





# BMJ Open PediaTrac V.3.0 protocol: a prospective, longitudinal study of the development and validation of a web-based tool to measure and track infant and toddler development from birth through 18 months

Renee Lajiness-O'Neill <sup>1,2</sup>, Seth Warschausky <sup>2</sup>, Alissa Huth-Bocks,<sup>3,4</sup> H Gerry Taylor,<sup>4,5</sup> Judith Brooks,<sup>6</sup> Angela Lukomski,<sup>7</sup> Trivellore Eachambadi Raghunathan,<sup>8</sup> Patricia Berglund,<sup>8</sup> Angela D Staples <sup>1</sup>, Laszlo Erdodi <sup>9</sup>, Stephen Schilling<sup>8</sup>

**To cite:** Lajiness-O'Neill R, Warschausky S, Huth-Bocks A, *et al.* PediaTrac V.3.0 protocol: a prospective, longitudinal study of the development and validation of a web-based tool to measure and track infant and toddler development from birth through 18 months. *BMJ Open* 2021;**11**:e050488. doi:10.1136/bmjopen-2021-050488

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-050488>).

posthumous co-author

Received 21 February 2021  
Accepted 18 November 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

#### Correspondence to

Dr Renee Lajiness-O'Neill;  
[rlajines@emich.edu](mailto:rlajines@emich.edu)

## ABSTRACT

**Introduction** The need for an efficient, low-cost, comprehensive measure to track infant/toddler development and treatment outcomes is critical, given the importance of early detection and monitoring. This manuscript describes the protocol for the development and testing of a novel measure, PediaTrac, that collects longitudinal, prospective, multidomain data from parents/caregivers to characterise infant/toddler developmental trajectories in term and preterm infants. PediaTrac, a web-based measure, has the potential to become the standard method for monitoring development and detecting risk in infancy and toddlerhood.

**Methods and analyses** Using a multisite, prospective design, primarcaregivers will complete *PediaTrac* V.3.0, a survey tool that queries core domains of early development, including feeding/eating/elimination, sleep, sensorimotor, social/sensory information processing, social/communication/cognition and early relational health. Information also will be obtained about demographic, medical and environmental factors and embedded response bias indices are being developed as part of the measure. Using an approach that systematically measures infant/toddler developmental domains during a schedule that corresponds to well-child visits (newborn, 2, 4, 6, 9, 12, 15, 18 months), we will assess 360 caregiver/term infant dyads and 240 caregiver/preterm infant dyads (gestational age <37 weeks). Parameter estimates of our items and latent traits (eg, sensorimotor) will be estimated by theta using item response theory-graded response modelling. Participants also will complete legacy (ie, established) measures of development and caregiver health and functioning, used to provide evidence for construct (discriminant) validity. Predictive validity will be evaluated by examining relationships between the PediaTrac domains and the legacy measures in the total sample and in a subsample of 100 participants who will undergo a neurodevelopmental assessment at 24 months of age.

## Strengths and limitations of this study

- PediaTrac is designed to be an efficient, low-cost, digital method to collect and track multidomain data on infant/toddler development, to enhance early identification of at-risk children and treatment outcomes.
- The PediaTrac method and survey tool have the potential to yield scales according to developmental domains and identify specific behaviours in infancy that could become common data elements for the study of specific neurodevelopmental conditions.
- Item response theory methods will be used to demonstrate PediaTrac's dimensionality and characterise unique developmental subgroups and trajectories.
- By involving caregivers in the digital reporting and child-centred monitoring of development, PediaTrac will improve access to care by decreasing the direct assessment burden on the clinician.
- The validity of parent report will be measured and monitored with the innovative inclusion of embedded response bias indices within PediaTrac.
- The PediaTrac instrument assesses infants/toddlers up to 18 months of age, and it will need to be extended beyond toddlerhood to capture additional critical phases of early development.
- Future projects will be necessary to adapt and validate PediaTrac into other languages.

**Ethics and dissemination** This investigation has single Institutional Review Board (IRB) multisite approval from the University of Michigan (IRB HUM00151584). The results will be presented at prominent conferences and published in peer-reviewed scientific journals.

## INTRODUCTION

The Eunice Kennedy Shriver National Institute of Child Health and Human Development 2012 Scientific Vision<sup>1</sup> stated that ‘within the next 10 years, scientists should be able to...fully understand the neurobiological bases, delineate the full developmental spectrum and trajectories, and identify the key biologic markers for five behavioural or cognitive disorders’. This laudable vision will not be fully actualised across the developmental spectrum until there is a: (1) consistent, systematic and universal measure of infant developmental behaviours and (2) multidimensional method of tracking early developmental trajectories. When applied systematically, such an approach to assessment would enable clinicians to better capture the full spectrum of infant behaviours, and developmental neuroscience researcher will be able to correlate these behaviours with molecular, cellular and brain-system level data. Thus, there is a critical need to develop an efficient, low-cost, yet comprehensive assessment tool to measure and track infant/toddler developmental status that can also be used in the study of behavioural correlates of brain structure and function.<sup>2</sup>

