on the reported risk factors in

current models (Kalimisetty, Wajih, Fay, and Khan, 2017). A third study aimed to recommend the study design for future development of delirium risk prediction models (Lindroth et al., 2018). Each of these studies contributed to background knowledge and aided in the identification of understanding of DRPMs.

The seven validated DRPM studies selected for this review were representative of level IV studies. The Kobayashi et al., 2013 and Wong et al, 2018 studies being retrospective cohort designs, and the other five are prospective cohort designs. Five of the studies were completed in individual university healthcare centers, one was conducted in two locations of a university hospital and validated at a Veterans Administration (VA) medical center. The final one was conducted at 118 VA system hospitals. The least number of participants in the studies was 100 patients and the most was 27,625 (Wong et al., 2018) with the median participant number at 308. All studies lacked randomization and control of variables. The definition of level of evidence can again be viewed in Table 11.

Appraisal and quality assessment cannot be completed with the use of tools created for systematic reviews, diagnostic studies, prognostic studies, RTCs, or qualitative studies. Discovery of the Newcastle Ottawa scale (NOS) allowed for the evaluation of the prediction models study quality. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS Checklist) provided a method to appraise these primary prediction model studies.

Newcastle-Ottawa Scale (NOS)

The quality of cohort and case-control studies can be assessed using the Newcastle-Ottawa Scale (NOS). It was developed to provide an instrument of ease for evaluating non-randomized cohort or case control studies to be used in systematic reviews (Peterson, et al.,

2011). There are not specific quality instruments for evaluation of prediction studies. Criteria included in this tool are selection of the study groups, comparability of the groups; and the outcome of interest for the studies and are rated based on a star system. The NOS aided assessing the quality of the studies and the creation of a comparative study table to outline relevant data from each risk assessment tool (Table 10). Newman et al. (2015) used the NOS in their systematic review noting two of the scale assessments were irrelevant to the DRPM studies. In the selection section it asks if the case definition is adequate, defined by the study being independently validated in a similar population, however, significant baseline differences are a study strength in prediction model validation, as this shows generalizability of a prediction model (Newman et al., 2015). There are no interventions employed in the research papers thus follow up of patients was not needed, therefore, 'adequate follow up after an intervention was employed' was also removed from the scoring. A quality rating of 7 was used instead of 9 for scoring.

CHARMS Checklist

The CHARMS checklist is designed to aid in reviewing and appraising of all types of primary prediction modelling studies (Moons et al., 2014). Additionally, it aids in data extraction of the individual studies in eleven domains: source of data (type of study), participants, predicted outcomes, candidate predictors, sample size, missing data, development, performance, evaluation, results, and interpretation/discussion. (Moons et al., 2014). Prior to this checklist there were no systematic methods for evaluating primary prediction modeling studies (Moon et al., 2014). Though this tool is intended for primary studies it was additionally used here to aid in data extraction from the systematic reviews noted in this paper. The systemic reviews included were used as guides to understand the primary model research and report the relevant themes

within this literature review. Understanding the currently published research is an important aspect of clarifying the gaps in research. The CHARMS checklist allowed a format to display the relevant study data (Table 13) found in the three systematic reviews since they are lightly discussed in this paper and highly relevant to future research or implementation of DRPMs.

Themes

It is evident in the literature that delirium results in detrimental effects for the healthcare systems, patients, and families. The rates and healthcare costs of delirium are unchanging with current practices in delirium risk prediction, prevention, and management and are expected to increase as the percentage of elderly population increases. The literature uncovers a lack of systematic methods to determine risk in current practice. The studies all report similar statistics and report their ability to stratify risk based on statistics. Lastly, model ease of use in clinical practice is a past barrier that must be addressed in the development of DRPMs.

The Impact on Health Systems, Patients, Families and Nursing Staff

The impact of delirium is far reaching for healthcare systems as the financial burden and resource utilization creates a strain on our health care systems. Implementation of delirium care pathways in acute care can decrease rates of onset, severity, and length of delirium in hospitalized patients (Inouye et al., 1999) and reduce overall costs of hospitalization (Brown et al., 2018). Increased length of hospitalization can lead to adverse outcomes, particularly in the elderly due to increased frailty (Vincent, Neale, & Woloshynowych, 2001). Falls, decubitus ulcers, feeding problems, urinary incontinence, urinary infections, and fractures are reported in literature to be a result of extended hospitalization (Groves & Huskin, 2011). The length of stay is a way organizational performance is measured. Preventing the onset of delirium or lessening its effects, by the recognition of risk and implementation of preventative measures, allows

healthcare providers to affect the length of stay and improve organizational performance measures.

Whitehorne, Gaudine, and Meadus (2015) reported that patients post-ICU delirium had poor or no recall of events, difficulty making connections and communicating simple needs, struggled with distinguishing reality from hallucinations and paranoia, and recalled feelings of being in imminent danger. Delirium creates a state of personal distress. The hallucinations are detrimental to the long-term mental health of a patient who believes they are real while recalling the experience (Whitehorne, et al. 2015). The following is a recollection of a recurrent hallucination experienced by a patient who recovered from delirium. It shows that patients experience long-term psychological trauma.

> The one that was most upsetting was the monkeys … up in the lights...You could hear them jumping up and down, and they were bawling like they were trying to get at me. They were on all the lights, not just at the one that was at my bed but all around the room...They were savages...I didn't know…if they wanted to get out or get at me.… I'm still afraid to look up at the lights...And I always…whisper because I'm afraid they'll hear me. (Whitehorne et al., 2015, p. 477)

The cognitive changes resulting from delirium may necessitate 24-hour caregivers, causing financial strain or long-term care center placement for safety. The consequences of cognition impairment are a loss of independence, acceleration toward dementia, and early mortality. The effects of delirium can persist for months in 20% of the cases (MacLullich & Hall, 2011).

Caregivers offer support and provide care to family members with persistent delirium whom are discharging from hospitals. Its noted that 80% of delirious hospitalized patients have

family members that are their caregivers, creating a burden of care (Abrantes & Racine, 2019). Three types of burdens laid on family members: symptom burden, emotional burden, and situational burden (Abrantes & Racine, 2019). Symptom burden is the experience and observation of the disorientation and personality changes that make the patient virtually unrecognizable. Greater than 70% of spouses reported stress related to the toll of caregiving and close family (Page & Ely, 2017). Friends and families worry that the cognitive changes are permanent (Page & Ely, 2017). Situational burden occurs as the result of feeling loss of control, safety concerns, lack of support, and the unpredictability of the course of delirium (Abrantes & Racine, 2019).

Page and Ely (2017) state nursing staff caring for delirious patients experience frustration and stress while trying to care for and comfort them (p.107). Nursing challenges occur whilst providing basic care to ensuring safety, protecting patients from causing harm to surgical sites, removing IV lines, Foley catheters, arterial and central venous lines; all while providing nursing care for the presenting illness.

Current practices for Risk Determination

In the current state of delirium management, healthcare providers individually use clinical judgment and their learned knowledge of delirium to assess a patient for delirium risk. A clinician must recognize and be cognizant of the predisposing and precipitating factors that trigger incident delirium. Delirium prevention bundles, clinical practice guidelines, organizational guidelines, and hospital-specific pathways list various risk factors, however, do not recommend specific stratification tools.

In ICU's across the United States, standard practice is to implement the 'ABCDEF bundle' to prevent delirium in the ICU, per the recommendation of the Society of Critical Care

Medicine (2018). This "bundle" provides nurses with interventions to reduce delirium, provide adequate pain management, and reduce the long-term consequences of delirium in adult intensive care unit (ICU) patients (Society for Critical Care Medicine [SCCM], 2018). This bundle called the **A**ssess, prevent, and manage pain; **B**oth spontaneous awakening and breathing trials; **C**hoice of analgesia and Sedation; **D**elirium assess, prevent, and manage; **E**arly mobility and exercise; **F**amily engagement /empowerment, is implemented for every ICU patient, since all ICU patients are considered at high risk for delirium and poor clinical outcomes. A recent study by Pun et al. (2019) found that over 15,000 patients receiving the ABCDEF bundle interventions had decreased rates of incident delirium, restraint use, intubation length, mortality rate, readmission to ICU, and were more likely to discharge to home rather than long-term care centers. This study also noted a decrease in the length of mechanical ventilation, coma, and delirium (Pun et al., 2019). Due to the positive effects of the ABCDEF bundle, the SCCM recommends all ICU's to employ this delirium bundle as the rates of delirium can be diminished and healthcare quality can be improved (SCCM, 2018).

Bundles of care for the inpatient units are less frequently employed; thus, patients continue to develop delirium at uncontrolled rates on medical units. The Hospital Elder Life Program is a research-based comprehensive program for the prevention of delirium in hospitalized older adults (HELP, 2019). This program utilizes hundreds of volunteers to assist with activities to keep patients alert and awake during daytime hours, meet hydration and nutrition needs (feedings), and encourage movement. According to their website, there are 200 sites (in 32 states and 11 countries) that have employed this program, including Methodist Hospital in Minnesota (HELP, 2019). The study by Zaubler et al. (2013) showed that implementation of the HELP program resulted in a 40% relative reduction in episodes of

delirium, a two day decrease in the length of stay, and a nine-month cost-savings of $841,000.00 (Zaubler et al., 2013). Additionally, the HELP website also notes that HELP implementation studies have resulted in the prevention of cognitive decline, reduced nursing home placement, decreased hospital rates of falls, and a reduction in the use of 1:1 sitters (HELP, 2019).

Brown et al. (2018) published an impact and implementation study on the AWOL tool in practice as part of a multicomponent prevention pathway (AWOL, CAM, and interventions to those triaged as a score of ≥ 2). This study reported a decrease in length of stay of >2 days, decreased 30 day re-admission rates from 11% to 5.45%, less restraint and 1:1 sitter use (Brown et al., 2018). There were no reductions in hospital days which were explained by stating that the interventions employed were studied in general medicine populations, but the impact study had a majority of neurological patients and increased recognition and sensitivity to delirium in general may elevate diagnosis of delirium.

Supporting Evidence for Risk Stratification

Delirium risk determination and prevention strategies are recommended by organizations such as NICE Delirium Guideline, the Iatrogenic Delirium Change Package, Delirium: Guidelines for General Hospitals, and the Delirium Prevention Strategies by the ANA (2016). Each of these documents list the risk factors but not one of the guidelines describe a method to adopt for a consistent, system-wide assessment for risk stratification. Clinical practice guidelines were reviewed in an attempt to clarify current practice to determine risk and not to review them for the purpose of interventions and application to practice, thus they are not critically appraised in this literature review.

The iatrogenic delirium change package was the only guideline to state the following ideas for practice change:

- Adapt and adopt a risk assessment tool that examines the following risk factors: age, dementia, metabolic imbalance, hypertension, alcohol abuse, severity of illness, coma and benzodiazepine administration.
- Assess the risk for delirium upon hospital admission, transfer within hospital or change in patient behavior.
- Develop prompts to promote the completion of the assessment and include the assessment on the admission checklist or in charge nurse rounds (HRET, 2018, p. 6).

The American Nurses Association (2016) developed a delirium prevention pathway that again states that the recognition of risk is the first step. The pneumonic, MIND SPACES, was created to aid healthcare providers to recall the predisposing and precipitating factors. Since this technique has not been researched or reported in quality improvement studies, it is difficult to determine its effectiveness and assumes staff will recall this acronym. As with previous guidelines, the ANA did not create this to be a DRPM, but simply a list of risk factors (ANA, 2016). Of note, NICE Guidelines state that all adults whom are at risk for delirium and are newly admitted to a hospital or long-term care center should receive a range of tailored interventions to prevent delirium (NICE, 2010).

As evidenced by the steady rates of delirium, the current methods of determining risk for delirium are failing the at-risk population. Healthcare providers fail to recognize both the risk of delirium and the onset of delirium all together; thus, clinical judgment alone is not enough to change the trajectory and prevent delirium from occurring. Implementing DRPM's into practice may improve recognition of delirium and diagnosis rates of delirium, as well as prevent its onset by providing time for interventions to be applied, thus preventing modifiable triggering events

from occurring (Douglas et al., 2013). Implementation of preventative interventions in practice are recommended by the SCCM, ANA (2016), and clinical practice guidelines. Within the next sections, the reader will be introduced to the statistics of risk prediction models to obtain a baseline understanding of the accuracy of the DRPMs presented in this review.

Delirium Risk Prediction Models: The Statistics

In the following paragraphs studies are compared based on their ability to accurately predict and stratify the risk of developing incident delirium. Each study followed a similar path for determining the independent predictive factors. They collected baseline characteristics such as demographics, living situation, age, sex, comorbidities, cognitive status, and varied in their collection of baseline lab values, medications, vital signs, infection, fracture, and admitting diagnosis. One study included dependency with activities of daily living, presence of urinary catheters, IV therapy, oxygen, and pressure sores as characteristics (Martinez et al., 2012).

A total of seven models were selected for appraisal and comparison for this literature review. They are:

- 'Clinical Prediction Rule for Delirium' by Martinez et al. (2012);
- Chi-Square Automatic Interaction Detector (CHAID) decision tree model by Kobayashi et al. (2013);
- 'AWOL' by Douglas et al. (2013)
- Delirium Prediction Score (DPS) by Carrasco et al. (2014);
- 'E-NICE' by Rudolph et al. (2016)
- 'Delirium Susceptibility Score (DSS)' by Pendlebury et al., (2016a);
- Automated machine learning tools (Gradient Boosting Machine (GBM) by Wong et al., (2018).

An additional study updating four of the models in preparation for a new model (DSS) was included in the review:

- Delirium risk stratification in consecutive unselected admissions to acute medicine: validation of externally derived risk scores (Pendlebury et al. (2016a)

An implementation study completed by Brown et al. offers significant clinical support of DRPM implementation into practice:

- Predicting inpatient delirium: The AWOL delirium risk-stratification score in clinical practice (Brown et al., 2017)

Another study included shows evidence in favor of DRPMs in practice is an implementation and clinical impact study:

- Evaluation of a multicomponent pathway to address inpatient delirium on a neurosciences ward (Brown et al., 2018)

For statistical comparison, all seven prediction model studies and the additional three supporting studies reported the area under the curve (AUC) statistics as either the area under the receiver operating curve (AUROC), AUC, receiver operating curve (ROC), or the concordance statistic (c-statistic or c-Index), thus enabling ease of comparison of discrimination. A model's ability to differentiate between those at high risk and those at low risk is its ability to discriminate. Attention should focus on the sensitivity (true positive rate) more than the specificity (true negative rate) when choosing tools to predict delirium risk, allowing for stratification by including nearly all that developed delirium (Ho, et al., 2019). The calibration of a model determines if the observed risk matches the predicted risk. Thus, both calibration and discrimination are essential to prediction modeling. Calibration is often not reported in

prediction models; in these models, calibration results as percentages of positive delirium patients per risk level (Table 14).

The models included in this review are those with moderate to good performance range like the AWOL risk score developed by Douglas et al. (2013), which reported an AUC of 0.69 in the development study, an AUC of 0.73 in the Pendlebury et al. (2016a) update study, and a AUC 0.69 in the comparison study by Brown et al. (2017). Wong et al. (2018) compared the AWOL to newly developed electronic DRPM's. The AWOL models' discrimination in the external validation cohort resulted in an AUC of 0.678 (showing consistency of the model from the original development by Douglas et al., 2013). The logistic regression model developed in the Kobayashi et al. 2013 study, was among the best performing of the non-electronic group with an AUC of 0.79 in the validation cohort. Interestingly, the validation cohort performed slightly better than the development cohort that had an AUC of 0.78, which can suggest over-fitting of the model. Over-fitting of the model to the sample population is possible when the AUC in a validation study is higher than that in the development cohort and means that the model was fitted to the validation population and thus may not be generalizable to other populations without further external validation studies to prove otherwise.

The future of risk prediction lies in machine learning or neural networking. The advent of the electronic medical record presents an interesting possibility for future prediction modeling as they record relevant patient information. These predictors are weighted (weights are commonly derived by logistic regression), which results in a score, then these scores are fitted into the models' predetermined levels. These models rely on accurate chart documentation, including complete and accurate history, diagnostic coding, and results of labs. The electronic versions included in this study are by Kobayashi et al. (2013) (CHAID decision-tree model),

Rudolph et al. (2011) (e-NICE), and the Wong et al. (2018) (machine learning models; "gradient boosting machine (GBM)", "Penalized logistic regression (PLR)" and "random forest" analysis (RF)). The CHAID had the lowest AUC in the validation cohort at 0.82. Wong's RF model had an AUC of 0.848, the PLR model is 0.854, and the GBM had an AUC at 0.855. The e-NICE model was the highest with an AUC of 0.91 (Brown et al., 2017).

The non-electronic high performer was Martinez et al. (2012), as the validation cohort had an AUC of 0.85, keeping in mind that the development cohorts AUC was 0.77 (considering overfitting of the model). The authors noted that that additional work to rule out overfitting is not needed as it was of no clinical significance. The validation study had the same patient population and setting with a higher incidence in delirium diagnosis (25% vs. 13%) in the development cohort and more dependence on others for assistance with ADL's. A second non-electronic model with high discrimination was found in the external validation study that Pendlebury et al. (2016a) completed, the model created by Isfandiaty found an AUC of 0.83. (validation of this developed tool was not completed; therefore, it is not included in the literature review). Statistical comparisons of the seven studies, along with the model update study by Pendlebury et al. (2016a) are entered into a table 14.

The capacity to stratify risk. The ability of a model to stratify risk is evaluated based on the accuracy of the proportion of each population's allocation into risk levels (Steyerberg, et al. 2010). The best models will effectively place subjects at both extremes of the risk distribution, thus enabling clear implications for future actions. Perfect models assign into only the highest risk and the lowest risk levels with no in-between, leaving no room for error of missing an event. Cook notes that there are no perfectly calibrated models (2008). A useless model will assign the same risk to the entire population, similar to the flip of a coin, which is equivalent to an AUC of

0.50 (Steyerburg et al., 2010). A model's calibration, or capacity to stratify the population into risk categories, and the accuracy of the classifications are the critical attributes of a model.

