Development of an Abbreviated Symptom Score for the Neonatal Abstinence Syndrome

I Chervoneva[1], **SC Adeniyi-Jones**[1,2], **F Blanco**[1], **WK Kraft**[1]

[1]Sidney Kimmel Medical College of Thomas Jefferson University, Philadelphia, PA

[2]Nemours DuPont at Jefferson, Philadelphia, PA

Abstract

Objective—We sought a shortened MOTHER NAS and Finnegan score that would retain comparable performance characteristics of the full instrument.

Study Design—Retrospective cohort

Results—124,170 MOTHER NAS scores between August 2007 and May 2016 from 775 infants (36 weeks) were examined. Classification and regression tree model identified the most important subsets of the scored variables. A 9 element shortened scale yielded > 90% sensitivity and specificity to predict clinical endpoints based on the full 19 element MOTHER NAS score. Conversion of the data sets to the Finnegan score, and applying the same procedure resulted in a 9-element score with similar performance characteristics.

Conclusion—Shortened scoring instruments were identified with high predictive power for clinical endpoints based on the 19-element full MOTHER NAS score. There were no substantial variation in performance for age, supporting the current practice of utilizing a single scoring tool regardless of post-natal age.

Introduction

The neonatal abstinence syndrome (NAS) is characterized by a pattern of signs resulting from the cessation of maternal transfer of certain xenobiotics. While the FDA and others have used the term neonatal opioid withdrawal syndrome (NOWS) to link the condition to an etiologic cause, the term NAS continues to be used elsewhere, as it encompasses more broadly the in-utero exposures seen in clinical practice. For this reason we have retained the use of NAS in this report. Manifestations of NAS are driven primarily by opioid withdrawal, though other exposures such as benzodiazepines and selective serotonin reuptake inhibitors can worsen the severity of symptoms. Several scoring instruments have been developed to standardize symptom assessment. By far the most commonly used instrument is the one developed by Loretta Finnegan in the 1970s (1,2). There has been drift in the specific

Corresponding Author Walter K. Kraft, MD, Department of Pharmacology and Experimental Therapeutics, Thomas Jefferson University, 1170 Main Bldg., 132 S. 10th St. Philadelphia, PA 19107-5244.

Conflict of Interest. None of the authors have a conflict of interest.

elements used to make up a "Finnegan instrument" at some sites, but the most commonly used version of the Finnegan neonatal abstinence scoring system (FNAS) with 21 scored elements is based largely on a 1979 NIDA monograph (3–5). There is debate of the accuracy of a canonical "modified Finnegan score" that is often described in the literature (5). That local variants of this published version have emerged suggests practitioners have not found the FNAS optimized. Formal psychometric assessment demonstrates evidence of both under and overweighting of certain scored elements (6). One variant is the MOTHER NAS Scale (MNAS), which was developed for use in the landmark MOTHER trial (7) and has been utilized in other randomized controlled trials (8,9). MNAS has 19 scored elements and has a very high degree of overlap with the FNAS. There are few direct comparisons between the two instruments, but Gomez-Pomar demonstrated a high concordance between FNAS with MNAS in the same hospital over two consecutive years. (10) In the Gomez-Pomar study, there were 12,847 observations using FNAS and 17,150 observations using MNAS, with a Pearson's correlation of 0.86 using a score ≥ 9. Sensitivity was 96% and specificity 80% to predict scores ≥ 9 on the MOTHER NAS scale.

Previous attempts to simplify the Finnegan scoring system focused on eliminating items found to be redundant while trying to maximize the correlation between candidate short version scores and the full FNAS scores. (Table 1) Maguire used a factor analysis approach to select the items that have highest associations (loadings) with the first two most important factors identified using the standard factor analysis (11). Conversely, Gomez-Pomar used step-wise elimination of the items in the FNAS with the smallest contributions to the Pearson's correlation to create a short version (10). Devlin (12) has reported in a poster abstract a multi-site analysis that adds two institutions to those reported in Gomez Pomar.

Using retrospective clinical data of MNAS scoring, we sought to identify shortened FNAS and MNAS systems with high sensitivity and specificity for matching of the key dichotomized FNAS/MNAS scores ≥ 8 and ≥ 12. The cut offs of ≥ 8 and ≥ 12 are commonly used to identify infants needing pharmacotherapy and guide dose adjustments. The goal of this exercise was to generate a system that would reduce the time and complexity burden of NAS assessment while retaining the diagnostic utility of the currently used instrument. Our explicit goal was not to evaluate the appropriateness of a specific score to initiate, intensify or de-escalate therapy, but instead to take the existing scores and treatment thresholds as the local "gold standard" against which a simplified score could be developed. Furthermore, we sought to evaluate how the performance of shortened MNAS and FNAS instruments may depend on postnatal age since each is used not only to identify the need for pharmacotherapy but also to adjust dosing. NAS signs and treatment may persist beyond the first month of life, a time when developmental changes in infant behavior could impact the utility of the instrument.