This type of assessment tool will significantly improve paediatric care by helping clinicians more fully evaluate developmental status, enhance early identification of risk and monitor intervention outcomes. Infancy is a critical stage of development, yet during this stage, clinicians have limited methods of tracking development. Developmental tracking of physical growth (head circumference, weight and height) has been a standard of paediatric care in the USA for the past five decades<sup>3</sup> and has provided insight into normal and atypical growth and developmental trajectories. In other developmental domains (eg, self-regulation and social functioning), something similar to a growth chart does not exist. Moreover, caregiver reports of infant abilities can be valid and reliable, although specific factors can affect the accuracy of caregivers’ perceptions of abilities.<sup>4</sup> Currently, clinicians primarily rely on cross-sectional, fragmented interview data to assess an infant’s developmental status in areas such as feeding, sensorimotor and communication skills. These data are difficult to synthesise in their present form. Clinicians do not always know when to consider a specific milestone delayed, given normal variability and lack of referenced percentiles.<sup>5</sup> There is an urgent need to identify risk of developmental deviations and neurodevelopmental disorders in the first year of life when prevention could alter a child’s trajectory, enhance quality of life and exponentially decrease lifetime costs of intervention.<sup>6</sup> There is a critical need to collect data through longitudinal methods of healthy and at-risk cohorts to better understand the ways in which disorders emerge early in life and identify factors contributing to these outcomes.<sup>1</sup>

Identifying risk in infancy could be achieved significantly earlier than is common in present-day practice. Only 19% of children with autism spectrum disorder (ASD), for example, are identified by age 3 years, with milder forms of ASD not identified until an average of

5.6 years.<sup>7</sup> According to the Centers for Disease Control and Prevention, fewer than half (42%) of the children with ASD undergo a developmental evaluation by 3 years of age.<sup>8</sup> Attention Deficit Hyperactivity Disorder, recently reported to affect up to 10% of USA school-age children,<sup>9</sup> is diagnosed at an average age of 7 years.<sup>10</sup> Developmental Language Disorders typically are undiagnosed until the third or fourth year of life<sup>11</sup> and affect 5%–8% of preschool children.<sup>12</sup> Similarly, while cerebral palsy is often detected within the first 2 years of life, milder forms may not be diagnosed until 4 or 5 years of age.<sup>13 14</sup> The economic burden of these disorders can be enormous; for example, the lifetime cost of support for an individual with ASD has recently been estimated at 2.4 million dollars.<sup>6</sup>

## Barriers to measuring neurodevelopment in infancy

Several clinical and research barriers obstruct progress towards achieving the goal of comprehensive measurement of developmental trajectories: (1) in-person standardised assessment instruments are resource intensive and there is a paucity of qualified professionals available to administer these tests.<sup>15 16</sup> In the USA, there are almost 1 million children living in areas where there is no local child physician. An additional 15 million children live in areas where there is an average of 22 paediatricians per 100 000 children.<sup>17</sup> Lack of access to qualified professionals hinders the ability to monitor early child development. (2) Methods for prospective (vs retrospective), systematic data gathering from caregivers are limited. Retrospective data gathering results in caregiver ‘telescoping’,<sup>18</sup> which refers to the reporting of developmental milestones earlier or later than they were actually achieved. This adversely affects the ability of practitioners and researchers to create an accurate picture of a child’s development and to identify when that child veers off track. Existing measures fail to concurrently assess such key facets of infant development as early relational health, social communication, sensorimotor development, feeding and sleep. Measures to assist clinicians exist, but have limitations. For example, the Child Health and Development Interactive System, a web-based data collection system that can be used by clinicians for distributing, scoring and graphing data using a repository of health questionnaires, may be cost prohibitive and does not plot developmental trajectories or predict risk while assessing multiple domains simultaneously.<sup>19</sup> The Patient-Reported Outcomes Measurement Information System (PROMIS)<sup>20</sup> lacks measures for infancy and toddlerhood. Ages and Stages uses risk cut-offs rather than standard scores, which precludes the ability to define developmental trajectories.<sup>21</sup> In addition, most of these tools fail to assess infant regulatory functions such as sleep or eating, or the content is embedded within other domains (eg, Ages and Stages); functions that are vital indicators of health.<sup>19–21</sup> None of the aforementioned tools assesses early relational health, one of the earliest markers of social emotional health. Formal standardised scores are

available for more in-depth *direct* assessment (eg, Bayley Scales of Infant Development, Mullen Scales of Early Learning, Developmental Assessment of Young Children) than screening questionnaires can provide.<sup>16 22</sup> Yet these measures are resource intensive and there are few qualified professionals available to administer them. In addition, while the National Institutes of Health (NIH) has established a standard for the use of common data elements in clinical research and patient registries, no data elements exist for infancy. (3) New methods are needed for analysis of multidimensional risks. Practitioners and researchers lack a system to collect and organise developmental and behavioural information in a way that measures developmental status in real time. The decentralisation of data limits both analyses and collaborative efforts to clarify the dimensions of heterogeneity across the population. The present protocol employs some of the methods used by the NIH PROMIS to improve overall quality of care and outcomes.<sup>23</sup> (4) Methods also are needed to systematically track early developmental trajectories, such as socio-emotional development.<sup>24</sup> Recent literature suggests that the ability to regulate emotions in infancy is a vital marker for early socioemotional development.<sup>24</sup> Clinician-administered measures such as Modified Checklist for Autism in Toddlers, Revised (M-CHAT-R) can be given when concerns are raised, but no methods have been developed to routinely track and plot development within domains or across domains over time.<sup>25</sup> Indeed, there has been a recent call to refine current cross-sectional screening methods (eg, M-CHAT-R) to include a broader range of skills to better identify patterns of deficits that emerge in early life among those subsequently diagnosed with ASD but escape detection using current screening algorithms due to substantial false-negative rates.<sup>26</sup>

## Objectives

To address these needs, our long-term goal is to develop PediaTrac. PediaTrac is a web-based measure designed to engage families in the gathering of longitudinal, prospective, multidomain data on infant/toddler development from birth to 18 months.<sup>27</sup> We anticipate that this instrument could become a standard for measuring developmental trajectories, monitoring development across domains and detecting precursors of neurodevelopmental disorders in the primary clinical care setting.

