Journal of Perinatology (2017) 00, 1–4

www.nature.com/jp

ORIGINAL ARTICLE

Analysis of the factors that influence the Finnegan Neonatal Abstinence Scoring System

E Gomez-Pomar[1], A Christian[2], L Devlin[3], KT Ibonia[1], VA Concina[1], H Bada[1] and PM Westgate[4]

OBJECTIVE: To evaluate factors that can influence the Finnegan Neonatal Abstinence Score (FNAS).
STUDY DESIGN: Retrospective analysis of 367 patients admitted to two level IV neonatal intensive care units. Linear mixed effects models were developed to evaluate daily census, time of the day, and day of the week as fixed effect predictors. The degree of influence that nurses had on FNAS variability was also estimated.
RESULTS: Bivariate analyses showed that daily census and the time of day have significant influence on the FNAS in institution 1, with minimal clinical significance. The proportion of variation in the FNAS attributable to differences in nurses was of 9.8% and 5.1% for institutions 1 and 2, respectively ($P < 0.0001$).
CONCLUSIONS: The minimal influences of extraneous factors on the FNAS support the clinical utility of the scoring system in the assessment and management of infants with Neonatal Abstinence Score.

Journal of Perinatology advance online publication, 6 April 2017; doi:10.1038/jp.2017.40

INTRODUCTION

Neonatal Abstinence Syndrome (NAS) refers to a constellation of signs in newborn infants resulting from the cessation of exposure to maternal opioids used during pregnancy. NAS is characterized by gastrointestinal, respiratory, autonomic, and central nervous system disturbances, which may have detrimental effects on the newborn.[1–8] Since the first case of NAS reported in 1892,[5] the incidence of NAS has dramatically increased.[3,9–12] The rate of NAS cases increased from a rate of 7/1000 neonatal intensive care admissions in 2004 to 27/1000 admissions in 2013.[13]

The Finnegan Neonatal Abstinence Score (FNAS), developed in 1975 to quantify the severity of withdrawal signs in the neonate[3,4] is a validated assessment tool with strong inter-user reliability (mean inter-rater reliability coefficient of 0.82 (0.75–0.96)).[3,4] It provides a diagnostic score which serves as a basis for initiation and adjustment of drug dosage. Finnegan *et al.* have proposed to assess infants every 3 or 4 h according to the feeding schedule and to use three consecutive scores $\geqslant 8$ or two scores $\geqslant 12$ for the initiation of pharmacological treatment.[2–4] Contiguous scores < 8 are often used as a measure of readiness for weaning pharmacologic therapy.[4] The threshold for treatment was noted to be adequate by Zimmermann-Baer *et al.*[14] by applying the FNAS to non-opiate exposed term infants for up to 5 to 6 weeks and finding that the 95th percentile of the scores never exceeded 8.

The 21-item FNAS allows a thorough assessment of infants, but the tool is lengthy and requires training and experience for its correct administration.[15] Some FNAS items have a subjective component that may influence the tool's validity.[5,16] To our knowledge, there are no studies that evaluated the variability of the FNAS due to inconsistency in scoring or to other external factors. Therefore, we examined the factors that may influence the FNAS that may result in significant variability in its administration in the clinical setting.

METHODS

This retrospective study was conducted using electronic medical records data from two level IV neonatal intensive cares; both are regional referral centers. The study was approved by the Institutional Review Board of each institution. Inclusion criteria included gestational age >37 0/7 weeks, admission to the neonatal intensive care for withdrawal manifestations after any *in-utero* opioid exposure, and signs of NAS assessed using FNAS. Exclusion criteria included gestational age <37 weeks, exposure to psychoactive substances without opioid exposure, and/or major congenital malformations. At each institution, the nurses who performed the FNAS had training with experienced scorers (video, demonstration by trainer and reverse demonstration by trainee) and testing for reliability (trainer and trainee assess and score same infant) before being assigned care of infants with NAS; retraining in scoring with FNAS was done annually.

Since the total FNAS score is utilized as the basis to initiate or change therapy, we used the total score at each evaluation for analysis. The protocol for the evaluation of NAS in infants between the two institutions is similar except that in institution 2 the interval of FNAS assessment is increased from every 3 to every 6 h once the infant is 'captured', that is, FNAS scores consistently <8 for a period of up to 24 h after initiation of therapy. The FNAS was designed to be used during the neonatal period and has not been tested to be used beyond. Some infants in our cohort were tested beyond the neonatal period. We decided to include them since our objective was to evaluate the factors that may affect the variability of the FNAS and not the ability of the score to diagnose NAS in infant.