Methods

This was a single center, retrospective analysis of infants with in utero opioid exposure. An electronic medical record search was conducted for data between August 8, 2007 and May 5, 2016 for term newborns with an ICD-9 code 779.5. MNAS was used to assess NAS in infants for the entire period of data extraction. Nineteen scored elements are summed for a

score every 3–4 hours by trained nurses. An additional set of unscored elements is recorded but not included in the final score. These non-scored elements are primarily more severe grades of scored items, such as the non-scored "projectile vomit" associated with the scored "vomiting or regurgitation". The protocol at Thomas Jefferson University Hospital defines the sum of three consecutive scores ≥ 24 (mean ≥ 8), or a single score ≥ 12 as the threshold to initiate or intensify pharmacologic therapy. All infants were admitted to the neonatal intensive care unit for NAS treatment. During the study period, morphine was used exclusively to treat NAS and phenobarbital was added as adjunct therapy for severe cases. The dose of morphine was continued for at least 48 hours after stabilization of symptoms, with weaning of dose when there was an average score < 8 in preceding 24 hours with no single score ≥ 8 in previous three scores. All infants were treated as inpatients until weaned off morphine. Reliability of nurse scoring was fostered by a program of periodic in-service sessions and observation of individual nurse scoring with feedback by a local nurse champion.

The data points were excluded if the MNAS score recorded was not equal to the sum of the individual item scores. MNAS scores were converted into FNAS equivalents using the modified FNAS description in Maguire (11). Since the scored and non-scored elements in MNAS jointly capture the information required to obtain all scores for the modified FNAS, the conversion algorithm was created to compute each item score in the modified FNAS using the values of scored and non-scored elements in MNAS. MNAS and FNAS scores were subsequently analyzed separately. The infants were randomly divided into a training set and a test set using a realization of the Bernoulli random variable with p=0.5. Recursive partitioning with 10-fold cross-validation was used to fit the classification and regression tree (CART) model(13) to the training set data and evaluate the importance of the scored variables for predicting FNAS/MNAS ≥ 8 (vs. FNAS/MNAS < 8), FNAS/MNAS ≥ 12 (vs. FNAS/MNAS < 12), or both, since these cutoffs are used in clinical decision making. In order to increase sensitivity, the observations with MNAS/FNAS ≥ 8 were weighted inversely to the proportion of observations with MNAS/FNAS ≥ 8 (19% for MNAS, 40% for FNAS) and observations with MNAS/FNAS ≥ 12 were weighted inversely to the proportion of observations with MNAS/FNAS ≥ 12 (2.5% for MNAS, 5.5% for FNAS). This choice of weights ensures that observations with MNAS/FNAS ≥ 12 (or ≥ 8) have the same total contribution to the CART model loss function as the much large proportion of observations with MNAS/FNAS < 12 (or < 8). In this way, minimizing the loss function defined with such case weights provides balance between the false negative and false positive errors. The approach also eliminated bias toward models with low false positive errors but low sensitivity for predicting MNAS/FNAS ≥ 12 (or ≥ 8). The resulting tree models were used only to identify the candidate optimal subsets of items based on Gini variable importance measure. Instead of considering complex decision tree models as candidate short scales, we studied the sums of scores for the identified optimal subsets of items. This provide simple-to-use candidate short scales for decision cut offs with one numeric total score analogous to MNAS/FNAS. Additional candidate short scales were considered as the sums of scores for other combinations of items that had the highest importance in multiple fitted CART models. The candidate short MNAS (sMNAS) and short FNAS (sFNAS) scales were analyzed using the receiver operating characteristic curves to obtain cutoffs for dichotomizing the total score

to predict MNAS/FNAS ≥ 8 and MNAS/FNAS ≥ 12. The sensitivity and specificity of the dichotomized sMNAS and sFNAS to predict both MNAS/FNAS ≥ 8 and MNAS/FNAS ≥ 12 was computed in the entire training set and in the subsets of MNAS/FNAS evaluations in the training set made only during post-natal week 1, 2, 3, 4, or ≥ 5. The optimal sMNAS and sFNAS were selected to maximize sensitivity and specificity for all post-natal weeks. Finally, the performance of the proposed optimal sMNAS and sFNAS were evaluated in the independent test set both overall and for evaluations made only during post-natal week 1, 2, 3, 4, or ≥ 5. The confidence intervals for sensitivity and specificity estimates were computed using the bootstrap method implemented in the R package 'pROC' (14). The performance of previously published short FNAS scales of Gomez-Pomar and Maguire was evaluated the same way but in the entire data set as the decision cut points were taken from the corresponding manuscripts. Since these short forms were developed using FNAS, we evaluated the sensitivity and specificity to predict FNAS ≥ 8 and FNAS ≥ 12. Statistical analyses were performed in SAS 9.4 (SAS Institute, Cary, NC) and R (R Foundation for Statistical Computing, Vienna, Austria). This study was approved by the Thomas Jefferson University Institutional Review Board.

