**Title**

# A Pilot Validation Study of the Newborn Behavioral Observations System: Associations with Salivary Cortisol and Temperament.



**Journal**
Journal of Developmental and Behavioral Pediatrics, 41(9)

**ISSN**
0196-206X

**Authors**
Congdon, Jayme L
Nugent, J Kevin
McManus, Beth M
et al.

**Publication Date**
2020-12-01

**DOI**
10.1097/dbp.0000000000000842

Peer reviewed

HHS Public Access
Author manuscript
*J Dev Behav Pediatr*. Author manuscript; available in PMC 2021 December 01.

Published in final edited form as:
*J Dev Behav Pediatr*. 2020 December ; 41(9): 716–723. doi:10.1097/DBP.0000000000000842.

# A Pilot Validation Study of the Newborn Behavioral Observations System: Associations with Salivary Cortisol and Temperament

**Jayme L. Congdon, MD, MS**[a], **J. Kevin Nugent, PhD**[b], **Beth M. McManus, PT, MPH, ScD**[c], **Michael Coccia, MS**[d], **Nicole R. Bush, PhD**[e]

[a]Division of General Pediatrics, Department of Pediatrics, University of California San Francisco; San Francisco, California

[b]Division of Developmental Medicine, Department of Pediatrics, Boston Children's Hospital, Harvard Medical School; Boston, Massachusetts

[c]Department of Health Systems, Management, and Policy, University of Colorado; Denver, Colorado

[d]Division of Developmental Medicine, Department of Psychiatry, University of California San, Francisco; San Francisco, California

[e]Division of Developmental Medicine, Department of Psychiatry and Department of Pediatrics, University of California San Francisco; San Francisco, California

## Abstract

**Objective—**There are few standardized neonatal neurobehavioral instruments available for longitudinal child development research. We adapted the established clinical tool, the Newborn Behavioral Observations (NBO) system, for research by standardizing the administration protocol and expanding the three-point coding scale to five points.

**Methods—**We administered the five-point NBO to 144 racially/ethnically diverse late-preterm or term infants born to low-income women (average age five weeks). Cronbach's alphas were calculated to determine internal consistency reliability of Autonomic, Motor, Organization of State, and Responsivity subscales. We examined concurrent validity using subscale associations with infant salivary cortisol reactivity to the NBO and maternally reported infant temperament.

**Results—**Two of the four NBO subscales, Organization of State and Responsivity, had excellent (0.91) and good (0.76) reliability, respectively, and were retained for further analyses. Infants with higher Organization of State scores (more optimal regulation) demonstrated lower cortisol reactivity ($r = -0.30$, $p < 0.01$) and temperamental negativity ($r = -0.16$, $p < 0.05$). Responsivity was unrelated to cortisol reactivity or temperament.

**Conclusions—**State regulation, as measured by the five-point NBO, was associated with a biologic marker of infant stress response to the NBO administration and reported temperament.

**Address correspondence to:** Jayme Congdon, Division of General Pediatrics, Department of Pediatrics, University of California San Francisco; 550 16th Street, San Francisco, California 94158, United States; Phone 415-476-1000, Fax 415-467-5360.

**Disclosure Statement:** Dr. Nugent is an author of the proprietary Newborn Behavioral Observations (NBO) system and receives associated royalty income. The other authors report no financial disclosures or conflicts of interest.

Poor reliability of the NBO's three-item Autonomic and seven-item Motor subscales suggests that further psychometric research in other samples, and likely refinement, are needed. Given the paucity of neurobehavioral assessment tools for infants, these findings justify such research as next steps in the incremental progression toward the development of a practical, reliable, and predictive measure of early neurobehavioral development.

### Keywords

child development; newborn infants; patient outcome assessment; temperament; Biomarkers

## INTRODUCTION

A complex web of environmental and genetic factors influence development and growth from the prenatal period into adulthood.[1] Advances in neurobiology have elucidated the importance of identifying early risk and clinical signs of suboptimal development to maximize the impact of interventions aimed at helping children achieve their developmental potential.[2] Early prediction of childhood developmental outcomes is thus paramount to optimizing a child's trajectory.[3] Clinical neurologic exam and neuromotor assessment tools are valuable for the early detection and prognostication of preterm or otherwise neurologically at-risk neonates.[4,5] However, for relatively healthy, late preterm and term neonates for whom we wish to detect subtler developmental differences, there are few validated options, hindering progress in the field of early intervention for at-risk infants.