We examined the following factors for their potential influence on the variability of the FNAS: unit daily census (number of infants in the unit each day); day of the week (weekday vs weekend); time of the day (day shift: 0700–1900 hours and night shift: 1900–0700 hours); number of scores given by a nurse (during the study period); and nursing ID. Nursing ID allows us to assess if any variance in the FNAS exists due to inconsistency across nurses when scoring.

[1]Division of Neonatology, Department of Pediatrics, University of Kentucky, Lexinton, KY, USA; [2]College of Medicine, University of Kentucky, Lexinton, KY, USA; [3]Division of Neonatology, Department of Pediatrics, University of Louisville, Louisville, KY, USA and [4]Department of Biostatistics, College of Public Health, University of Kentucky, Lexinton, KY, USA. Correspondence: Dr E Gomez-Pomar, Division of Neonatology, Department of Pediatrics, University of Kentucky, 138 Leader Avenue, Room 018, Lexington 40508-9983, KY, USA.
E-mail: Enrique.gomez@uky.edu
Received 17 November 2016; revised 28 February 2017; accepted 7 March 2017

Statistical analysis

Descriptive statistics and linear mixed effects models were used to assess unadjusted associations. Primary analyses used linear mixed effects models, one for each institution, in which the FNAS was the outcome of interest and daily census, time of day, and day of the week were fixed effects predictors. A random nurse effect using nursing ID was included to determine the degree of influences nurses have on scoring. Statistical correlation among repeated measurements over time from the same infant was accounted for via an exponential power decay structure. Two-sided *t*-tests were utilized for fixed effects, whereas likelihood ratio tests were used to test for between-nurse variability in scoring. All tests were conducted at the 0.05 level and were performed in SAS version 9.4 (SAS Institute, Cary, NC, USA).

RESULTS

The data comprised of a whole year's (2014) observations from 185 subjects of institution 1 (University of Louisville) for a total of 27 447 score observations, and from 182 subjects from institution 2 (University of Kentucky) for a total of 12 759 observations. The number of observations per infant in institution 1 ranged from 1 to 605, with a mean and s.d. of 148 (116) and 25th, 50th and 75th percentiles of 38, 153 and 210, respectively. The number of observations per infant in institution 2 ranged from 1 to 310, with a mean (s.d.) of 70 (59) and 25th, 50th and 75th percentiles of 18, 60 and 106, respectively.

The distribution of the scores in institution 1 showed a mean (s.d.) of 6.23 (2.5) with a variance of 6.34 and 25th, 50th and 75th percentiles of 5, 6 and 8, respectively; 83.5% of the observations were below 8. The FNAS for institution 2 showed a mean (s.d.) of 7.4 (3.8) with a variance of 14.34 and 25th, 50th and 75th percentiles of 5, 7 and 10, respectively; 64.6% of the observations were below 8.

Figure 1 shows the daily census for both institutions. The mean (s.d.) was 97.4 (8.8) with 25th, 50th and 75th percentiles of 92, 96 and 101, respectively, for institution 1, whereas the mean (s.d.) was 53.9 (9.4) with 25th, 50th and 75th percentiles of 48, 56 and 60, respectively, for institution 2. The total number of observations, or scores, were comprised of 13 859 (50.5%) daytime scores (0700–1900 hours) from institutions 1 and 6, 484 (50.8%) from institution 2.

Analyses of the individual factors that we hypothesized could affect the mean FNAS are presented in Table 1. For institution 1, the mean FNAS scores at nighttime were significantly higher that mean scores at daytime; mean scores did not differ between weekday and weekend. For institution 2, time of day and day of the week had no effect on the mean FNAS.

