Published in final edited form as:

J Perinat Neonatal Nurs. 2015 ; 29(4): 315–344. doi:10.1097/JPN.0000000000000135.

A Systematic Review: The Utility of the Revised Version of the Score for Neonatal Acute Physiology among Critically Ill Neonates

Shannon Morse, MS, ARNP [Doctoral Student],

University of South Florida, College of Nursing

Maureen Groer, RN, PhD [Gordon Keller Professor],

University of South Florida, College of Nursing

Melissa M. Shelton, PhD, RN [Assistant Professor],

University of South Florida, College of Nursing

Denise Maguire, PhD, RN, CNL [Assistant Professor], and

University of South Florida, College of Nursing

Terri Ashmeade, MD

University of South Florida, College of Medicine

Abstract

The revised version of the Score for Neonatal Acute Physiology (SNAP-II) has been used across all birth weights and gestational ages to measure the concept of severity of illness in critically ill neonates. The SNAP-II has been operationalized in various ways across research studies. This systematic review seeks to synthesize the available research regarding the utility of this instrument, specifically on the utility of measuring severity of illness sequentially and at later time points. A systematic review was performed and identified 35 research articles that met inclusion and exclusion criteria. The majority of the studies used the SNAP-II instrument as a measure of initial severity of illness on the first day of life. Six studies utilized the SNAP-II instrument to measure severity of illness at later time points and only two studies utilized the instrument to prospectively measure severity of illness. Evidence to support the use of the SNAP-II at later time points and prospectively is lacking and more evidence is needed.

Keywords

Neonatal severity of illness; SNAP-II; physiologic instability

Corresponding Author: Shannon Morse, MS, ARNP, University of South Florida, College of Nursing, 12901 Bruce B. Downs Blvd, Tampa, FL 33612, smorse@health.usf.edu.

Conflicts of Interest

All other authors report no conflicts of interest or funding received for the creation of this manuscript.

INTRODUCTION

Critically ill neonates born every day require the progressively high technological care that is available in the neonatal intensive care unit (NICU). The demand for neonatal critical care is on the rise. Halpern and Pastores found that neonatal intensive care beds in the United States have increased by eight percent¹ and the infant mortality rate still remains high compared to other developed countries.² Specifically, the 2010 US infant mortality rate of 6.1 deaths per 1,000 live births was the highest infant mortality rate among 26 developed countries included in the Organization for Economic Co-operation and Development.² Although NICU care has improved over the past 30 years and survival rates are increasing,³ advances are needed for the improvement of critical care for the most vulnerable newborns.

Clearly decreasing the infant mortality rate is a critical goal; however, neonatal survival is no longer the ultimate goal of neonatal intensive care. Vulnerable neonatal populations, specifically the youngest and smallest babies are at greatest risk of mortality as well as a lifetime of morbidity following a NICU admission.^{4,5} Researchers and clinicians are working diligently to optimize care for these critically ill neonates and thereby decrease the risk of mortality as well as reduce the risk of lifetime morbidity. Utilization of a tool to operationalize the degree of illness severity for neonates has potential for use as a bedside clinical tool as well as a research tool. Current trends utilizing electronic medical records (EMR) in the NICU could allow for automation of an illness severity score calculation by using computer code to pull chart data to assign an illness severity score. Operationalizing illness severity scores based on physiologic data points has the potential to help the bedside clinician establish risk at birth and monitor illness severity throughout the patient's admission. It is also important to consider the role of illness severity scores in neonatal research. An illness severity score has the potential to assist researchers in accounting for the level of illness severity at birth and to account for their illness severity when they evaluate for the effectiveness of the interventions. Clearly well designed research studies are needed to generate new knowledge and illness severity scores have the potential to improve the data that is obtained, thereby helping researchers to continue to improve the care that is provided for these critically ill newborns.

When designing research studies for this vulnerable population, it is important to consider the concept of neonatal illness severity. Neonatal severity of illness is a concept that is two fold: 1) determination of how acutely ill or physiologically unstable the newborn is at that time, as well as 2) assisting in the prediction of future risk of morbidity and mortality for the neonate because of this initial risk.⁶

Derivation of the score for neonatal acute physiology

One of the most common research instruments available to measure the concept of neonatal severity of illness is the revised version of the Score for Neonatal Acute Physiology (SNAP-II). The original Score for Neonatal Acute Physiology (SNAP) was developed by Richardson, Gray, McCormick, Workman, and Goldmann.⁷ One of the greatest reasons for this tool was the need to accurately compare outcomes within and among NICUs. Richardson and colleagues noted that an additional source of variance in NICU cohorts is neonatal illness severity; and without controlling for this initial variation of risk at birth,

outcome comparisons would be inaccurate.⁷ Therefore the SNAP measurement tool was created. The SNAP was specifically designed to measure the physiologic instability of the newborn across body systems that are present in the first 24 hours of life. These physiologic measurements change overtime and as such the SNAP instrument was designed to allow for sequential measurement.⁷

Severity of illness is related to risk of mortality; however, there are also perinatal risk factors that influence the risk of mortality that are independent of illness severity. Therefore the Perinatal Extension (SNAP-PE) was added to the SNAP score to quantify physiologic instability and perinatal mortality risk in one instrument.⁸ This instrument contains the full SNAP score and then adds in three additional perinatal parameters including: birth weight, small for gestational age, and the five minute APGAR score which accounts for additional perinatal risk factors of mortality.⁸

The creators of the original SNAP measurement tool generated initial evidence of validity by demonstrating that increasing SNAP scores were associated with increasing mortality rates ($p < 0.001$). Additional evidence of validity of the SNAP score was demonstrated by looking for positive relationships between other variables associated with increased illness severity. For example, positive relationships were noted between SNAP scores and additional parameters including: the length of stay ($R^2 = 0.59$), nursing acuity ($r = 0.59$), and increased need for therapeutic interventions ($r = 0.78$).⁷ Clearly when babies are more ill, they spend more days as an inpatient in the NICU, they require more interventions in general which includes a need for a higher level of nursing care. Finally, a positive correlation was noted between SNAP scores and the physician's estimation of mortality risk ($r = 0.65$).⁷ The SNAP-PE also demonstrated the ability to predict the combined physiologic and perinatal mortality risk in neonates. The SNAP-PE instrument was a better predictor than SNAP or birth weight alone (Area Under Curve (AUC) ranged from 0.91 to 0.93 ± 0.2). Furthermore, the Hosmer-Lemeshow goodness of fit (HL) test was poor for the birth weight alone, but good for other models ($p = 0.2-0.99$).⁸

The SNAP instrument had the potential to assist clinicians and researchers to quantify the concept of illness severity. However, this tool was extensive and required up to 15 minutes to evaluate many parameters including: blood pressure, heart rate, respiratory rate, temperature, PO_2 , PO_2/FIO_2 ratio, PCO_2 , oxygenation index, hematocrit, white blood cell count, immature to total ratio, absolute neutrophil count, platelet count, blood urea nitrogen, creatinine, urine output, indirect bilirubin, direct bilirubin, sodium, potassium, ionized calcium, total calcium, glucose, bicarbonate, pH, seizures, apnea, and stool guaiac.^{6,7} The authors acknowledged that the SNAP scoring system was cumbersome and their stated goal from the initial instrument development was to eventually make the tool more parsimonious.⁷ However, they needed a large cohort of neonates with SNAP data in order to further analyze and reduce the number of parameters, while still retaining the validity of the illness severity score.

Richardson et al.⁶ published the initial derivation and validation report of the SNAP-II and SNAPPE-II instruments from a cohort of 25,429 neonates across 30 sites from three neonatal networks in the US and Canada. The revised SNAP-II instrument achieved the

authors' goal of creating a more parsimonious tool that could be assessed within two to four minutes. The SNAP-II retained only six items and each item was weighted based on the beta weight from the logistic model. The parameters and possible point values were as follows: lowest mean blood pressure, lowest temperature, PO_2/FIO_2 , lowest serum pH, presence of multiple seizures, and low urine output. Just like the SNAP, the SNAP-II is a summative rating scale. The highest possible score is 115. The SNAPPE-II is also the summative rating scale as listed above and it adds three additional parameters including: birth weight, small for gestational age, and the five minute APGAR score. The highest possible SNAPPE-II score is 162. The higher the SNAP-II or SNAPPE-II score is, the more severely ill and physiologically unstable is the neonate. As described by the instrument authors, the SNAP-II was designed to measure the mortality risk due to physiologic instability and the SNAPPE-II was designed to measure the combined physiologic and perinatal mortality risk. Because perinatal factors will not change over time, the SNAPPE-II was only designed to be measured once with data from the first 12 hours following birth. However, the SNAP-II instrument was designed to assess and quantify the physiologic signs of illness that can be assessed clinically. These physiologic derangements can change over time; therefore the SNAP-II may be useful to measure severity of illness over time.

Parsimony was a strength of this revised tool, however, equally important is that the points assigned to each parameter were empirically derived from the beta weights from the logistic model for SNAP-II.⁶ While some tools have been designed to measure illness severity in specific birth weights of neonates, the SNAP-II instrument was designed as a universal tool for critically ill newborns of all gestational ages and birth weights. The SNAPPE-II demonstrated good sensitivity, specificity, and Hosmer-Lemeshow goodness of fit (HL) across all birth weights. When stratifying the cohort via birth weight, the values were as follows: 1) across all birth weights (AUC 0.91 ± 0.01 ; HL $p = 0.9$), 2) less than 1500 grams (AUC 0.85 ± 0.01 ; HL $p = 0.86$), and 3) greater than or equal to 1500 grams (AUC 0.87 ± 0.03 ; HL $p = 0.63$).⁶ No reliability data such as a Cronbach's alpha were reported.

PURPOSE

Specifically, this literature review will synthesize all research studies that have utilized the SNAP-II instrument since its creation. As designed, this instrument measures illness severity as physiologic instability of the neonate. While most studies used this instrument to measure illness severity in the first 12 hours of life, some studies used this instrument at later time points and for sequential measurements. While the authors of the tool are clear that the SNAPPE-II is an admission score only, the original SNAP instrument was designed for sequential measurements. However the utility of the SNAP-II at later time points and for measurements over time is unclear. This literature review seeks to describe how the SNAP-II instrument has been used across neonatal populations within the neonatal intensive care unit. Additionally, this literature review describes current evidence that is available to support the use of the SNAP-II tool at later time points and for prospective measurements of illness severity beyond the first day of life.

METHODS

All research studies that have utilized the SNAP-II instrument since it was originally derived and validated by Richardson, Corcoran, Escobar, Lee ⁶ will be reviewed. The primary author used three search engines to gather potential articles for review including: PubMed, Cumulative Index to Nursing and Allied Health Literature (CINAHL) and Web of Science. The search terms “neonatal severity of illness” and “SNAP” were used to retrieve relevant articles. The term “SNAP” was used instead of “SNAP-II” to ensure that no relevant studies were overlooked during the initial data retrieval process. Date ranges were omitted since it was feasible to evaluate older papers to understand the full history of the SNAP-II instrument. The inclusion criteria for this comprehensive review included: 1) articles available in English, 2) utilization of the SNAP-II instrument, 3) and original research. The exclusion criteria included: 1) duplication of the study from another search engine, 2) articles not available in English, 3) non-research articles including review articles, poster abstracts, conference abstracts, letters to the editor, and commentaries, and 4) studies that did not utilize the SNAP-II instrument in the study.

Once studies were identified, the primary author reviewed each study and data were extracted for synthesis. A table compiled relevant data extraction and included: 1) author and year of the publication, 2) study purpose, 3) study location (country), 4) sample size, 5) sample characteristics, 6) research design, 7) utilization of the SNAP-II instrument, 8) research findings, and 9) limitations and potential for bias (see Table 1). Data in the table were compiled in order to explore the complete utilization of the SNAP-II instrument since the revised instrument was created in 2001.