For comparative purposes, healthcare providers can review the true positives and negatives, false positives and negatives, as well as percentages of definite diagnosis, and ensure that the rates of delirium are increasing with the higher risk categories. Martinez et al. (2012), Douglas et al. (2013), Kobayashi et al. (2013), and Rudolph et al. (2016) models compare each risk level based on percentages of those with and without delirium in each level. Martinez et al. (2012) has the highest percentage of patients at their highest level; 64% of those with delirium are included in this level. Of note 44.4% of those with a score of ≥1 (highest stratification score) developed delirium, and only 7% of those classified as low risk developed delirium. The Martinez et al. model also has the potential to limit necessary interventions to 53% of the total population, making rationing of interventions possible, which is the goal of stratification of risk.

The Douglas et al., (2013), AWOL score sets a score of ≥ 2 as high risks. Eleven patients in this cohort are positive for delirium or 13.5% of the 165 patients included, of importance is that only 3.5% of all those said to be low risk developed delirium showing good calibration. The score of ≥ 2 captured 11 positives out of fourteen, which results in 79% accurately stratified. At this score, interventions would be limited to only 49% of the population allowing for an improved resource utilization. Interestingly, at zero factors one still developed delirium and a score of four, none developed delirium. The sample size of this study was just 165 patients, which could account for less reliable results. The Pendlebury et al. study (2016a) updated the AWOL model and improved the capacity to stratify risk as evidenced by an improved AUC of 0.78 and the Se still increases with each risk level and PPV increasing to 0.70 for three factors.

Comparatively, the CHAID decision-tree model, is unique in its application and statistical reasoning (Kobyashi, 2013). It can be used both electronically and on paper as an algorithm as demonstrated in Table 4. There is no cutoff score for risk stratification; instead, it is a model that identifies the presence of risk factors and follows a decision-tree to determine risk levels. Those noted to be moderate to highest risk levels are of two categories. The first split is those with a known history of delirium. The study notes that those over age 75 have an increased risk and account for 7.9% of the incidence of delirium. This decision-tree does not give the compounding risk of a patient whose age is ≥ 75 with malignancy and impaired ADL's thus, the reader presumes very high risk.

In the Rudolph et al. (2016) validation study (e-NICE), the rates of incident delirium increase significantly with increasing risk scores. With the addition of the Modified Richmond Agitation Sedation Scale (mRASS), there is an increased ability to stratify the high-risk category. In this model, risk levels that were high and very high-risk combine capturing 27 positive patients out of 246 total patients, but misses 16 of those delirious or 6.5%, which is unacceptably elevated since the goal is to prevent the highest number of those who developed incident delirium. The e-NICE model performs better in the developmental retrospective cohort than the prospective cohort as the development cohort has lower percentages of missed delirium in those categorized ≤5. For scores ≥6, the true positive rate (TPR) is 63% (27/43), and the false positive rate (FPR) is 33% (60/182), showing that more patients were correctly classified as a high risk rather than falsely classified. Eighty-seven patients were high risk out of 246, therefore only 35% (81 patients) of the total population would require interventional pathways.

The researchers in the consolidated e-NICE, Rudolph et al. (2016) model, offered additional mental status assessments for possible addition, such as the mRASS. Inclusion of the

mRASS would increase the number to treat to 108 with 40.7% of them delirium positive. If the score decreased to ≥3, it would capture 33 more positively delirious patients and increasing treatment to 52.4% of the total population, still allowing for triaging, thus the allocation of resources.

The Pendlebury et al. (2016a), Pendlebury et al. (2016b) and Carrasco et al. (2014) studies display the Se, Sp, PPV, and NPV of each level of prediction. In the Pendlebury (2016a) study, the models compared and updated were Inouye et al., 1999, Martinez et al., 2012, Rudolph et al., 2011, and Douglas et al., 2013. The updated tools contained very similar predictors while their Se, Sp, PPV, and NPV also resulted similarly. This study reported that no model statistically performs significantly better than another. As a result of this update study, researchers developed a Delirium Screening Scale (DSS) model which improved on the Se, Sp, PPV, and NPV as evidenced by an AUC of 0.81 (Pendlebury et al., 2016b). This is one of the few studies that reports odds ratio which for a score of 5-7 vs a score of ≤2 is 25 with a relative risk of 13 proving that the higher score is related to greater risk. The odds ratios for all the scores for this model and other models are noted in Table 14.

The Carrasco et al. (2014) study is unique with the use of a mathematical equation formulated out of lab values (BUN/Creatinine ratio) and the Barthel Index. Any score > -240 predicts high risk, and interestingly, 99% of those that are low risk (<-240) did not develop delirium. This tool was exceptional at stratifying those at high risk for delirium; however, the Barthel Index adds complexity to an assessment, and it may have less merit clinically.

The Wong et al. (2018) study is unique as it compares and externally validates the AWOL tool while it integrates hundreds of predictors (796 variables and the GBM contains 345 variables) into three machine e-learning models resulting in very high predictive values and

AUCs. All of their models use chart abstraction methods to calculate and stratify the risk score. The strength of this model is the ability to adjust the sensitivity and specificity to the desired level of the individual facilities. The statistical power this tool has allows for improved accuracy, which results in a narrowed number of patients targeted for interventions as evidenced by the number needed to screen of 4.8.

The DRPM's must have the ability to stratify risk levels appropriately, allowing for allocation of the preventative interventions for a targeted population. Outcomes of clinical application studies, such as Browns' (2017), are among the best clinical evidence to support or negate the use of a model in clinical practice (Brown et al., 2017). The statistics are represented in Table 14 which report the values as they relate to the predictive power of each stratified risk level, from which critical appraising and evaluations for clinical application can consider all scenarios.

Model Feasibility in Practice

To understand model feasibility in practice it is necessary to investigate the barriers to adopting DRPMs in current practice. Three studies were discovered that reported the barriers to clinical practice through the perception of a physician or a nurse. The Newman et al (2015) and Kappen et al. (2016) studies reported on physician perceptions and the Brown et al. (2015) study reported on the nurses perceptions. The systematic review by Newman et al. (2015) reported the barriers to clinical implementation of DRPMs. One obstacle is that health care healthcare providers perceive their use as complicated and time-consuming (Newman et al., 2015). Some of the variables included in the models were not available or tested upon admission (Newman et al., 2015). The overall theme in this review of studies was that the complexity of the predictors limited the use of prediction models.

Since the study by Newman et al. (2015), there was a study reporting the implementation and impact of a DRPM on a medical ward as part of a nurse-driven delirium care pathway by Brown et al. (2015). Nurses are responsible for completing the AWOL tool; unfortunately, the reported completion rate was only 48.6%. The researchers, Brown et al., then followed up with nursing in regard to what were the barriers to completion. The obstacles published are: (a) Lack of nurses time to complete (b) perceived lack of training, (c) the documentation was not required, (d) nurses were frequently disrupted in their workflow, and (e) nurses stated that it wasn't a unit priority (2017). The research team addressed all barriers and found an improved completion rate of 90%; thus, they suggest investing in more resources before and during implementation, supporting the use of the model (Brown et al., 2015). Additionally, Brown et al. found that the AWOL score could not be completed in patients who had a language barrier (somnolence, aphasia, or a non-English primary foreign language). They note modifications of this tool include translation and alternative assessments for cognition in aphasic patients similar to those used in intubated patients (Brown et al., 2017).

Kappen et al. (2016) studied physician perceived barriers to implementing a risk prediction model on postoperative nausea and vomiting. They noted that physicians state the outcome is not the main area of attention, their decision-making process is intuitive rather than analytical, knowledge of the risk level should be accompanied by corresponding management recommendations (knowledge of risk itself is insufficient), and prediction models do not weight benefits and harm of the interventions (Kappen et al., 2016).

To combat these barriers actionable interventions based on risk need to accompany the risk model, risk stratification should be automated into the workflow, reasoning explained with evidence to back them including how the risk is determined, and relevance to the physicians

direct practice will result in improved perception (Kappen et al., 2016). Knowledge of the barriers to use of risk prediction models will aid in creation of structured implementation of them in practice.

There are two DRPMs currently utilized in practice, the AWOL model and e-NICE (Rudolph et al., 2016; Douglas et al., 2013). Bedside RNs complete the AWOL upon patient admission and the e-NICE is a completely automated tool alerting healthcare providers to the risk level. The e-NICE electronic abstraction tool provides the Veterans Administration (VA) hospital system healthcare providers with a daily list of those inpatients at the highest risk. Of note, the AWOL study is researched in three additional reviews: Brown et al., 2017; Pendlebury et al., 2016a; and Wong et al., 2018. The Brown et al. (2017) study they found the AWOL has AUC of 0.73 with 4% of the delirious with a score of 0, 6% with a score of 1, 42% with a score of 2, and 57% with a score of 3, and none of the incident delirious patients in this cohort had a score of 4. In the Brown et al. (2017) study, they had an AUC of 0.73, the Pendlebury et al. (2016a) study resulted in an AUC of 0.78, and the Wong et al. (2018) found an AUC 0.678. The AWOL tool is available on the software application MdCalc (2017), improving its ease of use and availability.

Barriers created by DRPMs prevent the implementation of these systematic methods to predict risk in clinical practice. The more recent DRPMs researchers have answered this with risk prediction models that include predictors that are available upon admission. Models that include complex assessments such as the Barthel index or a mRASS are being rejected by healthcare providers because they are too time consuming. Automated risk scores may be the answer to the feasibility concerns as they allow a hands off assessment of risk, with predictors drawn from the electronic medical record. The models that are automated are the Wong et al.,

2018 and Rudolph et al., 2016. Unfortunately, some healthcare systems are not ready for automation. The AWOL (Douglas et al., 2013) contains predictors to be assessed upon admission and an online calculator for scoring on MdCalc. The DSS model (Pendlebury et al., 2016b) contains predictors that are readily available upon admission.

The ideal time to assess for delirium risk is upon admission before delirium develops, and at a point that preventative interventions implemented are effective (Douglas et al., 2013; Wong et al., 2018). Thus, the ideal model contains predictors commonly obtained on admission or readily available in the chart. The following studies by Douglas et al. (2013); Martinez et al. (2012); Pendlebury et al. (2016b); Rudolph et al. (2016) & Wong et al. (2018) included predictors that were available upon admission or shortly thereafter, increasing their clinical merit.

The first DRPM included four predictors: the Acute Physiology and Chronic Health Evaluation II score (APACHE II), history of cognitive impairment, presence of dehydration (BUN/Creatinine), and visual impairment (glasses or blindness) (Inouye et al., 1993). The APACHE II score created complexity by requiring assessments including the Glasgow coma scale score (GCS), temperature, Mean Arterial Pressure (MAP), heart rate, respiratory rate, FIO2, and the lab values of a PaO2, arterial pH, bicarbonate, sodium, potassium, creatinine, and hematocrit (Knaus et al., 1985). Not all of these tests (arterial blood gas to assess the PaO2, pH, and bicarbonate levels) or assessments (Glasgow coma scale) are clinically necessary for evaluation of all patients admitted resulting in ordering additional tests to complete the risk model. Extra testing increases healthcare costs, pain, and exposure to the risk of procedures such as the arterial blood gas (ABGs) collection. This model is not adopted in clinical practice due to the complexity of obtaining all the added clinical information necessary for the scoring to occur

(Douglas et al., 2013; Pendlebury et al., 2016a). Refer to Tables 1-7 for a simplified way to discover the predictors included in each of the seven models.

As evidenced by Inouye and colleagues' model of 1993, DRPM must contain predictors that are likely to be obtained upon admission (laboratory studies or assessments), carry a low burden to collect, and quickly calculate the risk levels. As Rudolph et al. (2011), Pendlebury et al. (2016a), and Wong et al. (2018) point out, tools for future practice need to be simple, credible, and externally validated. Additional specialized assessments such as the MoCA (Montreal Cognitive Assessment), MMSE (Mini-Mental Status Exam), APACHE II Score (Acute Physiology And Chronic Health Evaluation), Barthel Index, or the Charlson Comorbidity Index are too cumbersome to include in a bedside risk prediction model (Newman et al., 2015). These assessments require additional training for staff and are of less clinical use due to low completion rates (Carrasco et al., 2014). In the updated Pendlebury (2016a) study, modifications to the models for ease of use without impacting the accuracy of prediction.

Perhaps the most straightforward yet technically complex score is the machine learning models in the study of the Wong et al. (2018). All tools reported in this study have published superior statistics in comparison to the non-machine learning tools. This study resulted in the Gradient Boosting Machine (GBM) model, which contains 345 predictors electronically collected from the chart allowing for the automated extraction of the health data and alerting the clinical teams to elevated risk scores. The authors compared the model to the AWOL and noted that the GBM could target those at the highest risk with improved accuracy, thus limiting focused interventions to less than half that of the AWOL (Wong et al., 2018).

The Wong et al. (2018) electronic models are very complex computerized programs and require technical builds for integration into digital charting systems. It provides real-time

calculations of risk without additional assessments by healthcare providers (Wong et al., 2018). In the current healthcare environment, it is desirable because it frees up healthcare providers time.

Summary of Research Findings

The statistical evidence confirms validated delirium risk prediction models have the ability to stratify the risk for delirium. Each of the models report both discrimination and calibration as the AUC, sensitivity and specificity, or percentages of positive outcomes in relation to risk scores. A model that can discriminate has the ability to categorize high versus low risk. A well calibrated model effectively determines higher risk levels in correlation with increasing true positive rates. However, a model cannot be implemented simply based on its statistics. Models for clinical practice must also be feasible in practice (easy to use, consume very little clinical time to preform), utilized at the appropriate time, and must stratify the risk accurately (minimal false negative cases).

The systematic review by Newman et al. (2015) studied the barriers to implementation were that healthcare providers perceived too much time to assess and score models, models are complex and not understood, limited supporting evidence of the models clinical impact, and there were not recommendations for clinical decision-making based on the level or risk. Interestingly the automated computerized models (e-NICE and GBM) report both the highest ability to stratify risk with both a high Se and Sp, limiting the interventions to a narrow group of patients with minimal miss classification of positive cases of delirium. Because these model results are computer generated, they require no further assessments for providers, making them a very attractive option.

As for as implementation and impact studies, Newman et al. (2015) was unable to discover any published studies and noted they did not find evidence that any models were employed in practice prior to the studies publication. New evidence has emerged reporting successful implementation of a delirium prevention care pathway using the AWOL model to triage the use of interventions to those assigned a risk score of ≥ 2 (Brown et al., 2015; Brown et al., 2017). This evidence shows that use of a DRPM, as part of a delirium prevention bundle, can diminish the severity of delirium mitigating some secondary effects. It also relays a benefit of decrease in length of stay of > 2 days (Brown et al., 2017). The Brown et al. (2017) study is the only impact study discovered during the literature search employing the use of a DRPM as part of a multicomponent interventional pathway. This evidence suggests that DRPMs should not be implemented alone, but within a care pathway to impact the consequences of delirium.

DRPMs allow resources to be allocated to those at greatest risk, decreasing use of limited resources such as volunteers, physical and occupational therapy, psychology consults, pharmacy consults, or involvement of a geriatrician. The preventative strategies implemented must proceed the triggering factor of modifiable precipitating factors to be effective (Brown et al., 2017).

What methods are recommended for clinicians to triage preventative interventions for those at highest risk for development of delirium? Clinical practice guidelines and associations such as the ANA, SCCM, and HRET support determination of risk as the first step to multicomponent prevention pathways. Each of these organizations or guidelines discuss the risk factors for delirium and provide a long list or a pneumonic of them to be recalled during clinical practice. The ANA and SCCM do not discuss DRPM or suggest any systematic methods for determining risk. The HRET (2018), however, discusses risk and recommends employing a

delirium risk prediction model. Interestingly this published "package of change" is the most recently published recommendations for iatrogenic delirium management.

Gaps in Literature

Clinical decision-making is aided by risk prediction models in many settings of medical care. There are no studies reporting or focusing on what effects DRPMs have on the clinical decision-making of clinicians. The recommended delirium preventative care pathways, unlike risk stratification for stroke (CHA_2DS_2-VASc, which directs clinicians on use of anti-coagulation therapy based on calculated risk score (January et al., 2014), do not provide stepped levels of interventions based on calculated scores. Studies additionally stated that determining patients at risk and alerting providers to the elevated risk may improve recognition and diagnosis of delirium risk (Douglas et al., 2013; Rudolph et al., 2016). There is no clear evidence to support the claim that DRPMs enhance diagnosis of delirium by clinicians. Reporting evidence to address these gaps may improve support of DRPMs use in clinical practice.

Some models being developed are aimed at a very narrow population of focus, such as models for ICU, surgical patients, or patients with fractures rather than a broader, generalized population like the general acute care hospital admission population (Lindroth et al., 2018). Lindroth et al. (2018), with a focus on older, found 23 prediction models, 11 medical, 3 medical/surgical, and 9 for various surgical procedures. Another study reported finding 37 DRPMs, 16 focused on cardiovascular surgery, six on orthopedic surgery, and the other 15 from various hospital unit settings (van Meenen, L., van Meenen, D., Rooij, & Riet, 2014).

Interventions suggested by guidelines for management of delirium were created based on general medical care patient data sets (Brown et al., 2017), thus it is a mixed message that risk prediction is not generalized, but the prevention is generalized. The effectiveness of

interventions in other settings such as ICU or post CV surgery have not been studied, leaving an additional gap in the literature.

Clinical practice guidelines and organizations such as the American Nursing Association (2016) note that assessment of risk is the first step to preventing delirium, yet not one recommends a delirium risk prediction model. It may be attributed to the fact that there only two published implementation studies (Brown et al., 2015; Brown et al., 2017) and one impact study (Brown et al., 2017). Impact and implementation studies report on the findings during implementation and show what impact it has on the population. Significant gaps in literature are few studies have been externally validated, more models exist than validation studies, and it is rare to find reporting of impact studies or clinical improvement projects with the use of DRPM. It is possible that support for implementation of DRPMs in practice would increase if all the gaps in literature were studied and published.