Functional developmental outcomes of generally healthy neonates are perhaps better predicted using a neurobehavioral approach than the established clinical neurologic assessments. Neurobehavior encompasses state and emotion regulation, which predict a variety of developmental outcomes, including psychopathology and social functioning later in childhood.[6] The Neonatal Behavioral Assessment Scale (NBAS)[7] was one of the early instruments that utilized a neurobehavioral model to evaluate neonates, marking an important step forward in the understanding of and ability to measure early infant development. The NBAS has been widely studied and shown to predict later developmental outcomes.[4,5,8] There have been various attempts to improve the psychometric properties of the NBAS,[9,10] with inconsistent gains in the subscales' internal consistency reliability.[4] Brazelton collaborated with Tronick and Lester to incorporate many of the NBAS items into an updated comprehensive neurobehavioral assessment tool: the NICU Network Neurobehavioral Scale (NNNS).[11] A standardized administration protocol and published norms established the NNNS as a useful clinical and research tool. However significant time required for training and administration of both the NNNS and NBAS limit their practicality and accessibility to most clinicians and researchers, leading NBAS author Nugent and colleagues to create a shorter clinical tool, the Newborn Behavioral Observations (NBO) system.[12]

The NBO is a family-centered, developmentally supportive instrument that was designed to elucidate newborn neurobehavior for the purposes of crafting interventions that promote optimal newborn self-regulation and caregiver-infant interaction. The NBO is based on an appreciation of the richness and complexity of the newborn's behavioral repertoire and the agency of the baby in shaping parent-infant interactions.[7,13] NBO items focus on the baby's behavior and communication cues, which are a window into the baby's mental state. The twenty neurobehavioral items are designed to capture the infant's visual, auditory, perceptual, and self-regulatory abilities as the infant attempts to stabilize autonomic, motor, and state behavior across the first weeks and months of life. Since its inception in 2007, the NBO has been widely used in a number of postnatal inpatient (e.g. well baby and special care nurseries), outpatient (e.g. well baby visits), and home visit (e.g. early intervention and nurse home visiting) settings. The NBO has been shown to improve caregiver-infant interactions[13,14] and provider confidence.[15]

While the NBO and NNNS were both inspired by the NBAS and measure similar domains, there is a subtle yet important distinction in their overall intent, aside from the aforementioned differences in scope and time for training and administration. The NNNS is highly standardized to optimize its psychometric properties, including inter-rater reliability. In contrast, the NBAS and NBO are strengths-based and are intended to capture an infant's optimal performance over a variable period of observation, an approach that may be more susceptible to differences in examiner skill and the quality of interactions between the examiner, infant, and caregiver. These aspects of the NBO that make it a valuable relational tool may also present a challenge in settings where standardization is important (e.g. clinical research).

In an effort to expand the utility of the NBO to address this important gap in the availability of a brief, validated neurodevelopmental assessment, co-authors Drs. Nugent and McManus used an iterative process to adapt the original three-point scale into a five-point scale, whereby they revised the scoring scheme based upon feedback from content experts. The resultant five-point scale was formulated to increase variability, limit floor and ceiling effects, and improve specificity to the description of infant neurobehaviors and an infant's risk profile. With an eye toward providing early and more individualized interventions, greater detection of subtle differences in neurobehaviors can advance the understanding and more precise categorization of infants' needs for facilitation and support. Although the five-point scale has been used clinically, its psychometric properties are not yet well established, as it has not previously been published or used in research. This lack of psychometric data is significant, as the NBO utilizes a subset of items from the NBAS and retained its theory-derived factor structure, which has not consistently demonstrated adequate internal consistency reliability.[4,9,10]

For the present study, we sought to maintain the clinical applicability and practicality of the NBO, while strengthening its utility as a clinical decision-making and outcomes research tool. We developed a standardized procedure (see Addendum) and refined coding descriptions to capture linear incremental shifts in the construct of interest with replicable results. Here we pilot the five-point NBO in a relatively healthy sample of infants, describing the psychometric properties and concurrent validity for two measures of clinical interest: neonatal salivary cortisol reactivity and maternal report of infant temperament. As the downstream product of the hypothalamic-pituitary-adrenal axis, cortisol has become an increasingly utilized biomarker for individual stress response.[16] While the stress response system develops according to early environmental cues, temperament refers to relatively stable individual differences in emotionality, attention, activity, and self-regulation.[6] Early childhood cortisol and temperament have both been shown to predict social functioning and risk for psychopathology later in childhood and adulthood,[17] thus offering clinically relevant measures of concurrent validity for the NBO. We hypothesized that infants who demonstrated more optimal neurobehavioral development on the NBO would be rated by their mothers as temperamentally less reactive and better regulated and would have lower levels of cortisol reactivity to the developmentally appropriate behavioral challenges of the NBO.

## METHODS

### Participants

Participants were drawn from the Maternal Adiposity, Metabolism, and Stress (MAMAS) study, a controlled trial of a mindfulness-based small-group intervention to reduce stress and prevent excess weight gain during pregnancy in predominantly low-income women. This sample was recruited from obstetric clinics and community centers throughout the San Francisco Bay Area in California, United States. Due to the MAMAS study focus on weight gain in populations experiencing high stress levels, participants were overweight or obese and predominantly low-income; of note, given that approximately 66% of women in the U.S. are either overweight (body mass index (BMI) = 25 to < 30 kg/m$^2$) or obese (BMI ≥ 30 kg/m$^2$), and rates are even higher among women of color (82% of African Americans, 77.1% of Latinas),[18] this sample was fairly consistent with the U.S. childbearing population in terms of weight. Inclusion criteria were age 18–45 years, 8–23 weeks of pregnancy with a singleton gestation, BMI > 25 kg/m$^2$, and household income less than 500% of the federal poverty level, a U.S. indicator of low to middle-income. Medical conditions that may interfere with baseline body composition (e.g. polycystic ovarian syndrome, preexisting diabetes, active substance abuse) were exclusionary.