RESULTS

Derivation of included studies

Studies were evaluated based on the inclusion and exclusion criteria. The process of study selection is outlined below and is also depicted in Figure 1. The literature search terms “neonatal severity of illness” and “SNAP” yielded a total of 193 articles. Seventy-eight articles were duplicates and therefore eliminated. Six studies were excluded due to being published in foreign languages including: Chinese,⁹ Bulgarian,¹⁰ French,¹¹ Portuguese,^{12,13} and Polish.¹⁴ This yielded a total of 109 abstracts for review. Review of the abstract revealed if the measurement tool for the study included the initial Score for Neonatal Acute Physiology (SNAP) or the revised Score for Neonatal Acute Physiology (SNAP-II). Only studies that included the SNAP-II measurement tool were included, therefore an additional 59 studies were excluded. The final step in article selection was to exclude any non-research articles. Fourteen studies were excluded including: twelve-conference meeting or poster abstracts, one commentary, and one letter to the editor. This left a total of 36 papers for full review; however, one additional study was eliminated because the SNAP-II instrument was not used as an instrument for the study. Therefore, a total of 35 articles remained for inclusion in this review.

Characteristics of included studies

The SNAP-II is one of the most widely used measurement instruments used to operationalize the concept of neonatal illness severity. It has been used internationally in countries including: United States of America (USA),^{15–19} Canada,^{20–32} Spain,^{33–35} Italy,^{36,37} Brazil,^{12,38} India,^{39,40} France,⁴¹ Thailand,⁴² Netherlands,⁴³ Ireland,⁴⁴ China,⁴⁵ and Iran⁴⁶. Two studies had cohorts that were derived from populations across two countries including the USA and Canada.^{6,47} It is important to note that out of the 35 studies meeting inclusion criteria, 20 of these studies were conducted in Canada, the USA, or both. This includes the initial study of derivation and validation of the SNAP-II measurement tool.⁶

Research studies were almost equally divided between multicenter and single center studies. Three large neonatal networks participated in several studies including the Canadian Pediatric Surgery Network (CAPSNet),^{20,21,24–26} Canadian Neonatal Network (CNN),^{6,22,27,29–32} and the Vermont Oxford Network (VON).^{6,16,17,47} The sample sizes varied greatly among the studies as shown in Table 2. Samples ranged from the smallest sample of 32²⁸ to the largest cohort of 25,429.⁶ Nine studies included sample sizes less than 100^{15,28,33,34,36,39,42–44}; however there were eight larger studies with sample sizes over 1,000^{6,16,17,27,30–32,47}. Newborns in the studies varied greatly across gestational age and birth weight. While the two large validation studies included newborn cohorts across the entire continuum of gestational ages and birth weights,^{6,47} most studies focused on preterm, late preterm, or term newborns. Some studies specifically focused on very low birth weight newborns^{19,44,48} and extremely low gestational age newborns^{16,17} which are premature populations with expected risk of illness severity. Not only was the SNAP-II used across all gestational ages and birth weights, it was also used across many specific morbidities including: congenital diaphragmatic hernia,^{15,21,24,25,28,29} chorioamnionitis,^{27,48} sepsis,^{33–36,39} gastroschisis,^{20,26} patent ductus arteriosus,^{19,22} persistent pulmonary hypertension,⁴² and respiratory distress.⁴⁵

Illness severity as a predictor of mortality

Illness severity is closely tied with the concept of mortality risk. Clearly, as neonatal severity of illness increases, physiologic instability increases, and therefore the risk of neonatal death also increases. Not all of the included studies included analysis for this specific point, but much evidence is available to support that higher illness severity does lead to a higher mortality rate. Many of the included studies evaluated and reported the ability of the SNAP-II instrument to predict mortality in their study sample. Some researchers used the area under the receiver operating characteristic curve (AUC) to demonstrate the sensitivity and specificity of the SNAP-II. A result of one would be perfect discrimination, while a result of 0.5 would be completely random. The AUCs reported in these studies included: 0.76,²⁹ 0.77,¹⁵ 0.81,⁴¹ 0.82,^{39,47} and 0.88,^{12,40} Zupanic⁴⁷ further divided the study sample into birth weight above and below 1500 grams and found that the AUC was more accurate for the smaller babies (AUC 0.82 vs 0.79). Each of the results in the studies would be considered acceptable to good. It is interesting to note that the seminal article for the SNAP-II and SNAPPE-II instrument did not report an AUC for the SNAP-II, but only reported an AUC of 0.91 for the SNAPPE-II instrument across all birth weights.⁶

Some researchers used odds ratios to describe how the risk of mortality increased along with increasing illness severity. Nakwan et al.⁴² found that in a sample of neonates with persistent pulmonary hypertension (PPHN), the odds of death increased by 1.04 for every one-point increase in the SNAP-II score (OR 1.04, 95% CI 1.01–1.07, $p < 0.01$). Another study described the odds ratio of mortality for newborns with respiratory distress (OR 1.071; 95% CI 1.040–1.103, $p < 0.01$). The third and final study described the increased risk of mortality in a cohort of neonates with congenital diaphragmatic hernia (CDH) (OR 1.057, 95% CI 1.019–1.097).²⁹ Risk of mortality was also described as a Hazard ratio (HR 1.09, 95% CI 1.07–1.12, $p < 0.0001$)²⁵ as well as the relative risk of mortality (RR=1.07, 95% CI 1.0–1.1).²⁶ Nasr and Langer²⁴ demonstrated that higher SNAP-II scores are associated with mortality ($p = 0.005$). Wilson et al.²¹ demonstrated that illness severity scores (SNAP-II) predicted neonatal mortality ($p < 0.001$). Sundaram et al.³⁹ found that median SNAP-II scores were higher in the non-surviving cohort [43 (36–53.5) vs 18 (16–37) ($p < 0.001$)]. Some studies used cutoff values to define mortality risk. For example, Dammon et al.¹⁶ found that a SNAP-II score greater than 30 demonstrated that the newborn was 3.5 times more likely to die. Furthermore, when they adjusted for gestational age, the risk increased to 5.9.¹⁶

Illness severity and morbidity

Illness severity and morbidity are also closely related. Neonates that are sicker manifest increased illness severity scores and greater organ dysfunction.^{34,39} Neonates that are extremely low gestational age are also very much at risk for increased illness severity due to immaturity.^{16,17} When considering the relationship between illness severity and morbidity for neonates, researchers are also interested in the possibility of illness scores helping to predict future morbidity, future treatment needs, and treatment outcomes.

Prediction of future morbidity

Critically ill newborns have common illnesses that occur perhaps due to immaturity or illness severity such as intraventricular hemorrhage (IVH) and necrotizing enterocolitis. Dammon et al. found that SNAP-II scores greater than 30 were predictive of IVH and ventriculomegaly and echodense lesions in the brain; however they did not see a relationship with cerebral palsy, autism, or microcephaly at 24 months corrected age of life.¹⁷ Chien et al.³¹ found that the SNAP-II performed similarly to the original SNAP instrument and was able to predict IVH (AUC 0.73 ± 0.02 , $\chi^2 = 219$) and chronic lung disease (AUC 0.78 ± 0.01 , $\chi^2 = 470$).

Treatment needs

One study looked at severity of illness as a means to predict treatment needs for severely ill neonates. Specifically, Coleman et al.¹⁵ found that 78% of CDH patients with SNAP-II scores greater than 25 required extracorporeal membrane oxygenation (ECMO). Furthermore, SNAP-II scores were able to predict the use of ECMO (AUC 0.76, $p = 0.003$).

Treatment outcomes

A major area of focus is the relationship between illness severity and treatment outcomes. Wilson et al.²¹ evaluated surgical treatment outcomes of CDH patients. He used the SNAP-II scores to compare the groups of patients offered surgical treatment and those who were not. Since more than 80% of the surgical cases with SNAP-II scores between 30 and 39 survived, and the potential surgical candidates had a 100% mortality rate, the authors were advocating for more specific guidelines of when to withhold surgical treatment from the newborn.²¹ Nasr and Langer²⁴ worked to compare illness severity among CDH neonates who were inborn versus out born, indicating whether the neonate was born in that facility or if the newborn was born in another facility and later transported to the NICU. Although the inborn group was sicker as compared to the out born group [21(IQR, 7–32) vs 5(IQR, 9–12), ($p = 0.0001$)]; out born status was still associated with increased risk of mortality (OR 2.8, $p=0.04$).²⁴ This seems to indicate that perhaps the quality of care in referral centers is not as optimal as within a tertiary center that routinely cares for the most critically ill infants. Capasso et al.³⁶ focused on outcomes of septic newborns who did and did not receive immunoglobulin along with antibiotic therapy. The SNAP-II was used to demonstrate that the groups were homogeneous for severity of illness prior to comparison of the outcomes with varying treatments. The weakness of this study is that they did not clearly report when SNAP-II scores were assessed.³⁶ Maden et al.¹⁹ sought to discover if severity of illness was related to successful patent ductus arteriosus (PDA) treatment, however, she discovered that treatment success or failure was related to gestational age. As gestational age increased: PDA closure rates increased (OR 1.51 per week, 95% CI 1.14–2.01, $p = 0.004$) and gastrointestinal complication rates decreased (OR 2.41, 95% CI 0.52–0.84).¹⁹

Illness severity as a control

As noted above, neonatal illness severity is a predictor of mortality and morbidity and because of this influence illness severity must be controlled when assessing outcomes between cohorts within and among NICUs. As the need for evidence based practice continues, staff nurses will be called on to participate in more performance improvement projects as well as research studies. Accounting for and controlling for illness severity when evaluating outcomes in the NICU is a critical step to ensure accurate data interpretation. Eight of the included studies used the concept of illness severity, operationalized by the SNAP-II scores, as a means to establish group homogeneity or heterogeneity for illness severity prior to assessing outcomes between the groups. Included studies used illness severity as a control between septic newborns receiving antibiotics and immunoglobulin versus antibiotics alone,³⁶ newborns receiving surgical ligation for a PDA at a cardiac versus non-cardiac pediatric surgical center,²² newborns that had documented specific blood gas targets versus those that did not have specific blood gas targets established,²⁵ neonates admitted to the NICU during the daytime versus nighttime,³⁰ and finally two studies focused on inborn versus out born status for neonates.^{24,32} One study used illness severity scores along with birth weight to match the control group to a group of neonates that were diagnosed with CDH. Since the purpose of the study was to compare costs between CDH and non-CDH patients, it was important to ensure homogeneous samples were recruited for the comparison to be accurate.²⁸ Finally, the extremely low gestational age newborn

(ELGAN) study used severity of illness, operationalized by SNAP-II scores, to adjust for NICUs that had a greater percentage of sicker newborns as this would skew the mortality rates for comparison.¹⁶

SNAP-II scores have also been used as a control when evaluating relationships between other variables. For example, Soriasham et al.²⁷ found a relationship between chorioamnionitis and early sepsis (OR 1.6; 95% CI, 1.16–2.21) and IVH (OR, 1.6; 95% CI, 1.16–2.21). After controlling for illness severity, the relationships remained (Sepsis: OR, 5.54; 95% CI, 2.87–10.69) (IVH: OR 1.62; 95% CI, 1.17–2.24).

Illness severity on the first day of life

Most of the studies focused on assessing initial illness severity on the first day of life. Clinically this makes sense as one is establishing initial risk at birth. Specifically, this review seeks to establish the utility of the SNAP-II instrument, but it is important to note that many times when the researchers desire to report initial illness severity, they will report the SNAP-II with the perinatal extension (SNAPPE-II). This instrument accounts for three additional perinatal risk factors that will remain constant. That is why this is only assessed at one time point and not sequentially. Furthermore, most of the studies used a 12-hour data collection window for the SNAP-II score per the initial instrument design.^{6,16,18,20,25–32,41,42,45,47,48} However, three studies expanded the data collection period to 24 hours.^{15,23,43} Finally, three studies did not clearly report when the assessment was done and how many hours were used for the data collection window.^{21,22,36}

Illness severity at later time points

Conceptually, illness severity is not only appropriate to evaluate on the first day of life. In fact, illness severity is something that will change over time. This was a primary reason why the author of the instrument separated the SNAP-II from the SNAPPE-II. The SNAP-II measures the physiologic state of the newborn and has the ability to capture the continuum of illness severity as it relates to physiologic instability. Therefore, it makes sense that a measurement instrument is needed that can quantify illness severity across the entire NICU length of stay. However, only eight studies included severity of illness measurements after the first day of life. Two of those studies included more than one measurement; therefore they will be discussed in the following section.