Conceptual Framework: Stetler Model of Research Utilization

The Stetler Model of research utilization is the guiding framework for this literature review. It directs practitioners to develop common standards, tools, and policies that are supported by evidence-based research. It guides clinicians in critically reviewing and reflecting on practice to understand the relationship between research use and evidence-based practice. This model sets criteria and sets standards to view a problem (Stetler, 2001).

The five phases start with the literature search that guides the structure of evidence reported within this review. The subsequent sections are outlined based on these three phases from the Stetler model (2001). Phases four and five of the model relate to implementation, which is not the purpose of this literature review thus were not used for this project. Figure 1 was created by this author to show the three phases used in this paper using a circular figure.

Phase one begins with preparations including defining the purpose of the research, the context of the studies, and levels of evidence (Gray, Grove & Burns, 2017). Phase two is the validation phase which analyzes of the overall credibility, applicability, and operational details of the studies and aids in the evaluation of the fit of each research study to the purpose of the inquisition. Phase three is the comparative evaluation or decision-making phase, the evaluator organizes and displays the research findings based on their similarities and differences. Each phase builds on the findings from the previous phase adding depth of understanding of the studies. Phase three ends with recommendations for practice based on the evidence presented.

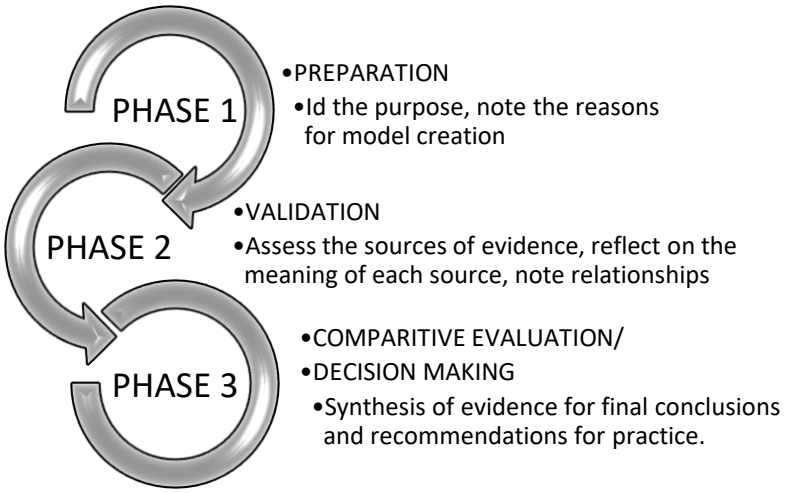

Figure 1. The Stetler Model of Evidence-based Practice figure for phases one through three (2001).

Phase One

Preparation for this integrated literature review consisted of identifying the purpose, context, and sources of evidence to include in this literature review/comparison study (Gray, Grove & Sutherland, 2017). During this phase the criteria was developed for inclusion: level of evidence a model contains, the research method (retrospective or prospective cohorts), the

inclusion of developed non-validated studies vs. validated studies, and the population of focus for the DRPMs. This led to the purpose of this review and the guiding questions.

Phase Two

In phase two, the overall credibility, applicability, and operational details were assessed. Each study was evaluated for the level of evidence and quality of reporting of the study with the use of the CHARMs checklist and the Newcastle-Ottawa scale. During this phase reflection of the meaning of each study was done as this author reviewed the variables included in the prediction model and its ability to perform its intended purpose. For ease of comparison literature tables were created (Table 12).

Phase Three: Review of Studies

In phase three a comparative evaluation between the selected studies was completed. This comparison guided the final recommendations for practice based on the evidence presented by the authors of each DRPM. A table of statistical comparison and the tables outlining the factors in each DRPM were created during this phase (Table 14 and Tables 1-7 respectively). Each phase of the Stetler Model builds upon the previous stages, the figure was developed to show this relationship and the ability to step from level 1 to level 2 or back up to level 1 again as the direction of the literature changes with discovery or new evidence.

Conclusions

The risks of complications are high for our fragile elderly patients being admitted into hospitals. Reactionary clinical practices are no longer valid in the prevention and treatment of delirium. The costs associated with delirium are not limited only to financial losses, as delirium affects the quality and quantity of a person's life. Cost and quality of care are leading healthcare clinicians to look for innovative ways to deal with challenges such as delirium. The goal of

medical care is changing from treatment of conditions to prevention. Today, clinical decision making for delirium prevention is not aided by DRPMs. This gap in preventative care exists because past models were not feasible to complete in clinical practice. More attention is needed to employ more recently developed models and develop preventative protocols for delirium across all acute care hospital settings.

The seven models included in this review have statistically proven their ability to stratify risk. What they have not proven is clinical effectiveness of DRPMs in practice. The recently published impact and implementation studies show promise in both the use of DRPM in general and their use within a delirium prevention care pathway (Brown et al., 2016; Brown et al., 2017). DRPM are suggested for use as a clinical decision tool to triage implementation of delirium prevention pathway or bundles of care (Douglas et al., 2013).

Clinical guidelines recommend assessing risk factors however they do not offer systematic methods to determine the risk of developing delirium. The DRPM's presented in this literature review are evidence that stratifying the risk of delirium is possible using validated models. The ability to stratify risk is the key to triaging resources to implement preventive interventions that are resource-intensive and expensive. Stratifying risk and applying interventions to those at greatest need has been shown to be cost-effective in implementation studies (Brown et al., 2017).

The conclusions drawn from the appraisal and synthesis of the models and supporting literature guide the recommendations for practice. The conclusions are:

- The literature shows evidence that DRPMs could be used in clinical practice as part of a multi-component interventional pathway

- Without further published impact studies or quality improvement studies that use DRPMs for triaging of interventions the guidelines cannot recommend a tool for use
- Delirium is a serious medical and psychiatric problem, leading to adverse health events, for which preventative measures are stated to reduce the rates, given the 2050 projected increase in elderly adults, the time to prevent is now
- Interventions for delirium should be studied in the population for which the DRPM is aimed
- There are no known risks of the preventative interventions, thus it is assumed that implementing preventative pathways can only provide benefit
- Healthcare needs standardization of the processes of preventing, managing, and treating delirium
- Adding actionable recommendations to a care pathway may provide clinicians with a reason to implement interventions and promote acceptance of a model by staff
- While complex to implement, automated models provide consistency and liberates time normally is spent by practitioners calculating a risk score

Recommendations

Implementing standardized healthcare processes are best accomplished by the development of care pathways or care bundles. According to the Institute for Healthcare Improvement a bundle is, "a structured way of improving the processes of care and patient outcomes: a small, straightforward set of evidence-based practices....that, when performed collectively and reliably, have been proven to improve patient outcomes." (Evidence Based Care Bundles, para. 1).

A recommended delirium preventative pathway will include three main elements; (a) an evidenced based delirium screening/diagnostic tool with high accuracy rates, such as the Confusion Assessment Method (CAM), (b) an automated DRPM, (c) evidence-based delirium prevention interventions as recommended by the NICE guideline (NICE, 2010) (or a similar guideline) with the addition of stepped interventions that increase in intensity of resource utilization as the level of delirium risk increases. During a personal communication with Dr. Douglas, a neurohospitalist from the University of California San Francisco hospital, he stated that two mistakes his team made when implementing the AWOL into practice was not using the CAM as a diagnostic tool and not spacing the implementation of the NuDESC for delirium screen adequately before implementation of the AWOL into practice. He suggested that prior to any implementation of a DRPM or preventative care pathway, a delirium screening tool such as the confusion assessment method (CAM) be employed for a minimum of three months (V. Douglas, personal communication, May 17, 2018). Adoption of an accurate diagnostic tool as a first step will allow for gathering of clinical data on current delirium rates.

Educating all staff before the implementation process on the evidence and predictive ability of the model may improve the perceived value, increase acceptance, and improve belief in the care pathway (Kappen et al., 2018). Education is priority prior to implementation of any of the five elements of a delirium prevention care pathway. Additional Education on the epidemiology, pathophysiology, diagnosed criteria, risk factors, methods to screen for onset, and evidence-based interventions for prevention of delirium need to be provided to staff.

Recommendation for a Specific DRPM

Automated risk scoring allows for provider notifications of risk levels without increasing their workload; therefore, the use of automated models is recommended as a way to break the

barrier to implementation. The GBM (Wong et al., 2018) or the e-Nice (Rudolph et al., 2016) are the recommended models, they are automated scores and have the highest Se and Sp of any tool. Of note the e-NICE is currently implemented without a care pathway at 118 VA hospitals (J. Rudolph, personal communication, June 14th, 2018).

Recommendations for Research

Decisions to implement evidence-based clinical practices would ideally be supported by a large multicenter pragmatic randomized control trial (RCT). A study of this level assesses the strengths of the model, limitations, and its effectiveness as a clinical decision-making tool. However, RCTs are difficult to conduct on the effectiveness of care pathways because of operational and ethical considerations such as the withholding of effective evidenced-based interventions from the control population (Cheah, 2000). The impact of a DRPM cannot be studied independent of prevention strategies because the prediction models use is only to detect if there is a risk of delirium and does not provide interventions to effect patient outcome.

Quality improvement (QI) projects are an integral part of good clinical practice and are designed to implement existing evidence-based knowledge to bring about improvements at the local level (Kappen et al., 2018). The AWOL impact study by Brown et al., (2018) is evidence that reporting of QI projects produce subject matter knowledge. The recommendation for future practice is that hospitals execute a QI project with the intent to decrease delirium rates and to decrease the negative effects on healthcare systems, patients, and families by mitigating the severity of delirium. The QI project would be a unit based project with the implementation of a delirium screening tool, a delirium risk assessment model, and evidence-based interventions. The study results ideally would be published to allow other healthcare teams to learn from the experiences and learn of the impact that the interventions in the QI produce.

The Stetler Model of Evidence Based Practice was practically employed with the vision that this literature review is a launching point to implementation. Progressing this evidence into the fourth and fifth phases of the model could be done as doctoral student or healthcare provider to implement a delirium care pathway into clinical practice. The last two phases of the Stetler Model can guide putting research into clinical practice. Future doctoral students should note that the Wong et al., (2018) computer model can be requested for the goal of clinical application.

Implications for Nursing

The Institute of Medicine's report in 2000 states that it is the responsibility of every healthcare worker to enact evidence-based principles of care to prevent patient harm and most clinical risks originate directly from defects or insufficiencies in the healthcare system (Adibi, Khalesi, Ravaghi, Jafari, & Jeddian, 2012). Systematic methods of preventing and managing delirium can prevent harm. Clinicians are the advocates for patient safety and can advocate for a systematic delirium pathway with a DRPM used to triage preventative interventions. This would require development of the care pathway, policy changes, education to staff, and data collection and analysis on implementation effects. If DRPMs are used, clinicians, nurses, and hospital administrators will be part of the creation of a new paradigm, a shift in the care of a hospitalized at-risk patient, with the potential to improve patient outcomes and decrease the cost of delivering healthcare.

For clinicians delirium has many consequences like the patients inability to consent for procedures, learn about personal healthcare interventions, participate in therapies, or any cognitive interaction. Family members need to be contacted for consent and illness education which may result in a delay in care or extended hospitalizations. Patients with delirium are unable participate in meaningful activities in therapy, provide self-care, or to comply with

medical management, increasing their risk of adverse outcomes. Delirious patients may inadvertently cause personal trauma while pulling at or removing medical devices such as intravenous catheters, Foley catheters, oxygen assistive devices, and monitoring equipment increasing the risk for infections, bleeding, and urinary incontinence; thus, nursing care increases as well as the use 1:1 sitters. Hallucinations and delusions set a patient up for unintentional self-harm or caregiver harm, for which physical or medicinal restraints may be applied for safety, again increasing nursing care, sitters, and added workload documenting safety.

Requiring another assessment tool would increase workloads that are already heavy, but employment of an automated delirium risk prediction model would mitigate additional work for assessments. The risk level will alert nurses to tailor interventions to meet a patient's health needs based on patient specific risk factors. Bedside nurses tailor care to meet patient needs as part of their nursing processes. They are in a special position to assess for delirium risk, discover onset of delirium, notify providers of the onset, and intervene with the tailored evidence-based preventative strategies. An example is enacting a tailored plan to treat a disturbed circadian rhythm due to nursing activities over-night. The nurse modifies timing of assessments and interventions, matching a patients sleep cycle; this may require calling clinicians to allow for decreased checks on vital signs or overnight nursing assessments. A second example is mitigating the effect of sensory deficits by encouraging the use of hearing aids or glasses during the day.

A clinicians role varies from a nurses role in that in response to level of risk for delirium a clinician would weigh the risk and benefit of procedures, medical tests, and medications against the potential triggering of iatrogenic delirium. If the onset of delirium is reported or discovered, clinicians must respond with a full medical work up, to determine differential

diagnoses and choose the appropriate medical interventions to treat the underlying cause. Clinicians can consult geriatricians, psychiatry, physical therapy, and pharmacy for their recommendations on preventative strategies and management of patients delirium. Directing nursing non-pharmacological management and ordering frequent assessments on mentation and ability to perform ADLs.

This integrated literature review has far-reaching implications for healthcare's ability to prevent the harmful effects of delirium by improving recognition of a patients risk for delirium, understanding that it is often preventable through the implementation of delirium prevention interventions, and prevention of delirium improves the quality of care to every hospitalized patient by decreasing adverse events associated with it. The World Health Organization (WHO) defines health as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity ("Constitution", para 1.). The World Health Organization defines health promotion as the process of "enabling people to increase their control over and improve their health" (What is health promotion, para. 1).

According to the IHI, the triple aim for healthcare is an approach to optimize healthy system performances by improving the patient experience, health of populations, and reducing the per capita cost of care (IHI, 2018). The role of healthcare providers is to maintain health, thus preserving the quality of life. Delirium negatively impacts patient and family healthcare experiences as it results in poor health outcomes and increases the cost of healthcare.

Summary of Conclusions, Recommendations, and Implications for Nursing

The aim of using a DRPM in clinical practice is to stratify the risk for incident delirium allowing clinicians to target those at the most risk with the preventative interventions. One impact study showed implementation of a delirium prevention pathway, including a DRPM to

triage the interventions, allowed for efficient use of resources. This implementation study reported a decrease in length of stay, 30-day readmission rates, and severity of delirium. The literature shows that DRPMs are able to stratify the risk of delirium and when included in a preventative care pathway effectively mitigate some repercussions of incident delirium.

Automated models have been developed in response to the barriers of adoption and implementation of DRPMs. After integration in computerized charts, automated models are the most feasible models because the result is available to the provider without additional time added to their workflows. The electronic versions of the e-NICE by Rudolph et al. (2016) and the Wong et al. (2018) models were statistically superior to all of the other models; they are technically complex, with the GBM containing hundreds of predictors mined from the computerized charting system (Wong et al., 2018). The Wong et al. study notes that if requested other researchers or healthcare systems seeking quality improvement of delirium care can request the computer program that the researchers developed and allow the use by another healthcare system. The DSS (Pendlebury et al., 2016b) shows promise for use in healthcare facilities where integration into an electronic medical record is not possible with its feasibility and accuracy of stratification.

The AWOL and e-Nice tools are the only tools reported in research that are clinically employed at this time. The AWOL tool was employed as part of a care pathway and was studied in three follow up studies, the Brown et al. (2017) impact study, Pendlebury et al., (2016a), and the Wong et al. (2018). It is available in the computer application called MdCalc, but this is not tied into the computer systems and clinicians still need to actively seek the risk score, again creating unreliability. The e-NICE tool is in current use in clinical practice at the VA medical centers, alerting clinicians of the patients at highest risk.

The Gradient boosting machine by Wong et al. (2018), is a tool with the highest sensitivity and specificity of all delirium risk stratification models. It is an electronic model that automatically uses data abstraction to electronically stratify every admission for delirium risk. It contains over 300 predictors, thus is the most comprehensive, and has the highest AUC of any delirium risk prediction model at .855, however this model is very complex. Neither the GBM nor the e-NICE have been implemented in tandem with a delirium prevention bundle.

Gaps in practice include lack of implementation and impact studies to provide clinical evidence of the effectiveness of the models. There are no studies showing which interventions to employ at each risk level which may increase the effectiveness and decrease use of unnecessary resources. Generalization of DRPMs are difficult related to their narrow focus of population and the fact that some are not externally validated. DRPMs are created to aid clinical decision making of healthcare providers, however, no studies examining their effect on clinical decision-making exist.

The primary studies of the DRPMs show validated statistical proof that they have the ability to stratify risk. The impact and implementations studies provide literature supporting their use within a care pathway. HELP interventions have shown to decrease delirium rates and NICE guidelines recommend similar interventions without the use of hundreds of volunteers. The main recommendation is to implement a full delirium prevention pathway. A DRPM will allowing triaging of the interventions within the pathway to deliver necessary preventative care to decrease risk, decrease adverse outcomes associated with delirium, and ultimately improve the quality of care given to patients. The process and result of implementation should be studied and published to begin to close the gap in evidence of clinical effectiveness of their ability to stratify risk, triage interventions, and prevent delirium cases and other negative implications of delirium.

In summary, optimal models must have the ability to discriminate, calibrate, and are validated in the clinical practice setting adopting the model. To be useful in practice (feasible), predictors must be readily available at the time of admission, or shortly after (same day), and cannot require additional medical testing or complex assessments. The more recently created and validated primary delirium risk prediction models have proven feasibility in their less complicated predictors. The risk assessment must be completed as early as possible after admission to allow for the implementation of preventative measures before insults occur that further increase the risk of delirium. Delayed risk assessment results in a lost opportunity to preventative delirium. Improvement in provider reception of a risk prediction model may follow the recommendation for an automated risk prediction score because of the liberation of a provider's time. These automated models also show a statistical improvement in stratification accuracy compared to the non-automated models.

Delirium is a medical emergency with consequences of death and disability, similar to a stroke or a myocardial infarction. There are no cures for delirium; healthcare providers must take action in the fight to prevent its onset. We cannot wait for the creation of the perfect model or care pathway. The healthcare system continues to dismiss evidence that a care pathway (such as the ABCDEF bundle or the HELP) can prevent delirium, which results in patient harm, family burdens, and rising healthcare costs.