While other validated tools exist and some have online versions (eg, Ages and Stages), few validated tools exist to assess development in infancy, and none, to our knowledge, assess a range of functions in an integrated, efficient and longitudinal manner with applications for clinical care and research. The innovative methodology of PediaTrac, which has potential to become part of children's electronic medical record, will provide: (a) a measure that is designed to be a low cost, efficient and adjunctive method for tracking developmental status in order to identify the need for clinical assessment earlier in development, (b) prospective data collection by caregivers who would eliminate reliance on inaccurate retrospective

reporting, (c) multidomain (comprehensive) data that would help clarify sources of heterogeneity in development and assist in targeting developmental interventions, (d) a tool from which to obtain referenced percentiles or standardised scores as well as developmental trajectories that could constitute common data elements for studies of infants and toddlers and (e) a method for assessing the behavioural effects of perinatal disorders and early interventions and the behavioural correlates of early neural development. No methods have been developed to routinely track and plot development within or across domains over time.<sup>25</sup>

In addition, broader application of the PediaTrac approach could yield a public repository of developmental data that would be accessible to researchers and would assist caregivers and clinicians in using and interpreting results from PediaTrac in caring for infants and toddlers. The anticipated clinical value of PediaTrac is that it will facilitate prevention and early intervention by measuring developmental patterns within social environments, before the canalisation of dysfunctional development. By involving caregivers in the monitoring and digital reporting of their infants' development, PediaTrac will greatly improve access to care by decreasing (but not replacing) the direct assessment burden on the clinician while providing more comprehensive developmental monitoring. The validity of parent report will be measured and monitored with the innovative inclusion of embedded validity indices within PediaTrac.

For this research protocol, there are three primary aims. Aim 1 is to refine the item bank and scales of PediaTrac (now PediaTrac V.3.0) and evaluate psychometric properties at the item, scale and full instrument levels. This will include examining reliability, construct and predictive validity of parent/caregiver PediaTrac ratings of ~510 term and preterm infants who are being followed across repeated assessments from birth to 18 months of age. Construct validity includes assessment of the dimensionality of the tool as well as convergent and discriminant validity. Relationships of PediaTrac domain scores with legacy measures of infant/toddler development and behaviour and caregiver functioning are also examined. For aim 1, we hypothesise that: (1) estimates of reliability (precision) of the PediaTrac domain scales using item response theory (IRT)<sup>28</sup> test information curves (TIC) will be in the acceptable to good range ( $>0.70$ ), (2) IRT analysis will provide evidence for PediaTrac's multidimensionality and (3) moderate to strong associations ( $r>0.5$ ) will be evident between PediaTrac domains and legacy measures assessing similar constructs, supporting convergent validity, while weak associations ( $r<0.3$ ) will exist between PediaTrac domains and legacy measures assessing theoretically unrelated constructs (discriminant validity).

Aim 2 is to characterise unique developmental subgroups (eg, typical/atypical trajectories) using PediaTrac data from the parents/caregivers of 360 term (uncomplicated deliveries) and 240 preterm infants

(gestational age <37 weeks) using growth mixture modelling. Initially, we will examine group (term, preterm) differences at each developmental time point. We hypothesise that modelling will reveal at least two distinct growth trajectories that distinguish development over the first 18 months of life in children with term and preterm status (ie, typical and atypical growth trajectories).

Aim 3 is to examine the ability of the PediaTrac domains at each sampling period to individually and cumulatively predict overall development at 24 months in a subsample of 100 participants. We hypothesise that an algorithm based on items from multiple domains and time periods of PediaTrac will be a better predictor of functioning on performance-based and observational methods at 24 months than either a single domain of the PediaTrac tool or any of the existing legacy measures.

## METHODS

### Study design and overview

The PediaTrac study will use a prospective, longitudinal design with repeated measures in a proposed sample of 600 caregiver/infant dyads from three sites: 240 from site number 1, University of Michigan/Michigan Medicine (UM), 240 from site number 2, Case Western Reserve University/University Hospital (UH) and 120 from site number 3, Beaumont Pediatrics/Corner Health Center/Eastern Michigan University (EMU). Assuming 15% attrition, the number of recruited and enrolled families will allow us to reach a final sample size of 510, which would provide adequate power to support data analytic plans. American Psychological Association ethical guidelines will be followed and a Reliant Institutional Review Board (IRB) approval already has been obtained from the University of Michigan.