Two of the studies focused on illness severity assessment after a transport.^{38,40} Although the assessment of severity of illness could occur on the first day of life, often times the assessment occurred later (gestational age at birth 35.3 versus gestational age at transport 35.7). Both studies also compared another instrument for severity of illness against the SNAP-II score. One study demonstrated that the SNAP-II, SNAPPE-II and TRIPS (The Transport Risk Index of Physiologic Stability) performed equally well ($p = 0.625$).³⁸ The goodness of fit was also tested across models and demonstrated a good fitting model using the Hosmer-Lemeshow Goodness of Fit (SNAP-II $p = 0.29$; SNAPPE-II $p = 0.88$; TRIPS $p = 0.49$).³⁸ The second study split the sample into surviving and non-surviving neonates. The mean age of transport and subsequent admission to the NICU was 7.01 days and 5.57 days respectively, meaning that these assessments took place sometimes 5 and 7 days after birth.

It is important to note that both instruments performed well. The AUC for the SNAP-II and TOPS (Temperature, Oxygenation, Proxy for Perfusion – Capillary Refill, and Sugar – Blood Sugar) was 0.88 and 0.89 respectively.

The remaining four studies dealt with the assessment of illness severity with septic newborns. Sundaram et al.³⁹ sought to measure severity of illness for 12 hours after the onset of septicemia. The median age of these study participants was four days with an interquartile range of three to six. Although the severity of illness assessment (SNAP-II) scores were delayed by four days for most babies, this study still demonstrated the ability to predict mortality (AUC 0.82; 95% CI 0.68–0.95).³⁹ This study also demonstrated that with SNAP-II scores over 40, there was more organ dysfunction (3.8 ± 0.4 vs 2.9 ± 0.8 , $p = 0.001$).³⁹ This supports that the SNAP-II instrument was able to discriminate between varying levels of severity of illness. The final three articles sought to measure severity of illness at the highest level of clinical illness.^{33–35} The participants in the studies ranged in age between 1–15 days,³⁵ within two days of sepsis diagnosis (range of diagnosis=3–264 hours)³⁴, and between 1 hour and 32 days.³³ There is much diversity within the actual day of assessment for these studies. Not only is there variation among the three studies, but also there is variation within each study. The main purpose of these three studies was to assess relationships between SNAP-II scores and other biomarkers that could potentially provide additional information about illness severity.

Illness severity as a sequential measure

The main purpose of this paper was to synthesize the available evidence measuring severity of illness prospectively. Out of 35 studies, this review only found two studies that measured severity of illness over time. The first study measured the concept at two time points. Specifically, the author sought to determine if there was a relationship with illness severity and outcomes associated with the treatment of PDA closure and indomethacin treatment.¹⁹ SNAP-II scores were measured at birth and then again with the first dose of indomethacin administration. No relationship was found between SNAP-II scores at either time point (birth and with first dose of indomethacin) related to: failed closure of the PDA ($p = 0.22$, $p = 0.82$) or gastrointestinal complications ($p = 0.24$, $p = 0.61$).¹⁹ The relationships that they found indicated the critical influence that gestational age plays when considering treatment success and treatment complications for this population.

The final study that measured severity of illness over several time points used initial severity of illness scores with the SNAP-II and SNAPPE-II for the first 12 hours of life and then continued to measure daily SNAP-II scores until the baby was discharged or died. The authors were specifically evaluating the relationship between severity of illness as measured by the SNAP-II and sepsis, necrotizing enterocolitis and death. The study failed to demonstrate a relationship between SNAP-II scores as an accurate predictor of these later adverse events. In fact, results demonstrated that the majority (66%) of the SNAP-II scores greater than 10 were not related to any adverse events.¹⁸ And the majority of patients that experienced a complication had a score of zero within five days of the event. Furthermore, 92% of the total SNAP-II scores were zero, thus the median SNAP-II score was zero.¹⁸

DISCUSSION

Clearly the SNAP-II instrument has evidence to support its' utility in neonatal research. Quantifying illness severity is helpful in research studies because it allows: 1) quantification of the baseline risk of mortality and morbidity for the neonate 2) quantification of how physiologically ill or unstable the newborn is at the time it is measured, 3) evaluation for homogeneous versus heterogeneous samples within and among research studies, 4) stratification of patients based on level of risk, and 5) the ability to determine initial risk at birth and then subsequent improvement or decline in the health of the neonate.

Neonatal severity of illness as operationalized by the SNAP-II has the majority of evidence available to support its sensitivity, specificity, and utility as a measure of illness severity measured during the first 12 hours of life or upon admission to the NICU. Although there is a recommended 12-hour data collection period for the SNAP-II,⁶ subsequent researchers have expanded this back to the original 24-hour window utilized by the original SNAP instrument.^{15,23,43} There is concern that treatment effects may be measured as well as initial illness severity with such a long data collection window. Although the data supported the validity of the SNAP-II instrument across all gestational ages and birth weights, Zupancic et al.⁴⁷ demonstrated that the SNAP-II is better at discriminating illness severity in very low birth weight neonates (< 1500 grams).

The specific research interest of this review was to determine if the SNAP-II instrument could be used to measure neonatal illness severity at a later time point (after the initial day of birth) and as a sequential measure. Two studies evaluated severity of illness after transport and provided evidence of validity and discrimination.^{38,40} The four remaining studies evaluated severity of illness in relationship to neonatal sepsis. Each of the studies found relationships that provided evidence of measurement of severity of illness. Sundaram et al.³⁹ found that SNAP-II scores were higher in babies that died and specifically SNAP-II scores > 40 were associated with more organ dysfunction.³⁹ The other three studies were conducted by Figueras-Aloy et al.³³ and examined relationships between illness severity and various biomarkers associated with sepsis. Each of these studies demonstrated positive relationships to indicate that the SNAP-II does have some utility as an adequate measure after the first day of life. However, these were all single center studies and only involved transported and septic newborns. There is potential that these newborns were sick enough to pass a threshold of discovery by the SNAP-II instrument.

Only two studies evaluated neonatal severity of illness over time in a prospective manner. Both Madan et al.¹⁹ and Lim and Rozycki¹⁸ were not successful at finding hypothesized relationships. Only one study measured severity of illness daily and specifically sought to evaluate the ability of the SNAP-II instrument to measure illness severity over time.¹⁸ An interesting phenomenon surfaced concerning null values. Lim and Rozycki¹⁸ found that 92% of the SNAP-II scores were zero and that null values often surrounded major morbidity events. One possible explanation is that if a specific test, for example an arterial blood gas, has not been performed, then the assumed value is normal and therefore zero points are earned for the neonate. Treating missing data as normal values is questionable.

Strengths and weaknesses

A major strength of this review is that it included all of the research studies that utilized the SNAP-II instrument and was not limited to a specific age or date range. This is the first review article evaluating the utility of the SNAP-II instrument. However, the major limitation of this review is that it does not contain a review of all instruments that have been used to operationalize the concept of illness severity.

The strength of the evidence is supported by several large studies conducted across multiple hospitals. For many of the studies, this included neonatal networks that are working to promote research across the hospitals within specific geographic locations. The greatest benefit of this collaboration includes recruiting patients from different NICUs, so that the recruited sample might be more representative of the general population, thus more generalizable research findings. However, when recruiting patients across multiple hospitals, it is important to consider the effects that the environment may have on each subset of neonates. If the environmental effect is not controlled for, an ecological fallacy error can be made when interpreting results and population level findings are mistakenly attributed to individuals.⁴⁹ As a minimum standard, the interclass correlations (ICC) should be reported and provide justification as to why multilevel modeling approaches were not used. Perhaps the greatest weakness of the studies is that they were observational in nature with mostly convenience samples. This study type limits the ability of researchers to assess for cause and effect relationships among the study variables. However, the studies provide useful data within the ethical constraints of human subject research on vulnerable neonatal populations. Finally, although the studies provide measures of validity for the SNAP-II instruments, no reliability measures were reported.

Implications for research and practice

The SNAP-II instrument has been used successfully to quantify the concept of illness severity upon admission to the NICU; however, more evidence is needed concerning the sensitivity and specificity of the SNAP-II instrument when measuring illness severity sequentially and at later time points. Caution is advised when utilizing the SNAP-II for sequential measurements.¹⁸ Although there is evidence to support the validity of the SNAP-II instrument, reliability data are still lacking. Future studies using the SNAP-II instrument should report measures of reliability such as Cronbach's alpha for this purpose.

To date, the SNAP-II instrument has only been used to generate population level data for research purposes.¹⁶ One author mentioned that illness severity scores can be used to counsel parents¹⁸; however, other authors have advised caution and recommended that SNAP-II scores not be used to guide decisions for individual patients.^{6,47} However, there is a need for further development of this tool so that it can provide individual patient level data and it can be utilized as a clinical decision making tool. Furthermore, researchers need to work with informatics specialists to automate the scoring process using data that is already routinely entered by neonatal nurses into an electronic medical record.⁴⁷

CONCLUSION

Neonatal severity of illness is a key concept when caring for critically ill neonates. The SNAP-II has assisted researchers and clinicians in quantifying this concept in order to control varying severity of illness when evaluating outcomes at the population level. Although severity of illness is something that changes over time, more evidence is needed to determine if the SNAP-II instrument is able to accurately measure this at a later time point and as a sequential measure. Furthermore, the need for a precise measurement tool that can measure this concept at the individual patient level as well as the population level is the ultimate goal for improving neonatal critical care.

Acknowledgments

Source of Funding: Dr. Maureen Groer received NIH funding (R21NR3094).

References

1. Halpern NA, Pastores SM. Critical care medicine in the United States 2000–2005: an analysis of bed numbers, occupancy rates, payer mix, and costs. *Crit Care Med*. 2010; 38(1):65–71. [PubMed: 19730257]
2. MacDorman MF, Matthews TJ, Mohangoo AD, Zeitlin J. International comparisons of infant mortality and related factors: United States and Europe, 2010. *Natl Vital Stat Rep*. 2014; 63(5):1–6. [PubMed: 25252091]
3. Doyle LW. Evaluation of neonatal intensive care for extremely low birth weight infants in Victoria over two decades: I. Effectiveness. *Pediatrics*. 2004; 113(3 Pt 1):505–509. [PubMed: 14993541]
4. Mathews TJ, MacDorman MF. Infant mortality statistics from the 2008 period linked birth/infant death data set. *Natl Vital Stat Rep*. 2012; 60(5):1–28. [PubMed: 24974588]
5. Mathews TJ, MacDorman MF. Infant mortality statistics from the 2009 period linked birth/infant death data set. *Natl Vital Stat Rep*. 2013; 61(8):1–28. [PubMed: 24979974]
6. Richardson DK, Corcoran JD, Escobar GJ, Lee SK. SNAP-II and SNAPPE-II: Simplified newborn illness severity and mortality risk scores. *J Pediatr*. 2001; 138(1):92–100. [PubMed: 11148519]
7. Richardson DK, Gray JE, McCormick MC, Workman K, Goldmann DA. Score for Neonatal Acute Physiology: a physiologic severity index for neonatal intensive care. *Pediatrics*. 1993; 91(3):617–623. [PubMed: 8441569]
8. Richardson DK, Phibbs CS, Gray JE, McCormick MC, Workman-Daniels K, Goldmann DA. Birth weight and illness severity: independent predictors of neonatal mortality. *Pediatrics*. 1993; 91(5):969–975. [PubMed: 8474818]
9. Lin HJ, Ma XL, Shi LP, Luo F, Du LZ. Predicting outcome in necrotizing enterocolitis with the score for neonatal acute physiology: a retrospective study of 62 cases. *Zhonghua Er Ke Za Zhi*. 2013; 51(5):326–330. [PubMed: 23941836]
10. Vakrilova V. Scoring systems for assessing illness severity and predicting outcome in very low birth weight infants. *Akush Ginekol (Sofiia)*. 2011; 50(1):37–41. [PubMed: 21695942]
11. Smeesters PR, Johansson AB, Coppens S, Blum D, Vanderlinden R, Kahn A. The pain of the newborn: between reality and perception. *Arch Pediatr*. 2005; 12(9):1332–1337. [PubMed: 16023841]
12. Zardo MS, Procionoy RS. Comparison between different mortality risk scores in a neonatal intensive care unit. *Rev Saude Publica*. 2003; 37(5):591–596. [PubMed: 14569334]
13. Bastos G, Gomes A, Oliveira P, da Silva AT. A comparison of 4 pregnancy assessment scales (CRIB, SNAP, SNAP-PE, NTSIS) in premature newborns. *Clinical Risk Index for Babies. Score for Neonatal Acute Physiology. Score for Neonatal Acute Physiology-Perinatal Extension. Neonatal Therapeutic Intervention Scoring System*. *Acta Med Port*. 1997; 10(2–3):161–165. [PubMed: 9235848]