The current practice of assessing risk by clinical intuition and experience allows for vast variations in practice. As evidenced by the high delirium rates in our acute care facilities it is also very inefficient. Therefore, a systematic method needs to be employed to consistently stratify the risk of every patient admitted to a hospital. The recommendation for application in clinical practice settings is to develop and implement complete delirium prevention, treatment, and

management care pathway. A DRPM can be employed to triage the hospital's limited resources to those of greatest need.

References

Abrantes, T., & Racine, A. (2019). *Family caregiving for delirious patients: What it's like, what we know, and what's next*. Retrieved from the Network for Investigation of Delirium: Unifying Scientist website: https://deliriumnetwork.org/family-caregiving-for-delirious-patients/

Adibi, H., Khalesi, N., Ravaghi, H., Jafari, M., & Jeddian, A. (2012). Development of an effective risk management system in a teaching hospital. *Journal of Diabetes and Metabolic Disorders*, 11(1), 1–7. doi: 10.1186/2251-6581-11-15

Agyeman, A. A., & Ofori-Asenso, R. (2015). Perspective: Does personalized medicine hold the future for medicine? *Journal of Pharmacy & Bio Allied Sciences*, *7*(3), 239–244. doi: 10.4103/0975-7406.160040

Ackley, B. J., Swan, B. A., Ladwig, G., & Tucker, S., (2008). *Evidence-based nursing care guidelines: Medical-surgical interventions. (p. 7). St. Louis, MO: Mosby Elsevier*

Alagiakrishnan, K. (2019). Delirium. *Medscape*. Retrieved from https://emedicine.medscape.com/article/288890-overview

Avelino-Silva, T. J., Campora, F., Curiati, J., & Jacob-Filho, W. (2018). Prognostic effects of delirium motor subtypes in hospitalized older adults: A prospective cohort study. *PloS one*, *13*(1),1-18. doi: 10.1371/journal.pone.0191092

American Nurses Association. (2016). *Delirium: Prevent, Identify, Treat.* Retrieved from https://www.nursingworld.org/practice-policy/work-environment/health-safety/delirium/.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Washington DC: American Psychiatric Association.

AWOL Score for Delirium (2017). In MdCalc for Apple iOS (Version 2.2.7) [Mobile application software]. Retrieved from https://www.mdcalc.com/awol-score-delirium

Bernard, A. (2017). Clinical prediction models: a fashion or a necessity in medicine? *Journal of Thoracic Disease, 9*(10), 3456–3457. doi: 10.21037/jtd.2017.09.42

Brown, E. G., & Douglas, V.C. (2015). Moving beyond metabolic encephalopathy: An update on delirium prevention, workup, and management. *Seminars in Neurology, 35*(6), 646-655. doi: 10.1055/s-0035-1564685

Brown, E. G., Josephson, S. A., Anderson, N., Reid, M., Lee, M., & Douglas, V. C. (2017). Predicting inpatient delirium: The AWOL delirium risk-stratification score in clinical practice. *Geriatric Nursing, 38*(6), 567-572. doi:10.1016/j.gerinurse.2017.04.006

Carrasco, M. P., Villarroel, L., Andrade, M., Calderon, J., & Gonzalez, M. (2014). Development and validation of a delirium predictive score in older people. *Age and Ageing, 43*(3), 346-351. doi:10.1093/ageing/aft141

Cerejeira, J., & Mukaetova-Ladinska, E. B. (2011). A clinical update on delirium: From early recognition to effective management. *Nursing Research and Practice, 2011*, 1-12. doi:10.1155/2011/875196

Cheah, J. (2000). Development and implementation of a clinical pathway programme in an acute care general hospital in Singapore. *International Journal for Quality in Health Care, 12*(5), 403-412. doi:10.1093/intqhc/12.5.403

Columbia University Mailman School of Public Health. (2019). *Risk Prediction*. Retrieved from https://www.mailman.columbia.edu/research/population-health-methods/risk-prediction

Cook, N. R. (2008). Statistical evaluation of prognostic versus diagnostic models: Beyond the ROC curve. *Clinical Chemistry, 54*, 17-23. doi: 10.1373/clinchem.2007.096529

Douglas, V. C., Hessler, C. S., Dhaliwal, G., Betjemann, J. P., Fukuda, K. A., Alameddine, L. R.,. . . Josephson, S. A. (2013). The AWOL tool: Derivation and validation of a delirium prediction rule. *Journal of Hospital Medicine, 8*(9), 493-499. doi:10.1002/jhm.2062

Ford, A. H. (2016). Preventing delirium in dementia: Managing risk factors. *Maturitas, 92,* 35-40. doi:10.1016/j.maturitas.2016.07.007

Gaudreau, J. D., Gagnon, P., Harel, F., Tremblay, A., & Roy, M. A. (2005). Fast systematic, and continuous delirium assessment in hospitalized patients: The nursing delirium screening scale. *Journal of Pain and Symptom Management, 29*(4), 368-375. doi: 10.1016/j.jpainsymman.2004.07.009

Gonzalvo, C. M., de Wit, H. A., van Oijen, B. P., Deben, D. S., Hurkens, K. P., Mulder, W. J., . . ., Van Der Kuy, P. M. (2017). Validation of an automated delirium prediction model (DElirium MOdel (DEMO)): An observational study. *BMJ Open, 7*(11). doi:10.1136/bmjopen-2017-016654

Gray, J., Grove, S. K., & Sutherland, S. (2017). *Burns and Groves the practice of nursing research: appraisal, synthesis, and generation of evidence*. St. Louis, MO: Elsevier.

Groves, M., and Huskin, P.R. (2011). The American Psychiatric publishing textbook of psychosomatic medicine. In J. L. Levenson (2nd Ed.), *Psychologic responses to illness* (pp. 67-90). Washington, D.C: American Psychiatric Publishing.

Halladay, C. W., Sillner, A. Y., & Rudolph, J. L. (2018). Performance of electronic prediction rules for prevalent delirium at hospital admission. *Journal of the American Medical Association Network Open*, 1(4). doi: 10.1001/jamanetworkopen.2018.1405

Health Research & Educational Trust (2018). *Preventing iatrogenic delirium*

change package: 2018 update. Retrieved from https://patientcarelink.org/wp-content/uploads/2018/09/preventing-and-managing-iatrogenic-delirium-change-package.pdf

Ho, M.H., Chen, K.H., Montayre, J., Liu, M. F., Chang, C.-C., Traynor, V., … Chiu, H.-Y. (2019). Diagnostic test accuracy meta-analysis of PRE-DELIRIC (PREdiction of DELIRium in ICu patients): A delirium prediction model in intensive care practice. *Intensive and Critical Care Nursing. 57*, 1-7. doi: 10.1016/j.iccn.2019.102784

Hospital Elder Life Program (HELP). (2013). *AGS CoCare: HELP background and impact.* Retrieved from https://help.agscocare.org/chapter-abstract/HELP%20Background%20&%20Impact/H00102/H00102_PART004_001

Inouye S. K. (2018). Delirium: A framework to improve acute care for older persons. *Journal of the American Geriatrics Society*, *66*(3), 446–451. doi: 10.1111/jgs.15296

Inouye, S. K., Bogardus, S. T., Charpentier, P. A., Leo-Summers, L., Acampora, D., Holford, T. R., & Cooney, L. M. (1999). A multicomponent intervention to prevent delirium in hospitalized older patients. *New England Journal of Medicine, 340*(9), 669-676. doi:10.1056/nejm199903043400901

Inouye, S. K., Viscoli C.M., Horwitz R.I., Hurst, L. D., & Tinetti, M. (1993). A predictive model for delirium in hospitalized elderly medical patients based on admission characteristics. *Annals of Internal Medicine, 119*(6), 474-81. doi: 10.7326/0003-4819-119-6-199309150-00005

January, C. T., Wann, S., Alpert, J. S., Calkins, H., Cigarroa, J.E., Conti, J. B., ...Yancy, C. W. (2014). AHA/ACC/HRS Guideline for the Management of Patients With Atrial

Fibrillation: Executive Summary. *The Journal of the American College of Cardiology, 64*(21), 2246-2280. doi: 10.1016/j.jacc.2014.03.021.

Kalimisetty, S., Wajih, A., Fay, B., and Khan, A. (2017). Models for predicting incident delirium in hospitalized older adults: A systematic review. *Journal of Patient-Centered Research and Reviews, 4*, 69-77. Doi: 10.17294/2330-0698.1414

Kappen, T. H., van Klei, W. A., Van Wokfswinkle, L., Kalkman, C. J., Vergouwe, Y., & Moons, K. G. M. (2018). Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. *Diagnostic and Prognostic Research, 2*(11), 1-11. doi: 10.1186/s41512-018-0033-6

Kappen, T. H., van Loon, K., Kappen, M. A., van Wolfswinkel, L., Vergouwe, Y., van Klei, W. A., ...Kalkman, C. J. (2016). Barriers and facilitators perceived by physicians when using prediction models in practice. *Journal of Clinical Epidemiology, 70*, 136-145. doi: 10.1016/j.jclinepi.2015.09.008

Knaus W.A, Draper, E. A., Wagner D. P., & Zimmerman (1985). APACHE II: A severity of disease classification system. *Critical Care Medicine, 13*(10), 818-829

Kobayashi, D., Takahashi, O., Arioka, H., Koga, S., & Fukui, T. (2013). A prediction rule for the development of delirium among patients in medical wards: Chi-square automatic interaction detector (CHAID) decision tree analysis model. *The American Journal of Geriatric Psychiatry*, *21*(10), 957–962. doi: 10.1016/j.jagp.2012.08.009

LaHue, S.C., Douglas, V.C., Kuo, T., Conell, C.A., Liu, V.X., Josephson, S.A., … Brooks, K. B. (2019). Association between inpatient delirium and hospital readmission in patients ≥ 65 years of age: A retrospective cohort study. J*ournal of Hospital Medicine,* 14(4), 201–206. doi:10.12788/jhm.3130

Lee, Y. H., Bang, H., & Kim, D. J. (2016). How to establish clinical prediction models. *Endocrinology and Metabolism*, *31*(1), 38-44. doi:10.3803/EnM.2016.31.1.38

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. *Archives of Internal Medicine,168*(1), 27-32. doi:10.1001/archinternmed.2007.4

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,* 1-22. doi: 10.1136/bmjopen-2017-019223

Martinez, J. A., Belastegui, A., Basabe, I., Goicoechea, X., Aguirre, C., Lizeaga, N., Urreta, I., & Emparanza, J. I. (2012). Derivation and validation of a clinical prediction rule for delirium in patients admitted to a medical ward: an observational study. *BMJ Open*, *2*(5). 1-7. doi: 10.1136/bmjopen-2012-001599

Maclullich, A. M., Ferguson, K. J., Miller, T., de Rooij, S. E., & Cunningham, C. (2008). Unravelling the pathophysiology of delirium: a focus on the role of aberrant stress responses. *Journal of psychosomatic research*, *65*(3), 229–238. https://doi.org/10.1016/j.jpsychores.2008.05.019

MacLullich, A. M., & Hall, R. J. (2011). Who understands delirium? *Age and Ageing*, *40*(4), 412-414. doi:10.1093/ageing/afr062

Michaud, L., Büla, C., Berney, A., Camus, V., Voellinger, R., Stiefel, F., & Burnand, B. (2007) Delirium Guidelines Development Group. Delirium: guidelines for general hospitals. *Journal of psychosomatic research*, *62*(3), 371–383. doi.org/10.1016/j.jpsychores.2006.10.004

Mehta, S., Cook, D., Devlin, J. W., Skrobik, Y., Meade, M., Fergusson, D., . . . Burry, L. (2015). Prevalence, risk factors, and outcomes of delirium in mechanically ventilated adults. *Critical Care Medicine, 43*(3), 557-566. doi:10.1097/ccm.0000000000000727

Moons, K. G. M., Groot, J. A. H. D., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D. G., … Collins, G. S. (2014). Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The CHARMS checklist. *PLoS Medicine*, *11*(10). 1-12. doi: 10.1371/journal.pmed.1001744

Newman, M. W., O'Dwyer, L. C., & Rosenthal, L. (2015). Predicting delirium: a review of risk-stratification models. *General Hospital Psychiatry. 37*(5). 408-413. doi: 10.1016/j.genhosppsych.2015.05.003

National Institute for Health and Care Excellence (2019). *Delirium: Prevention, diagnosis, and management*. (Clinical guideline [CG103]) Retrieved from www.nice.org.uk/guideline /cg103

Page, V. J., & Ely, E. W. (2017). *Delirium in critical care* (2nd ed.) [Kindle Edition]. Cambridge: Cambridge Univ. Press. https://www.amazon.com/Delirium-Critical-Care-Core-ebook/dp/B008LW25DO

Pendlebury, S. T., Lovett, N., Smith, S. C., Cornish, E., Mehta, Z., & Rothwell, P. M. (2016a). Delirium risk stratification in consecutive unselected admissions to acute medicine: Validation of externally derived risk scores. *Age and Ageing*, *45*(1), 60-65. doi:10.1093/ageing/afv177

Pendlebury, S. T., Lovett, N. G., Smith, S. C., Wharton, R., & Rothwell, P. M. (2016b). 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, 46*(2), 226-231. doi:10.1093/ageing/afw198

Peterson, J., Welch, V., Losos, M., & Tugwell, P. (2011). *The Newcastle-Ottawa scale (NOS) for assessing the quality of non-randomized studies in meta-analyses*. Ottawa: Ottawa Hospital Research Institute.

Pun, B. T., Balas, M. C., Barnes-Daly, M. A., Thompson, J. L., Aldrich, J. M., Barr, J., … Ely, E. W. (2019). Caring for critically ill patients with the ABCDEF bundle: Results of the ICU liberation collaborative in over 15,000 adults. *Critical Care Medicine, 47*(1). 3-14. doi: 10.1097/CCM.0000000000003482

Reston, J. T., & Schoelles, K. M. (2013). In-facility delirium prevention programs as a patient safety strategy: A systematic review. *Annuals of Internal Medicine.* https://doi.org/10.7326/0003-4819-158-5-201303051-00003

Rudolph, J. L., Doherty, K., Kelly, B., Driver, J., Archambault, E.,(2016). Validation of a Delirium Risk Assessment Using Electronic Medical Record Information. *Journal of the American Medical Directors Association, 17*(3), 244-248.

Rudolph, J. L., Harrington, M. B., Lucatorto, M. A., Chester, J. G., Francis, J., & Shay, K. J. (2011). Validation of a medical record-based delirium risk assessment. *Journal of the American Geriatrics Society, 59*. 289–S294. doi:10.1111/j.1532-5415.2011.03677.x

Society of Critical Care Medicine. (2018). *ICU Liberation Bundle (A-F). Retrieved from* https://www.sccm.org/ICULiberation/ABCDEF-Bundles

Solà-Miravete, E., López, C., Martínez-Segura, E., Adell-Lleixà, M., Juvé-Udina, M. E., &

Lleixà-Fortuño, M. (2017). Nursing assessment as an effective tool for the identification of delirium risk in older in-patients: A case-control study. *Journal of Clinical Nursing, 27*(1-2), 345-354. doi:10.1111/jocn.13921

Stetler, C. (2001). Updating the Stetler model of research utilization to facilitate evidence-based practice. *Nursing Outlook, 49,* 272-279. doi: 10.1067/mno.2001.120517

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. *Epidemiology, 21*(1), 128-138. doi:10.1097/ede.0b013e3181c30fb2

Taft, T., Nelsen, S. D., Slager, S., & Weir, C. (2018). Physician Information Needs in Managing Delirium. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, *2017*, 1675–1684.

The Institute for Healthcare Improvement. (n.d.). Triple aim for populations: *IHI*. Retrieved from http://www.ihi.org/Topics/TripleAim/Pages/default.aspx

Thurber, S., Kishi, Y., Trzepacz, P. T., Franco, J. G., Meagher, D. J., Lee, Y., . . . Leonard, M. (2015). Confirmatory factor analysis of the delirium rating scale revised-98 (DRS-R98). *The Journal of Neuropsychiatry and Clinical Neurosciences, 27*(2), 122-7. doi:10.1176/appi.neuropsych.13110345

U S Census Bureau. (2010, May). *THE NEXT FOUR DECADES. The Older Population in the United States: 2010 to 2050*. Retrieved from https://www.census.gov/prod/2010pubs/p25-1138.pdf

Van Meenen, L., Van Meenen, D., Rooij, S., and Riet, G. T. (2014). Risk prediction models for postoperative delirium: A systematic review and meta-analysis. *Journal of the American Geriatrics Society, 62*(12), 2383-2390. doi: 10.1111/jgs.13138

Vasilevskis, E., Ely, E. W., Speroff, T., Pun, B., Boehm, L., & Dittus, R. (2010). Reducing iatrogenic risks: ICU-acquired Delirium and Weakness-- Crossing the quality chasm. *CHEST,* 13*8*(5), 1224-33. doi:10.1378/chest.10-0466

Vasilevskis, E., Pandharipande, P., Girard, T., & Ely, W. (2010). A screening, prevention, and restoration model for saving the injured brain in intensive care unit survivors: Erratum. *Critical Care Medicine, 38*(10), S683-S691. doi: 10.1097/CCM.0b013e3181f245d3

Vincent, C., Neale, G., & Woloshynowych, M. (2001). Adverse events in British hospitals: preliminary retrospective record review. *British Medical Journal, 322*, 517-519. doi: 10.1136/bmj.322.7285.517

Wang, X. T., Lyu, L., Tang, B., Wang, C., & Liu, D. (2017). Delirium in Intensive Care Unit Patients. *Chinese Medical Journal, 130* (20). 2498-2502. doi: 10.4103/0366-6999.216405

Wassenaar, A., Schoonhoven, L., Devlin, J.W., van Haren, F. P, Slooter, A.C, Jorens, P. G., ... & can de Boogaard, M., (2018). Delirium prediction in the intensive care unit: comparison of two delirium prediction models. *Critical Care 22*, 1-9. doi.org/10.1186/s13054-018-2037-6

Wass, S., Webster, P. J., & Nair, B. R. (2008). Delirium in the elderly: a review. *Oman Medical Journal, 23*(3), 150–157. http://omjournal.org/ReviewArticle/PDF/200807/Delirium%20in%20the%20Elderly.pdf

Whitehorne, K., Gaudine, A., & Meadus, R., & Solberg, S. (2015). Lived experience of the intensive care unit for patients who experienced delirium. *American Journal of Critical Care, 24*(6). 474–479. doi:10.4037/ajcc2015435.