Results

The entire data set included 160,382 MNAS scores (total score and each item score) for 822 infants. After exclusion of infants < 36 weeks gestation, 775 infants were included in the analysis. The median gestational age was 39 weeks (range 36–42 weeks, interquartile range 38–39 weeks), and the median birth weight was 2.95 kg (range 1.79–5.24 kg, interquartile range 2.66–3.29 kg). The training set included 402 infants with 61,026 MNAS scores, and the test set included 373 infants with 63,144 MNAS scores. Initially, four candidate sMNAS scores were identified with the best overall performance for the given number of items included in the training set. The 9-item shortened score (sMNAS-9) and the 11 item shortened score (sMNAS-11) yielded both sensitivity and specificity higher than 85%, while the best 6-item and 7-item scales exhibited unsatisfactory sensitivity to predict MNAS ≥ 8. (Supplementary Table 1) Furthermore, when performance was evaluated by post-natal week, the sensitivity of 6-item and 7-item short scales to predict MNAS ≥ 8 was even lower for the first 3 post-natal weeks. (Supplementary Figure 1) Meanwhile, sMNAS-9 had the highest sensitivity to predict MNAS ≥ 8 and otherwise similar performance to sMNAS-11. Thus, sMNAS-9 was identified as the optimal shortened MOTHER NAS instrument. sMNAS-9 includes Crying, Sleep, Undisturbed Tremors, Increased Muscle Tone, Fever > 37.3 C, Tachypnea, Poor Feeding, Loose Stools, and Vomiting/Regurgitation (Table 2). The performance characteristics in the test cohort were essentially unchanged from those in the training set, with both sensitivity and specificity higher than 90% (Table 3 and Supplementary Table 2). The same analysis conducted for converted FNAS scores in the training set. A 9-item sFNAS scale was identified as the optimal (Supplementary Tables 3 and 4, and Supplementary Figure 2). The optimal sFNAS –9 includes Crying, Excessive Sucking, Poor Feeding, Vomiting, Projectile Vomiting, Stools, Sleep, Tremors, and Fever (Table 4). Utilizing the sMNAS-9 with fewer scored elements implies lower numerical cut-off points. A sMNAS score of ≥ 5 and ≥ 7 would replace ≥ 8 and ≥ 12 used for the full

MNAS 19-element scale. Similarly, lower cut-off values of 4 and 7 for sFNAS would replace 8 and 12 used for the full FNAS-21 scale.

An analysis of age effects on performance of sMNAS-9 and sFNAS-9 scales is reflected in Figures 1 and 2. For the selected optimal short scale, there was relatively small difference of score test performance between any of the weeks of life. In contrast, the performance of shorter candidate subscales with 6–8 items varied between different weeks of life (Supplementary Figures 1 and 2). These data indicate that despite reported neurobehavioral changes over the first weeks of life,(15) both sMNAS-9 and sFNAS-9 maintained predictive power for clinically important cutoffs FNAS/MNAS 8 and FNAS/MNAS 12 based on the full length parent scales (MNAS and FNAS).

The performance of previously published short FNAS scales in the entire data set is reported in Table 5. The performance characteristics using Jefferson Hospital data were remarkably similar to those published by Gomez-Pomar. The only divergent measure was that sensitivity of this scale to predict FNAS 12 (65%, 95%CI: 64–66%) was considerably lower than corresponding sensitivity of sFNAS-9 (88%, 95%CI: 87–89%) (Table 3). Meanwhile, Gomez-Pomar (24) short scale included all items in sFNAS-9 scale except fever, plus three additional items (Increased Muscle Tone, Tachypnea and Nasal stuffiness) (Table 4). The Maguire (11) proposed short FNAS scale had performance characteristics that were lower by 4% to 13% as compared to performance of sFNAS-9 scale (Table 5).

Discussion

Attempts to improve NAS therapeutics include optimization of symptom scoring instruments. The conceptual goal is to quantify the predictive power of individual scored elements in the full FNAS. Removing less predictive elements would ideally reduce nursing effort, as well as simplify and focus training on high-yield assessments. This exercise involves value judgements that weigh the costs of brevity with losses in test performance. Our finding of a high specificity that was relatively resistant to loss of predictive power is consistent with that seen by others. Specificity reflects the power of the instrument to correctly identify true negatives (i.e., MNAS that is < 8). High specificity protects against misidentifying a non-threshold infant as an infant with symptom severity requiring therapy initiation or intensification. Low specificity would result in the potential overtreatment of infants. Sensitivity on the other hand is a reflection of the ability to identify a true positive. Loss of sensitivity with reducing the number of evaluated items means fewer infants who would have triggered a dose change at MNAS 8 would have this change in treatment with a shorter scale. In this case reduction in the number of test elements would lead to under rather than over treatment of NAS. For scales sMNAS-4 and sMNAS-7 (supplemental Table 1), sensitivity fell <82%. A backstop to this limitation of low sensitivity is that disease severity is likely to be progressive, and thus potentially self-correcting as subsequent scores rise in response to failure to initiate treatment earlier or suboptimal drug dosing. The functional consequences of under-treatment would be manifested primarily by lack of proper weight gain, poor state control, lack of sleep and excessive crying resulting in patient, parent and staff discomfort. The long-term impact of delayed or under treated NAS is unknown.