14. Maruniak-Chudek I, Swietlinski J. The intensity of metabolic acidosis and serum lactate concentrations in relation to illness severity evaluated by SNAP in newborns admitted to intensive care unit. *Przegl Lek.* 2002; 59(Suppl 1):78–82. [PubMed: 12108082]
15. Coleman AJ, Brozanski B, Mahmood B, Wearden PD, Potoka D, Kuch BA. First 24-h SNAP-II score and highest PaCO₂ predict the need for ECMO in congenital diaphragmatic hernia. *J Pediatr Surg.* 2013; 48(11):2214–2218. [PubMed: 24210188]
16. Dammann O, Shah B, Naples M, et al. Interinstitutional variation in prediction of death by SNAP-II and SNAPPE-II among extremely preterm infants. *Pediatrics.* 2009; 124(5):e1001–1006. [PubMed: 19858146]
17. Dammann O, Naples M, Bednarek F, et al. SNAP-II and SNAPPE-II and the risk of structural and functional brain disorders in extremely low gestational age newborns: the ELGAN study. *Neonatology.* 2010; 97(2):71–82. [PubMed: 19672122]
18. Lim L, Rozycki HJ. Postnatal SNAP-II scores in neonatal intensive care unit patients: relationship to sepsis, necrotizing enterocolitis, and death. *J Matern Fetal Neonatal Med.* 2008; 21(6):415–419. [PubMed: 18570120]
19. Madan J, Fiascone J, Balasubramanian V, Griffith J, Hagadorn JI. Predictors of ductal closure and intestinal complications in very low birth weight infants treated with indomethacin. *Neonatology.* 2008; 94(1):45–51. [PubMed: 18196930]
20. Stanger J, Mohajerani N, Skarsgard ED. Practice variation in gastroschisis: factors influencing closure technique. *J Pediatr Surg.* 2014; 49(5):720–723. [PubMed: 24851755]
21. Wilson MG, Beres A, Baird R, Laberge J-M, Skarsgard ED, Puligandla PS. Congenital diaphragmatic hernia (CDH) mortality without surgical repair? A plea to clarify surgical ineligibility. *J Pediatr Surg.* 2013; 48(5):924–929. [PubMed: 23701761]
22. Wong C, Mak M, Shivananda S, et al. Outcomes of neonatal patent ductus arteriosus ligation in Canadian neonatal units with and without pediatric cardiac surgery programs. *J Pediatr Surg.* 2013; 48(5):909–914. [PubMed: 23701758]
23. Zwicker JG, Grunau RE, Adams E, et al. Score for neonatal acute physiology-II and neonatal pain predict corticospinal tract development in premature newborns. *Pediatr Neurol.* 2013; 48(2):123–129. e121. [PubMed: 23337005]
24. Nasr A, Langer JC. Influence of location of delivery on outcome in neonates with congenital diaphragmatic hernia. *J Pediatr Surg.* 2011; 46(5):814–816. [PubMed: 21616232]
25. Brindle ME, Ma IW, Skarsgard ED. Impact of target blood gases on outcome in congenital diaphragmatic hernia (CDH). *Eur J Pediatr Surg.* 2010; 20(5):290–293. [PubMed: 20509108]
26. Mills JA, Lin Y, Macnab YC, Skarsgard ED. Perinatal predictors of outcome in gastroschisis. *J Perinatol.* 2010; 30(12):809–813. [PubMed: 20357809]
27. Soraisham AS, Singhal N, McMillan DD, Sauve RS, Lee SK. A multicenter study on the clinical outcome of chorioamnionitis in preterm infants. *Am J Obstet Gynecol.* 2009; 200(4):372. e371–376. [PubMed: 19217596]
28. Lam JC, Claydon J, Mitton CR, Skarsgard ED. A risk-adjusted study of outcome and resource utilization for congenital diaphragmatic hernia. *J Pediatr Surg.* 2006; 41(5):883–887. [PubMed: 16677875]
29. Skarsgard ED, MacNab YC, Qiu Z, Little R, Lee SK. SNAP-II predicts mortality among infants with congenital diaphragmatic hernia. *J Perinatol.* 2005; 25(5):315–319. [PubMed: 15716986]
30. Lee SK, Lee DS, Andrews WL, Baboolal R, Pendray M, Stewart S. Higher mortality rates among inborn infants admitted to neonatal intensive care units at night. *J Pediatr.* 2003; 143(5):592–597. [PubMed: 14615728]
31. Chien LY, Whyte R, Thiessen P, Walker R, Brabyn D, Lee SK. Snap-II predicts severe intraventricular hemorrhage and chronic lung disease in the neonatal intensive care unit. *J Perinatol.* 2002; 22(1):26–30. [PubMed: 11840239]
32. Chien LY, Whyte R, Aziz K, Thiessen P, Matthew D, Lee SK. Improved outcome of preterm infants when delivered in tertiary care centers. *Obstet Gynecol.* 2001; 98(2):247–252. [PubMed: 11506840]

33. Figueras-Aloy J, Gomez-Lopez L, Rodriguez-Miguel JM, et al. Serum soluble ICAM-1, VCAM-1, L-selectin, and P-selectin levels as markers of infection and their relation to clinical severity in neonatal sepsis. *Am J Perinatol.* 2007; 24(6):331–338. [PubMed: 17564956]
34. Figueras-Aloy J, Gomez-Lopez L, Rodriguez-Miguel JM, et al. Plasma endothelin-1 and clinical manifestations of neonatal sepsis. *J Perinat Med.* 2004; 32(6):522–526. [PubMed: 15576275]
35. Figueras-Aloy J, Gomez L, Rodriguez-Miguel JM, et al. Plasma nitrite/nitrate and endothelin-1 concentrations in neonatal sepsis. *Acta Paediatr.* 2003; 92(5):582–587. [PubMed: 12839289]
36. Capasso L, Borrelli AC, Parrella C, et al. Are IgM-enriched immunoglobulins an effective adjuvant in septic VLBW infants? *Ital J Pediatr.* 2013; 39:63. [PubMed: 24098953]
37. De Felice C, Toti P, Parrini S, et al. Histologic chorioamnionitis and severity of illness in very low birth weight newborns. *Pediatr Crit Care Med.* 2005; 6(3):298–302. [PubMed: 15857528]
38. Lucas da Silva PS, Euzebio de Aguiar V, Reis ME. Assessing outcome in interhospital infant transport: the transport risk index of physiologic stability score at admission. *Am J Perinatol.* 2012; 29(7):509–514. [PubMed: 22495897]
39. Sundaram V, Dutta S, Ahluwalia J, Narang A. Score for neonatal acute physiology II predicts mortality and persistent organ dysfunction in neonates with severe septicemia. *Indian Pediatr.* 2009; 46(9):775–780. [PubMed: 19430085]
40. Mathur NB, Arora D. Role of TOPS (a simplified assessment of neonatal acute physiology) in predicting mortality in transported neonates. *Acta Paediatr.* 2007; 96(2):172–175. [PubMed: 17429899]
41. Iacobelli S, Bonsante F, Quantin C, Robillard P-Y, Binquet C, Gouyon J-B. Total Plasma Protein in Very Preterm Babies: Prognostic Value and Comparison with Illness Severity Scores. *Plos One.* 2013; 8(4)
42. Nakwan N, Nakwan N, Wannaro J. Predicting mortality in infants with persistent pulmonary hypertension of the newborn with the Score for Neonatal Acute Physiology-Version II (SNAP-II) in Thai neonates. *J Perinat Med.* 2011; 39(3):311–315. [PubMed: 21488808]
43. ter Horst HJ, Jongbloed-Pereboom M, van Eykern LA, Bos AF. Amplitude-integrated electroencephalographic activity is suppressed in preterm infants with high scores on illness severity. *Early Hum Dev.* 2011; 87(5):385–390. [PubMed: 21419583]
44. Miletin J, Pichova K, Doyle S, Dempsey EM. Serum cortisol values, superior vena cava flow and illness severity scores in very low birth weight infants. *Journal of Perinatology.* 2010; 30(8):522–526. [PubMed: 20336081]
45. Ma XL, Xu XF, Chen C, et al. Epidemiology of respiratory distress and the illness severity in late preterm or term infants: a prospective multi-center study. *Chin Med J (Engl).* 2010; 123(20):2776–2780. [PubMed: 21034581]
46. Kadivar M, Sagheb S, Bavafa F, Moghadam L, Eshtrati B. Neonatal mortality risk assessment in a neonatal intensive care unit (NICU). *Iranian Journal of Pediatrics.* 2007; 17(4):325–331.
47. Zupancic JA, Richardson DK, Horbar JD, Carpenter JH, Lee SK, Escobar GJ. Revalidation of the Score for Neonatal Acute Physiology in the Vermont Oxford Network. *Pediatrics.* 2007; 119(1):e156–163. [PubMed: 17158947]
48. De Felice C, Del Vecchio A, Criscuolo M, Lozupone A, Parrini S, Latini G. Early postnatal changes in the perfusion index in term newborns with subclinical chorioamnionitis. *Arch Dis Child Fetal Neonatal Ed.* 2005; 90(5):F411–414. [PubMed: 15863488]
49. Hox, J. *Multilevel Analysis: Techniques and Applications.* 2. New York, NY: Routledge; 2010.

Figure 1.