MAMAS participants with live births were contacted postpartum for recruitment into the Stress, Eating, and Early Development (SEED) follow-up study of offspring (details published elsewhere).[6] Of the 215 MAMAS participants, 13 were not eligible for enrollment in SEED (five dropped out of the MAMAS study, three miscarriages, one fetal death, one moved out of the area, and three were lost to follow up), resulting in 202 potentially eligible participants. Of the 202 eligible dyads, 162 (80%) enrolled in SEED. Of the 162 women enrolled in SEED, 144 completed in-person postnatal assessments by 12 weeks (adjusted for gestational age at birth) due to the recommended age range for the NBO (nine were enrolled late, nine assessments were completed by phone and lacked in-person exam data). There were no differences in baseline characteristics between the women who consented to postnatal follow-up compared to those who declined or who were lost to follow-up. Table 1 provides a description of the study's sample demographic and obstetric characteristics.

This study was approved by the Institutional Review Board at University of California, San Francisco. Written informed consent was obtained from all maternal participants.

## Measures

**NBO—**The NBO[12] was used to assess newborn behavior and neurodevelopment. The version piloted in this study utilized the same 20 observation items as the three-point NBO, requiring approximately 15–20 minutes to administer. The items were administered using a standardized protocol adapted for this study, and observations were coded by the trained NBO administrator using a five-point scale (see Addendum). The 20 NBO items are subcategorized into four domains (Autonomic, Motor, Organization of State, and Responsivity) based upon how the NBO is currently administered and scored in clinical practice. The Autonomic subscale includes three items: tremors, startles, and skin color changes. Motor includes rooting, sucking, hand grasp, crawling reflexes, neck and shoulder muscle tone, extremity muscle tone, and a rating of the optimality of overall activity during the session. The third subscale, Organization of State, consists of five items: habituation to light, habituation to sound, crying, soothability, and overall state regulation throughout the session. Finally, the five-item Responsivity subscale includes the ability to: track a face, track a face plus a voice, track an inanimate object (i.e. red ball), locate a voice, and locate a rattle. For each item, possible scores range from one (lack of response) to five (robust response with little facilitation). A midpoint score of three indicates minimal response and increased need for support.

To maximize inter-rater reliability, two NBO administrators were trained extensively by NBO author, J. Kevin Nugent, PhD, and supervised by child psychologist and study principal investigator, Nicole Bush, PhD. All NBO sessions were video recorded, and in instances in which behavioral coding was unclear, both NBO administrators and Dr. Bush reviewed the video and discussed coding to agree on a final code. However, precise inter-rater reliability is not calculable as many items require direct scoring by the NBO administrator in person (e.g. sucking, reflexes, tone). Assessments occurred at a target timeframe of two to six weeks of age and no later than 12 weeks (adjusted for gestational age at birth). To increase participation across a broad population of mothers, participants were given the option of completing the infant assessments at the University of California, San Francisco pediatric research unit ($n = 115$) or in their homes ($n = 29$). For all participants, the NBO was completed prior to anthropometric assessments (details not presented here).

A mean score was created for each of the four NBO subscales. The 3-item Autonomic subscale was calculated with complete data ($n = 144$). Missing data for the Responsivity subscale was expected since the NBO specifies that infants must be in an “available state” (i.e. not fussing or crying) in order to assess their ability to track or locate animate and inanimate visual and auditory stimuli. The five-item Responsivity and seven-item Motor subscale means were created with all participants missing no more than two items ($n = 127$ and $n = 144$, respectively). There was considerable missing data for two items within the 5-item Organization of State subscale: habituation to light and habituation to sound. These two items require that the infant be asleep at some point during the encounter; these data were therefore missing for the majority of infants and were excluded from all subjects’ mean Organization of State scores, resulting in a three-item scale for which all infants had complete data ($n = 144$).

**IBQ—**The Infant Behavior Questionnaire-Revised (IBQ-R) Short[19] assesses infant temperament by parent-report. The IBQ-R has been well validated in infants ages three to 12 months, though in its original publication, the authors noted that its reliability, convergent validity, and stability have been demonstrated in infants as young as two weeks of age. The IBQ-R is comprised of 91 items that elicit the frequency of various behaviors reported on a seven-point Likert scale, which mothers completed in person during the visit. Validated measures of temperament in early infancy are rare, and although we did utilize one lesser-known instrument simultaneously in this study, we found that it did not demonstrate adequate scale reliability for research. Thus, to make the IBQ-R more developmentally appropriate for this age group, mothers completed a modified version that omitted irrelevant items (i.e. items related to smiling and laughing). The modified sub-selection of IBQ items and subscales showed very good reliability (Table 2) despite being validated on infants slightly older than those assessed here. Items were classified into two composite scales: Infant Negativity (derived from subscale items tapping sensitivity to unfamiliar caregivers, crying/fussiness, distress to limitations, and falling reactivity) and Infant Regulation (derived from items tapping low intensity pleasure and soothability).