**Table 1** PediaTrac domains and number of items sampled at each time period per domain

Domain	Number of items per domain by time period								Number of unique items
	NB	2	4	6	9	12	15	18	NB-18
Demographic	32–37								32–37
Medical	77–90	35–41	34–40	35–41	35–41	37–43	37–43	41–45	104–115
Early relational health (previously called attachment)	34	34	41	42	51	55	55	51	65
Feeding/eating/elimination	23–43	22–42	24–54	26–56	12–36	36–45	36–45	36–45	56–87
Sensorimotor	19	19	36	38	48	44	33	26	68
Social/communication/cognition	26	27	46	61	71	73	65	48	105
Social/sensory information processing	25	25	25	25	25	25	25	25	25
Sleep	24	24	24	24	24	24	24	24	24
*Embedded response bias items	32	32	32	32	32	32	32	32	32
<b>Items sampled by time period</b>									
NB	292–330								
2 month		218–244							
4 month			262–298						
6 month				283–319					
9 month					298–328				
12 month						326–341			
15 month							307–322		
18 month								283–296	
<b>Total number of unique items</b>									511–558

\*Thirty-two repeated items to be used in the development of embedded response bias indices are also included in the number of questions asked at each time period.  
NB, newborn.

PediaTrac V.3.0 will be administered in blocks of ~220–340 items per sampling period, with a total of 511–558 unique items across the 18-month study period (table 1). Information describing the original item bank and domain development, expert panel reviews, cognitive interviews with parents and the pilot validation results of PediaTrac V.2.0 have been previously published, along with the items comprising the initial item bank.<sup>27</sup> Previous classical test theory and IRT analyses of PediaTrac V.2.0<sup>27</sup> suggested that the Sensorimotor, Social-Emotional/Communication (now called Social/Communication/Cognition), Feeding/Eating/Elimination and Attachment (now called Early Relational Health) domains for the PediaTrac instrument had the potential to produce reliable and valid estimates of infant development, although results indicated the need for a greater number and variety of items at each age group, a greater number of common items across adjoining assessments and a study sample that included infants with clear developmental risks (ie, preterm birth).

The current version (V.3.0) of PediaTrac extended the previous version by adding 15-month and 18-month assessments and including duplicate items across time points to ensure a sufficient sampling of the range of abilities across development and to allow for modelling of developmental trajectories. To provide for more precise estimates of the domains used in PediaTrac, binary choices were replaced by ordinal response options (ie, 5-point Likert scales) in the current version. The domains assessed in the current version have remained unchanged, though the names have been slightly modified for clarity (see table 2 for the domains assessed and descriptions of those domains). The PediaTrac V.2.0 item bank (newborn (NB) through 12-month items) has been substantially revised, with items eliminated that performed poorly statistically. To more effectively model the development of latent traits (eg, sensorimotor) at each developmental time point in the current study, PediaTrac items are duplicated across earlier and later time periods (ie, two prior consecutive younger ages and the next older age), except at the NB and 2-month period, where only items reflecting later development could be included. This is reflected in the shading of items in table 1. The purpose for this was to ensure that items were sufficiently sampling the

range of abilities/traits across development and to allow for a method of yoking consecutive sampling periods in the modelling of developmental trajectories. Due to branching logic, a range of items are possible at a single assessment period for some of the domains (eg, medical).

### Participants

The term infants (n=360) participating in this investigation will have a gestational age of  $\geq 37$  weeks at birth, minimum birth weight of 2500g and no history of prenatal or intrapartum complications, brain injury, neurological illnesses or disease or known genetic disorders. Preterm infants (n=240) will have a gestational age of  $< 37$  weeks and will be excluded from the study if they are diagnosed with neonatal abstinence syndrome or Down syndrome. Randomisation will be used to enrol infants in cases of multiple births. We will attempt to recruit an equal number of male and female infants into this study. Primary caregivers of term infants will complete study materials soon after birth or when infants born at gestational ages 37–38 weeks reach a postmenstrual age of 39 weeks. Caregivers of preterm infants will complete these materials at a postmenstrual age of 39 weeks. All subsequent data collection time points will be based on corrected age for preterm infants. Parents/caregivers will be required to be a minimum of 18 years old and to have access to a personal device such as a smartphone, tablet or computer. English-language competence will be required for participation given that PediaTrac is not yet available in other languages.