The high specificity observed in all conditions tested limits the utility of generating a receiver operator curve AUC as a discriminator of instrument performance.

sMNAS-9 contains elements that make up neurologic, autonomic, respiratory and gastrointestinal domains, which are helpful due to the heterogeneity between infants in expression of NAS signs which could be lost information with shorter scales. There is minimal loss of specificity, protecting against over treatment. Misclassification error (known also as total prediction error, a sum of false negative and false positive predictions divided by the total number of predictions) for both 8 and 12 cut points are similarly low. We posit that the performance of sMNAS-9 relative to the full 19 element Mother NAS score has an acceptable loss in the ability to identify infants who would have required a treatment alteration for several reasons. It is anticipated that the loss of sensitivity would be greatest near the threshold border. An infant with very severe signs of NAS would be unlikely to be misclassified as having moderate or low symptomatology. Secondly, NAS has been recognized as having signs that can vary over the course of a day. The variability in symptoms over a day may be less than that a shorter instrument introduces. Lastly, NAS symptoms are assessed every 3–4 hours which allows for a quick correction of under treatment. The progressive nature of undertreated withdrawal means that that an under-scored infant at one time point would be recognized at a subsequent evaluation. Test specificity was excellent in all the shortened scores when a MNAS cut off of >12 was used. This likely reflects a clinical situation in which there are severe NAS signs, for which the key elements reliably predict symptom severity. Given the relative infrequency of data points with MNAS scores 12 (2.5% of all measured scores) inverse weighting was implemented in the regression tree models to improve sensitivity to detect MNAS 12. To illustrate the problem with any unweighted approach (such as used by Maguire), if a shortened score is set to predict all MOTHER scores < 12 then for such a shortened score, the specificity is 97.5% and the total prediction error is only 2.5%, but the sensitivity is 0%.

Our examination using the same approach after conversion of the Mother NAS scores to their equivalent Finnegan scores also generated the option of 9 element shortened form (sFNAS-9). This shortened version had only 2–3% lower sensitivity and specificity to predict the full score compared to sMNAS-9 (Table 3). Limitations to this conversion are that the original data set and treatment rules were based on Mother NAS. However, given the large overlap of items and general similarity of results suggests a high internal validity. External validity is supported by the high degree over overlap with the specific items identified by other analyses (Table 4). In addition, our data generated essentially the same test characteristics as seen by Gomez-Pomar, suggesting that our scoring system is generalizable to other institutions (Table 5).

We investigated an age dependency for scores, with the first week potentially differing from later ages as developmental changes in neurobehavior and withdrawal signs occur. Lower gestational age at birth has generally been associated with decreased need for pharmacologic treatment for NAS and decreased intensity of treatment (16–18). Post-conceptual age (gestational age plus postnatal age) reflects advancing maturity similar to higher gestational age. We demonstrate for the first time the ability of a shortened scales to perform well for

the gestational age range 36 – 42 weeks and post- conceptual 37 - 47 weeks included in this analysis (Figures 1 and 2).

Gomez-Pomar has characterized the generation of scoring elements as formative rather than reflective (19). The reflective approach is characterized by the psychometric method, in which each scored element is generally considered to be reflective of NAS and thus also highly correlated. The formative approach is the one used to develop the Finnegan score, in which signs were identified, weighted by severity and subsequently interpreted with clinical judgement about the cut off and need for treatment. The current approach does not attempt to incorporate new items, such as the pharmacogenetics or gestational age of the infant, but instead uses the MNAS/FNAS as a base of existing items with a goal of removing those with less predictive power relative to the reference. This is the approach used by Gomez-Pomar, who proposed a 10 element score with a cut-off of 6 and 10 (instead of 8 and 12) which compared to the full Finnegan instrument provided a Pearson's correlation of 0.914 (10). Maguire described a factor analysis with 7 elements that was correlated with the total score on the 21-item modified Finnegan version used at that institution (r = 0.917; P < .001).

There have been efforts to define simplified instruments with a goal of easily screening for NAS related outcomes (20–22). These approaches differ from our current approach in that they do not seek to replace the scope of the current Finnegan score paradigm. Jones proposed an initial three element instrument of hyperactive Moro reflex, mild tremors when undisturbed, and increased muscle tone, which was able to distinguish between infants with in utero opioid exposure and those without (22). Jones applied this approach to the MOTHER study data to address a different question of differentiating opioid exposed infants who required pharmacologic therapy from those who did not. A revised five element index (tremors, muscle tone, excoriation, tachypnea, and irritability) had an AUC of the receiver operator curve of 0.90 compared to 0.94 for the full MOTHER NAS score (20). Of note, the Moro reflex, which was an element of the three-score instrument, fell out in the revised 5 score. In both screening projects, the ability to effectively reduce the scoring elements is consistent with the high specificity seen in all of the shortened versions generated in the current exercise. Isemann similarly demonstrated relatively high predictive power to predict need for pharmacologic therapy based upon muscle tone, tremors when disturbed and excoriation, with further refinement based upon type of in utero exposure (21). Given the predictive power of the three element scores, the use of our shortened sMNAS score solely for screening NAS patient who may require treatment is not likely to be ideal. However, it has robust performance for our goal of developing a tool with dual function of identifying NAS infants requiring pharmacotherapy and as a guide for dose adjustments.