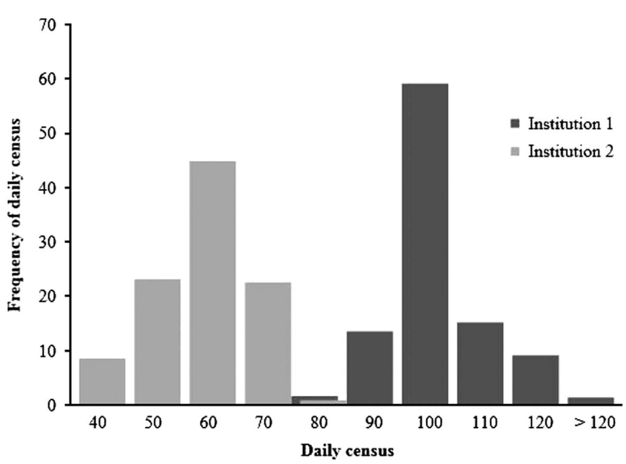

Figure 1. Histogram representing the daily census distribution in both NICUs. NICU, neonatal intensive care.

Nurse experience was analyzed using the number of scores that the nurse performed during the study period. For institution 1, 328 nurses performed the FNAS and the number of FNAS scores done by each nurse ranged from 1 to 565 with a mean (s.d.) of 83.7 (138.5) and 25th, 50th and 75th percentiles of 12, 31.5 and 64.5, respectively. For institution 2, 320 nurses performed the FNAS and the number of FNAS scores ranged from 1 to 346 with a mean (s.d.) of 39.9 (48.5) and 25th, 50th and 75th percentiles of 5, 25 and 58, respectively. To analyze the association between the number of scores given per nurse and the mean FNAS, a linear model was assumed showing an estimated slope of 0.0008 ($P < 0.001$) and −0.0007 ($P = 0.09$) for institutions 1 and 2, respectively. This result implies that each additional score done by a nurse is estimated to increase the FNAS by 0.0008 in institution 1.

The results from the linear mixed effects models are presented in Table 2. Daily census was a statistically significant factor for institution 1, resulting in an estimated small decrease of 0.015 in the mean FNAS score for each additional increase in the census by 1; this estimate is very small and may only have a small clinical impact. The impact of daily census for institution 2 was not significant. Furthermore, the time of day and day of the week were not significant predictors of FNAS for both institutions.

Table 1. Effects of day of the week and time of day by institution on mean FNAS

	Institution 1		*Institution 2*	
	Scores	*Mean ± s.d. median (IQR)*	*Scores*	*Mean ± s.d. median (IQR)*
All scores	27447	6.2 ± 2.5 6 (5, 8)	12416	7.4 ± 3.8 7 (5, 10)
Weekday	15713	6.2 ± 2.5 6 (5, 8)	7190	7.4 ± 3.8 7 (5, 10)
Weekend	11734	6.2 ± 2.5 6 (4, 8)	5569	7.4 ± 3.8 7 (5, 10)
0700–1900 hours	13588	6.1 ± 2.5[a] 6 (4, 8)	6275	7.4 ± 3.8 7 (5, 10)
1900–0700 hours	13859	6.3 ± 2.5 6 (5, 8)	6484	7.4 ± 3.8 7 (5, 10)

Abbreviations: FNAS, Finnegan Neonatal Abstinence Score; IQR, interquartile range. [a]Statistically significant difference in mean scores comparing time of the day in institution 1, $P < 0.0001$.

Table 2. Linear mixed effect models analysis of factors (daily census, time of day and weekday versus weekend) by institution affecting FNAS[a]

	Estimate	*95% CI*	P-*value*
Institution 1			
Daily Census	−0.015	(−0.019, −0.011)	< 0.0001
1900–0700 hours	Reference	(−0.153, 0.115)	0.78
0700–1900 hours	−0.019		
Weekday	Reference	(−0.049, 0.109)	0.46
Weekend	0.030		
Institution 2			
Daily census	0.01	(−0.01, 0.02)	0.26
1900–0700 hours	Reference	(−0.16, 0.18)	0.94
0700–1900 hours	0.01		
Weekday	Reference	(−0.08, 0.30)	0.26
Weekend	0.11		

Abbreviations: CI, confidence interval; FNAS, Finnegan Neonatal Abstinence Score. [a]Daily census is significant factor only for institution 1.