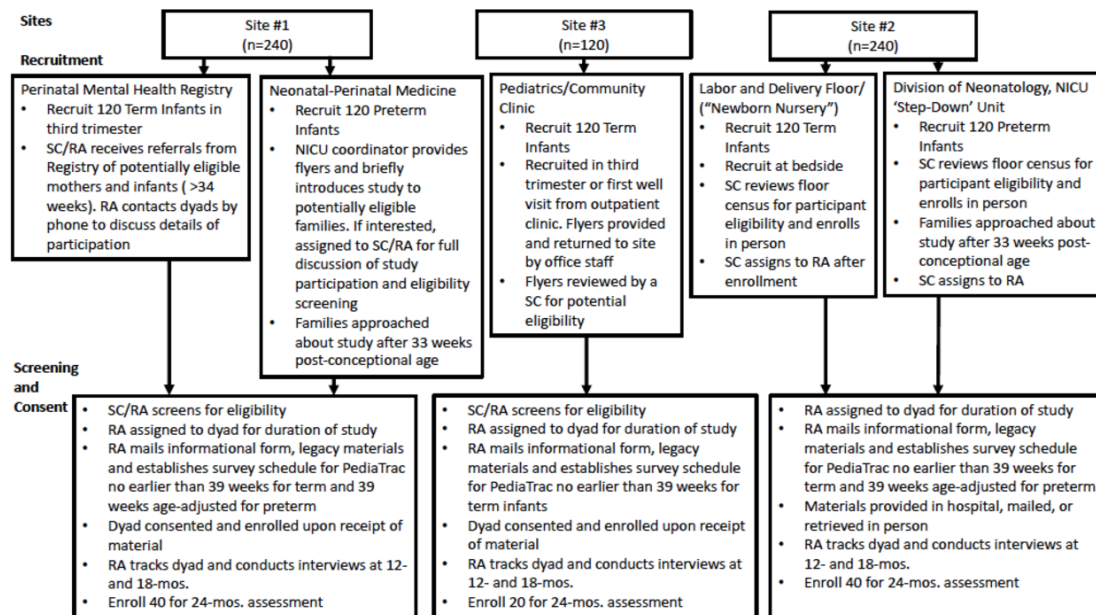
### Study procedures

#### Participant recruitment and informed consent

Women will be recruited in their last trimester of pregnancy, after their infants' birth in the hospital or at their first NB visit from three large metropolitan academic hospital systems and a community health centre in the Midwest of the USA. Participant recruitment, screening and consent processes for all caregiver–infant dyads are outlined in figure 1. An informational sheet will be provided to caregivers, outlining the study procedures and expectations. Caregivers will verbally consent to the study and they will be formally enrolled once they have completed and returned the NB study measures.

**Table 2** Domains assessed by PediaTrac and description of each domain

Domains	Descriptions
Early relational health	Perceptions of parenting, the infant/toddler's social-emotional status and attachment behaviours
Feeding/eating/elimination	Breast/formula feeding intake and behaviours, solid food eating and drinking patterns, nutrition and elimination
Sensorimotor	Head, trunk, limb movement and functional sensorimotor skills
Social/communication/cognition	Response to stimulation, expressing emotion, communication and acquiring knowledge
Social/sensory information processing	Identifies who calms the infant/toddler through touch and vocalisation, with whom the infant toddler makes eye contact
Sleep	Sleep pattern, sleep behaviours and sleep difficulties



**Figure 1** Description of PediaTrac recruitment and screening methods by site. NICU, neonatal intensive care unit; RA, research assistant; SC, study coordinator.

### Developmental assessments at 24 months

At enrolment, participants will be offered the opportunity for a follow-up developmental evaluation at 24 months of age to examine the ability of PediaTrac to predict objectively assessed outcomes. One hundred toddlers, randomly selected from those who expressed interest and are still in the study at the 18-month assessment, will be scheduled for this clinic-based evaluation. Assessments will be proportionate to the numbers initially enrolled (40 infants from site number 1 (UM), 40 infants from site number 2 (UH) and 20 infants from sites number 3 (EMU)).

### Data collection and tracking

Caregivers will complete subsets of the PediaTrac survey ranging from ~220 to 340 items, depending on the time period of the assessment. PediaTrac queries multiple developmental domains (Feeding/Eating/Elimination, Sleep, Sensorimotor, Social/Communication/Cognition, Early Relational Health (referred to as Attachment in prior versions of PediaTrac)) and Social/Sensory Information Processing at each of eight sampling periods (NB, 2, 4, 6, 9, 12, 15 and 18 months). Survey questions about demographics as well as family and perinatal medical characteristics will be completed during the NB period, with information on the family environment and infant medical status provided as part of all subsequent assessments.

At each assessment, caregivers also will complete paper-based legacy measures and at both 2 and 12 months, questionnaires assessing response bias. These materials will be organised in a study binder with instructions about how to complete the questionnaires at each time point. Caregivers will be provided with postage paid, preaddressed envelopes to return

the paper questionnaires to the respective study sites. Participants will be asked to return the completed questionnaires within 1 week of each assessment period but will be given 30 days to complete the study materials for that time period. Rigorous tracking procedures will be employed to minimise missing data and attrition, including: (1) scheduled reminders at 1 and 2 weeks prior to when the materials are due, (2) REDCap<sup>29</sup> automated reminder emails by 3 days past the opening of the time period, with outreach from the RA by phone, text or email at 5 days and as needed thereafter until all materials are received or day 30 is reached; (3) Zoom, Skype, Google Hangout, phone contact, text or email are used at 6, 12 and 18 months to further maintain family engagement, (4) letters, texting or home visits will be employed as needed and (5) communication with 'alternate contacts' (with participant permission) if necessary. To further encourage participation and minimise attrition, caregivers will receive scheduled participant payments of up to a total of US\$410 over the course of 2 years. As a benefit for participation in the study, a brief report of the neurodevelopmental test results will be provided to the 100 families undergoing assessment at 24 months.