Our study is strengthened by a large data set of >124,000 individual observations. However, there are several limitations in the proposed approach. As such, the study population consisted almost entirely of in utero methadone exposure and all infants were managed at a single center. Gaalema identified differences in affected domains between infants with in utero methadone and buprenorphine exposure (23). The current approach does not examine the impact of inter-rater reliability, though this has been estimated at two hospitals to account for only 5–10% of score variability,(24) and Westgate and Gomez-Pomar have suggested subjective NAS elements as valuable to management even if the inter-rater

reliability is less than 90% (19). There is also little empiric evidence of variation by day of the week or time of the day (25). Our hospital protocol was to closely follow a decision rule based upon the sum of three MNAS score 24 or a single score 12 for initiation of treatment or dose advancement of therapy. The data set collected was not assessed to verify the degree of adherence to the protocol, nor the impact of the additional recorded but unscored elements. We did not question the validity of the current approach of a clinical decision point (starting or adjusting pharmacologic treatment) and instead assumed the scores to be the gold standard. The appropriateness of the cut points of 8 or 12 for the Mother or Finnegan NAS scores have not been established, or much debated. The data set also did not contain rich clinical information outside of scores. It is possible there were infants with in utero exposure to opioids who were not captured by an ICD-9 code. This would be more likely in those infants who did not require pharmacologic treatment, and as such had less severe symptomatology. We have not tested the proposed sMNAS-9 with cut-off values of 5 and 7 prospectively. This would require extensive training, validation of inter-observer reliability and dual scoring that are beyond the scope of the current paper. Lastly, the data collected and therapeutic decisions were based upon MNAS cut-off points. Though we transformed MNAS to FNAS scores for generation of shortened FNAS scores, it is possible that the actual cut points for dose changes may have varied for a small number of infants.

The use of non-pharmacologic treatment for NAS evolved at Thomas Jefferson University over the collection period. These measures included, but were not limited to, a strong push to encourage breast feeding, rooming in, and the development of specialized quiet rooms for infants with in utero opioid exposure. Similarly, "back-to-sleep" positioning has made excoriations less common and reduced its contribution to overall FNAS scores. These changes likely contributed to drift in the symptomatology demonstrated by infants. Obstetrical care has evolved with the goal of optimizing maternal methadone doses and treating coexisting psychiatric illness to avoid maternal relapse and return to use of street drugs during pregnancy. As a result more infants are co-exposed to opioid and prescribed psychotropic agents. Lastly, we have not estimated the difference in time it would take for a nurse to administer the shortened score, nor how much time it would take to train using this instrument. However, compared to our gold standard MNAS, the sMNAS-9 would result in 10 fewer items being assessed, scored and documented at each time point or 60 to 70 fewer occurrences daily for 3 or 4 hourly scoring.

In summary, we have evaluated several potential approaches to shorten the NAS MOTHER scoring system to increase ease of use while maintaining utility. We identified a 9 item shortened MNAS score that maintained excellent discriminative properties and has the potential to reduce nursing burden. We have generated a shortened Finnegan Score version with similar test characteristics. Furthermore, our proposed 9 element FNAS score (sFNAS-9) has better test characteristics than the 7 element score proposed by Maguire. Our 9 element score was similar in performance to the 10 element of score of Gomez-Pomar, except in improved sensitivity for predicting >12 decision points in the FNAS score. Lastly, we demonstrated no evidence of significant age effects on scoring performance of the shortened tools.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

(1). Finnegan LP, Connaughton JF Jr, Kron RE, Emich JP. Neonatal abstinence syndrome: assessment and management. Addict Dis 1975;2(1–2):141–158. [PubMed: 1163358]

(2). Finnegan LP, Kron RE, Connaughton JF, Emich JP. Assessment and treatment of abstinence in the infant of the drug-dependent mother. Int J Clin Pharmacol Biopharm 1975 7;12(1–2):19–32. [PubMed: 1100537]

(3). D'Apolito K, Finnegan L. Assessing signs and symptoms of neonatal abstinence using the Finnegan Scoring Tool: an inter-observer reliability program. 2010; Available at: https://www.neoadvances.com. Accessed 01/16/2020

(4). Finnegan LP, Kaltenbach K. Neonatal abstinence syndrome In: Hoekelman RA, Friedman SB, Nelson N, Seidel HM, editors. Primary Pediatric Care. Second ed Mosby: St. Louis; 1992 p. 1367–1378.