Winona State University Library. (n.d.). Research hub: Doing a Literature Review. Retrieved from https://libguides.winona.edu/litreview

Wong, A., Young, A. T., Liang, A. S., Gonzales, R., Douglas, V. C., & Hadley, D. (2018). Development and validation of an electronic health record-based machine learning model to estimate delirium risk in newly hospitalized patients without known cognitive impairment. *Journal of American Medical Association Network Open*, *1*(4). 1-12. doi:10.1001/jamanetworkopen.2018.1018

Zaubler, T. S., Murphy, K., Rizzuto, L., Santos, R., Skotzko, C., Giordano, ... & 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.org/10.1016/j.psym.2013.01.010

Appendix

Table 1

e-NICE: Rudolph et al., 2016

Risk Factor	Abstraction Terms	Score
Cognitive Impairment (positive if one term present)	Dementia, Alzheimer's Disease, Poor Historian, Memory loss, Unarousable, Uncooperative, Demented, Delirium, Change in mental status, confused, Encephalopathic, Disoriented, Lethargic, Obtunded, Stuporous, Combative,	
Sensory Impairment (positive if one term present)	Visual loss, Blindness, wears glasses, Hearing impairment, Hard of Hearing, Wears hearing aids, Presbycusis	
Dehydration (Positive if BUN/Creatinine ≥18)	BUN, Creatinine	
Severity of illness (positive if 2 terms present)	Age>60, Metastatic Cancer, Lymphoma, Leukemia, AIDS, RR > 25, Systolic blood pressure <100 mmHg, Pulse >120, Creatinine >2.0, Albumin < 2.5, total Bilirubin > 2.9	
Delirium Risk	***Risk Factors***	
Low	0	
Medium	1-2	
High	3-4	

Table 2

Martinez et al.,2012

Variable	Total
Age ≥85	
Dependent in 5+ ADLs • Grooming • Dressing • Toileting • Ambulation • Bowel/bladder control • Feeding	
Psychotropic medications • Antidepressants • Antidementia drugs • Antipsychotropics • Anticonvulsants • Benzodiazepines	
Total >1 High risk	

Table 3

AWOL: Douglas et al., 2013

Variable:	Score
Age ≥ 80 =1	
Ability to spell WORLD backward=1	
Disorientation to place = 1	
Nurse rated illness severity = 1	
Total >2	

Table 4

CHAID Decision Tree Model: Kobayashi et al., 2013

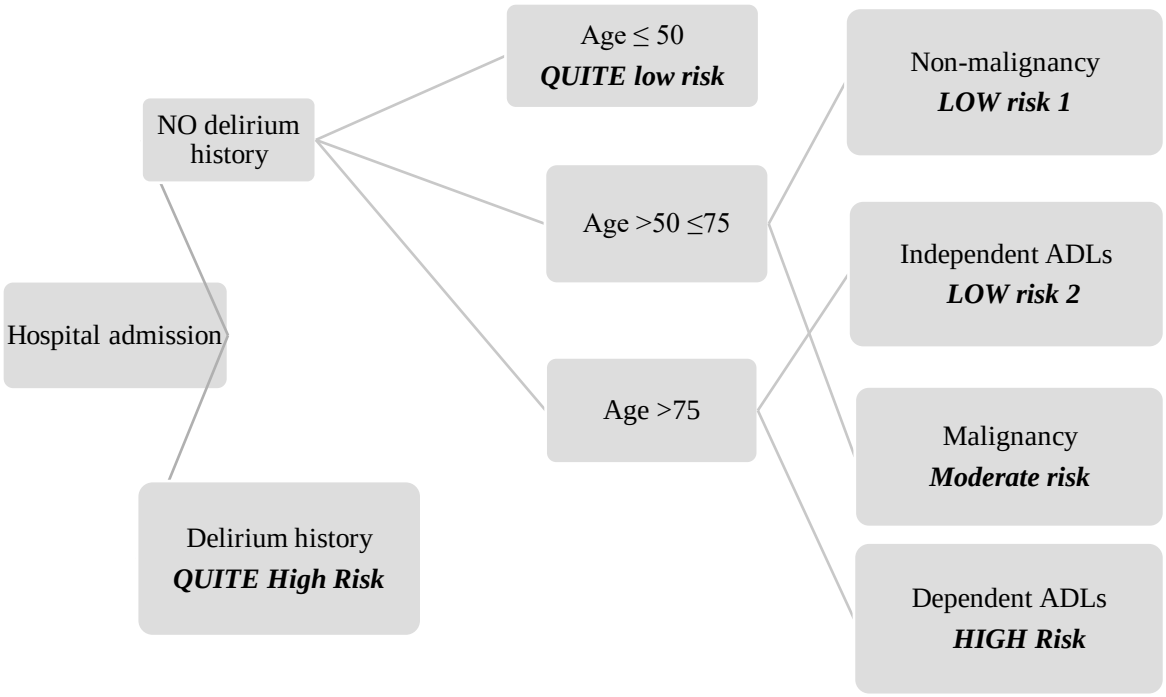

Table 5

Delirium Prediction Score (DPS): Carrasco et al., 2014

DPS= (5 X BUN (mg/dl)/Creatinine (mg/dl) ratio) – (4 x Barthel Index) Note: the cut point is -160

Table 6

DRPM Comparison and Model Update: Pendlebury et al., 2016a

Original author: Martinez et al.	*Variable:*	Score
	Age ≥85	
	Functional Dependence • Living at a care center or at home with homecare	
	Diagnosis of Dementia	
	Total >1	
Isfandiaty et al.	Cognitive Impairment =3	
	Functional Dependence = 2 • Living at a care center or at home with homecare	
	Infection without sepsis =1 Infection with sepsis (SIRS≥2)= 2	
	Total >3	
Douglas et al. 2013 (AWOL)	Age ≥ 80	
	Dx of Dementia or low cognitive score as defined by this study (2 points)	
	Illness severity (nurse assessment)	
	Total ≥2	

Table 7

Delirium Susceptibility Score: Pendlebury et al., 2016b

Variable:	Score
Age ≥ 80 =2	
Cognitive Impairment = 2 (MMSE <24 or AMTS<9 or known dementia)	
Infection =1 Infection with sepsis (SIRS≥2)= 1	
Visual Impairment= 1	
Total ≥ 5	

Table 8

Delirium Prevention Links

Resource	Website
American Nurses Association Delirium Prevention Strategies 2016	https://www.nursingworld.org/~4afecf/globalassets/practiceandpolicy/innovation--evidence/prevention-best-practices-wg10272016.pdf
National Institute for Health and Care Excellence (NICE)	https://www.nice.org.uk/guidance/cg103/evidence/appendix-a-summary-of-evidence-from-surveillance-pdf-6594316238
Health Research & Educational Trust	https://patientcarelink.org/wp-content/uploads/2018/09/preventing-and-managing-iatrogenic-delirium-change-package.pdf
The American Geriatrics Society (post-op delirium)	https://geriatricscareonline.org/
American College of Critical Care Medicine/ Society of Critical Care Medicine (SCCM)	https://www.sccm.org/search?searchtext=delirium&searchmode=anyword
The American Association of Critical Care Nurses (AACN)	https://www.aacn.org/clinical-resources/practice-alerts/assessment-and-management-of-delirium-across-the-life-span

Table 9

Newcastle-Ottawa Scale for Quality Assessment

Study	Selection (3)	Comparability (2)	Outcome/Exposure (1)	Total score
Martinez et al., 2012	**(method to determine prevalent delirium not stated)	**(Confounding-IM MD to dx delirium, low rate of delirium)	*(Follow up not necessary in this study type)	5/7
Douglas et al., 2013	***	**(potential tautology-parts of CAM for risk factors)	*	6/7
Carrasco et al., 2014	**(refers reader to a previous study for baseline characteristics	**(Confounding-lower rates due to delays in admission)	*	5/7
Kobayashi et al., 2013	***	**	*	6/7
Rudolph et al., 2016	***	**	*	6/7
Pendlebury et al, 2016b	***	**(Cofounding-both a prognostic and diagnostic model)	*	6/7
Wong et al., 2018	**(ages not rep. of typical age18+ 46/114 delirious ages 18-65)	**(Confounding-low rate of delirium r/t ages 18+ all included)	*	5/7

Table 9, Cont.

Newcastle Ottawa Scale for Quality Assessment Explanation.

Selection One max: 4 stars
1) Representativeness of the exposed cohort
 a. Truly representative of the average delirium study in the community*
 b. Somewhat representative of the average in the community*
 c. Selected group of users (e.g., nurses, volunteers)
 d. No description of the derivation of the cohort
2) Selection of the nonexposed cohort
 a. Drawn from the same community as the exposed cohort*
 b. Drawn form a different source
 c. No description of the derivation of the nonexposed cohort
3) Ascertainment of exposure
 a. Secure record*
 b. Structured interview*
 c. Written self-report
 d. No description
4) Demonstration that outcome of interest was not present at start of study
 a. Yes*
 b. No
Comparability (max: 2 stars)
1) Comparability of cohorts on the basis of the design or analysis
 a. Study controls for _____ [select the most important factor]*
 b. Study controls for any additional factor*
Outcome (max: 3 stars)
1) Assessment of outcome
 a. Independent blind assessment*
 b. Record linkage*
 c. Self-report
 d. No description
2) Was follow-up long enough for outcomes to occur
 a. Yes*
 b. No
3) Was there adequacy of follow-up of cohorts
 a. Complete follow-up = all subjects accounted for*
 b. Subjects lost to follow-up unlikely to introduce bias (e.g., small number or percentage lost)
 c. No statement

Table 10

Levels of Evidence Description

Level of Evidence	Description
I	Evidence from a systematic review or meta-analysis of all relevant RCT;s (randomized controlled trial) or evidence-based clinical practice guidelines based on systemic reviews or RCTs or three or more RCTs of good quality that have similar results.
II	Evidence obtained from at least one well-designed RCT (e.g. large multisite RCT)
III	Evidence obtained from a well-designed controlled trial without randomization (i.e. quasi-experimental).
IV	Evidence from well-designed case control or cohort studies.
V	Evidence from systematic reviews of descriptive and qualitative studies (meta-synthesis).
VI	Evidence from a single descriptive or qualitative study.
VII	Evidence from the opinion of authorities and /or reports of expert committees.

Note: This level of effectiveness rating scheme is based on the following: Ackley, B.J., Swan, B. A., Ladwig, G., & Tucker, S., (2008). *Evidence-based nursing care guidelines: Medical-surgical interventions. (p. 7). St. Louis, MO: Mosby Elsevier*

Note: In the literature review tables (Table 12), the single asterisks * indicate studies of the models critiqued and compared in this literature review. The double asterisks ** indicate supporting studies for the DRPM models (Table 12).

Table 11

Delirium risk prediction model and supporting studies: Level of Evidence and Study Design

Delirium Risk Prediction Model	Supporting evidence	Level of Evidence	Study design
AWOL	Douglas et al., 2013	IV	Prospective cohort
	Wong et al., 2018	IV	Retrospective cohort
	Pendlebury et al., 2016a	IV	Prospective cohort
	Brown et al., 2017	IV	Retrospective cohort
e-NICE	Rudolph et al., 2016	IV	Prospective cohort
	Halladay et al., 2018	IV	Retrospective cohort
Gradient Boosting Machine	Wong et al., 2018	IV	Retrospective cohort
CHAID Decision Tree	Kobayashi et al., 2013	IV	Retrospective cohort
Delirium Susceptibility Score	Pendlebury et al., 2016b	IV	Prospective observational audit
Isfandiaty et al.	Isfandiaty et al., 2014	IV	Retrospective cohort
	Pendlebury et al., 2016a	IV	Prospective cohort
Clinical Prediction Rule	Martinez et al., 2012	IV	Prospective cohort
	Pendlebury et al., 2016a	IV	Prospective cohort
Delirium Prediction Score	Carrasco et al., 2013	IV	Observational cohort

Table 12

Literature Review Tables

Citation	**Purpose/ Objectives**	**Study population/ Sample/ Setting**	**Study Design/ Methods/ Major Variables/ Instruments and Measures**	**Result(s)/ Main Findings**	**Implications /critique**	**Comments Themes**	**Level of Evidence**
*Rudolph, et al., 2011 *Validation of a medical record-based Delirium risk assessment*	Improve ID of pts at high risk for delirium, develop a chart abstraction tool and validate the tool against clinical expert diagnosis	100 VA patients, medical units, One center, age 65 and over	Prospective cohort based on the previous Inouye (1993) developed prediction rule that included cognitive impairment, sensory deficits, severity of illness (APACHE II), and BUN: Creatinine ratio. Variable-delirium as an outcome Instruments: MMSE-geriatrician performed on admission and daily to Dx based on DSM-IV, Charlson Comorbidity Index, (not APACHE due to increase need for labs costly) BUN:Creat ratio Chart review for sensory deficits and cognitive impairments	-Chart abstraction took 2 minutes 19 seconds -Delirium + in 23% with incident in 14% (prevalent 9%) -cognitive impairments correlated low MMSE score -Higher Charlson comorbidity index scores correlated with chart identified severe illness. 59% had sensory impairments. -more risk factors correlated with delirium dx was statistically significant (X^2 = *9.2, df=2, P=0.01*. *C* statistic 0.65, 95% CI= .54-.76 prevalent delirium removed then the rate was not significant (X^2 = *1.3 , df=2, P=0.53.* *C* statistic 0.56, 95% CI= .42-.74	statistically significant with prevalent delirium only - simplicity to use -automated chart abstraction tool to rate delirium risk for real-time decision support to prompt targeted interventions. -Limitations were sample size and setting	Statistically insignificant Dx of delirium by a geriatrician MMSE daily is time consuming rates of delirium correlate with the study by Inouye et al. (1993). Charlson Comorbidity Index not as effective as the APACHE II. Generalizability limited due to single setting, white male Predict prevalent delirium best.	IV

Citation	Purpose/ Objectives	Study population/ Sample/ Setting	Study Design/ Methods/ Major Variables/ Instruments and Measures	Result(s)/ Main Findings	Implications /critique	Comments Themes	Level of Evidence
Isfandiaty, et al., 2014 *Incidence and Predictors for Delirium in Hospitalized Elderly patients: a Retrospective cohort study.*	Determine incidence and predictors for delirium. Develop a prediction model for delirium in an Indonesian hospital population.	457 patients, aged 60 and over, in an internal medicine dept and acute geriatric ward in Indonesia.	Retrospective cohort -March thru April of 2011. -prevalent delirium excluded -Variables in the development cohort included: 12 predictors derived from previous studies: age, gender, hypoalbuminemia, anemia, anticholinergic drugs, decrease in functional status (Barthel index), stroke, metabolic disturbance, heart disease, infection with and without sepsis, and hypoxia -Md DX delirium- no screening tools noted. Data analysis via SPPS 17.0 with SD -Cox hazard: to determine indep predictors. -Missing data: imputed	87 pts experienced delirium out of 475 patients=18.8% --no delirium= 75.5%- mean survival is 11.8 days without -59.3% of delirium occurred within the first 3 days, and 81.4% within the first 7 days of admission. Missing variables dealt with by Estimation and Maximalization (EM) to impute missing data to maintain study power- this method is preferred to excluding data. Cox proportion hazard method showed infection, decreased function status and cognitive impairment the 3 indep predictors. Logistic regression used to determine model based on 3 predictors. Hosmer-Lemeshow test P value 0.066 and AUC 0.823: CI 95% 0.776-0.877)	-risk of delirium based on days in pt. -Simple prediction model for incidence delirium in the first 14 days of admission -this study shows infection is risk -Cognitive impairment as risk-this study backs this up. 20.2% of pts had cognitive impair. -Used 2 cognitive predictors could cause Tautology (cog andADLs)	Cannot locate model to apply to patient, no explanation of application No validation internal or external Nice use of statistical models to determine weights of predictors, however this is not explained nor is how we derive a score Indonesia study, unable to generalize to general population of mixed races Note that infection with sepsis was the greatest rated predictor.	IV

Citation	Purpose/ Objectives	Study population/ Sample/ Setting	Study Design/ Methods/ Major Variables/ Instruments and Measures	Result(s)/ Main Findings	Implications /critique	Comments Themes	Level of Evidence
*Marinez, et al., 2012. *Derivation and Validation of a Clinical Prediction Rule for Delirium in Patients Admitted to a Medical Ward: an Observational Study* (Clinical Prediction Rule)	Develop /validate a clinical prediction rule for patients admitted to an internal medicine unit based on easily identifiable measures available on admission	Hospital in Spain 397 total patients Derivation cohort: Mean age 75.9 years Incidence of delirium 13% 302 total patients Validation cohort: mean age 75.8 years, Incidence of delirium 25%	Single Prospective Cohort original study Outcome: delirium dx: CAM 18 or over, May 1st- June 30th, 2008 and 2009 Admitted to medical units. Independent researchers reviewed charts for delirium with dx by MD Validation model predictors: age 85 or greater, level of dependence (more than 5 areas of dependency) , psychotropic medication (2 or more)- each one point Final model: result positive if 1 or more is the score.	Used the Hosmer Lemeshow test ROC curve analysis 2x2 table sensitivity of 93.4% CI 85- 97.2% specificity 60.6% *95%CI: 54.4- 66.8%* PPV 44.4%, 95% CI 36.9-52.1% NPV 96.5% -95% CI-92-98.5% Those in the Validation cohort were significantly more dependent for ADL's validation cohort the AUC is 0.85 -classify around 53% of this cohort as high risk, limiting interventions to this 53%, covering 93.4% that did develop delirium. DX of delirium cases may still have been missed as an internal medicine md dx it/not a psychiatrist	3 independent risk factors were age 85 or older, dependent in 5 or more ADLs, and taking anti-psychotropic medications predisposing factors are mostly related to degenerative brain disease, but triggers are related to hospitalization insults, none of the later are included in this risk model. Limitation: Dx of delirium by IM MD and not a psychiatrist thus some patients may have been undiagnosed.	Ease of use would increase the Feasibility of use upon admission Predict prevalent delirium best. For resource utilization when applying costly interventions this model is too general and includes more than 50% of the population as high risk Basic model, all on predictors not precipitating factors. *Oversimplification causing decreased specificity and lower PPV*	IV