### COVID-19 protocol modifications

The protocol was formulated prior to the COVID-19 pandemic. In response to the pandemic, the following modifications will be implemented primarily centred on addressing the participants' need for greater flexibility and recognition of barriers such as restrictions on in-person recruitment and disrupted postal services. The modifications include creating electronic surveys in the REDCap database to capture legacy measures, completing legacy measures over the phone, and

**Table 3** Response type and anchors used for core domains

Domain	Response type	Anchors
Early relational health	5-point Likert	1, Never; 2, Rarely; 3, Sometimes; 4, Often; 5, Always
	5-point Likert	1, Strongly disagree; 2, Disagree; 3, Neither; 4, Agree; 5, Strongly agree
Feeding/eating/elimination	5-point Likert	1, Never; 2, Rarely; 3, Sometimes; 4, Often; 5, Always
	Categorical	Example—What is your child’s primary source of liquid nutrition? Please select all that apply? 1, breast milk through nursing; 2, breast milk through bottle; 3, breast milk through tube feeding; 4, infant formula through bottle; 5, infant formula through tube feeding
Sensorimotor	5-point Likert	1, Never; 2, Rarely; 3, Sometimes; 4, Often; 5, Always
Social/communication/cognition	5-point Likert	1, Never; 2, Rarely; 3, Sometimes; 4, Often; 5, Always
Social/sensory information processing	5-point Likert	1, Never; 2, Rarely; 3, Sometimes; 4, Often; 5, Always
Sleep	5-point Likert	1, Never; 2, Rarely; 3, Sometimes; 4, Often; 5, Always
	Categorical	Example—In what position does your child sleep most of the time? 1, on his/her belly; 2, on his/her side; three on his/her back
Embedded response bias items	5-point Likert	1, Strongly disagree; 2, Disagree; 3, Neither; 4, Agree; 5, Strongly agree

allowing return of legacy instruments via secure email attachments or text message photo attachments.

### Study variables and measures

Study variables will include item-level digital responses to the PediaTrac questions in each developmental domain at the eight assessments. [Table 3](#) provides information on the response type and anchors used for the core domains apart from General Medical, which uses a variety of categorical and continuous response options. Study variables also will include digital responses to PediaTrac questions about demographics, environmental/family factors and medical variables evaluated at each assessment. Six different ‘Orders’ of PediaTrac were developed to minimise order effects in the administration and data collection of the PediaTrac surveys. Approximately, five to six embedded validity items are interspersed between the developmental domains for each order.

[Table 4](#) lists the legacy measures that will be used to evaluate the validity of PediaTrac, and the assessment periods in which those measures are completed. The parent-report measures chosen were based on those that had good psychometric properties and/or acceptance within the field, sampled constructs that were conceptually aligned with the PediaTrac domains, correlated with the domains of interest in our pilot investigation, and based on expert opinion. Measures of the parent–child relationship and the emotional well-being of the caregiver will be administered to examine associations of PediaTrac with familial and environmental factors. Two external criterion measures of response bias are also listed in [table 4](#) (ie, Inventory of Problems (IOP-29); Personality Assessment Inventory (PAI)<sup>28</sup> Validity Indices), which will be administered at only the 2 and 12 month assessments to validate indices of response bias embedded within PediaTrac. The justification for these embedded

indices is outlined below. The PAI Affective Anxiety (ANX-A) and Anxiety-Related Disorder, Traumatic Stress (ARD-T) subscales, also will be administered to further examine parental mental health and/or trauma history as predictors of child development. Measures will be administered at the ages for which they were developed (PediaTrac) or have been standardised (legacy measures). To reduce participant burden, some of the questionnaires will be administered at only a subset of the assessments. The 24-month developmental assessment includes the Bayley Scales of Infant and Toddler Development-4,<sup>15</sup> Adaptive Behavior Assessment System-3,<sup>30</sup> and Modified Checklist for Autism in Toddlers, Revised with Follow-Up (M-CHAT-R/F).<sup>25</sup> The time required to complete PediaTrac is 20–23 min, while the time to complete the legacy instruments ranges from 30 to 90 min depending on the time period. See online supplemental document A for a description of each measure.

### Embedded response bias indices

Examining potential biases in caregiver ratings is mandated by multiple professional organisations<sup>31–34</sup> and essential to clinical interpretation.<sup>35–37</sup> Reasons for biased self-report data range from suboptimal engagement to limited reading skills, defensive response style, symptom exaggeration, malingering and fatigue. To address a threat to internal validity that might occur from fatigue, such as an order effect, six separate PediaTrac surveys will be used for which the order of the domains is rotated, with each participant completing the same order throughout the study. To identify threats to internal validity, most comprehensive symptom inventories (eg, PAI; Minnesota Multiphasic Personality Inventory [MMPI]) include multiple empirically derived symptom validity scales. Several test items were developed for inclusion in PediaTrac to assess these threats and help ensure that caregiver



**Table 4** Developmental, behavioural and caregiver self-report legacy measures

Measures and sampling periods	Birth	2 months	4 months	6 months	9 months	12 months	15 months	18 months	24 months (follow-up n=100)
PediaTrac									
<b>Standardised paediatric developmental and behavioural questionnaires</b>									
Infant Behavior Questionnaire-Revised <sup>52</sup>									
Brief Infant Sleep Questionnaire <sup>53</sup>									
Brief Infant-Toddler Social-Emotional Assessment <sup>54</sup>									
Ages & Stages Questionnaire-3 <sup>55</sup>									
Communication and Symbolic Behavior Scales-Developmental Profile Infant-Toddler Checklist <sup>56</sup>									
M-CHAT-R/F <sup>25</sup>									
Bayley Scales of Infant and Toddler Developmental-4 <sup>15</sup>									
Adaptive Behavior Assessment System-3 <sup>30</sup>									
<b>Standardised caregiver/parenting and mental health questionnaires</b>									
Edinburgh Postnatal Depression Scale <sup>57</sup>									
Brief Symptom Inventory <sup>58</sup>									
Postpartum Bonding Questionnaire <sup>59</sup>									
Parenting Stress Index-4-Short Form <sup>60</sup>									
Maternal Self-Report Inventory-Short Form <sup>61</sup>									
<b>Caregiver response bias/validity</b>									
Inventory of Problems (IOP-29) and PAI Validity Scales (INF, PIM and NIM) <sup>38 62</sup>									
<b>Total time per sampling period</b>	60	60	90	120	120	120	120	120	120