(5). Kaltenbach K Assessment of the newborn prenatally exposed to drugs: The history. Semin Fetal Neonatal Med 2019 4;24(2):111–114. [PubMed: 30744981]

(6). Jones HE, Seashore C, Johnson E, Horton E, O'Grady KE, Andringa K, et al. Psychometric assessment of the Neonatal Abstinence Scoring System and the MOTHER NAS Scale. Am J Addict 2016 8;25(5):370–373. [PubMed: 27287965]

(7). Jones HE, Kaltenbach K, Heil SH, Stine SM, Coyle MG, Arria AM, et al. Neonatal abstinence syndrome after methadone or buprenorphine exposure. N Engl J Med 2010 12 9;363(24):2320–2331. [PubMed: 21142534]

(8). Kraft WK, Adeniyi-Jones SC, Chervoneva I, Greenspan JS, Abatemarco D, Kaltenbach K, et al. Buprenorphine for the Treatment of the Neonatal Abstinence Syndrome. N Engl J Med 2017 6 15;376(24):2341–2348. [PubMed: 28468518]

(9). Agthe AG, Kim GR, Mathias KB, Hendrix CW, Chavez-Valdez R, Jansson L, et al. Clonidine as an adjunct therapy to opioids for neonatal abstinence syndrome: a randomized, controlled trial. Pediatrics 2009 5;123(5):e849–56. [PubMed: 19398463]

(10). Gomez Pomar E, Finnegan LP, Devlin L, Bada H, Concina VA, Ibonia KT, et al. Simplification of the Finnegan Neonatal Abstinence Scoring System: retrospective study of two institutions in the USA. BMJ Open 2017 9 27;7(9):e016176–2017-016176.

(11). Maguire D, Cline GJ, Parnell L, Tai CY. Validation of the Finnegan neonatal abstinence syndrome tool-short form. Adv Neonatal Care 2013 12;13(6):430–437. [PubMed: 24300963]

(12). Devlin L, Breeze JL, Terrin N, Gomez Pomar E, Lester B, craig A, et al. Simplified Finnegan Scoring System to Assess the Need for Treatment in Neonatal Abstinence Syndrome (NAS). Pediatric Academic Societies, 5 2019, Baltimore, MD 2019.

(13). Breiman L. Classificaiton and Regression Trees. 1st ed. Boca Raton: Chapman and Hall/CRC; 2017.

(14). Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011 3 17;12:77–2105-12–77.

(15). Provenzi L, Olson K, Giusti L, Montirosso R, DeSantis A, Tronick E. NICU Network Neurobehavioral Scale: 1-month normative data and variation from birth to 1 month. Pediatr Res 2018 6;83(6):1104–1109. [PubMed: 29938699]

(16). Gibson KS, Stark S, Kumar D, Bailit JL. The relationship between gestational age and the severity of neonatal abstinence syndrome. Addiction 2017 4;112(4):711–716. [PubMed: 27886650]

(17). Liu AJ, Jones MP, Murray H, Cook CM, Nanan R. Perinatal risk factors for the neonatal abstinence syndrome in infants born to women on methadone maintenance therapy. Aust N Z J Obstet Gynaecol 2010 6;50(3):253–258. [PubMed: 20618243]

(18). Kaltenbach K, Holbrook AM, Coyle MG, Heil SH, Salisbury AL, Stine SM, et al. Predicting treatment for neonatal abstinence syndrome in infants born to women maintained on opioid agonist medication. Addiction 2012 11;107 Suppl 1:45–52.

(19). Westgate PM, Gomez-Pomar E. Judging the Neonatal Abstinence Syndrome Assessment Tools to Guide Future Tool Development: The use of Clinimetrics as Opposed to Psychometrics. Front Pediatr 2017 9 20;5:204. [PubMed: 28979894]

(20). Jones HE, Seashore C, Johnson E, Horton E, O'Grady KE, Andringa K. Measurement of neonatal abstinence syndrome: Evaluation of short forms. J Opioid Manag 2016 Jan-Feb;12(1):19–23. [PubMed: 26908300]

(21). Isemann BT, Stoeckle EC, Taleghani AA, Mueller EW. Early Prediction Tool to Identify the Need for Pharmacotherapy in Infants at Risk of Neonatal Abstinence Syndrome. Pharmacotherapy 2017 7;37(7):840–848. [PubMed: 28500629]

(22). Jones HE, Harrow C, O'Grady KE, Crocetti M, Jansson LM, Kaltenbach K. Neonatal abstinence scores in opioid-exposed and nonexposed neonates: a blinded comparison. J Opioid Manag 2010 Nov-Dec;6(6):409–413. [PubMed: 21269001]

(23). Gaalema DE, Scott TL, Heil SH, Coyle MG, Kaltenbach K, Badger GJ, et al. Differences in the profile of neonatal abstinence syndrome signs in methadone- versus buprenorphine-exposed neonates. Addiction 2012 11;107 Suppl 1:53–62.

(24). Gomez-Pomar E, Christian A, Devlin L, Ibonia KT, Concina VA, Bada H, et al. Analysis of the factors that influence the Finnegan Neonatal Abstinence Scoring System. J Perinatol 2017 7;37(7):814–817. [PubMed: 28383539]