Table 1

Review of the relevant research articles for utilizing the score for neonatal acute physiology

Authors and Date of Publication	Purpose	Country of Origin	Sample Size (N)	Sample Characteristics	Research Design	SNAP-II	Findings	Limitations and Potential for Bias
1) Stanger et al. 2014	To correlate prognostic variables with the intended and actual abdominal closure technique (primary repair versus silo) and assess related outcomes in neonates with gastroschisis	Canada	N = 679	Dx gastroschisis from 5/05 to 12/11 GA mean 35.8 BW mean 2557g Canadian Pediatric Surgery Network (CAPSNet) 17 perinatal/ surgical centers	Observational	SNAP-II used to stratify risk for gastroschisis patients. SNAP-II used as a predictor of treatment method: PR vs silo. SNAP-II measured within 12 hours of NICU admission. SNAP-II was not used as a sequential measure. SNAPPE-II was not measured.	SNAP-II scores were higher for patient's undergoing primary repair (PR). (11.1 vs 7.21, $P < 0.001$) SNAP-II score was a predictor of closure type using multivariate analysis ($P = 0.04$) The authors questioned if the SNAP-II scores were higher due to the course of treatment versus initial severity of illness. Higher SNAP-II scores were not correlated to poor outcomes.	Observational Study Duration of silo placement was not obtained; therefore situational factors versus patient factors for closure decisions may be confounding. There is potential that the higher SNAP-II score were measuring treatment effect and not initial illness severity. (Should the SNAP-II be limited to a shorter window to decrease treatment effects?) Authors questioned a type 1 error? Authors were concerned of accuracy of SNAP-II as a predictor of treatment.
2) Coleman et al. 2013	To evaluate the 24 hour SNAP-II score and highest PaCO ₂ as predictors of ECMO use in neonates with congenital diaphragmatic hernia (CHD)	USA	N = 47 ECMO n = 24 No ECMO n = 23	Dx with CHD from 2007–2010 All newborn and tx to tertiary children's hospital ECMO vs No ECMO GA 38.5/38.0 BW 3.04/2.90kg	Retrospective cohort study	Used as a predictor for ECMO tx. SNAP-II measured with data from the first 24 hours of admission. SNAP-II was not used as a sequential measure. SNAPPE-II was not measured.	78% of neonates with SNAP-II > 25 needed ECMO SNAP-II predicted ECMO use (AUC = 0.76, $P = 0.003$) SNAP-II predicted mortality (AUC = 0.77 [0.63–0.91] $P = 0.002$)	Retrospective design Single center Small sample size
3) Capasso et al. 2013	To investigate the effectiveness of IgM-enriched immunoglobulins in reducing short-term mortality of neonates with late onset sepsis.	Italy	N = 79 Tx with antibiotic and IgM n = 40 Tx with antibiotic only n = 39	VLBW neonates diagnosed with late onset sepsis from 2008–2012 Cohorts split before and after 6/1/2010 IgM/No IgM GA 27±2.6/27.6±3.9 (NS) BW 924±277/951±362 (NS) SNAP-II 15±13/12±9 (NS)	Retrospective cohort study	SNAP-II used to assess illness severity/disease severity and establish the two groups were homogeneous for this measure. Timing of the SNAP-II score was not reported. SNAP-II was not used as a sequential measure. SNAPPE-II was not measured.	No statistically significant differences in SNAP-II scores between the two groups; therefore the groups were homogeneous for illness severity. Short term mortality was significantly less in the group that received both IgM and antibiotics (9 vs 18, OR 0.16, CI 0.3–0.7, $P = 0.005$)	Observational Retrospective Single center Cohorts came from two different time points. (Cohort receiving antibiotics alone. Subsequent cohort received both IgM and antibiotics after a practice change occurred routine tx of sepsis in this institution.)
4) Iacobelli et al 2013	To evaluate the association of “severe adverse outcome” (defined as inpatient death or severe neurological injury)	France	N = 761	Derived from a single site study over 10.5 years (January 2001 to June 2011). GA 24–31 BW 1203±346g Cases were excluded if any data was missing for the calculation of	Observational Cohort Study	Measures of illness severity for comparison included: SNAP-II, SNAPPE-II, CRIB, CRIB-II, and total plasma protein at 12 hours of life.	Total plasma protein was significantly better at predicting the adverse outcome as compared to SNAP-II and CRIB-II ($p < 0.05$). Total plasma protein had similar predictive capabilities as compared to CRIB and SNAPPE-II.	Observational Single center Blood draws for plasma protein levels were not consistent and extended between 12 – 20 hours of life.

Authors and Date of Publication	Purpose	Country of Origin	Sample Size (N)	Sample Characteristics	Research Design	SNAP-II	Findings	Limitations and Potential for Bias
	and hypoproteinemia (<40g/L), and hypoproteinemia (<40g/L). To compare the predictive ability of hypoproteinemia as compared to other illness severity measurement instruments.			the CRIB, CRIB-II, SNAP-II, or SNAPPE-II		Measures of illness severity were compared for their ability to predict poor outcomes including: mortality prior to discharge or severe neurological injury. SNAP-II and SNAPPE-II were assessed with data from the first 12 hours of life. SNAP-II was not used as a sequential measure.	Total plasma protein AUC = 0.849 HL $P = 0.73$ SNAPPE-II AUC = 0.822, HL $P = 0.26$ SNAP-II AUC = 0.810, HL $P = 0.47$ Plasma protein predicted severe adverse outcomes better than CRIB-II and SNAP-II ($P < 0.05$)	
5) Wilson et al. 2013	To compare patients with congenital diaphragmatic hernia (CDH) that did and did not undergo surgical correction.	Canada	N = 275 PSC/NS n = 11 NSC/NS n = 24 Surgery/Survivor n = 224 Surgery/NS n = 16	Derived from prospective CAPSNet database from 2005 – 2009 Dx with CDH GA 26–42 (Median = 38) BW 930–4930 (Median = 3197.5) LOS 0–135 days (Median = 18) SNAP-II 5–63 (Median = 21)	Retrospective cohort study	SNAP-II was used to assess mortality risk. Desired to establish a threshold SNAP-II score as a predictor of poor surgical outcome. Timing of the SNAP-II score was not reported. SNAP-II was not used as a sequential measure. SNAPPE-II was not measured.	SNAP-II predicted neonatal mortality. ($P < 0.001$) Non-survivors had higher SNAP-II scores as compared to survivors. ($P < 0.001$) SNAP-II scores were similar between operative non-survivors and potential surgical candidates (34 vs 29, $P = 0.431$). Median SNAP-II scores for these groups was similar at 34 vs 30, $P < 0.431$) More than 80% of surgical cases with SNAP-II scores between 30 – 39 survived. 100% mortality rate for potential surgical cases	Observational Retrospective The population of interest, PSC, was small. (N = 11) Potential for critical information to be omitted from the database including: 1 confirmed cause of death 2 Undocumented congenital anomalies.
6) Wong et al. 2013	To compare neonatal outcomes of patients receiving a patent ductus arteriosus (PDA) ligation in a pediatric cardiac center versus a pediatric non-cardiac care center	Canada	N = 990 (Represents 22.9% of total sample dx with PDA) PDA Ligation at CC n = 711 PDA Ligation at NCC n = 279	Dx with PDA and requiring surgical ligation. Derived from Canadian Neonatal Network (CNN) Database from 2005 to 2009. (18 NICUs with Cardiac surgeons/9 NICUs with pediatric surgeons) Combined cohort/CC/NCC (Cohort 33 weeks and < 1500 g). GA ($P = 0.01$) 25.8 (2.0) 25.7 (2.0) 26.0 (2.0) BW ($P < 0.01$) 846 (279)g 830 (270)g 889(300)g SNAP-II ($p = 0.23$) (Median/IQR) 14 (9,28) 16 (9,29) 14, (9,24)	Retrospective Cohort Study	SNAP-II was used as a predictor of mortality risk upon admission. SNAP-II was used as a control when assessing mortality rates across this sample. The SNAP-II score was assessed upon admission; however, the time window for measurement was not reported. SNAP-II was not used as a sequential measure. SNAPPE-II was not measured.	No difference between pediatric cardiac care centers and pediatric non-cardiac care center in the SNAP-II adjusted mortality rate (8.7% vs 10.7%, $P = .32$) Overall mortality rate was 9.1% in this study. Increased morbidity including cranial ultrasound abnormalities (OR = 0.55, CI = 0.40–0.77) and sepsis (OR = 0.53, CI = 0.39–0.71) noted for neonates treated at non-cardiac care centers.	Observational Retrospective
7) Zwicker et al. 2013	To examine the relationships of risk factors (antenatal,	Canada	N = 126	Derived from current prospective ongoing study (April 2006 – September 2010).	Observational	SNAP-II was used to measure early illness	Increased SNAP-II scores were associated with a slower rise in fractional anisotropy of the	Observational Single Site

Authors and Date of Publication	Purpose	Country of Origin	Sample Size (N)	Sample Characteristics	Research Design	SNAP-II	Findings	Limitations and Potential for Bias
	perinatal, and postnatal) for abnormal brain development of the corticospinal tract. perinatal, and postnatal) for abnormal brain development of the corticospinal tract. perinatal, and postnatal) for abnormal brain development of the corticospinal tract. perinatal, and postnatal) for abnormal brain development of the corticospinal tract. perinatal, and postnatal) for abnormal brain development of the corticospinal tract.			SNAP-II scores were used to stratify the neonate's risk: <10, 10–29, 30. SNAP-II was not used as a sequential measure. SNAPPE-II was not measured.		severity in first 24 hours of life. Early illness severity, operationalized by SNAP-II, was used as the perinatal risk factor in this study. SNAP-II measured with data from the first 24 hours of admission. SNAP-II scores were used to stratify the neonate's risk: <10, 10–29, 30. SNAP-II was not used as a sequential measure. SNAPPE-II was not measured.	corticospinal tract. ($P = 0.02$) This relationship remained significant ($P = 0.009$) even after controlling for GA and postnatal risk factors. SNAP-II predicted development of the motor pathways of the brain.	
8) Lucas Da Silva et al. 2012	To assess the validity of the transport risk index of physiologic stability score (TRIPS) score in a NICU sample in Brazil. To compare the performance of the TRIPS, SNAP-II and SNAPPE-II scores in a NICU sample in Brazil.	Brazil	N = 175	Single Regional NICU GA at Birth 35.3 (27–42) GA at Transport 35.7 (27–42) SNAP-II 10±15.6 (Median = 5, IQR = 0–14) SNAPPE-II 14.4±21 (median = 5, IQR = 0–24.7)	Observational Prospective	SNAP-II and SNAPPE-II were used as a comparison tool against the TRIPS. (Criterion validity) SNAP-II and SNAPPE-II were evaluated with data from the first 12 hours of admission to the NICU after transport. (GA at transport may have been different than GA at birth; therefore, the timing of SNAP-II assessment may vary.) SNAP-II was not used as a sequential measure.	SNAP-II, SNAPPE-II, and TRIPS ROC were not statistically different ($P = 0.625$) H-L showed good calibration of each instrument. SNAP-II ($P = 0.29$) SNAPPE-II ($P = 0.88$) TRIPS ($P = 0.49$) TRIPS - AUC ranged from 0.80 – 0.99 and was most accurate for younger gestations. Overall AUC = 0.80 (95% CI = 0.64 to 0.95) GA>32 weeks – AUC = 0.71 (95% CI = 0.63 to 0.78) GA <32 weeks – AUC = 0.99 (95% CI = 0.85 to 1) AUC for SNAP-II and SNAPPE-II was not reported.	Observational Single site Small sample size TRIPS score was only measured at one time point, whereas previous studies measured it twice.
9) Nasr et al. 2011	To determine if outborn status is associated with increased mortality for patients with an antenatal diagnosis of congenital diaphragmatic hernia (CDH).	Canada	N = 140 Inborn n = 75 Outborn n = 65	Dx with CDH during antenatal period. Derived from all neonates in the Canadian Pediatric Surgery network (CAPSNet) (16 pediatric surgical centers) between 2005–2008 Inborn/Outborn GA 37.7±1.09/37±6.5 ($p=0.3$) BW 3079±655/2959±789 ($p=0.3$)	Observational Retrospective Cohort Study	SNAP-II was used to determine severity of illness SNAP-II was used to compare severity of illness between both groups. SNAP-II was not used as a sequential measure. SNAPPE-II was not measured.	Higher SNAP-II scores were associated with mortality ($P = .005$) NBs in the inborn group [21 (IQR, 7–32)] were sicker than the outborn group. [5 (IQR, 9–12)]. ($P = 0.0001$) Outborn status increased the risk of mortality. (OR 2.8, $P = 0.04$)	Observational Retrospective Inborn and outborn groups were not homogeneous for SNAP-II scores. Potential for critical information to be omitted from the database including prenatal or perinatal risk factors influencing healthcare decisions.
10) Nakwan et al. 2011	To evaluate the ability of the SNAP-II instrument to predict mortality in neonates	Thailand	N = 41 Cohorts: S n = 27	Dx with PPHN Cohort derived from a single center from June 2008 – March 2010	Observational Prospective Cohort Study	The authors divided the SNAP-II scores into categories: Mild <23 Moderate 24–42 Severe 43	SNAP-II was able to predict mortality for neonates with PPHN. SNAP-II >43 greatly increased the risk of mortality	Observational Single center Small sample size