**Cortisol Reactivity—**Salivary cortisol was measured at three time points: (A) baseline at the start of the visit after mothers were consented, and mother-infant pairs had acclimated to the visit context and staff, (B) following the NBO stimulus and anthropometric measurements of length, weight, and skin fold thickness, and (C) 15 minutes after sample B was collected. Trained study personnel collected saliva samples using Salimetrics (Carlsbad, California) SalivaBio Infant's Swab, which was placed in the infants' mouths for approximately 30 seconds. Swabs were temporarily stored at −20°C before transport to University of California, San Francisco Langley Porter Psychiatric Hospital for long-term storage at −80°C. At the University of Dresden, samples were thawed and centrifuged, then assayed using Cortisol Luminescence Immunoassay manufactured by IBL-Hamburg (Hamburg, Germany). The detection limit of the assay was 0.179 nmol/L. The mean inter- and intra-assay variations were 7.1 – 9.0% and 4.0 – 6.7% respectively. Samples with a cortisol concentration greater than 40 nmol/L at baseline ($n = 2$) or greater than 100 nmol/L at any time point ($n = 1$) were excluded due to biologic implausibility. Cortisol concentration values were natural log transformed prior to conducting analyses. Cortisol reactivity was then calculated twice: as the difference from time point A-to-B and as the difference from time point A-to-C. Correlation analysis with cortisol concentration and reactivity were adjusted for sampling time-of-day and time-elapsed between samples (i.e., time from A-to-B and A-to-C). These methods have been reported in detail elsewhere.[20]

## Analysis

Analyses were performed using IBM SPSS Statistics Version 24 after data were assessed for normal distribution and outliers. We analyzed demographic and NBO descriptive statistics and assessed correlations between demographics and the NBO, as well as intercorrelations among all study variables. Internal consistency (Cronbach's alpha) was determined for each of the NBO and IBQ subscales, and subscales with acceptable internal consistency (i.e. Cronbach's alpha > 0.7) were retained for subsequent analyses. NBO concurrent validity was assessed using Spearman's rank partial correlations between the NBO and IBQ subscales and between the NBO subscales and cortisol reactivity, controlling for corrected infant age, which was computed as the sum of gestational age at birth and postnatal age on the day of assessment. A priori analyses utilized the four-subscale structure originally published for the NBO, which was informed by theory.[12] We also conducted post hoc exploratory factor analysis to ascertain empirically-derived subscales, first retaining factors with eigen values greater than 1 and next examining the factor structure resulting from forcing a two, three, and four-factor solution. We used oblique rotation given our expectation that there would be correlation among factors due to their physiologic underpinnings.[21] Analyses with the theory-derived subscales were repeated with empirically-derived NBO factors to discern similarities across use of those two subscale approaches in an effort to guide future use with the NBO.

# RESULTS

## Sample Description

Table 1 displays the sample characteristics. Infants were born at an average gestational age of 39.6 weeks ($SD = 1.3$) and completed the assessment at an average postnatal age of 5.1 weeks ($SD = 1.9$). Infants born earlier than 38 weeks were not assessed until at least four weeks postnatal age, and infants were not assessed if mothers did not report that they were healthy. There were 73 females and 71 males. Ninety-one percent were ethnic or racial minorities. Approximately 30% of mothers had completed high school or less, 51% had some college or vocational training, and 19% had earned a college degree. Annual household income was $0-$86,000 (median = $20,000), with the majority falling below the United States poverty level for a family of four at the time of data collection ($22,550 in 2013).[22]

## NBO Descriptive Statistics

NBO scores were moderate-to-high for the Motor, Organization of State, and Responsivity subscales (Table 2). Scores for the Autonomic subscale showed little variability and a substantial ceiling effect with a mean of 4.6 out of a possible 5, which may be because such items are most useful with premature or very young neonates, rather than the healthy, approximately one-month-old infants who we assessed.

Internal consistency fell within the acceptable range for the Responsivity subscale ($\alpha = 0.76$). The three Organization of State items had excellent internal consistency ($\alpha = 0.91$). The Motor and Autonomic subscales had poor internal consistency ($\alpha$s = −0.26 and 0.14, respectively) and were therefore excluded from subsequent analyses. Table 3 displays the intercorrelations among the two NBO scores with sufficient reliability for comparison ($r = 0.66$, $p < 0.001$).