(25). Kushnir A, Bleznak JL, Saslow JG, Stahl G. Nurses' Finnegan scoring of newborns with Neonatal Abstinence Syndrome not affected by time or day of the week. Am J Perinatol. 2020;37:224–30. [PubMed: 31600793]
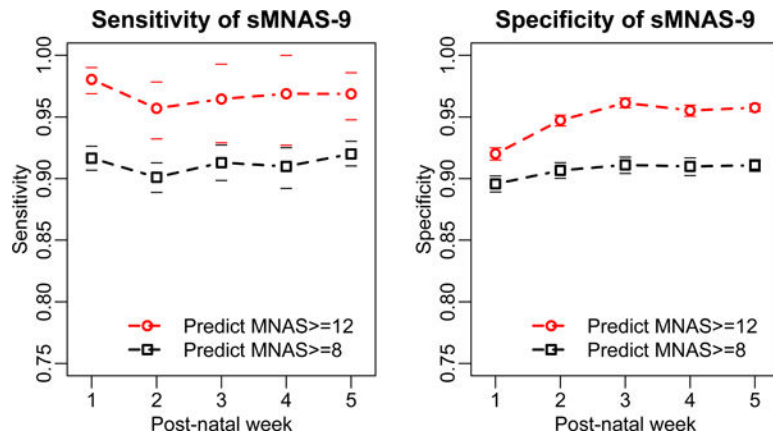
Figure 1.
Performance of the 9 element shorted MOTHER NAS scale (sMNAS-9) in the test set for postnatal weeks 1–5. Horizontal dashes of the same color indicate 95% confidence limits.
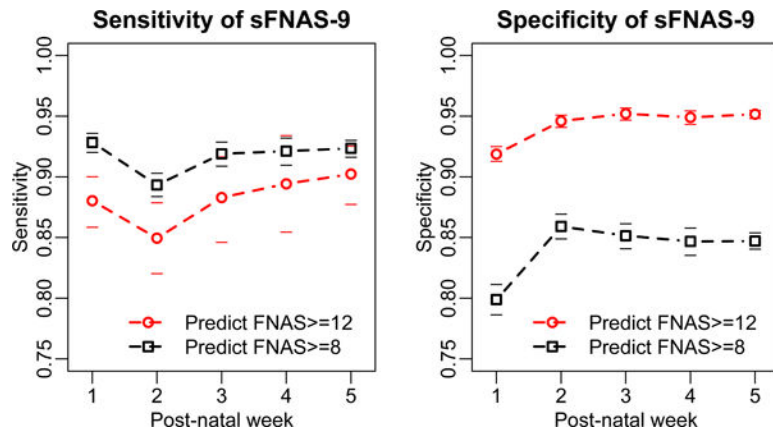
Figure 2.
Performance of the 9 element shorted Finnegan scale (sFNAS-9) in the test set for postnatal weeks 1–5. Horizontal dashes of the same color indicate 95% confidence limits.

Table 1.

Previously proposed short tools for evaluating NAS. Scoring systems were either the Finnegan(4) or the MOTHER NAS(7)

	Number of patients	Scoring system	Number of observations	Number of Scored Elements	Notes
Short scales for treatment decision making					
Gomez-Pomar(10)	367	Finnegan	40,294	10	Proposed shortened instrument threshold of 6 and 10 used to predict FNAS scores 8 and 12, respectively
Maguire(11)	171	Finnegan	33,856	7	Score of 8 highly correlated between proposed short form and standard long FNAS
Devlin(12)	424	Finnegan	NA	9	Presented in an abstract; prediction of highest score on the day pharmacotherapy initiated. Removed high pitch cry as highly variable among sites
Screening tools for initial NAS diagnosis					
Jones(22)	55	Finnegan	NA	3, 4, 5	Screening tool to distinguish opioid and non-opioid exposed
Jones(20)	131	Mother NAS	NA	5	Screening tool applied to pharmacologically treated or non-treated
Isemann(21)	264	Finnegan	NA	3	3 clinical signs, with further refinement based upon type of in utero opioid exposure

NA = not available

Table 2.

The full MOTHER NAS instrument (MNAS), which consists of 19 scored and 9 unscored items and proposed shortened instrument (sMNAS-9) with 9 scored items

Signs and Symptoms	Severity	Score	
		MNAS	sMNAS-9
Crying	Excessive high pitched Continuous high pitched	2 3	2 3
Sleeps	< 1 hours after feeding < 2 hours after feeding < 3 hours after feeding	3 2 1	3 2 1
Moro Reflex	Hyperactive Markedly Hyperactive	1 2	
Tremors: Disturbed	Hands or feet only, up to 3 seconds Arms or legs, over 3 seconds	1 2	
Tremors: Undisturbed	Hands or feet only, up to 3 seconds Arms or legs, over 3 seconds	1 2	1 2
Increased Muscle Tone	Difficult but possible to straighten arm and head lag present Unable to straighten arm and head lag absent	1 2	1 2
Fever > 37.3 C (99.2 F)		1	1
Tachypnea	Respiratory Rate >60/mm	2	2
Poor feeding	Takes >20 minutes, uncoordinated, takes small volume, frequent stops to breathe	2	2
Vomiting (or regurgitation)	Vomits whole feeds, or at least x 2/feed when not burping	2	2
Loose Stools	Diaper is half liquid/half solid ± water ring	2	2
Excoriation	Skin is red but intact or healing, no longer broken Skin not intact	1 2	
Generalized Seizure	Eyes staring, rapid involuntary eye movements, chewing, back arching, fist clenching, tonic-clonic movements	8	
Frequent Yawning	4 or more successive times	1	
Sweating	Wetness on forehead or upper lip	1	
Nasal Stuffiness	Any nasal noise	1	
Sneezing (4 or more successive times)	4 or more successive times	1	
Failure to thrive	Current weight 10% below birth weight	2	
Excessive Irritability	Consoling calms infant in 5 min or less	1	
	Consoling calms infant in 6 – 15 min	2	
	Consoling takes > 15 min or is unsuccessful. Baby is sensitive or aversive to sound, light touch or, unable to calm by self.	3	
	Summed Score	0 – 43	0 –19
Recorded but unscored elements			
Myoclonic jerks, Mottling, Convulsions, Fever 38.4 C (101.2 F), Retractions, Nasal flaring,			
Excessive Sucking, Projectile Vomiting, Watery Stools. All noted as present or absent			