Table 3. Stratified analysis by day of week and time of day for effect of nurses on FNAS variability by institution

	Weekday (day)	*Weekday (night)*	*Weekend (day)*	*Weekend (night)*
Institution 1	14.8%[a]	12.5%[a]	16.2%[a]	7.6%[a]
Institution 2	3.2%[a]	8.3%[a]	7.8%[a]	7.9%[a]

Abbreviation: FNAS, Finnegan Neonatal Abstinence Score. [a]Statistically significant difference in the variation of the FNAS attributable to nurses; $P<0.0001$.

Evaluation of the variability of the FNAS only showed a greater variability in FNAS scores in institution 2, as the models provide total variance (s.d.) estimates of 6.25 (2.5) and 14.31 (3.78) for institutions 1 and 2, respectively. However, nurses contributed similar amounts of variation to these total variances. Specifically, nurses were estimated to have contributed 0.61 (Δs.d. 0.13) and 0.73 (Δs.d. 0.09) units of variation in institutions 1 and 2, respectively (Δ indicates the increase in s.d. due to the variances arising from scoring inconsistencies). Although these were statistically significant contributions ($P<0.0001$), they only accounted for 9.8% (0.61/6.25 = 0.098) and 5.1% (0.73/14.30 = 0.051) of the total variation in institutions 1 and 2, respectively. In other words, if these slight inconsistencies in scoring across nurses had not occurred, the estimated variance (s.d.) would be 5.64 (2.37) and 13.58 (3.69) as opposed to estimates of 6.25 (2.5) and 14.31 (3.78) for institutions 1 and 2, respectively. Therefore, the clinical impact of differential scoring by nurses is weak. Further analyses of the FNAS variability due to nursing were performed by combining day of the week and time of day using a stratified approach. We encountered that the amount of variation in FNAS attributable to differences in nurses was statistically significant in all strata (Table 3).

DISCUSSION

The present study supports that the FNAS is an objective tool. Although some factors may have a statically significant effect on the scores, the effects are minimal and thus, not clinically significant. To our knowledge, this is the first systematic analysis of the common factors that can affect the performance of the FNAS as a diagnostic tool.

The authors of the original report of the FNAS reported a high variability in an infant's scores, particularly during the first day of life.[3] This finding was later confirmed and expanded by Zimmermann-Baer *et al.*[14] who found that in infants without NAS there was an increase in the variability of the scores on day 2 of life with a sustained decrease by day 3 of life. They also reported a higher daytime scoring compared to nighttime. This difference became significant at 5 to 6 weeks of life.

In our study, daily unit census, time of the day, and day of the week did not show a clinically significant impact on the FNAS. Although daily census had a statistically significant influence on FNAS in institution 1, the association was clinically minimal. Daytime scores were significantly lower than nighttime scores in institution 1, but the difference in mean FNAS was only 0.2. This difference in the day and nighttime scores is opposite to the result reported by Zimmermann-Baer *et al.*[14] who found that in infants without NAS, there is a decreased score during the nighttime compared to the daytime. We speculate that this difference in findings may be due to the alteration of the circadian rhythm reported in infants with NAS compared to infants without NAS.[2,10] The lack of significant difference in scores between weekday and weekend is a remarkable result. In both institutions, experienced nurses typically worked during the weekdays and while those with a few years of experience were assigned on weekends.

Overall, the clinical significance of the variability in the FNAS due to inconsistency in scoring across nurses is minimal. On one extreme, the largest proportion of variation in the FNAS due to inconsistent scoring that we found was 16% (weekend day/institution 1). This variability could occasionally make a non-clinically significant value become significant (or vice versa) but the decision to treat or modify treatment dosing is usually based in consecutive assessment scores rather than a single score. We speculate that variation in scores are due to some extent to an intrinsic variability of neonates in their scores, which was already identified when the original FNAS was developed.[3] The authors noticed that an infant can have a score higher than 8 at one evaluation but with subsequent lower score at the next assessment. This variability was therefore addressed by the recommendation of obtaining three consecutive scores of 8 or higher or two consecutive scores of 12 or higher in order to provide pharmacological treatment.[2–4,14] These subtle changes were noted in non-exposed infants having scores of more than 8 at certain times but these scores 'normalized' on subsequent assessment.[14]

Nursing competence related to scoring is a factor that we could not fully analyze. The FNAS has items with one to multiple levels of severity,[2] which leaves room for subjectivity that can lead to an increase in the variability of the reported scores.[17,18] This variability also led to question the objectivity of the FNAS.[15] Lucas *et al.*[18] evaluated the inter-rater reliability of the FNAS by comparing the nurses' scores with a perfect scorer pre and post re-education; they found that 90% of the participants showed 10% or more improvement in the reliability of the post training scores.