Authors and Date of Publication	Purpose	Country of Origin	Sample Size (N)	Sample Characteristics	Research Design	SNAP-II	Findings	Limitations and Potential for Bias
	with persistent pulmonary hypertension (PPHN) with persistent pulmonary hypertension (PPHN) with persistent pulmonary hypertension (PPHN)	hypertension (PPHN) hypertension (PPHN) hypertension (PPHN)	Non-S n = 14	<u>Cohorts:</u> <u>SNS</u> GA 39.4±1.7 39.5±1.0 (<i>P</i> = 0.89) BW 3042±470 3088±477 (<i>P</i> = 0.88) Males = 73.2% SNAP-II 40.6±18.5		SNAP-II measured with data from the first 12 hours of admission. SNAP-II was not used as a sequential measure. SNAPPE-II was not measured.	The subscales of mean blood pressure (<i>P</i> = 0.03), PaO ₂ /FIO ₂ (<i>P</i> < 0.01), and urine output (<i>P</i> = 0.03) were individual predictors of mortality risk in this sample. Raising the SNAP-II score by 1 point, raised the odds of mortality by 1.04 (95% CI = 1.01–1.07, <i>P</i> < 0.01) SNAP-II scores were significantly higher in non-surviving neonates (50.1±18.5 versus 35.7±16.8, <i>P</i> = 0.02).	Only 9 neonates had a confirmed dx of PPHN by echocardiography Male/Females not equally distributed in the sample. Treatments such as extracorporeal membrane oxygenation (ECMO) and nitic oxide were not readily available for use in this NICU.
11) Ter Horst et al. 2011	First aim – To assess if a calculation of aEGG amplitudes is consistent with pattern recognition. Second aim – To assess if there is a relationship between SNAP-II scores and aEGGs of premature infants.	Netherlands	N = 38	Cohort derived from a single center from 2/2006 to 2/2007 GA 29.7 (1.4) Range 26–31.8 BW 1340 (380)g SNAP-II 13.5 (7.8)	Prospective observational study	SNAP-II was used to assess if a relationship exists between illness severity and amplitude integrated EEG (aEEG). SNAP-II was used as a measure of initial severity of illness. SNAP-II was assessed with data from the first 24 hours of life. SNAP-II was not used as a sequential measure. SNAPPE-II was not measured.	SNAP-II scores were inversely correlated with the 5 th and 50 th aEEG (<i>r</i> = −0.34, <i>P</i> = 0.0001/ <i>r</i> = −0.27, <i>P</i> = 0.001) The strongest correlations between the SNAP-II and aEEG were on the first day of life (<i>r</i> = −0.5, <i>P</i> = 0.005/ <i>r</i> = −0.47, <i>P</i> = 0.018) Relationships were the strongest on day one and were non-significant by day 5.	Observational Single center Small sample size Equipment limitation created maximum of 3 infants daily that could have EEG testing completed, therefore all newborns were not included that met inclusion criteria. Authors state that infants were still selected at random. This study evaluated the first 5 days of life only. Long term follow-up is lacking.
12) Brindle et al. 2010	To evaluate the impact of blood gas targets on outcomes for neonates with congenital diaphragmatic hernia (CHD).	Canada	N = 147 Target n = 63 No Target n = 84	Dx with CHD (64% with a prenatal dx) Derived cohort from the Canadian Pediatric Surgery Network (CAPSNet) (11 centers) from May 2005 to November 2007. <u>Cohorts:</u> <u>Target/No Target</u> GA 38 (Median) 37–40 (IQR)/ 39 (Median) 37–40 (IQR) (<i>P</i> > .05) BW 3.0 (1.2–4.4)kg 3.3 (1.5–4.9)kg (<i>P</i> > .05)	Observational Cohort Study	SNAP-II scores were used to establish that the two groups were homogeneous for illness severity. SNAP-II was used to account for mortality risk. SNAP-II scores were obtained in the first 12 hours of the NICU admission SNAP-II was not used as a sequential measure. SNAPPE-II was not measured.	Infants with established targets were less likely to die (Hazard ratio = 0.27, 95% CI = 0.11 – 0.69, <i>P</i> < 0.05) SNAP-II predicted mortality with multivariate analysis (HR = 1.09, 95% CI = 1.07–1.12, <i>P</i> < 0.0001)	Observational Study Small sample size Willingness for the neonatologist to establish blood gas parameters may be a proxy for another variable that affects mortality.
13) Miletin et al 2010	To assess the relationship between serum cortisol and illness severity and organ dysfunction in VLBW neonates.	Ireland (Study conducted in Ireland, but one author was from Czech Republic.)	N = 38	Cohort derived from single site from November 2006 through June 2007. BW 1.08±0.3 kg GA 27.8±1.9	Prospective Observational Cohort Study	SNAPPE-II, CRIB and NEOMOD instruments were all used. SNAP-II was only measured with the perinatal extension. SNAPPE-II was used as a measure of illness severity to	Correlation noted between serum cortisol and NEOMOD (<i>p</i> =0.006, value of the correlation was not reported) No relationship was found between serum cortisol and mean blood pressure.	Observational Single center Small sample size Echocardiography timing varied between 12–24 hours of life.

Authors and Date of Publication	Purpose	Country of Origin	Sample Size (N)	Sample Characteristics	Research Design	SNAP-II	Findings	Limitations and Potential for Bias
	(SNAPPE-II, CRIB, and SNAPPE-II, CRIB, and	NEOMOD).		SNAPPE-II 16 (Median) 0–80 (Range)		look for relationships between illness severity and serum cortisol levels. SNAPPE-II was measured, but the time frame for data collection was not clearly reported. SNAP-II was not used as a sequential measure.	No relationship was found between serum cortisol and SVC blood flow (echocardiography)	Timing of the second dose of antenatal corticosteroids was not reported. Urinary free cortisol may be a better marker of adrenal function; however, serum cortisol was measured. Values for the SNAP-II were not reported.
14) Mills et al. 2010	To identify perinatal risk factors that may predict outcomes in neonates born with gastrochisis.	Canada	N = 239	Dx with gastrochisis Sample recruited from May 2005 to May 2008 period across 14 centers within the CAPSNet database. GA 36.2±2.1 BW 2566.3±564.7 SNAP-II 9.2±12.3	Observational study	SNAP-II scores calculated within 12 hours of birth SNAP-II was used to predict mortality. SNAP-II score was used as a continuous and categorical variable for statistical analysis. SNAP-II was used to predict course of treatment and treatment failure. SNAP-II was not used as a sequential measure. SNAPPE-II was not measured.	SNAP-II scores predicted the relative risk of mortality (RR = 1.07, 95% CI = 1.0–1.1) SNAP-II scores predicted mechanical ventilation ($P = 0.03$ as a continuous variable) and ($P = 0.001$) as a categorical variable for SNAP-II scores > 28. TPN need was predicted by SNAP-II scores ($P = 0.006$ for continuous variable and $P = 0.001$ for SNAP scores > 28.). LOS was predicted by continuous SNAP-II scores ($P = 0.01$). LOS was also predicted by SNAP-II scores between 17 – 28 ($P = 0.05$) SNAP-II scores > 28 ($P = 0.0003$) SNAP-II scores >28 has a relative risk of surgical closure failure of 4.9, 95% CI = 1.43 to 16.8	Observational Short term outcomes only
15) MA Xiaoli, et al. 2010	To explore respiratory distress in late preterm/term infants and compare severity of illness instruments (SNAP-II and ACoRN – Acute care of at-risk newborns)	China	N = 503 Mild RD n = 237 Mod RD n = 227 Severe RD n = 39	Dx with respiratory distress requiring CPAP or ventilator support Cohort derived from seven centers in China from November 2008 to October 2009. Cohorts: Mild, Mod, & Severe RD GA 36.7±2.1 36.7±2.4 37.7±2.1 ($P = 0.032$) BW 2648.1±573.3 2760.7±604.1 3080.6±643.5 ($P = 0.000$)	Prospective multi-center cohort study	ACoRN was used to stratify levels of respiratory distress. Then SNAP-II was used as a comparison. ACoRN is easier to obtain. It does not need 12 hours of data and does not need an arterial blood gas. SNAP-II assessed 12 hours after birth. SNAP-II was used to predict mortality. SNAP-II was not used as a sequential measure. SNAPPE-II was not measured.	SNAP-II scores were noted as a risk factor for mortality via univariate and multivariate analysis. (OR: 1.071; 95%CI = 1.040–1.103; $P < 0.01$) (Values for univariate analysis not reported.) ACoRN and SNAP-II were correlated ($r = 0.219$, $p < 0.01$), but ACoRN was not a predictor of mortality.	Observational
16) Dannon et al. 2009 ^d	To determine the performance of the SNAP-II and SNAPPE-II in extremely low	USA	N = 1467	Cohort derived from the ELGAN study across 14 centers and 5 states from 2002 through 2004. GA < 28 weeks 27% 23–24 wks 45% 25–26 wks	Prospective cohort study	SNAP-II and SNAPPE-II assed with data from the first 12 hours of life. SNAP-II and SNAPPE-II were used to predict mortality.	All infants received zero points for the SNAP-II subcategory for multiple seizures. When evaluating mortality risk by week of gestation, non-surviving infants always had higher SNAP-II	Observational

Authors and Date of Publication	Purpose	Country of Origin	Sample Size (N)	Sample Characteristics	Research Design	SNAPII	Findings	Limitations and Potential for Bias
	gestational age newborns gestational age newborns			29% 27 wks BW not reported All inborn status		SNAP-II and SNAPPE-II were used to compare mortality outcomes between institutions. SNAP-II and SNAPPE-II scores used as dichotomous variables. Arbitrary cutoffs were explored for SNAP-II and SNAPPE-II scores. SNAP-II was not used as a sequential measure.	and SNAPPE-II scores compared to surviving infants at the same gestational age. Newborns with a SNAP-II score of 30 were 3.5 times more likely to die and newborns with a SNAPPE-II score of 45 were 3.9 time more likely to die when controlling for GA. Without adjustment for gestational age, SNAP-II scores 30 were 5.9 times more likely to die and SNAPPE-II scores 45 were 6.8 times more likely to die. SNAP-II and SNAPPE-II are population level predictors of mortality. SNAP-II and SNAPPE-II scores demonstrated an inverse relationship with week of gestation.	
17) Dammon et al. 2010 ^d	To assess if SNAP-II and SNAPPE-II scores for Extremely Low GA Newborns (ELGANS) were predictive of neurological sequella.	USA	N = 1399	Cohort derived from the ELGAN study across 14 centers and 5 states from 2002 through 2004. GA < 28 weeks BW not reported All inborn	Observational Study	SNAP-II and SNAPPE-II assigned within the first 12 hours of life. SNAP-II and SNAPPE-II were used to predict neurological sequella. Dichotomies of SNAP-II (>30) and SNAPPE-II (>45) scores were used. SNAP-II was not used as a sequential measure.	High SNAP-II (>30) and SNAPPE-II (>45) scores were predictive of IVH, moderate to severe ventriculomegaly and echodense lesions in cerebral white matter. High SNAP-II and SNAPPE-II did not predict cerebral palsy, positive autism screen, and small head circumference at 24 months corrected age. SNAP-II and SNAPPE-II were inversely correlated with GA. Discussion concerning whether SNAP-II and SNAPPE-II are part of the causal chain to brain damage or if they are a marker for developmental vulnerability.	Observational Study The newborns in the ELGAN study that did not complete the 24 months follow-up had heterogeneous socioeconomic backgrounds as compared to the group that completed the follow-up. Missing data - Not all neonates in the study participated in every assessment, so the denominator changes between different outcomes of interest.
18) Sundaram et al. 2009	To evaluate the relationship between the SNAP-II score and mortality as well as organ dysfunction in a cohort of newborns with severe sepsis over a 14 day period.	India	N = 40 Survivors n = 15 Non-S n = 25	Dx with severe sepsis and < 28 days old GA 30.2 ± 2 BW 1188.3 ± 282.8 Day of Onset of Septicemia 4 (Median) 3–6 (IQR)	Prospective Cohort Study	SNAP-II scores were obtained with data for the 12 hours following the onset of severe septicemia. SNAP-II was used as a prognostic tool after the dx of sepsis. SNAP-II score was assigned arbitrary cut off values: Mild – 1–20 Moderate – 21–40 Severe - > 40 SNAP-II was not used as a sequential measure. SNAPPE-II was not measured.	SNAP-II scores > 40 were associated with more dysfunctional organs (3.8±0.4 vs 2.9±0.8, <i>P</i> = 0.001). Median SNAP-II score was higher in babies that died within 14 days of study enrollment (43 (36 – 53.5) vs 18 (16 – 37) <i>P</i> < 0.001). AUC = 0.82 (95% CI = 0.68–0.95, <i>P</i> < 0.001) SNAP-II scores < 40 were associated with fewer dysfunctional organs (3.8±0.4 vs 2.9±0.8, <i>P</i> = 0.001) Three category's on the SNAP-II score (MAP < 29 mmHg, Lowest	Observational Study Single Center Small sample size A large number of potential participants were excluded from the study. [Inclusion criteria met (n = 75), yet 45.9% (n = 34) were excluded due to various reasons.]