## Exploratory Factor Analysis

We examined the factorability of the 18 NBO items for which we had near complete data (excluding the two habituation items missing for most of the sample, as discussed in Methods). Ten of 18 items correlated at least 0.3 with at least one other item: all three Organization of State items, all five Responsivity items, one Motor item that is an overall impression of motor activity, and one Autonomic item measuring color changes. These 10 items all loaded on the first derived factor, with an eigen value of 5.6, explaining 31% of the variance. The subsequent five factors with eigen values greater than one only explained an additional 4%, 4%, 3%, 2%, and 2% of the variance, respectively, for a total of 47%. The factor analysis results did not significantly differ when forcing two, three, or four-factor solutions or after eliminating the items that were least correlated with other items. The two-factor solution resulted in the same 10 items loading in the first factor; the three- and four-factor solutions resulted in 9 of the 10 items (all except Responsivity to Voice) loading in the first factor. The two-, three-, and four-factor solutions explained 35%, 38%, and 41% of the variance. As no clearly superior factor structure resulted from exploratory factor analysis, primary analyses were conducted with the NBO subscales as structured in its original publication; secondarily, exploratory post-hoc analyses examined whether an NBO factor-score derived by averaging the scores from the 10 NBO items that loaded on the first factor produced similar associations.

### Concurrent Validity

Spearman rank partial correlation coefficients were used to examine associations between NBO subscale scores and infant temperament and cortisol reactivity, adjusted for corrected age (Table 3). Exploration of the need to additionally adjust associations with cortisol reactivity for basal cortisol levels showed that inclusion did not alter the magnitude or level of significance of associations with cortisol A-to-B and A-to-C, so basal cortisol was not retained in the partial correlation analyses. Infants with better Organization of State scores demonstrated lower cortisol reactivity ($r = -0.30$ and $r = -0.30$ for A-to-B and A-to-C reactivity, respectively, $p$s $< 0.01$). Higher Organization of State scores were also associated with lower maternal report of Temperamental Negativity on the IBQ ($r = -0.16$, $p < 0.05$), found in sub-analyses to be largely driven by the two items within the Negativity composite related to sensitivity to unfamiliar caregivers ($r = -0.29$, $p < 0.001$). Correlation analyses were also run with the primary factor derived through factor analyses. Results paralleled those found for the Organization of State subscale in direction of association and magnitude; the significance was slightly weaker for the association with cortisol ($r = -0.24$ and $r = -0.27$ for A-to-B and A-to-C reactivity, respectively, $p$s $< 0.01$) and Temperamental Negativity ($r = -0.15$, $p = 0.07$).

## DISCUSSION

The goal of this research was to explore the applicability of the NBO as a clinical research tool, given the existence of very few options for standardized measurement of neurobehavioral development in young infants. We studied the NBO’s reliability and concurrent validity, relating it to other well-validated measures of reactivity and regulation. The key finding was the ability to triangulate (1) an independently-coded behavioral observation with (2) a health-relevant biologic marker of stress reactivity, salivary cortisol,[16] as well as (3) maternal report of infant temperament, considered important to the parent-infant relationship[23–25] and to later developmental outcomes.[6]

By administering a brief behavioral challenge, we independently observed neonates' strengths and challenges. We found a small association between the NBO and maternal report of infant negativity, a key domain of temperament, suggesting that a brief, standardized clinical observation may be a useful and practical proxy for a mother's more extensive experience of her child's behavior in a naturalistic environment. Associations between the NBO Organization of State score and infant cortisol suggest that the coding captures not only observable changes in infants but also underlying biological reactivity and regulation. Such predictive utility has the potential for significant value in research and clinical domains, though additional research on the NBO's reliability and validity, particularly in clinically diverse populations, and likely further refinement of the NBO, are needed.

Analysis of the NBO's psychometric properties revealed some of its limitations in this study sample. Only two of the four subscales performed well enough to reliably use as predictor variables, and exploratory factor analysis did not uncover a superior factor structure in terms of concurrent associations with physiology or maternal perceptions of temperament. Associations between the empirically-derived primary factor-score and our outcomes of interest paralleled associations using the published, theoretically-derived Organization of State subscale. However, utilization of the Organization of State subscale is more parsimonious than the factor-score, aligns with the theoretical framework of the NBO developers, and led to slightly more robust associations with concurrent physiology and maternal perspective on behavior. The reliability of NBO subscales in our sample was not unexpected based upon previous literature showing inconsistent reliability of the NBAS subscales,[4] from which the NBO items and factor structure were drawn. While two of the NBO subscales had poor reliability, the items comprising those subscales should likely not be omitted from the assessment. The NBO scale that proved to be of most predictive value was the Organization of State subscale, which includes summative items that the NBO administrator codes based upon the overall impression of an infant's ability to organize. This rich observation would not be possible if the administrator had not observed the neonate through the challenge of sustaining attention and orientation to the entire range of stimuli presented. This challenge and the infant's visible physiologic and emotional reactivity to the challenge was what allowed for discrimination between infants. Moreover, the Motor and Autonomic subscales may have greater utility with younger or preterm neonates, who have been demonstrated to exhibit disproportionately less motor and autonomic organization at lower gestational ages relative to other domains of neonatal behavior and regulation.[26]

An important future direction of this research is to explore the NBO's psychometric properties among a sample of infants who are younger or born at earlier gestational ages relative to our sample, in particular to determine whether the motor and autonomic subscales demonstrate better reliability. Extension of the work presented here, within samples with a full range of socioeconomic status would also advance understanding of generalizability of the findings from this sample. If future studies similarly find poor reliability of the Motor and Autonomic subscales, refinement of these scales and/or items will be needed to determine if their inclusion is required to maintain the utility of the summative Organization of State subscale. One potential direction is to examine whether administration of a shorter subset of items (i.e. Responsivity) is adequate to reliably score the summative items.