We considered the number of scores that were given per nurse as a reflection of a nurse's experience in administering the FNAS. From our analysis, we found an increase in the mean FNAS of 0.08 and a decrease of 0.07 for each additional 100 scores per nurse for institutions 1 and 2, respectively. These findings indicate only a negligible coefficient and therefore such changes may not be clinically significant in practice.

We found that the estimated proportions of variability in the FNAS attributable to nurses were only 9.8 and 5.1% for institutions 1 and 2, respectively (when controlling for unit census, time of the day, and day of the week). Therefore, our findings are in support of the FNAS as an objective tool for the evaluation and management of infants with NAS. Because of the influence, although minimal, of external factors on the variability of FNAS and the inherent variability within the infant, we support the recommendation of obtaining consecutive scores to determine the need for treatment of infants with NAS.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1 Hudak ML, Tan RC, Committee On D, Committee On F, Newborn, American Academy of P. Neonatal drug withdrawal. *Pediatrics* 2012; **129**(2): e540–e560.

2 Finnegan LP, Kaltenbach K. Neonatal abstinence syndrome. In: Hoekelman RA FS, Nelson NM (eds). *Primary Pediatric Care*, 2nd edn. Mosby: St. Louis, 1992, pp 1367–1378.

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

4 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; **12**(1-2): 19–32.

5 Jones HE, Fielder A. Neonatal abstinence syndrome: historical perspective, current focus, future directions. *Prev Med* 2015; **80**: 12–17.

6 Desai RJ, Huybrechts KF, Hernandez-Diaz S, Mogun H, Patorno E, Kaltenbach K *et al.* Exposure to prescription opioid analgesics *in utero* and risk of neonatal abstinence syndrome: population based cohort study. *BMJ* 2015; **350**: h2102.

7 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; **6**(6): 409–413.
8 Hoegerman G, Wilson CA, Thurmond E, Schnoll SH. Drug-exposed neonates. *West J Med* 1990; **152**(5): 559–564.
9 Patrick SW, Schumacher RE, Benneyworth BD, Krans EE, McAllister JM, Davis MM. Neonatal abstinence syndrome and associated health care expenditures: United States, 2000-2009. *JAMA* 2012; **307**(18): 1934–1940.
10 Kaltenbach K, Jones HE. Neonatal abstinence syndrome: presentation and treatment considerations. *J Addict Med* 2016, **10**: 224–228.
11 Davies H, Gilbert R, Johnson K, Petersen I, Nazareth I, O'Donnell M *et al*. Neonatal drug withdrawal syndrome: cross-country comparison using hospital administrative data in England, the USA, Western Australia and Ontario, Canada. *Arch Dis Child Fetal Neonatal Ed* 2016; **101**(1): 26–30.
12 Turner SD, Gomes T, Camacho X, Yao Z, Guttmann A, Mamdani MM *et al*. Neonatal opioid withdrawal and antenatal opioid prescribing. *CMAJ Open* 2015; **3**(1): E55–E61.
13 Tolia VN, Patrick SW, Bennett MM, Murthy K, Sousa J, Smith PB *et al*. Increasing incidence of the neonatal abstinence syndrome in U.S. neonatal ICUs. *N Engl J Med* 2015; **372**(22): 2118–2126.
14 Zimmermann-Baer U, Notzli U, Rentsch K, Bucher HU. Finnegan neonatal abstinence scoring system: normal values for first 3 days and weeks 5-6 in non-addicted infants. *Addiction* 2010; **105**(3): 524–528.
15 Bagley SM, Wachman EM, Holland E, Brogly SB. Review of the assessment and management of neonatal abstinence syndrome. *Addict Sci Clin Pract* 2014; **9**(1): 19.
16 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; **25**(5): 370–373.
17 Orlando S. An overview of clinical tools used to assess neonatal abstinence syndrome. *J Perinat Neonatal Nurs* 2014; **28**(3): 212–219.
18 Lucas K, Knobel RB. Implementing practice guidelines and education to improve care of infants with neonatal abstinence syndrome. *Adv Neonatal Care* 2012; **12**(1): 40–45.