Authors and Date of Publication	Purpose	Country of Origin	Sample Size (N)	Sample Characteristics	Research Design	SNAP-II	Findings	Limitations and Potential for Bias
						SNAPPE-II was not measured. SNAPPE-II was not measured.	Male newborns were more likely to experience GI complications. (OR = 2.41, 95% CI = 1.07–5.45) An association was found between gestational age and ductal closure with indomethacin. (AUC = 0.71, 95% CI = 0.63 – 0.79)	
22) Figueras-Aloy et al 2007	To determine if intercellular adhesion molecule (ICAM)-1, vascular cell adhesion molecule (VCAM)-1, L-selectin, and P-selectin can be used to help diagnosis sepsis. To determine if there is a relationship between these markers and severity of illness in septic neonates.	Spain	N = 37 HC– n = 10 HC+ n = 15 Control n = 12	Dx with R/O sepsis Cohort derived from consecutive admissions in a single center over a 21 month period. (Actual dates not reported) Age of the newborns ranged from within 1 hour of birth to 32 days. <u>Cohorts:</u> Control HC– HC+ GA 38.5 (37–40) 40 (36–40) 37 (32–39) ($p=0.113$) BW 3037 (2625–3835) 3365 (2840–3612) 2350 (1390–3120) ($P = 0.124$) Day of Life at initial blood draw (Median/IQR) 1 (1–1) 2.5 (2–3) 10 (3–14) ($P = 0.001$)	Prospective Observational Cohort study	SNAP-II was measured at the moment of highest clinical illness. The exact timeframe of data collection for SNAP-II measurement was not clearly reported. SNAP-II was used to quantify severity of illness for potential septic patients. SNAP-II was used as a comparison tool to assess for relationships between severity of illness and ICAM-1, VCAM-1, L-selectin, and P-selectin. Baseline SNAP-II was not measured upon admission for comparison. SNAPPE-II was not measured.	Soluble ICAM-1 demonstrated a positive correlation ($r = 0.537$, $P = 0.006$) with SNAP-II scores. ICAM-1 Levels > 274µg/L were associated with HC + sepsis Positive correlation between sICAM-1 and gestation age ($r = 0.623$, $P = 0.041$) The day of life for the initial blood draw was significantly different between the three groups ($P < 0.001$). Mean day of life was 1 for the control group, 2.5 for the HC– group and 10 for the HC+ group.	Observational Single center Small sample size
23) Kadirvar et al 2007	To determine if the SNAPPE-II instrument is a predictor of neonatal mortality among the neonates admitted in a single NICU in Iran.	Iran	N = 213	Recruited September 2003 to August 2004 Mean SNAPPE-II = 21.6 Mean BW = 2479.8g Mean 5 minute APGAR = 7.71 Day of Life 7.6 (0.5) GA 35.8 (0.2) BW 2479.8 (29.4) SNAPPE-II 21.6 (1.1)	Prospective Observational	SNAP-II was only measured with the perinatal extension. SNAPPE-II was measured within 12 hours of NICU admission. (This was not always the day of birth.) SNAP-II was not used as a sequential measure.	Factors associated with mortality: SNAPPE-II ($P = 0.04$) 5 minute APGAR ($P = 0.01$) GA ($P = 0.03$) BW ($P = 0.02$) Logistic regression demonstrated that SNAPPE-II and APGAR at 5 minutes were significant predictors of mortality. (Chi Square results not reported.)	Observational Single site study
24) Mathur et al 2007	To assess the ability of new tool (TOPS) to assess for illness severity in neonates and to compare the TOPS instrument to SNAPPE-II.	India	N = 175 NS n = 60 Survivor n = 115	Cohort derived from outborn infants transported to a tertiary NICU in India from a single center from March to December 2003. <u>Cohorts:</u> NS	Prospective Observational Cohort study	SNAP-II was determined with data from the first 12 hours of admission SNAP-II was used as a predictor of mortality.	TOPS and SNAP-II performed similarly. TOPS AUC=0.89 SNAP-II AUC = 0.88 Goodness of Fit Test was 0.75 (TOPS)/0.80 (SNAP-II)	Observational Single center

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Authors and Date of Publication	Purpose	Country of Origin	Sample Size (N)	Sample Characteristics	Research Design	SNAPII	Findings	Limitations and Potential for Bias
				Survivor Age at admission 5.57 (0.89) Weight at admission 2045 (693) 2193(709) ($P = 0.29$) GA 37.4 (4.83) 36.5 (4.25) ($P = 0.25$)		SNAP-II was used as a comparison tool. (Criterion validity) SNAP-II was not used as a sequential measure. SNAPPE-II was not measured. TOPS score was assessed upon NICU admission. TOPS Temperature Oxygenation Proxy for Perfusion Blood Sugar	TOPS is a simplified score and eliminates the need for expensive equipment	
25) Zupancic et al 2007	To revalidate the performance of the SNAP-II and SNAPPE-II	USA Canada	N = 9897 1500 g n = 5617 >1500g n = 5280	The cohort was derived from 58 self-selected from the Vermont Oxford Network from January to December 2002. Cohorts: 1500g >1500 g Total BW 1021 ±305 2826 ±783 2375 ±1048 GA 28 ±3 36 ±3 34 ±5 SNAPPE-II 28 9 13 SNAP-II (total sample) 11.7 (Mean) 5 (Median)	Prospective Observational Study	SNAP-II was calculated in the first 12 hours of admission. SNAP-II was used as designed as a measure of illness severity. SNAP-II was being compared to the VON-RA. SNAP-II was not used as a sequential measure.	Discrimination of SNAP-II and SNAPPE-II was best for neonate’s 1500g. Discrimination of the instrument improved when congenital anomalies were accounted for. Combined Weight SNAP-II AUC = 0.86 SNAPPE-II AUC = 0.89 1500g SNAP-II AUC = 0.82 SNAPPE-II AUC = 0.86 >1500 g SNAP-II AUC = 0.79 SNAPPE-II AUC = 0.82 H-L P values ranged from 0.184 – 0.965	Observational No reliability data reported.
26) Lam et al 2006	To describe the cost for hospital care for neonates diagnosed with CDH. To compare costs between neonates with CDH vs non-CDH in two cohorts with similar SNAP-II scores upon admission. To identify any correlations between the SNAP-II score and cost for NICU care.	Canada	N = 32 CDH n = 32 Control n = 27	Cohort derived from a single center NICU/PICU between 1999 and 2003. The cohort of interest included patients diagnosed with CDH. The control group consisted of patients with similar SNAP-II scores. Mean BW 3.1kg Mean GA 38.2 weeks SNAP-II 12.5 (Mean) 13 (Median) 0–34 (Range)	Retrospective Cohort Study	SNAP-II scores were completed within 12 hours of NICU admission. SNAP-II was used as a control to ensure the two groups were homogeneous for severity of illness. SNAP-II scores were also used to group surviving infants into three terciles of risk. Group 1 (0–5) Group 2 (6–23) Group 3 (>23)	Admission SNAP-II score correlates to the total cost of care for the neonate with CDH. (The actual correlation is not reported.) Mean total cost per CDH survivor = \$54,102 (range \$9,608–\$244,734) Mean total cost per non-survivor-CDH was \$9,761 (range \$943–\$25,888) Mean total cost for control group (non-CDH) was \$13,722 (range \$1,544–\$83,945) CDH cohort vs the control group had significantly greater costs per	Single center Small sample size Underestimation of costs due to: 1) unavailability of some provider costs, 2) estimation of nursing costs based on units of service versus actual salary, 3) cost of continuing care after discharge is unavailable.

Authors and Date of Publication	Purpose	Country of Origin	Sample Size (N)	Sample Characteristics	Research Design	SNAP-II	Findings	Limitations and Potential for Bias
27) De Felice et al 2005	To evaluate the relationships among histologic chorioamnionitis (HCA), illness severity, and outcomes in a sample of very low BW (VLBW) infants.	Italy	N = 116 HCA + n = 67 HCA – n = 49	Cohort derived from 2 centers in Italy. Cohorts: HCA+ HCA– BW 928 ± 322g 1044 ± 264 (<i>P</i> = 0.001) GA 26.91 ±2.61 29.63 ±2.29 (<i>P</i> = 0.0001)	Prospective Observational Study	SNAP-II was not used as a sequential measure. SNAP-II was not used as a sequential measure.	SNAP-II was not used as a sequential measure. SNAP-II was not used as a sequential measure.	Observational Two study centers Small sample size
28) Skarsgard et al 2005	To validate the SNAP-II score as a predictor of mortality in a sample of neonates diagnosed with CDH. To compare the SNAP-II with the new predictive equation developed by the Congenital Diaphragmatic Hernia Study Group (CDHSG).	Canada	N = 88	Dx with CDH. The cohort was derived from the 19,507 neonates admitted through the Canadian Neonatal Network (CNN) (17 centers) from January 1996 through October 1997. GA 37.4 ± 3.3 BW 2939 ± 735	Observational Retrospective	SNAP-II was measured with data collected during the first 12 hours of admission. SNAP-II was used as a measure of illness severity. SNAP-II was compared with the performance of a new research instrument (CDHSG). SNAP-II measured as a continuous variable. SNAP-II was not used as a sequential measure. SNAPPE-II was not measured.	SNAP-II (OR 1.057, 95% CI = 1.019–1.097) and GA (OR = 0.838, 95% CI = 0.703 – 1.000) were predictive of mortality. AUC/H-L *Gestational Age 0.68/0.75 *SNAP-II 0.76/0.7 *Combined Model with GA and SNAP-II 0.81/0.88 *CDHSG Model 0.83/0.06	Observational Retrospective Small sample size CDH was defined by the ICD-10 code 456.6, which could be used for a variety of defects.
29) Figueras-Aloy et al 2004	To determine if plasma endothelin-1 (ET-1) is associated with signs and symptoms of sepsis in newborns.	Spain	N = 35 HC– n = 13 HC+ n = 22	Dx with sepsis Cohort derived from single center admissions from March 1998 to March 2000. Cohorts: HC– HC+ GA (Median/IQR) 40 (39–40) 32 (30–38) (<i>P</i> = 0.002) BW (Median/IQR) 3450 (3120–3660) 1630 (1230–28000) (<i>P</i> < 0.001)	Observational Prospective Cohort Study	SNAP-II was measured at the moment of highest clinical illness. The exact timeframe of data collection for SNAP-II measurement was not clearly reported. SNAP-II score was used to assess severity of illness specifically for the septic newborn. SNAP-II was used as a comparison tool to assess for relationships between	SNAP-II scores were higher in the HC+ newborns compared to newborns that were HC–. The total SNAP-II scores were correlated with ET-1 (<i>r</i> = 0.373, <i>P</i> = 0.027). Concentrations of ET-1 rose significantly as SNAP-II subcategories: blood pressure (<i>P</i> = 0.030), pH (<i>P</i> = 0.048), Seizures (<i>P</i> = 0.010), and urine output (<i>P</i> = 0.013) increased. The most accurate subcategories of the SNAP-II were blood pressure	Observational Single center Small sample size