Alternatively, items with poor reliability could be administered for the purpose of scoring summative items rather than rating individual items or subscales. Future work to refine the NBO will need to consider the overall objective of the assessment – for example, whether detection of motor deficits is an essential and appropriate component of the NBO toolkit, in the context of its evolving role in clinical and research settings.

An additional limitation of the present study is the lack of a gold standard to establish construct validity of the five-point NBO. Examination of associations between the NBO, NBAS, and NNNS would be useful in establishing the NBO as a shorter and more easily administered proxy or screener for these more extensive neurobehavioral assessments. The ultimate goal of this work will be the ongoing refinement toward a practical and reliable assessment that can be used in longitudinal research for the prediction of emotion regulation development. While the magnitude of the associations between the NBO, cortisol, and temperament were modest, these findings suggest that further research to this end is merited.

Aside from the primary objective being the refinement of the five-point NBO as valid and reliable longitudinal clinical research tool, an additional line of research could focus on the potential clinical role of the NBO in individual treatment planning. To this end, future studies could replicate this work in clinically diverse samples, which may present with more varied motor and autonomic system fragility that could be captured by the NBO and used for individualized care plans.

Demonstrating the value of longitudinal research, previous studies have investigated childhood developmental outcomes predicted by neonatal neurobehavioral assessments in substance exposed,[4,27] preterm,[5] and low birth weight infants.[4] A few studies have examined the predictive value of neonatal neurobehavioral assessment in healthy, term infants.[4,8,28,29] While these data are limited, they consistently show that the neonatal period constitutes an important early opportunity for the screening and identification of children who may have suboptimal developmental trajectories. In some longitudinal studies, neonates who performed better on neurobehavioral assessments later exhibited fewer behavioral problems,[29] more optimal mental and psychomotor development,[4,8,29] increased school readiness,[8] higher intelligence,[8] and better social communication.[30] Although such findings have not been consistently demonstrated across studies, and effect sizes have generally been small, this evidence suggests that the neonatal period may be a clinically useful time point to assess characteristics that have relevance to later childhood functioning. Further refinement of the NBO system and study of the specific long-term outcomes predicted by neonatal neurobehavioral development at various timepoints is needed. Such investigation is crucial for the most effective and efficient allocation of early intervention resources.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

The authors thank Marialma Gonzales, Julissa Cabrera, Katie Blackburn, Nancy Adler, Elissa Epel, Barbara Laraia, study staff, and study participants. This work was supported by the National Institutes of Health [NHLBI 5 R01

HL116511-02], the Albert Brodie Smith and Margaret Gretchen Smith Scholarship, and the Schoeneman Endowment Fund.

**Funding Source**: National Institutes of Health [NHLBI 5 R01 HL116511-02], the Albert Brodie Smith and Margaret Gretchen Smith Scholarship, and the Schoeneman Endowment Fund.