Citation	Purpose/ Objectives	Study population/ Sample/ Setting	Study Design/ Methods/ Major Variables/ Instruments and Measures	Result(s)/ Main Findings	Implications /critique	Comments Themes	Level of Evidence
*Douglas, et al., 2013. *The AWOL Tool: Derivation and Validation of a Delirium Prediction Rule*	Develop and validate a tool to predict likelihood of developing Delirium	2 academic and one VA medical center-strength. May 2010-Nov 2010 and October 2011-March 2012 Adults over the age of 50 admitted to medical units without delirium at time of admission (no prevalent delirium included in this study) Derivation cohort-209 Validation cohort- 165 Medical and neuro pts	Prospective cohort study -excluded prevalent delirium after adjustments CAM to assess for outcome of delirium within 6 days of admit 4 items were assessed: *A*ge ≤80, Failure to spell *W*ORLD backwards, Dis*O*rientation to place, high nurse rated i*LL*ness severity (AWOL) Statistical significance was derived in both cohorts (P <0.001 in derivation and in validation cohort P<0.025 AUC derivation cohort was 0.81 (95% CI 0.73-0.90) and Validation cohort 0.69 (95% CI 0.54-0.83)-showing moderate clinical usefulness.	Predictors entered into stepwise logistic regression analysis and ID'd 4 indep. Predictors of delirium in the derivation cohort. Each assigned a value of 1 pt. The higher the score the higher the rates for delirium in the derivation cohort. 40% of the patients in risk category 3-4 developed (P<0.001) delirium, and 0 in the 0 score. Validation cohort occurred at the VA- more male than the derivation cohort. Completed in <2 minutes by RN staff- developed for bedside RN use! This tool characterizes medical pts at risk at the time of admission and could be used in trials of delirium prevention, that will include impact studies.	- research assistants and a 4th year medical student screened VA- more males than derivation cohort. The validation in a VA setting is also clinically significant as the AUC remained adequately high proving clinical usefulness. -external validation of a larger cohort is needed to determine why 0 were delirious in level 4, is the tool sensitive and specific – should show higher % with delirium in this level.	Ease of use Bedside RNs assess in daily routine No additional labs or intricate assessments CAM to assess for outcome	IV

Citation	Purpose/ Objectives	Study population/ Sample/ Setting	Study Design/ Methods/ Major Variables/ Instruments and Measures	Result(s)/ Main Findings	Implications /critique	Comments Themes	Level of Evidence
*Kobayashi, et al., 2013 *A Prediction Rule for the Development of Delirium among Patients in Medical wards: Chi-square Automatic Interaction Detector Decision Tree Analysis Model (CHAID)*	Predict delirium in pts in medical wards by the CHAID Model	2400 pts in derivation group 1170 in validation group Age≥18	-Retrospective cohort study 2009-10 at St. Luke's community hospital Tokyo Japan. Internal med unit -Predictor variables for CHAID(5): hx of delirium, dementia, aged, underlying malignancy, impaired ADL's, ETOH -Predictor variables for logistic(5) Age, hx of delirium, dementia, malignancy, EtOH abuse, and ADL impairment -Providers and nurses monitored for delirium Dx of outcome made by DSM IV -Data mining technique-CHAID and a logistic regression were compared to find the best model. -SPSS software used for analysis except for CI computed by Stata version10	CHAID Validation AUC 0.82 (95% CI:0.77-0.86) Divided into 6 levels of risk, quite low, low 1, low 2, moderate, high, and quite high- when broken down the delirium rates climbed as each level increased The logistic regression model showing all variables included to be significant Validation AUC =0.79 (95% CI:0.72-0.86) Ease of use noted. 3.8% developed delirium in derivation group, and 4.2% in the validation cohort	Significance level can be modified to fit number of comparisons. Similar sample to the derivation group for validation can lead to elevated rate of reliability Necessary to externally validate a tool such as this	Pts monitored by RN and MDs for delirium on set, however no tools used like the CAM or the NuDESC to determine changes. Combine levels to absorb the 2 highest risk levels and the 2 lowest risk levels CHAID can be broken down to ages and a score for children can be derived too. This tool is noted to be easy and flexible with use of data One of the only tools to include dementia!	IV

Citation	Purpose/ Objectives	Study population/ Sample/ Setting	Study Design/ Methods/ Major Variables/ Instruments and Measures	Result(s)/ Main Findings	Implications /critique	Comments Themes	Level of Evidence
Gonzalvo, et al., 2017 *The development of an automated ward independent delirium risk prediction model. (DEMO)*	Develop and automated ward independent delirium risk prediction model- exclusively from electronically available risk factors and increase the ability to be part of the clinical decision support system to aid providers in initiation of delirium prevention	Age ≥60 Admitted to hospital, no dx of delirium within the first 24 hours (no prevalent delirium) -Only one setting for derivation of this model- hospital in Netherlands Control group= 1066 pts Delirium group 225/646 pts included-!	Retrospective cohort study -Logistic regression analysis with a $p<0.05$ included in multivariable model as -Outcome- MD documentation of Dx. Use of Delirium Rating Scale and Delirium Obs scale to aid in Dx. No use of DSM criteria. Medication and age model preformed as well as the full model to suggested validation of the medication model as it requires no labs.	Compared 2 developed prediction scores. Addition of clinicals was irrelevant to the accuracy of the prediction model, thus they opted for the simplest model containing only age and Medications AUC of the full predictive model-0.78 AUC of the medication model-0.76- addition of lab values did not provide additional benefit and many lab values missing-state that this shows the model is not overfit	Some studies state prevalent delirium is onset within 48 hours here it is 24 hours. Not generalizable due to study setting. Not a validation study Medications used in this country may vary greatly than others thus may not be predictors of delirium in other settings Over simplistic models can cause overfitting equaling elevated AUC score and predictive capability	Study done in Netherlands in one setting. EMR used to evaluate risk prediction in these models. Both appear relative; however, they note that risk factors that are typically significant in other studies are not significant indicators in this study. Which is concerning for accuracy of data and were pts accurately dx with delirium? Dx based on DRS and DOS, whereas most studies use CAM or CAM-s	IV

Citation	Purpose/ Objectives	Study population/ Sample/ Setting	Study Design/ Methods/ Major Variables/ Instruments and Measures	Result(s)/ Main Findings	Implications /critique	Comments Themes	Level of Evidence
Pendlebury, et al., 2016a. *Delirium risk stratification in consecutive unselected admissions to acute medicine: validation of externally derived risk scores* *Published 1st and the next article adds to this.*	Determine reliability of 4 externally derived risk scores in a consecutive cohort of older acute care medicine patients externally validate and compare clinically applicable risk scores within the same data group in an external center during usual routine clinical care.	308 Adults aged ≥65, admitted in and 8-week period in 2010, 2012.	Prospective cohort- Oxford University Hospital -Model update study of 4 existing tools - CAM and DSM IV criteria. -Cohort 1 used MMSE within 24 hours <24 positive cog impairs Cohort 2 used AMTS <9 positive cog impairs -Delirium rates noted : 28/95 incident. -prevalent and incident delirium -Gathered data on demographics, admit complaint, potential risk factors, hx of dementia, vision/hearing deficits, VS, SIRS score. -acute medicine pts. -Sensitivity, specificity, PPV, NPV were measured as well as the AUC.	<u>Studies included:</u> Inouye et al., 2007 Martinez et al., 2012 Isfandiaty et al., 2012 Douglas et al., 2013 Unable to include Carrasco et al, 2014, no ability to complete Barthel index Results: AUC 0.73-0.83 for incident delirium. All scores performed better than chance, -no superior tool is found all relatively equivalent. Noted that these can facilitate targeting of multicomponent interventions. May help recognize risk, improve dx of delirium.	Superior study in the framing and reasoning for use of risk assessment tools: decreasing missed dx, focused interventions Noted some assessments required non-routinely avail. Data which required simplification (i.e. WORLD spelt backward eliminated from AWOL).	Some tools altered to fit data obtained by this patient group such as no spelling world backward instead if confusion present score here was a 1. Externally validated 4 tools in one study. More validation in differing cohorts is needed. Interested to see if there is an impact study to follow this study.	IV

Citation	Purpose/ Objectives	Study population/ Sample/ Setting	Study Design/ Methods/ Major Variables/ Instruments and Measures	Result(s)/ Main Findings	Implications /critique	Comments Themes	Level of Evidence
*Pendlebury, et al., 2016b. *Delirium risk stratification in consecutive unselected admissions to acute medicine: validation of a susceptibility score (DSS) based on factors identified externally in pooled data for use at entry to the acute care pathway.* *Delirium Susceptibility Score (DSS) developed here*	Derivation and validation of a simplified delirium susceptibility score for use on admission in clinical practice on older acute medicine patients. Builds on previous studies from this main author including a study on risk factors in 2015.	Age ≥65 308 patients 31% dx with delirium, (67/95 with prevalent and 28/95 incident)	-Prospective cohort -September to Nov 2010 and again April-June 2012 -Screened for outcome with CAM, dx with DMS-IV by MD -Tripod guidelines followed for development of a prediction tool. -NICE Guidelines for factors Reliability determined by: AUC was 0.81 (0.70–0.92), for incident delirium; odds ratios (ORs) for risk score 5–7 versus <2 were 17.9 (5.4–60.0), $P < 0.0001$ for any delirium, 8.1 (2.2–29.7), $P = 0.002$ for prevalent delirium, and 25.0 (3.0–208.9) $P = 0.003$ for incident delirium, with corresponding relative risks of 5.4, 4.7 and 13.	Detects prevalent and incident delirium Pragmatic/Simple tool. DSS had higher AUC than any other previously tested model (previously tested in this cohort), but once AUC accounted for the correction for multiple comparisons the AUC was generally comparable. Simplified form of the previous scores- derivation of new prediction score in this article. Each risk factor was removed, and AUC was analyzed to determine necessity of the risk factors. Vision is the only one that is non-significant, however removal of age >80 was removed AUC improved to 0.80 (0.74-0.86) for any, 0.74 (0.67-0.81	Higher scores indicative of increased frailty, high care needs, and poor outcomes indicating good face validity of this tool Preformed as well as other prediction scores previously compared in the study above. New score developed based on multi center risk factors and previous studies derived based on data obtained by the cohorts. Statistical analysis in this study was robust due to TRIPOD.	Simplicity could be automated in the EMR. Developed as a diagnostic and prognostic model- could it dx delirium? Pulled in relevant studies, guidelines, and framework for building a prediction model	IV

Citation	Purpose/ Objectives	Study population/ Sample/ Setting	Study Design/ Methods/ Major Variables/ Instruments and Measures	Result(s)/ Main Findings	Implications /critique	Comments Themes	Level of Evidence
**Brown, et al., (2017). *Predicting inpatient delirium: The AWOL delirium risk-stratification score in clinical practice*.	efficacy of the AWOL score in clinical practice with bedside RNs assessing at admission on all patients over the age of 50. AWOL vs normal care 2014-15.	University medical center Neuro and medical patient population Admit dates April 2014-March 2015 Retrospective cohort data pulled from charts from Nov 2013 -800 charts were randomly selected, 797 were included in final. Ages 50 & > scoring 2 or more considered high risk and prevention plan initiated.	Retrospective cohort / IMPACT study CAM to screen for delirium outcome- every shift Variable was addition of prevention plan for a score of ≥2. Compared outcomes (delirium) before and 6 mo. after implementation of the delirium care pathway which included interventions aimed to prevent delirium. ROC curve analysis completed, sensitivity and specificity noted.	Those with AWOL score of 0 =5.45% delirious with 3.11% with incident delirium. Score of ≥2 60.5% delirious, with 25% having incident delirium. AUC for incident delirium only was 0.73 (95% CI 0.60-0.85). incident delirium only group sensitivity 50.0% and specificity 89.2%. PPV-25% NPV-96.1% Only 46% of patients were scored! Need for increased education to staff prior to next IMPACT study.	Based on the AUC and specificity and sensitivity scores this model has successfully stratified patients into high and low risk- resources can be allocated when needed. Due to less sensitivity CAM or assessment monitoring should continue to be assessed to prevent missing diagnosis. Predicts both incident and prevalent delirium thus needing to eval more closely those ID'd @ higher risk to ensure delirium is not present.	Less sensitive due to the factors do not account for all the risk factors of delirium or account for all cases. Possibly seeing decreased rates of incident delirium because of the implementation of the care pathway. Not applicable to surgical populations (would need validation and re-calibration) Also validated below in Pendlebury study.	IV

Citation	Purpose/ Objectives	Study population/ Sample/ Setting	Study Design/ Methods/ Major Variables/ Instruments and Measures	Result(s)/ Main Findings	Implications /critique	Comments Themes	Level of Evidence
*Carrasco, et al., 2014. *Development and validation of a Delirium Predictive Score in older people* *(DPS)*	Develop and validate a predictive score upon admission for *INCIDENT* delirium in medically admitted patients over the age of 65.	Inclusion: adults hospitalized ≥65, admitted to general medicine unit University affiliated hospital Inclusion: Exclude aphasia, coma, or inability to perform cog. Eval. Validation cohort 104 patients, 12 developed delirium. Predictive performance AUC was 0.78 (95% CI: 0.66-0.90)	Observational prospective cohort- to develop and validate a score. CAM within 48 hours and every 48 hours thereafter to assess for delirium Data collected included: comorbidity, illness severity, functional status and laboratory data. Final tool included: Barthel Index used for functional status and BUN to assess for dehydration. No other risk factors were found to be statistically significant to add to this design.	The authors devised a formula to result the Delirium Predictive risk Score (DPS) = (1370 X BUN (mmol/l)/creatinine (µmol/l) ratio)- (4 x Barthel Index). Or conventional BUN and Creatinine measures the DPS= [5 x BUN (mg/dl)/creatinine (mg/dl)]- (3 x Barthel Index) with cutoff point -160 AUC for development cohort 0.86 (95% CI: 0.82-0.91) For cut off value of -240 & -160 due to its high sensitivity and specificity. AUC for validation cohort 0.78 (95% CI: 0.66-0.90)- higher than many tools noted thus far! + LR 3.4 -LR 0.16	Use of CAM should be increased to minimally daily with best practice once per shift. Language was not a barrier for enrollment! Barthel Index is noted to be based on the patients function status 2 weeks prior to admission, RN could assess with patient or family. Noted that its negative LR is 0.16 showing it is very good a predicting who will not develop delirium allowing focused interventions to those at greater risk.	Simple measurements, statistically sound. Needs further larger population external validation in various centers	IV

Citation	Purpose/ Objectives	Study population/ Sample/ Setting	Study Design/ Methods/ Major Variables/ Instruments and Measures	Result(s)/ Main Findings	Implications /critique	Comments Themes	Level of Evidence
Gonzalvo, et al., 2017. *Validation of an automated delirium prediction model (DEMO)); an observational study*	Validate the DEMO model that was developed at Zuyderland Medical Center and uses only electronically available data to predict occurrence of delirium.	All patients admitted to hospital over the age of 60 in Sittard and Heerlen, Netherlands. from Jan 2016-Oct 2016 450 patients included Zuyderland Medical Center which was the origination location for the DEMO tool.	-Observational study -over the age of 60 assessed every 24 hours. -Retrospective chart review was done to chart check for delirium. -Variables: age, polypharmacy and use of antidementia medications, antidepressants, anti-Parkinson's agents, anti-diabetic drugs, analgesia and sleeping tables -Applied hospital wide -Original model AUC 0.770 (95% CI: 0.736-0.804) Sensitivity 78.2% (tested positive) Specificity 63.7% (non-delirious tested negative) -DOSS (delirium observation Screening Scale) used to determine delirium-	Excluded all prevalent delirium Sensitivity≥ 0.827 Specificity≥0.779 (better than reported in original study) PPV from 43.2% (day 1) to 64.8% on day 5 NPV from 96.9 (day 1) to 93.4% day 5. AUC- was not tested Simplified models can result in overfitting Medication classes were included not specific medications- which may decrease overfitting	Dx of delirium based on chart reviews and audits for key words- could cause false positive and skew this data. Needs prospective study that includes clearer dx and onset of delirium Needs further validation outside of this developing facility and outside of the Netherlands. Focus is mainly on medications on MAR the day prior onset, some medications have a cumulative effect such as anticholinergics	Weakness, pts is asked 3 questions prior to any screening for delirium is done, if these questions are all negative then screening for delirium is not done! These questions are not validated instruments. Single center study in Netherlands not generalizable without external validation Medications listed not the same as other countries Automated daily score- Ease of use!	IV