INF, infrequency; NIM, negative impression ; PAI, Personality Assessment Inventory; PIM, positive impression .



ratings are relatively free from response bias and/or that response biases are accounted for in prediction models. Included items target three sources of distortion: random responding (RND<sub>PediaTrac</sub>), positive (PIM<sub>PediaTrac</sub>) and negative impression management (NIM<sub>PediaTrac</sub>). Random responding is operationalised as unusual endorsement of statements that have an obvious answer. The logic behind the RND<sub>PediaTrac</sub> scale is that all *bona fide* examinees who are literate, proficient in English and attend to item content should be able to choose the one correct option. The NIM<sub>PediaTrac</sub> scale consists of items that provide an evaluative statement of the infant in the negative direction (ie, indicating harsh judgement or an overly pessimistic outlook on the child's future) that were rarely endorsed in the normative sample. Conversely, the PIM<sub>PediaTrac</sub> scale consists of items that provide an evaluative statement of the infant in the positive direction (ie, indicating unrealistically positive opinion or an overly optimistic outlook on the child's future) that were rarely endorsed in the normative sample.

The validity of these three scales will be evaluated by examining their sensitivity to response bias as identified by four previously established measures. These include the IOP-29<sup>38</sup> and three validity scales embedded within the PAI, referred to as infrequency (INF<sub>PAI</sub>), positive impression management (PIM<sub>PAI</sub>) and negative impression management (NIM<sub>PAI</sub>).

## Data analyses

### IRT modelling of PediaTrac domains

IRT modelling will be applied to PediaTrac ratings to assess infant development. The IRT analysis will be performed using SAS (SAS V.9.4, SAS Institute, 2016). Because PediaTrac item response options are primarily ordinal (ie, ordered categorical responses), graded response models (GRM) will be used to model item parameters.<sup>39</sup> IRT models the probability of each item response category as a function of item parameters and the infants' latent abilities or traits in a given domain.<sup>28</sup> Key item and subject parameters in IRT models include: (1) item difficulty (location)= $b$ ; (2) item discrimination= $a$  (*slope*) and (3) subject trait score (developmental ability)= $\theta$ . In the simplest binary case,  $b$  is defined as the location on the latent trait, where the probability of endorsing an item is 50%. Items with higher  $b$  values are those endorsed at higher trait levels. Item discrimination,  $a$ , describes how well an item can discriminate or differentiate between examinees at different trait levels. Theta parameters will be estimated for each individual for each relevant PediaTrac domain (eg, Sensorimotor, Social/Communication/Cognition) at each assessment.

### Descriptive analyses

Descriptive statistics will be computed for all demographic and dependent variables, including IRT-based thetas for PediaTrac domains and raw or standard scores (where available) for legacy measures. Group differences stratified by sex, term status and site will be computed for

demographic and dependent variables using parametric and non-parametric tests. Missing data in the item variables will be handled through constructing the observed data log-likelihood, and the missing covariates will be handled through multiple imputation, depending on the analyses. For the legacy instruments, we will follow the established rules for identification and estimation of missing items. Covariates for aim 2 and 3 will include, at minimum, biological sex, term status, site, time since birth and other demographic variables that differ by group (eg, maternal education, maternal age). PediaTrac item characteristics will be examined using IRT methods.<sup>28 40</sup> These methods include category characteristic curves (CCC) (also referred to as item characteristic curves when item choices are binary). CCC express the likelihood of raters at different levels of the trait choosing a given item category (eg, from 1 to 5). Apart from the IRT analysis, all other analyses will be conducted using R (R Core Team, 2020)<sup>41</sup> or IBM SPSS Statistics, V.24.0.

## Primary analyses (by aim)

### Aim 1

The psychometric properties of the PediaTrac items will be examined using IRT analyses of reliability, model and item fit, and construct and discriminant validity. One of the major innovations of IRT is the extension of the concept of reliability, which in IRT is assessed by the test information function, indicating the degree of precision of a test across the range of a latent trait (theta,  $\theta$ ). Item and test information (I) functions will be reported as item information curves (IIC) and the TIC. IICs display the contribution of each item to the ability estimation at points along the ability range. Item information is typically highest in the region of the trait near the location of parameter  $b$ . The TIC displays the total information provided by the sum of all of the items along the ability continuum assessed by the domain. The reliability estimate is based on  $Rho = (\text{information} / (\text{information} + 1))$ .<sup>42</sup> The fit of the PediaTrac data to the GRM will be examined using item-level fit statistics.<sup>43</sup> Exploratory factor analysis will also be conducted on the items in each PediaTrac domain to further assess the assumption of unidimensionality.