## REFERENCES

1. Halfon N, Larson K, Lu M, et al. Lifecourse Health Development: Past, Present and Future. Matern Child Health J. 2014;18(2):344–365. [PubMed: 23975451]
2. Britto PR, Lye SJ, Proulx K, et al. Nurturing care: promoting early childhood development. The Lancet. 2017;389(10064):91–102.
3. Council on Children With Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying Infants and Young Children With Developmental Disorders in the Medical Home: An Algorithm for Developmental Surveillance and Screening. Pediatrics. 118(1):405–420. [PubMed: 16818591]
4. Noble Y, Boyd R. Neonatal assessments for the preterm infant up to 4 months corrected age: a systematic review. Developmental Medicine & Child Neurology. 2012;54(2):129–139. [PubMed: 22142216]
5. Craciunoiu O, Holsti L. A Systematic Review of the Predictive Validity of Neurobehavioral Assessments During the Preterm Period. Physical & Occupational Therapy In Pediatrics. 2017;37(3):292–307. [PubMed: 27314272]
6. Bush NR, Jones-Mason K, Coccia M, et al. Effects of pre- and postnatal maternal stress on infant temperament and autonomic nervous system reactivity and regulation in a diverse, low-income population. Development and Psychopathology. 2017;29(5):1553–1571. [PubMed: 29162167]
7. Brazelton T. Berry, Nugent J. Kevin. Neonatal Behavioral Assessment Scale. 4th ed. Mac Keith Press; 2011.
8. El-Dib M, Massaro AN, Glass P, et al. Neurodevelopmental assessment of the newborn: An opportunity for prediction of outcome. Brain and Development. 2011;33(2):95–105. [PubMed: 20494536]
9. Costa R, Figueiredo B, Tendais I, et al. Brazelton Neonatal Behavioral Assessment Scale: A psychometric study in a Portuguese sample. Infant Behavior and Development. 2010;33(4):510–517. [PubMed: 20800286]
10. Lester BM, Als H, Brazelton TB. Regional Obstetric Anesthesia and Newborn Behavior: A Reanalysis toward Synergistic Effects. Child Development. 1982;53(3):687–692. [PubMed: 7094677]
11. Lester BM, Tronick EZ, Brazelton TB. The Neonatal Intensive Care Unit Network Neurobehavioral Scale procedures. Pediatrics. 2004;113(3 Pt 2):641–667. [PubMed: 14993524]
12. Nugent JK, Keefer C, Minear S, et al. Understanding Newborn Behavior and Early Relationships: The Newborn Behavioral Observations (NBO) System Handbook. Baltimore, Md: Brookes Publishing; 2007.
13. McManus BM, Nugent JK. A Neurobehavioral Intervention Incorporated into a State Early Intervention Program is Associated with Higher Perceived Quality of Care Among Parents of High-Risk Newborns. J Behav Health Serv Res. 2014;41(3):381–389. [PubMed: 22529036]
14. Nugent JK, Bartlett JD, Von Ende A, et al. The Effects of the Newborn Behavioral Observations (NBO) System on Sensitivity in Mother-Infant Interactions. Infants and Young Children. 2017;30(4):257–268.
15. McManus BM, Nugent JK. Feasibility study of early intervention provider confidence following a neurobehavioural intervention for high-risk newborns. Journal of Reproductive and Infant Psychology. 2011;29(4):395–403.
16. Slopen N, McLaughlin KA, Shonkoff JP. Interventions to Improve Cortisol Regulation in Children: A Systematic Review. Pediatrics. 2014;133(2):312–326. [PubMed: 24420810]
17. Stifter C, Dollar J. Temperament and Developmental Psychopathology. In: Developmental Psychopathology. D. Cicchetti; 2016:1–62.

18. Flegal KM, Kruszon-Moran D, Carroll MD, et al. Trends in Obesity Among Adults in the United States, 2005 to 2014. JAMA. 2016;315(21):2284–2291. [PubMed: 27272580]
19. Gartstein MA, Rothbart MK. Studying infant temperament via the Revised Infant Behavior Questionnaire. Infant Behavior and Development. 2003;26(1):64–86.
20. Jones-Mason KM, Coccia M, Grover S, et al. Basal and reactivity levels of cortisol in one-month-old infants born to overweight or obese mothers from an ethnically and racially diverse, low-income community sample. Psychoneuroendocrinology. 2018;88:115–120. [PubMed: 29223002]
21. Osborne J Best Practices in Quantitative Methods. 2455 Teller Road, Thousand Oaks California 91320 United States of America: SAGE Publications, Inc; 2008.
22. Health and Human Services Department. Annual Update of the HHS Poverty Guidelines.; 2013.
23. Braungart-Rieker JM, Hill-Soderlund AL, Karrass J. Fear and anger reactivity trajectories from 4 to 16 months: the roles of temperament, regulation, and maternal sensitivity. Dev Psychol. 2010;46(4):791–804. [PubMed: 20604602]
24. Stams G-JJM, Juffer F, van IJzendoorn MH. Maternal sensitivity, infant attachment, and temperament in early childhood predict adjustment in middle childhood: The case of adopted children and their biologically unrelated parents. Developmental Psychology. 2002;38(5):806–821. [PubMed: 12220057]
25. Stright AD, Gallagher KC, Kelley K. Infant Temperament Moderates Relations Between Maternal Parenting in Early Childhood and Children’s Adjustment in First Grade. Child Development. 2008;79(1):186–200. [PubMed: 18269517]
26. Mouradian L, Als H, Coster W. Neurobehavioral Functioning of Healthy Preterm Infants of Varying Gestational Ages. Journal of Developmental & Behavioral Pediatrics. 2000;21(6):408–416. [PubMed: 11132791]
27. Liu J, Bann C, Lester B, et al. Neonatal Neurobehavior Predicts Medical and Behavioral Outcome. Pediatrics. 2010;125(1):e90–e98. [PubMed: 19969621]
28. Bedford R, Pickles A, Sharp H, et al. Reduced Face Preference in Infancy: A Developmental Precursor to Callous-Unemotional Traits? Biological Psychiatry. 2015;78(2):144–150. [PubMed: 25526972]
29. Sucharew H, Khoury JC, Xu Y, et al. NICU Network Neurobehavioral Profiles Predict Developmental Outcomes in a Low Risk Sample. Paediatr Perinat Epidemiol. 2012;26(4):344–352. [PubMed: 22686386]
30. Bowers K, Khoury J, Sucharew H, et al. Early Infant Attention as a Predictor of Social and Communicative Behavior in Childhood. International Journal of Behavioral Development. 2018:0165025418797001.