Citation	Purpose/ Objectives	Study population/ Sample/ Setting	Study Design/ Methods/ Major Variables/ Instruments and Measures	Result(s)/ Main Findings	Implications /critique	Comments Themes	Level of Evidence
**Halladay et al., (2018) *Performance of Electronic Prediction Tools for Prevalent Delirium at Hospital Admission* *Name: "consolidated NICE rule"*	Develop and assess a prediction rule for delirium using 2 populations of veterans and compare to previously created rules.	Veterans Affairs at 118 VA medical centers Inpatient facilities Admitted with CHF ACS, Community acquired PNA, COPD D=Oct 1, 2012- Sept.30th 2013 V=Oct 1, 2013- March 31, 2014 Delirium within 24 hours of admit.	Retrospective cohorts, x2 derivation and validation cohorts. Outcome: delirium within 24 hours of admission 27625 patients included in derivation cohort and 11752 in validation Compared to the e-NICE and Pendlebury DSS (2016b) Predictors: Cognitive impairment, infection, sodium level, and age 80 or greater. Predictors obtained through the NICE guidelines and developed too from the most independent risk factors	Developed and validated this tool. Compared this tool to the eNICE and Pendlebury DSS 2016b with higher discrimination in this tool then them. AUROC, 0.91; 95% CI:0.91-0.92;p<0.001-high discriminatory value Cognitive impairment was the most important factor followed by infection, sodium level, then age. Delirium upon admission in devel. Cohort=8.5% and validation cohort=7.0% Increasing score was correlated to increased rates of delirium in all tools. 3 levels of risk, low (0-2) Intermediate (3-4) and high (5-6)	4 factors for simplicity Based on previously studies risk factors- NICE Stratify risk to id those needing more cog. Assessments. Decrease unrecognized delirium to prevent poor outcomes Could be added to EMR, flags for teams could be instituted 9however be aware of alert fatigue) Development and use of a risk assessment tool are recommended by guidelines.	Compares 2 other previously developed tools, this tool has higher AUC but indicates recommendation of additional RASS or MYBW tests for mentation with may require more charting and training of staff= poor compliance etc... Used random forest test for predictors Use of NICE meta-analysis increases validity and generalizability. HIGH proportion of men due to study at VA.	IV

Citation	Purpose/ Objectives	Study population/ Sample/ Setting	Study Design/ Methods/ Major Variables/ Instruments and Measures	Result(s)/ Main Findings	Implications /critique	Comments Themes	Level of Evidence
Sola-Miravete, et al., 2017 *Nursing assessment as an effective tool for the identification of delirium risk in older in-patients: a case-control study*	Evaluate the use of comprehen sive nursing assessment to determine risk of delirium in older in-patients from a model of care needs based on variables measured easily by nursing staff. 2nd ID predictors that are easily measured by nurses	≥65, 2013-14, admitted to surgical and medical units in Catalonia's hospital in Spain. 150 patients with incident delirium and 304 without it were studied Minimum 3 day admit Focus in predicting INCIDENT delirium	-Case- control study -CAM-S, assessed daily -Virginia Henderson's' needs model was used -Indep predictors were age, incontinence, urinary catheter, ETOH abuse, hx dementia, ability to get OOB, insomnia, and social risk. (all included in final model). -Univariate logistic regression for associated variables -Lasso technique to ensure no overfitting and generalization -cross validation x9 was for validity of model -Final model, goodness of fit ($p<0.001$) by Hosmer-Lemeshow test.	increase identification of risk factors through the nursing assessment for a care team focus to personalize plans for vulnerable patients to prevent or manage delirium Highly sensitive, specific, and high AUC showing reliability AUC- the AUC was 0.945 (95% CI: 0.922-0.970) AUC when applied to test set was 93.3. Sensitivity for predicting- 94.6% and specificity to predicting absence 89.4%. Ease of use as RNs already preform bedside assessments. Needs to be externally validated in similar populations. Populations included surgical and medical patients (GS Trauma Urology IM neurology and other medical specialties.	Use of RNs typical assessment could be expanded to an EMR tool that would signal a best practice alert to staff and clinicians – alerting to high risk for delirium. Used NICE guidelines and HELP protocols to show relationships between the Henderson model to delirium prevention interventions. Needs external validation and prospective cohort design.	Interesting variant of a DRPM As Dependency of care increases so does the risk of delirium, therefor any measure of dependency may prove useful in risk prediction models. Increased dependency is also an indication of fragility. Would this cross to CV populations and ICU?	IV

Citation	Purpose/ Objectives	Study population/ Sample/ Setting	Study Design/ Methods/ Major Variables/ Instruments and Measures	Result(s)/ Main Findings	Implications /critique	Comments Themes	Level of Evidence
*/**Wong, et al., 2018 *Development and Validation of an electronic Health record-based machine learning model to estimate delirium risk in newly hospitalized patients without known cognitive impairment.*	Develop and validate a machine learning model to predict delirium risk in patients without known cognitive impairment whom are hospitalized Compare AUC to AWOL tool currently used at this facility.	-UCSF University hospital -18 and over -non-ICU units -delirium in the first 24 hours excluded. -Devel set: 14, 227 patients non-ICU admissions 1/17-8/17 -Training set: 3996 patients, 8/2017-11/ 2017. Studied incident delirium only.	Retrospective Cohort -5 machine learning algorithms to predict delirium using 345 clinical variables available in EMR upon admission: Demographics, dx, nursing records, labs, and medications BASELINE comparison AWOL: AUC 0.678 Exclusions: AMS or confusion, ICU/ ICU admit, GCS verbal <4, ICD9 code for delirium or psychosis, NU-DESC positive CAM≥1.Any intervene. to prevent or treat delirium excluded Outcome: delirium by NuDESC or CAM ICU by nurses every 12 hours.	Gradient boosted Machine model: AUC 0.855 SET Specificity of 90% (95% CI:89-90.9%) Sensitivity: 59.7 (95% CI:52.5-66.7%) PPV=23.1 (95% CI:20.5-25.9%) NPV=97.8% (95% CI:97.4-98.1% Number to screen 4.8% or 191 cases was missed. 4 models compared including the AWOL	This tool can be adjusted based on clinical need for specificity or sensitivity- would be great to code this in a program to change recommendations of which interventions to initiate due to cost/staff resources. This center uses AWOL already to screen so use of it to compare this tool is unique! Simple, computerized assessment! Needs external validation and possibly impact studies.	Complex computer analysis. Code provided on another link site for others to validate. Ease of changing settings of S/s	IV

Table 13
Table of Charms Checklist Data

Domain	Key items	page #
SOURCE OF DATA "SYSTEMATIC REVIEW OF PREDICTION MODELS FOR DELIRIUM IN THE OLDER ADULT INPATIENT" Lindroth et al., 2018	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data): *PubMed, CINHAL, PsychINFO, SocINFO, Cochran, Web of science and Embase were searched from January 1990 to December 31st 2016.* *AIM: Through a systematic review, provide important recommendations on study design for future delirium prediction models while integrating knowledge gained from the study of both medical and surgical populations.*	1
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria) *Inclusion: age >60 years, inpatient hospital setting, developed/validated a prognostic delirium prediction model, publication dates of 1 January 1990–31 December 2016* *Exclusion: alcohol related delirium, sample size less than 50, population noted as Emergency room, palliative care and hospice, ICU, skilled nursing facilities.* *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.*	1
	Participant description *Inclusion criteria: age >60 years, inpatient, developed/validated a prognostic delirium prediction model. Exclusion criteria: alcohol-related delirium, sample size ≤50.* *Twenty-three delirium prediction models were identified, 14 were externally validated and 3 were internally validated. The following populations were represented: 11 medical, 3 medical/surgical and 13 surgical, area under the receiver operating curve range from 0.52 to 0.94*	1

	Variables Extracted*: study characteristics (study design, population and sample size), outcome measure (method of identification and diagnosis, frequency and length of screening), model performance information including the diagnostic accuracy of the delirium prediction models, calibration metrics and events per variable (EPVs), characteristics of the models (variables used in model and scoring/stratification system), cognitive measures used in the study and statistical methods applied for analysis.*	
	Details of treatments received, if relevant-NA	
	Study date *January 1990 to December 31st 2016.*	1
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome *Outcomes in this study were: Study characteristics, outcome measure (method to dx, frequency and length of follow up), diagnostic accuracy of the delirium prediction models, calibration metrics and events per variable (EPVs), characteristics; variables used in model and stratification system, cognitive measures used in the study and statistical methods applied for analysis*	2
	Was the same outcome definition (and method for measurement) used in all patients? ***Calibration:*** *Goodness of fit/ Hosmer-Lemeshow test- agreement between observed outcomes and predictions* ***Discrimination****: AUC* ***Clinical utility****: Sensitivity, specificity, positive/negative predictive values, OR's, relative risk, AUC or clinical utility curve noting model cut off values* *IF any cognitive assessments and predictive variable use per model.*	2
	Type of outcome (e.g., single or combined endpoints) *Data extraction and comparison between tools- is there 23 or 27 models?* *The average NOS quality ranking for included cohort studies was seven; six studies received the maximum of nine stars. Risk of bias was assessed using the CHARMS checklist,*	2
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)? *No mention of blinding*	
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)? *No, the candidate predictors were not part of the outcome of this review, this review focused on content of each study*	2
	Time of outcome occurrence or summary of duration of follow-up	

	No follow up needed in this study as they were evaluating studies for recommendations for future studies.	
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics) *outcome measure (method to dx, frequency and length of follow up)* *diagnostic accuracy of the delirium prediction models, calibration metrics and events per variable (EPVs)* *characteristics which included variables used in model and stratification system, cognitive measures used in the study and statistical methods applied for analysis*	2
	Definition and method for measurement of candidate predictors *Outcomes measures:* *Characteristics of studies were reported and compared, as was tool to dx delirium, Model design or statistical methods, variables used in each study (most common Pre-existing impaired cognition, sensory impairment, old age, impair ADL's, Illness severity, Infection, history of alcohol use,* *Predictive ability: reported as the AUC with a table showing each* *Model calibration: Chi-square statistics and if they had calibration plots or slopes* *EVPs to determine overfitting-Models estimating more parameters than events in a 1:10 ratio are at risk of statistical overfitting, potentially leading to overly optimistic model performance* *Clinical Utility: OR's, RR, Sensitivity and specificity, ROC curves, R^2 and integrated discrimination improvement indices as well as the clinical utility curve statistic and the decision curve analysis*	3-16
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment initiation) *Note discussed as each of these are measured within each study. Predictors not critically evaluated in this systematic review.*	
	Were predictors assessed blinded for outcome, and for each other (if relevant)? *NA-Not relevant*	
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorized) *NA- not relevant*	
SAMPLE SIZE	Number of participants and number of outcomes/events *27 or 23 studies were reviewed, medical and surgical included, all had 50 or more for the same size to prevent overfitting.*	2
	Number of outcomes/events in relation to the number of candidate predictors (Events Per Variable) *Varied between each study*	

MISSING DATA	Number of participants with any missing value (include predictors and outcomes) *Each study reported their own missing data; however, this study did not comment on this which would be a benefit to recommend how to deal with missing data for future studies*	Not addressed
	Number of participants with missing data for each predictor-NA	NA
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)-NA	NA
MODEL DEVELOPMENT	Modeling method (e.g., logistic, survival, neural network, or machine learning techniques) *Model design and methods between the models were compared and they noted how many studies used which techniques i.e.: 12 used univariate or bivariate analyses, 5 of these used bootstrapping technique to address low sample size and event size. Noted it was common to have narrow validation studies, in which external validation is needed for risk of bias is possible.*	4-5
	Modelling assumptions satisfied-N/A	
	Method for selection of predictors **for inclusion** in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome) *This review does not create a model it simply recommends inclusions for future research* *Newcastle Ottawa Scale was used to determine quality ranking*	3-16
	Method for selection of predictors **during multivariable modeling** (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike Information Criterion) *This was compared between models included in this study*	NA
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation) N/A	NA
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) Discrimination (C-statistic, D-statistic, log-rank, AUROC) measures with confidence intervals *Discussed in context of combined models, Table B9 is a comparison of the AUROC for each model included. Provides a nice visual comparison*	17
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used *Discussed in context of combined models*	
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	

<table>
<tr><td></td><td>Discussed in context of combined models- model performance was assessed through calibration and classification metrics
Clinical utility statistics such as sensitivity, specificity, positive predictive values, negative predictive values, ORs, relative risk statistics and use of decision curve analysis or clinical utility cure analysis were also collected from each delirium prediction model in reference to the model's reported cut-off value.
<table><tr><td>TP</td><td>FP</td><td></td></tr><tr><td>FN</td><td>TN</td><td></td></tr><tr><td>with delirium total</td><td>Without delirium</td><td></td></tr></table></td><td></td></tr>
<tr><td></td><td>In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)

Not discussed, though would be important to discuss for future research recommendations.</td><td>NA</td></tr>
<tr><td>RESULTS</td><td>Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)
This was also not discussed but pertinent to future research when creating or assessing usability of these prediction models.</td><td>NA</td></tr>
<tr><td></td><td>Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance; NA</td><td>NA</td></tr>
<tr><td></td><td>Comparison of the distribution of predictors (including missing data) for development and validation datasets
Missing data not discussed</td><td>NA</td></tr>
<tr><td>INTERPRETATION AND DISCUSSION</td><td>Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)
moderate predictive ability (AUROC 0.52–0.94) in 14 externally validated delirium prediction models with 8 out of 14 models using narrow validation</td><td></td></tr>
</table>

	Comparison with other studies, discussion of generalizability, strengths and limitations. *Limitations: Study design, application and reporting of statistical methods appear inadequate, low EPV result in overfitting and over optimism of the tools and increase risk of bias.* *Variable definitions too broad: functional and cog abilities- overlap data.* *And last the outcome variable, delirium, was largely non-systematic frequency of measurement ranged based on time of day and was avoided weekends- delirium is fluctuating and requires increased screening.* *Generalizability: each model tends to focus on a specific population and is difficult to generalize to all populations, some studies are also focused on prevalent vs incident delirium* *Strengths: interprofessional team, multiple perspectives for recommendations for future studies. Systematic review that was prospectively developed. Comprehensive literature search was completed. This is the first to identify study and model design issues and discusses the paucity of measurements sensitive to the spectrum of cognitive impairment.* *Implications for future research:* 1. *Model aggregation* 2. *Develop and broad validation simplifying cognitive tests that would include MCI and be sensitive to cognition* 3. *Develop dynamic models using Bayesian Networks, artificial intelligence, machine learning* 4. *Build predictors based on only data available prior to delirium onset* 5. *Twice daily assessments of delirium-screen* 6. *Include variables that are commonly available in clinical practice* 7. *Follow rigorous methods outlined by Steyerberg and Vergouwe that allow for strategies to counter low EPV, use of Akaike info criterion and Bayesian information criterion to assess model fit.* 8. *Broad Validation* 9. *Adhere to TRIPOD for reporting methods.* 10. *Focus on building two models, one elective and one emergent.*	

Domain	Key items	page #
SOURCE OF DATA "PREDICTING DELIRIUM: A REVIEW OF RISK-STRATIFICATION MODELS" Newman et al., 2015	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data): *PubMed MEDLINE (1940-present), EMBASE (1947- present), PsychINFO (1800-present), CINAHL (1981-present), and Cochrane since inception. NO date or language exclusions for studies through December 5 2013. Grey literature search also completed on studies found. PRISMA flow diagram included on pg. 409* *AIM: review studies of validated risk stratification models and discuss **barriers to use** and future research directions. Qualitative study.*	408-9
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria) *Inclusion: Original research, adult population, acute medical inpatients and presence of validation. 10 articles met this criteria and were included.*	409
	Participant description- participants here are studies of development and validation of risk prediction models for delirium in the hospital setting *Inclusion: 10 studies were included, and quality assessed by a modified version of the Newcastle Ottawa Scale for cohort studies, noting that due to nature of the research papers they discounted 2 of the scores so that the highest rating is 7 not 9.*	410
	Details of treatments received, if relevant-NA	
	Study date *retrospective literature review, inception to Dec. 5th 2013*	408
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome *Outcomes in this study were: **Variables Extracted**: each studies variables, risk-stratification model- linear vs otherwise, All models tested on a validation cohort, population, delirium outcome assessment tool, risk factors included in prediction model, model structure, statistics typically AUC LR OR or %, and compared to the validation cohort.*	410

	Was the same outcome definition (and method for measurement) used in all patients? *Outcome tool was listed and compared for each assessment model* *The outcome of this systematic review is for research in the future and identify barriers for implementation.*	408
	Type of outcome (e.g., single or combined endpoints) *Review and compare models, find barriers*	408
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)? *No mention of blinding*	
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)? *No, the candidate predictors were not part of the outcome of this review, this review focused on content of each study, though they were discussed and compared in a table form.*	410
	Time of outcome occurrence or summary of duration of follow-up *No follow up needed in this study as they were evaluating studies for recommendations for future studies.*	NA
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics) *Each predictor was discussed and noted to be included in how many of the studies included out of 10. Age, Illness severity, Cognitive impairment, BUN, ADL impairment, Model performance (validation cohorts less AUC than development), AUCs were compared, Carrasco tested +LR and -LR. Use in clinical practice was also a topic-noting no model was found to be actively used in clinical practice.*	411-412
	Definition and method for measurement of candidate predictors *Outcomes measures: AUC or LRs*	412
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment initiation) *8/10 used clinical data available at time of admission to predict subsequent delirium*	410
	Were predictors assessed blinded for outcome, and for each other (if relevant)? *NA-Not relevant*	NA
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorized) *Noted that most handled them linearly*	410

SAMPLE SIZE	Number of participants and number of outcomes/events *10 models included. AUC/C-statistic/LRs compared. This study did not reveal rates of delirium in each study or % of positive prediction*	410
	Number of outcomes/events in relation to the number of candidate predictors (Events Per Variable) *This study did not discuss EPV which would be a significant finding as EPV is essential to building prediction modeling*	NA
MISSING DATA	Number of participants with any missing value (include predictors and outcomes) *Each study reported their own missing data; however, this study did not comment on this which would be a benefit to recommend how to deal with missing data for future studies*	Not addressed
	Number of participants with missing data for each predictor-NA	NA
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)-NA	NA
MODEL DEVELOPMENT	Modeling method (e.g., logistic, survival, neural network, or machine learning techniques) *Model design and population, delirium assessment, risk factors all compared. The specific model development methods were not discussed in this review.*	
	Modelling assumptions satisfied-N/A	
	Method for selection of predictors **for inclusion** in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome) *This review does not create a model it simply recommends inclusions for future research* *Newcastle Ottawa Scale was used to determine quality ranking*	410
	Method for selection of predictors **during multivariable modeling** (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike Information Criterion) *This was not discussed*	NA
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation) N/A	NA

MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) Discrimination (C-statistic, D-statistic, log-rank, AUROC) measures with confidence intervals *Discussed in context of combined models, table B9 compares AUC and other statistics*	410
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used *These were not mentioned in this study, which is a significant limitation to this study if they are looking for clinical limitations*	
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators) *Discussed in context of combined models- model performance evaluated by AUC or LRs. Clinical utility statistics such as sensitivity, specificity, positive predictive values, negative predictive values, ORs, relative risk statistics were not discussed.*	412
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added) *Noted validation studies decreased in AUC, poor to good accuracy.*	NA
RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals) *This was also not discussed but pertinent to future research when creating or assessing usability of these prediction models.* *Reasons for lack of implementation is cited as: complex or time consuming, variables difficult to measure on admission like the APACHE II, additional cognitive tests if not done right may incorrectly diagnose. Intensive staff education would be needed to implement, compliance teams to follow, and assessing reliability.*	412
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance; *NA*	NA

	Comparison of the distribution of predictors (including missing data) for development and validation datasets *Missing data not discussed*	NA
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed) *moderate predictive ability (AUROC 0.52–0.94) in 14 externally validated delirium prediction models with 8 out of 14 models using narrow validation*	
	Comparison with other studies, discussion of generalizability, strengths and limitations. *Limitations: possible to miss some relevant studies. Studies may have been incorrectly screened out; this is a qualitative review thus no attempt to pool data from the studies.* *Strengths: First study of its kind, comparing quality and defining reasons for lack of implementation in clinical practice. Notes future focuses for research.* *They noted the heterogeneity of results and methods used to develop the models and use of risk factors.* *Implications for future research: Replicate results of current studies and compare them, develop new prediction tools focusing on fast reliable assessments with well-supported risk factors.* *Future hope: Lower cost of care and decreased mortality may be a result of timely identification, prevention, and treatment of delirium.*	412

Domain	Key items	page #
SOURCE OF DATA: "Models for Predicting Incident Delirium in Hospitalized Older Adults: A Systematic Review" Kalimisetty et al., 2017	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data): *Medical librarian customized and conducted a search for all published medical articles on delirium prediction models. Ovid MEDLINE, CINAHL, Cochran Database of systematic Reviews, EMBASE, and PsycINFO were searched using PICO- based inquiry. Terms used were variants of delirium, AMS, Acute confusional stated Acute brain syndrome, acute brain failure metabolic encephalopathy, predict, predictive, models, modeling, scores, tests testing, scoring, rules, index, and indices. Grey literature review was also completed.* ***AIM:** Summarize risk prediction models and identify the most prevalent factors of incident delirium in the older in-patient populations (65 or greater), for future build of a risk prediction model to be used by the Hospital Elder Life Program (HELP) to reduce incident delirium cases.*	69
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria) *Inclusion: PRISMA Guidelines used. English language only, older population, original research to development models with derivation and validation cohorts. excluded systematic reviews and meta-analysis. 12 articles included.*	70
	Participant description- participants here are studies of development and validation of risk prediction models for delirium in the hospital setting *Inclusion: 12 studies were included, and quality assessed by a modified version of the Newcastle Ottawa Scale for cohort studies, original scale of 1-9 was used*	70
	Details of treatments received, if relevant-NA	
	Study date *Dates not noted, nor noted in inclusion criteria*	
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome ***Variables Extracted**: study description, population, delirium assessment method, incidence of reported delirium rate, and risk factors for delirium*	70

	Was the same outcome definition (and method for measurement) used in all patients? *Quality was defined as the NOS scale.* *All other variables were defined equally*	70
	Type of outcome (e.g., single or combined endpoints) *Single endpoints, all compared in tables.*	70-4
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)? *No mention of blinding-NA*	
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)? *Yes, this study was assessing what predictors commonly are used in the RPMs. Comparative table on page 74 – 75 listing all the co-efficient and statistics of each predictor.*	74
	Time of outcome occurrence or summary of duration of follow-up *No follow up needed in this study as they were evaluating studies for recommendations for future studies.*	NA
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics) *All candidate predictors were compared that were found in the studies (20)*	74
	Definition and method for measurement of candidate predictors *Outcomes measures: RR, OR, B, HR (hazard ratio), SE (standard errors)* *Noting the most common risk factors used dementia, decreased functional status, high blood urea nitrogen-to-creatinine ratio, infection and severity of illness*	74-5
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment initiation) *Any time during admission*	71
	Were predictors assessed blinded for outcome, and for each other (if relevant)? *NA-Not relevant*	NA
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorized) *Uncertain as the models and stratification levels were not a focus in this review*	NA
SAMPLE SIZE	Number of participants and number of outcomes/events *12 models compared for predictors.* ***Aim was also to summarize models, the only summary done was comparing retro vs prospective and the predictors***	70
	Number of outcomes/events in relation to the number of candidate predictors (Events Per Variable) *This study did not discuss this.*	NA

MISSING DATA	Number of participants with any missing value (include predictors and outcomes) *Not addressed*	NA
	Number of participants with missing data for each predictor-NA	NA
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)-NA	NA
MODEL DEVELOPMENT	Modeling method (e.g., logistic, survival, neural network, or machine learning techniques) *Model design and population, delirium assessment, risk factors all compared. The specific model development methods were not discussed in this review.*	NA
	Modelling assumptions satisfied- N/A	NA
	Method for selection of predictors **for inclusion** in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome) *This review does not create a model it simply reviews predictors used in RPM for a future build of an RPM for use with HELP interventions* *Newcastle Ottawa Scale was used to determine quality ranking*	70
	Method for selection of predictors **during multivariable modeling** (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike Information Criterion) *This was not discussed*	NA
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation) N/A	NA
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) Discrimination (C-statistic, D-statistic, log-rank, AUROC) measures with confidence intervals *Discussed in context of combined models, table 4 compares statistics of the specific predictors*	74-5
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used *These were not mentioned in this study, as they were looking for predictors and to summarize the current models*	70

MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators) *Not discussed.*	412
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added) *Noted that these models were scored based on the NOS and noted to be fair to good rating.*	72
RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals) *NA*	NA
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance; *NA*	NA
	Comparison of the distribution of predictors (including missing data) for development and validation datasets *Missing data not discussed, comparison of predictors was the aim of this study*.	70-5
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed) *Very little data summarized for these RPM, the aim was to summarize them however the only summary given was a comparison of how many studies were prospective vs retrospective, rates of delirium, number of risk factors included in each was 2-6, and the most/least common risk factors.*	70

	Comparison with other studies, discussion of generalizability, strengths and limitations. *Limitations: variation in the studies in the way they assessed for delirium. Decreased NOS due to studies did not include follow up data. Varied incidence of delirium between retrospective and prospective cohorts (common finding to be less rated as delirious in a retrospective study).* *Strengths: consistent with previous research on predictors-high number of predictors compared to study numbers (delirium is multifactorial in nature).* *Implications for future research: application of these predictive variables to a future tool for implementation with HELP interventions, making an automated tool to be used in the EMR.* *Future hope: The authors noted that they have already implemented a tool to mark pt. as at risk for delirium based on the Acute Care for Elders Tracker (ACE). However, the risk factors for HELP may not be the best sued in the HER due to missing data. Use of the HER with simplified variables may aid in more accurate RPMs.*	70-3

Table 14

Statistical Comparison of the Validated Delirium Risk Prediction Models

Study	Δ£≠œ: Type of study Validation Sample	Delirium diagnosis (identified by)	Delirium cases identified	Risk factors included in the tool (weighting)	Modelling approach/ Discrimination values	Reported risk levels	Observed delirium cases, by risk level	Discrimination and events per variable (EPV)
Martinez, 2012 Clinical prediction rule for delirium	£ n = 302 Mean age= 76.8 Medical	CAM-s (researchers)	**76 (25.5%)** Incident	•Age ≥85 †DADLs *Drugs 1 point each Cut off ≥1	Regression AUC 0.85 (0.80-0.90) Se=93.4% Sp=60.6% PPV=44.4% NPV=96.5%	0 Factors 1 Factor ≥2 Factors	3.5% (5/142) 23% (18/77) 64% (53/83) **Writer conclusions made:** 44.4% with a score of ≥1 or high risk, developed delirium and only 7% of those low risk developed delirium Limiting interventions to 53% of this population, making rationing interventions possible	EPV 25

Study	Δ£≠œ: Type of study Validation Sample	Delirium diagnosis (identified by)	Delirium cases identified	Risk factors included in the tool (weighting)	Modelling approach/ Discrimination values	Reported risk levels	Observed delirium cases, by risk level	Discrimination and events per variable (EPV)
Douglas, 2013 The AWOL Tool	£ n=165 Mean age =70.72 Medical /Neuro Time=<2 minutes	CAM-s (physicians)	**14 (8%)** Incident	Age ≥80 ¡WORLD Orientation RN rated Severity of illness 1 point each Cut off ≥2	Logistic Regression AUC 0.69 (0.54-0.83)	0 Factors 1 Factor 2 Factors 3 Factors 4 Factors	4% (1/25) 3% (2/59) 10% (5/49) 21% (6/28) 0% (0/4) **Writer Conclusions made:** 78% of those dx with delirium were correctly categorized as elevated risk, thus 21% of those with delirium would have been missed and not included in the intervention group. Interventions would be limited to 81 out of 165 or 49% of the total population making risk stratification possible. Including all the factors 1 or more would include 99% of all delirious but increase the number treated to 140, which is nearly 85% of the population making triaging less effective.	EPV 3.5

Study	Δ£≠œ: Type of study Validation Sample	Delirium diagnosis (identified by)	Delirium cases identified	Risk factors included in the tool (weighting)	Modelling approach/ Discrimination values	Reported risk levels	Observed delirium cases, by risk level	Discrimination and events per variable (EPV)
Kobayashi, et al., 2013 CHAID	£ N=1170 Mean age=78.5 delirious and 64.8 no delirium Medical	DMS IV TR (physician)	**51 (4.4%)** Incident	Delirium hx Age Underlying malignancy ADL impairment	Chi Square AUC 0.82 (0.76-0.88)	0 Factors 1 Factor 2 Factors 3 Factors 4 Factors 5 Factors	0.0% 1.8% 1.5% 2.5% 9.4% 46.4% Writer Conclusions: this is an algorithm, thus those in moderate to high risk would be those you would focus interventions on. 1170 pts included; 51 cases identified. If scoring included factors 3-5 only approx. 3% of cases would be missed. (factors 0-2). Of note zero patients with a zero score were delirious and the highest % delirious were included in all 5 factors!	EPV 10.2

Study	Δ£≠œ: Type of study Validation Sample	Delirium diagnosis (identified by)	Delirium cases identified	Risk factors included in the tool (weighting)	Modelling approach/ Discrimination values	Reported risk levels	Observed delirium cases, by risk level	Discrimination and events per variable (EPV)
Rudolph, 2016 ***E-NICE Rule*** A separate look at e-Nice plus mRASS	£ N=246 Mean age=72.1 (E-Nice-Time 8 seconds Electronically) Medical *Added m*	DSM IV TR Geriatric physician	**43 (19%)** Incident	Cognitive Impairment 4 Age≥65 2 Age>80 3 Fracture 4 Vision 1 Severe illness 2	Chi Square and Rank Sum AUC 0.68 (0.59-0.77) *+mRASS= AUC 0.72*	0-2 score 3-5 score 6-8 score ≥9 score add score (mRASS added)	10.4% (10/96) 14.3% (6/42) 23% (13/55) 43.8%(14/32) If you increase this to include positive mRASS would increase # to treat to 108 with 40.7% delirium positive and TMYB # to treat is 127 with 39.4% delirium positive. But if you drop the rate to include a score of ≥3 you capture 33 delirious out of 129 or 25.5% delirious rate and treating 129/246 or 52.4% of the total of people. Which is still allowing for triaging and allocation limitations but is decreasing those at risk that are missed. I recommend dropping the score to 3 or more.	4= EPV 16 5= EPV 12.8

Study	Δ£≠œ: Type of study Validation Sample	Delirium diagnosis (identified by)	Delirium cases identified	Risk factors included in the tool (weighting)	Modelling approach/ Discrimination values	Reported risk levels	Observed delirium cases, by risk level	Discrimination and events per variable (EPV)
Pendlebury 2016a External validation and update of DRPM's	œ N=308 Mean age=81 Medical	CAM DSM IV	**28 (9.1%)** Incident	Inouye 1993 updated Vision **SIRS≥2** MMSE<24 BUN/Cr rat. Martinez, 2012 Updated Age≥85 ∑**Functional dependence** Psychotropic *** drugs** Isfandiaty Cognitive impairment Functional dependency Infection no sepsis/w sepsis Douglas2013 Age≥80 WORLD **ΩDisoriented to place** RN-Illness severity	Inouye 1993 updated AUC=0.73 (0.62-0.84) Martinez, 2012 AUC= 0.78 (0.68-0.88) Isfandiaty AUC=0.83 (0.74-0.91) Douglas 2013 AUC=0.78 (0.68-0.88)	1 Factor 2 Factors 3 Factors 4 Factors 1 Factor 2 Factors 3 Factors 1 score 2 score 3 score 4 score 5 score 6 score 7 score 1 Factor 2 Factors 3 Factors 4 Factors	Se, Sp, ppv, npv .95, .34, .19, .98 .52, .80, .31, .91 .14, .96, .38, .87 (unreported) Se, Sp, ppv, npv .95, .36, .19, .98 .81, .68, .29, .96 .38, .88, .35, .90 Se, Sp, ppv, npv 1.0, .34, .20, 1.0 .95, .43, .21, .98 .90, .55, .25, .97 .81, .71, .31, .96 .57, .77, .29, .92 .48, .95, .59, .92 .33, .93, .44, .90 Se, Sp, ppv, npv .95, .18, .16, .96 .95, .66, .27, .94 .76, .66, .27, .94 .27, .93, .70, .68 Writer conclusion: lower NPV and higher Sp than it is to have a high PPV or low Se because including more patients in risk levels that MAY develop delirium is imperative to prevention.	Martinez EPV 9.33 All other studies EPV 7

Study	Δ£≠œ: Type of study Validation Sample	Delirium diagnosis (identified by)	Delirium cases identified	Risk factors included in the tool (weighting)	Modelling approach/ Discrimination values	Reported risk levels	Observed delirium cases, by risk level	Discrimination and events per variable (EPV)
Pendlebury 2016b Delirium Screening Score (DSS)	£ N=308 Mean age=81 Medical	CAM DSM IV	**28 (9.1%) Incident**	AGE ≥80 2 ≈Cognitive Impairment2 SIRS ≥2 1 Infection 1 Vision impairment 1 (Based on UK NICE guidelines and previous comparison study above)	AUC 0.81 (0.70–0.92)	1 score 2 score 3 score 4 score 5 score 6 score 7 score	Se, Sp, ppv, npv 1, .17, .16, 1 .95, .19, .16, .96 .86, .49, .21, .95 .81, .61, .25, .95 .71, .88, .50, .95 .29, .95, 0.5, .89 0, 1, 1, .86 Writer conclusion: this is a well-researched study with a very high AUC. The risk stratification levels correlate well with the actual diagnosed cases of delirium. Well researched risk factors contribute to the high accuracy of this test. This was a second study based on the same population as of their update model study 2016a. This model out preforms the others, as they learned much from the previous study and previous DRPMs.	EPV (28/5) = 5.6

Study	Δ£≠œ: Type of study Validation Sample	Delirium diagnosis (identified by)	Delirium cases identified	Risk factors included in the tool (weighting)	Modelling approach/ Discrimination values	Reported risk levels	Observed delirium cases, by risk level	Discrimination and events per variable (EPV)
Carrasco, 2014 Delirium risk prediction score	£ N=104 Mean age=75.5 Medical	CAM-Spanish version	**12 (11.7%) Incident**	Bun Creatine ratio Barthel index DPS Equation: (1370 X BUN (mmol/l)/creatine (µmol/l) ratio)- 4 X Barthel Index) Cut off -240 Or conventional units DPS= [5X BUN (mg/dl)/creatinine (mg/dl)} –(3 X Barthel index) Cut off -160	AUC 0.78 (0.66-0.90)	>-240	Sp-0.74 Se-0.88 NPV- 0.99 Writer conclusion: So, 99% of those that test negative do not develop delirium, thus including more of those at truly high risk. Here out of 104 people	EPV 4

Study	Δ£≠œ: Type of study Validation Sample	Delirium diagnosis (identified by)	Delirium cases identified	Risk factors included in the tool (weighting)	Modelling approach/ Discrimination values	Reported risk levels	Observed delirium cases, by risk level	Discrimination and events per variable (EPV)
Wong, 2018 Machine learning models **Four electronic tools compared**	£ N=3996 Mean age= Medical Computerized scoring. Automated, based on chart data	Nu-DESCCAM-ICU	**191 (4.8%)**	AWOL risk factors as stated above in table Hundreds of risk factors for each e learning program. GBM-345 LR- 111 RF- 114	Penalized logistic regression AWOL AUC-0.678 SE- 32.8% SP-90.5% GBM-AUC 0.855 Logistic regression model AUC-0.854 Random Forest AUC-0.843	4 Factors-see above AWOL Risk levels determined by setting sensitivity and specificity within the computer models.	Writer conclusion: easy to adjust the Se and Sp to your desired levels for these machine learning tools. Higher Sp and lower NPV would be the goal in order to balance interventions with less misses of those that do develop delirium. This method would also be flexible enough to continue to modify these values based on data obtained during say a QI project or external validation study.	AWOL EPV-20 Unable to calculate others as it is not directly stated the number of predictors used.

1 **Note.** Δ£≠œ Type of study: Δ Development and validation on same cohort, £ Development and validation using a different cohort, ≠ Validation only, Œ Model update with validation

• Age ≥85

† DADLs: dependence in five or more activities of daily living

* Drugs prescribed prior to admission, one point for antidepressants, antidementia, and anticonvulsants and two points for antipsychotics

¡ Ability to spell WORLD backwards

∑ researchers assess patient in six activities of daily living

Ω replaced by dx of dementia or cognitive score cut-off point

≈ MMSE