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Authors and Date of Publication	Purpose	Country of Origin	Sample Size (N)	Sample Characteristics	Research Design	SNAP-II	Findings	Limitations and Potential for Bias
30) Lee et al 2003	To examine the variation in mortality occurrences within 7 days of admission for neonates < 32 weeks gestation.	Canada	N = 5192 Day Admit n = 2061 Night Admit n = 3131	Preterm cohort derived from admissions to the Canadian Neonatal Network (CNN) NICU's between January 1996 and October 1997. Cohorts: Day Admit Night Admit GA Range (24–32) 29.2 ± 2.4 29.0 ± 2.5 (<i>P</i> > 0.05) BW 1340 ± 462 1330 ± 474 (<i>P</i> > 0.05) SNAP-II 12.0 ± 13 13.0 ± 13.5 (<i>P</i> < 0.01)	Prospective Observational Cohort Study	SNAP-II was measured with data collected during the first 12 hours of admission. SNAP-II scores were used as risk adjustments for the day versus night admission. SNAP-II was used to assess for homogeneous groups. SNAP-II was not used as a sequential measure. SNAPPE-II was not measured.	Mortality odds were 60% higher in NICU night admissions as compared to day time admissions. Factors that predicted early neonatal death included: male, outborn status, APGAR < 7 at five minutes of life, congenital anomalies, early gestations, high SNAP-II scores. Night admission (OR = 1.6) and in-house neonatologist at night (OR = 0.6) predicted early mortality. Mortality rates were significantly higher at night (5.4% vs 4%, <i>P</i> < 0.05). SNAP-II scores were significantly different between daytime (12.0 ± 13) and nighttime (13.0 ± 13.5) admissions.	Observational Study Delivery room deaths were not included; therefore the mortality rate is underestimated. Obstetrical care data was not available. This omission of information can skew the data as perinatal conditions will affect neonatal outcomes. Data on nursing experience between shifts was not available and controlled for. Authors questioned if the 12 hour data collection period could be confounded by treatments or quality of care during the initial hours of life instead of initial illness severity alone.
31) Figueras-Aloy et al 2003	To determine variation in plasma nitrite/nitrate and endothelin-1 in neonatal sepsis.	Spain	N = 128 Control n = 68 HC– n = 35 HC+– Mod n = 21 HC+ Severe n = 4	Dx with sepsis and control group without sepsis. Cohort derived from a single center from 1998 – 2000. Newborns ranged in age from 1–15 days. Cohorts: Control HC– HC+ Mod HC+ Severe GA 37.8 (2.0) 38.2 (2.8) 35.1 (4.8) 29.8 (2.1) (<i>P</i> < 0.001 for all except 1&2) BW 2995 (602) 3071 (717) 2343 (981) 1336 (439) (<i>P</i> < 0.001 for all except 1&2) SNAP-II None for Controls 1.1(3.8) 7.1(10.7)	Prospective Cohort Study	SNAP-II and SNAPPE-II were assessed when the newborn displayed the greatest severity of illness from sepsis. The exact timeframe of data collection for SNAP-II and SNAPPE-II measurement was not clearly reported. SNAP-II was used as a comparison tool to assess for relationships between severity of illness, nitrite/nitrate, and plasma endothelin-1 (ET-1). SNAP-II scores were used to stratify patients according to severity of illness levels. SNAP-II was not used as a sequential measure.	Endothelin-1 and SNAP-II were positively correlated (<i>r</i> = 0.305, <i>P</i> = 0.018). SNAPPE-II and endothelin-1 were positively correlated (<i>r</i> = 0.288, <i>P</i> = 0.026).	Observational study. Single center Control and study groups were not homogeneous for GA and BW. SNAP-II not assessed in the control group, so illness severity homogeneity/heterogeneity of the sample is unknown.

Authors and Date of Publication	Purpose	Country of Origin	Sample Size (N)	Sample Characteristics	Research Design	SNAP-II	Findings	Limitations and Potential for Bias
				73.7(31.8) ($P < 0.001$) SNAPPE-II None for Controls 1.4(4.2) 7.7(10.9) 73.7(31.8) ($P < 0.001$)				
32) Zardo et al 2003	To compare birth weight and illness risk scores for prediction of neonatal mortality.	Brazil	N = 494 Survivor n = 450 NS n = 44 (8.9%)	Cohort derived from NICU admissions in a single center from March 1997 to June 1998. Cohorts: Survivor NS BW 2354g (Mean for all neonates) 2418 (Median) 1775–3070 (IQR) 860 (Median) 643–1690 (IQR) GA 36 (Mean)	Cohort/Retrospective Cross Sectional Study	SNAP, SNAPPE, and birth weight were measured prospectively. Data was collected with the first 24 hours after admission. SNA-P-II, SNAPPE-II and CRIB, were measured retrospectively upon chart review. SNAP-II and SNAPPE-II were assessed using data collected within the first 12 hours after admission. All illness severity measures were used as comparisons to determine the best prediction of mortality. SNAP-II was not used as a sequential measure.	All risk scores performed better than BW alone (AUC = 0.81). AUCs SNAP0.85 SNAPPE 0.90 SNAP-II 0.88 SNAPPE-II 0.91	Observational Single Center
33) Chien et al 2002	To determine the ability of the SNAP-II to improve prediction of grade III or IV intraventricular hemorrhage (IVH) and chronic lung disease (CLD).	Canada	N = 4226 CLD N = 4226 IVH N = 3778 (77% had cranial U/S for evaluativ e)	GA 32 weeks Cohort derived from admissions to one of 17 Canadian NICUs in the Canadian Neonatal Network (CNN) between 1996 and 1997. Cohorts: IVH CLD BW 1190 ± 426 1390 ± 457 GA 29 ± 2 29 ±2 SNAP-II 14 ± 13 11 ±11	Prospective Observational	SNAP was used as the “criterion” to assess for criterion validity for the SNAP-II. SNAP-II was used to predict outcomes including IVH and CLD. SNAP-II was assessed using data collected within the first 12 hours after admission. SNAPPE-II was not measured.	Prediction of IVH: SNAP AUC = 0.73±0.01 $\chi^2 = - 222$ SNAP-II AUC = 0.73±0.02 $\chi^2 = - 219$ Prediction of CLD SNAP AUC 0.74±0.01 $\chi^2 = 381$ SNAP-II AUC = 0.78±0.01 $\chi^2 = 470$ In model with gestational age, small for gestational age and SNAP-II the AUC improved. (IVH 0.78/CLD 0.82). Goodness of fit was demonstrated with a H-L (IVH $P = 0.16$ /CLD $P = 0.51$)	Observational Limitations of the SNAP-II: *It takes 12 hours to complete severity and treatment response. Discrepancy between reported GA for IVH group (29 vs 31 weeks).
34) Chien et al 2001	To examine if newborns born outside of a tertiary care facility have an increased risk of	Canada	N = 3769 Outborn n = 605 (16%) Inborn n = 3164	Cohort was derived from preterm infants (32 weeks) admitted to a NICU in the Canadian Neonatal Network (CNN) from January 1996 – October 1997.	Prospective Observational Cohort Study	SNAP-II was measured within the first 12 hours of admission to the NICU. SNAP-II was used to measure illness severity.	Outborn infants had statistically higher incidence of mortality, IVH, CLD, PDA, respiratory distress, and hospital acquired infections. After controlling for initial illness	Observational Mortality is underestimated because delivery room deaths and transport deaths were not included in the study.

Authors and Date of Publication	Purpose	Country of Origin	Sample Size (N)	Sample Characteristics	Research Design	SNAP-II	Findings	Limitations and Potential for Bias
	morbidity and mortality after controlling for illness severity and perinatal factors for preterm infants morbidity and mortality after controlling for illness severity and perinatal factors for preterm infants morbidity and mortality after controlling for illness severity and perinatal factors for preterm infants morbidity and mortality after controlling for illness severity and perinatal factors for preterm infants			Factors for preterm infants Indicators for preterm infants Predictors for preterm infants Risk factors for preterm infants Risk factors for preterm infants 27–28 wk = 17%/18% 29–30 wk = 24%/23% 31–32 wk = 38%/32% SGA 6%/2% SNAP-II 0–9 points 53%/45% 10–19 points 23%/28% 20–29 points 12%/13% 30 points 12%/14%		SNAP-II was used to compare for homogeneity versus heterogeneity illness severity between the 2 groups.	severity, the statistical significance remained for everything above except CLD. Outborn infants had higher SNAP-II scores compared to inborn infants.	
35) Richardson et al 2001	To empirically derive and validate the second generation Score for Neonatal Acute Physiology (SNAP-II) and the second generation Score for Neonatal Acute Physiology-Perinatal Extension (SNAPPE-II)	USA Canada	N = 25,429 Derivation cohort n = 10,819 Canada n = 5588 California n = 5,530 New England n = 3492	Cohort was derived from newborns of all birth weights and gestational age admitted to the NICU in 30 different study sites from 3 neonatal networks (Canada, California, New England). Newborns were recruited from January 1996 to October 1997.	Prospective Observational	SNAP-II and SNAPPE-II were assessed with data from the first 12 hours of admission. Points for each subcategory of the SNAP-II were empirically derived from the Beta weights in the logistic model. SNAP was used as a control for group homogeneity. SNAP was used as the criterion for criterion validity of the SNAP-II.	Total sample AUC=0.91±0.01 Goodness of Fit = 0.90 Correlation between SNAPPE and SNAPPE-II = 0.91 Birth weights < 1500g AUC = 0.85±0.01 Goodness of Fit = 0.86 SNAPPE and SNAPPE-II (r = 0.86) Birth weights > 1500g AUC = 0.87±0.03 Goodness of Fit = 0.63 Correlation between SNAPPE and SNAPPE-II r = 0.88 1500g = 0.87±0.03 H-L = 0.90.	Observational Some values of zero were assigned due to the unavailability of test results. In the absence of a test result, tests were assumed to be normal (zero points).

^a =same study; multiple publications (ELGAN study)

Abbreviations: GA, gestational age in weeks; BW, birth weight; g, grams; LOS, length of stay; AUC, Area under the receiver operator characteristic curve; HL, Hosmer-Lemeshow goodness of fit test; Tx, treatment; Dx, diagnosed; Kg, kilograms; PSC, potential surgical candidates; NSC, Non-surgical candidate; NS, Non-survivor; SNAP-II, Score for Neonatal Acute Physiology-Perinatal Extension-II; SNAPPE-II, Score for Neonatal Acute Physiology-Perinatal Extension-II; CDH, congenital diaphragmatic hernia; Canadian Neonatal Network, CNN; CC, cardiac center; NCC, non-cardiac center; IQR, Interquartile Range; S, survivor; Non-S, Non-Survivor; aEEG, amplitude integrated EEG; IQR, Interquartile Range; VLBW, very low birth weight; CAPSNet, Canadian Pediatric Surgery Network; ACoRN, Acute Care of At Risk Newborns; Mod, Moderate; RD, respiratory distress; ELGAN, extremely low gestational age newborns; IVH, Intraventricular hemorrhage; E-Chorio, exposed to chorioamnionitis; NEC, necrotizing enterocolitis; IP, intestinal perforation; ICAM-1, intercellular adhesion molecule; VCAM-1, vascular cell adhesion molecule; HC-, hemoculture negative; HM+, hemoculture positive; R/O, rule out; VON-RA, Vermont Oxford Network risk adjustment; HCA+, histologic chorioamnionitis positive; HCA-, histologic chorioamnionitis negative; NTISS, Neonatal Therapeutic Intervention Scoring System; CDHSG, Congenital Diaphragmatic Hernia Study Group; ET-1, plasma endothelin-1; Mod, moderate; SGA, small for gestational age

Table 2

Sample Size for Included Studies using the SNAP-II

Sample size between 1–49	15, 28, 33, 34, 39, 42, 43, 44
Sample size between 50–99	29, 36
Sample size between 100–499	12, 18, 19, 21, 23, 24, 25, 26, 35, 37, 38, 40, 46
Sample size between 500–999	20, 22, 41, 45
Sample size between 1,000–4,999	16, 17, 27, 31, 32
Sample size between 5,000–9,999	30, 47
Sample size 10,000	6