**Table 1.**

Sociodemographic and Delivery Characteristics of SEED Study Participants (*N* = 144), California, 2010–2013

| | ***Mean (SD) or n (%)*** | ***Range*** |
|---|---|---|
| Birth weight (kg) | 3.4 (.4) | 2.35 – 4.76 |
| Gestational age at birth (wks)[a] | 39.6 (1.3) | 33.6 – 42.6 |
| Postnatal age at assessment (wks) | 5.1 (1.9) | 2.4 – 13.6 |
| Sex | | |
| Female | 73 (51%) | |
| Male | 71 (49%) | |
| Race/ethnicity | | |
| Black/African American | 43 (30%) | |
| Hispanic/Latinx | 42 (29%) | |
| Other/Multiple | 41 (28%) | |
| White | 13 (9%) | |
| Asian/Pacific Islander | 5 (4%) | |
| Mode of delivery | | |
| Vaginal | 107 (74%) | |
| Assessment location | | |
| Clinic | 115 (80%) | |
| Home | 29 (20%) | |
| Maternal Education | | |
| Less than high school | 13 (9%) | |
| High school graduate | 30 (21%) | |
| Some college/vocational | 73 (51%) | |
| College degree or higher | 28 (19%) | |
| Annual Household income | $24,425 (20,220) | $0–86,000 |

(a) Six (4%) delivered in the moderate to late preterm period (i.e. 33.5–37 weeks); the remainder were born term; SEED=Stress, Eating, and Early Development.

**Table 2.**

SEED Study Participants' Behavioral Observation Scores, Temperament Scores, and Salivary Cortisol (*N* = 144); California, 2010-2013

| | N | Mean | SD | Range | Cronbach's α |
|---|---|---|---|---|---|
| **NBO**[a] | 144 | | | | |
| Autonomic | 144 | 4.6 | 0.4 | 3.0 – 5.0 | 0.14 |
| Motor | 144 | 3.3 | 0.6 | 1.0 – 4.3 | −0.26 |
| Organization of State | 144 | 3.5 | 1.2 | 1.0 – 5.0 | 0.91 |
| Responsivity | 127 | 3.6 | 0.8 | 1.5 – 5.0 | 0.76 |
| **IBQ**[b] | 144 | | | | |
| Infant Negativity Composite | 144 | 3.5 | 0.7 | 1.6 – 5.5 | 0.79 |
| Sensitivity to Caregiver Changes | 140 | 2.1 | 1.4 | 1.0 – 7.0 | 0.82 |
| Distress to Limitations | 144 | 4.0 | 1.1 | 1.4 – 7.0 | 0.76 |
| Falling Reactivity | 144 | 4.5 | 1.2 | 1.5 – 7.0 | 0.78 |
| Crying/fussiness | 134 | 2.0 | 1.2 | 1.0 – 7.0 | N/A |
| Infant Regulation Composite | 144 | 5.5 | 0.6 | 3.5 – 6.9 | 0.72 |
| Low Intensity Pleasure | 144 | 5.9 | 0.8 | 3.7 – 7.0 | 0.69 |
| Soothability | 144 | 5.4 | 0.9 | 3.2 – 7.0 | 0.74 |
| **Cortisol Reactivity**[c] | 129 | | | | |
| A (Basal) | 129 | 1.9 | 0.9 | −1.6 – 3.7 | |
| A to B | 124 | 0.2 | 1.0 | −1.7 – 3.3 | |
| B to C | 115 | 0.5 | 1.3 | −2.1 – 3.7 | |

(a) Possible range 1–5

(b) Possible range 1–7

(c) After logarithmic transformation; SEED=Stress, Eating, and Early Development; NBO=Newborn Behavioral Observation; IBQ=Infant Behavior Questionnaire.

**Table 3.**

Spearman's Rank Correlation Coefficients between Infant Behavioral Observation Subscale Scores, Temperament, and Salivary Cortisol Age (*N* = 144)

| | | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|---|
| **NBO** | (1) Organization of State (N=144) | -- | | | | | |
| | (2) Responsivity (N=127) | **0.66** *** | -- | | | | |
| **IBQ**[a] | (3) Negativity Composite (N=144) | **−0.16** * | −0.04 | -- | | | |
| | (4) Regulation Composite (N=144) | 0.05 | 0.07 | **−.44** *** | -- | | |
| **Cortisol**[a,b] | (5) A (Basal) (N=129) | 0.15 | 0.09 | −0.05 | −0.10 | -- | |
| | (6) A-to-B (N=124) | **−0.30** ** | −0.09 | 0.01 | 0.03 | **−0.55** *** | -- |
| | (7) A-to-C (N=115) | **−0.30** ** | −0.16 | −0.01 | 0.09 | **−0.72** *** | **0.85** *** |

† $p < 0.10$

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$

[a] Partial correlations adjusted for corrected age at the time of assessment, computed as gestational age at birth plus postnatal age

[b] Correlations with Cortisol after adjusting for time-of-day and time elapsed between saliva collections (i.e., minutes from A-to-B or A-to-C).
NBO=Newborn Behavioral Observation; IBQ=Infant Behavior Questionnaire