Table 3.

Overall performance of optimized short NAS scales in the test set.

Scale	Threshold for Dose Escalation		Sensitivity	95% confidence limit		Specificity	95% confidence limit		Misclassification error
	MNAS	sMNAS		Lower	Upper		Lower	Upper	
sMNAS-9	8	5	0.911	0.906	0.916	0.907	0.904	0.909	0.092
	12	7	0.972	0.964	0.980	0.948	0.947	0.950	0.051

Scale			Sensitivity	95% confidence limit		Specificity	95% confidence limit		Misclassification error
	FNAS	sFNAS		Lower	Upper		Lower	Upper	
sFNAS-9	8	5	0.918	0.914	0.922	0.843	0.839	0.847	0.126
	12	7	0.882	0.870	0.894	0.945	0.943	0.948	0.058

sMNAS = shortened MOTHER NAS score, MNAS = 19 item full MOTHER NAS score, FNAS = Finnegan NAS scoring system, sFNAS = shortened Finnegan NAS score

Table 4.

Comparison of the full Finnegan neonatal abstinence scoring system (FNAS) and shortened FNAS proposals. The numbers in rows to the right of scored elements and severity refer to the number of points an infant would receive if the sign was present. Gomez-Pomar contained 10, Maguire 7, and proposed sFNAS-9 contained 9 scored items.

Scored Elements	Severity	FNAS	Gomez-Pomar	Maguire	sFNAS-9
Number of scored items		*21*	*10*	*7*	*9*
Trigger for clinical decision for FNAS ≥ 8		*8*	*6*	*8*	*5*
Trigger for clinical decision for FNAS __12		*__12*			*__7*
Scored Elements	**Severity**				
Crying	Excessive high pitched	2	2 #	2	2
	Continuous high pitched	3		3	3
Excessive sucking		1	1	1	1
Poor feeding		2	2		2
Vomiting or Regurgitation		2	2 #		2
Projectile vomiting		3			3
Stools	Loose	2	2 #		2
	Watery	3			3
Sleeps	< 1 hours after feeding	3	3	3	3
	< 2 hours after feeding	2	1 #	2	2
	< 3 hours after feeding	1		1	1
Tremors	Disturbed: Mild	1	1 #		1
	Disturbed: Moderate-Severe	2			2
	Undisturbed: Mild	3	5 #	3	3
	Undisturbed: Moderate-Severe	4		4	4
Increased Muscle Tone		2	2	2	
Tachypnea	Respiratory Rate >60/min	1	1 #	1	
	RR >60/min + retractions	2		2	
Fever	Fever > 37.3 C (99.2 F)	1			1
	Fever > 38.4 C (101.2 F)	2			2
Nasal Stuffiness		1	1		
Sweating		1		1	
Moro Reflex	Hyperactive	2			
	Markedly Hyperactive	3			
Sneezing (4 or more successive times)		1			
Excoriation		1			
Generalized Seizure (or convulsion)		5			
Frequent Yawning (>3 successive)		1			
Mottling		1			
Nasal flaring		2			

		FNAS	Gomez-Pomar	Maguire	sFNAS-9
Myoclonic jerks		3			

[#]= for all severity

Table 5.

Performance metrics of published shortened Finnegan Score Instruments. Threshold represents the score at which a dosage change would take place. 95% confidence intervals are given in parentheses. Author provided is the analysis presented in the referenced manuscript. Analysis Using TJUH Data are test characteristics of the published short forms using data from Thomas Jefferson University Hospital.

Predictor Model	Ability to Predict Finnegan Score	Shortened Instrument Threshold for Dose Escalation	Author Provided		Analysis Using TJUH Data	
			Sensitivity	Specificity	Sensitivity	Specificity
Gomez-Pomar (24)	8	6	0.888 (0.874–0.903)	0.883 (0.870–0.895)	0.906 (0.904–0.908)	0.822 (0.820–0.825)
	12	10	0.637 (0.587–0.686)	0.992 (0.990–0.994)	0.650 (0.640–0.660)	0.965 (0.964–0.966)
Maguire (11)	8	8	N/A	N/A	0.833 (0.830–0.836)	0.717 (0.714–0.720)
	12	12	N/A	N/A	0.845 (0.838–0.853)	0.812 (0.810–0.814)

sFNAS = shortened Finnegan Scores, TJUH = Thomas Jefferson University Hospital