Discriminant validity will be assessed by computing Pearson or Spearman Rho correlation coefficients to examine the relationships between PediaTrac domain scores (thetas) and the 13 developmental, behavioural and caregiver legacy measures. Controlling for age, we hypothesise that PediaTrac domains will be associated with legacy measures of infant development assessing related constructs. We also anticipate associations of more advantaged family environments with higher age-adjusted scores in PediaTrac domains.<sup>44</sup>

### Aim 2

To assess whether the constructs measured in aim 1 can discriminate between term and preterm infants, we will incorporate a regression relationship between the theta

parameters in the IRT model, used as the dependent variables, and a dummy variable indicating preterm status as the independent variable. We will extend this to a multiple regression model depending on the analysis. We will examine the beta coefficients from the regression models that predict theta to identify those that best predict group membership at each time period. To assess the temporal change in the measure of the construct and to assess term/preterm differences in the trends over time, we will modify the IRT model to include responses for items at all time points and formulate a longitudinal growth curve model for the theta parameter that will include time since birth, covariates (noted above) and interaction terms between time and some key covariates as predictors. For model estimation, we will use maximum likelihood or Bayesian methods.<sup>39 40</sup>

### Aim 3

The predictive validity of PediaTrac will be assessed by examining associations of domain scores at each sampling period and of changes in these scores across assessments to scores on the Bayley-4 at 24 months. We anticipate that PediaTrac surveys of multiple domains of infant/toddler development obtained at multiple time periods will better predict performance-based (Bayley-4) and caregiver-reported (Adaptive Behaviour Assessment System-3; ABAS-3 and M-CHAT) outcomes at 24 months than either a single domain of the PediaTrac tool or any of the established developmental, behavioural or caregiver survey measures. A series of multiple regression models will be computed to determine whether the PediaTrac domain scores and total score are better predictors of the Bayley-4, ABAS-3 or M-CHAT scores at 24 months than any of the existing developmental, behavioural or caregiver measures. Predictor variables for PediaTrac will be the feeding/eating, sleep, sensorimotor, social communication/cognition and attachment domain scores. Predictor variables for the legacy measures will be their respective summary score(s) (eg, Ages & Stages Questionnaire-3 (ASQ-3) Communication, Gross Motor, Fine Motor, etc). Criterion or outcome variables will be the Cognitive, Language, Motor, Social-Emotional and Adaptive Behavior scales of the Bayley-4, the General Adaptive Composite of the ABAS-3 and the M-CHAT total score. Descriptive statistics and regression diagnostics will be conducted prior to the main analysis. Results from analysis of variance (ANOVA) will be inspected to identify significant associations, and R and R<sup>2</sup> will provide information on the amount of variance accounted for in the outcome variables by predictor variables. In addition, logistic regression will be employed to determine if PediaTrac domain scores better predict criterion variables' class membership (low-risk vs high-risk groups) than legacy measures.

### Statistical power, sample size and attrition

While power refers to the probability of observing a significant result given a specified effect size, IRT models

**Table 5** Final possible sample sizes for various retention rates assuming recruitment of N=600

70% (n=420)	75% (n=450)	80% (n=480)	85% (n=510)
-------------	-------------	-------------	-------------

are not typically interpreted in terms of p values, and modelling is oriented towards prediction rather than the statistical significance of a given parameter estimate. With regards to analyses that address aims 1 and 2, the empirical literature has recommended a sample size of 500 for IRT GRMs for ordinal data.<sup>45 46</sup> For growth mixture modelling, Kim<sup>47</sup> conducted a Monte Carlo simulation that revealed a minimum of 300 participants was required to model data consistent with what is being proposed. Across several longitudinal studies with adults, retention rates vary from 45% to 88% (12% to 55% attrition);<sup>48</sup> however, we anticipate attrition at only ~15% given that caregivers are generally highly motivated in the care and raising of their infants, our rigorous follow-up and tracking procedures and retention rates in our prior longitudinal work. However, even if attrition ranges as high as 30%, with only 420 participants completing the study, the sample size will be sufficient to conduct the proposed analyses (see table 5). For aims 2 and 3, we estimated power using G\*Power with regression to test a model with a maximum of six predictors (eg, biological sex, term status, domain theta, maternal education, ASQ score) and one outcome. We have the ability to detect a medium to small effect ( $f^2=0.05$ ) with up to six predictors given a sample size of 461 (Aim 2) with 80% power and alpha=0.05. Based on our current attrition rate, we will have a sufficient sample to achieve this sample size. We have the ability to detect a medium effect ( $f^2=0.15$ ) with up to six predictors given a sample size of n=98 (aim 3) with 80% power and alpha=0.05, and smaller effect sizes with less predictors. Recruitment from multiple sites with demographically diverse populations will allow us to examine variations in findings across regions and socioeconomic strata.

### Patient and public involvement

While not directly involved in the current investigation, as noted above, the public was involved in the development of the original item bank. Based on modifications resulting from the Expert Panel Review, caregivers were asked to review the next version of the survey (PediaTrac V.1.1; 363 items) to identify items that were difficult to understand or could be misinterpreted. Specifically, Cognitive Interviews were completed with adult caregivers, all of whom had cared for an infant; fostered, biological or adopted. There were 10 caregivers (male=4) with an average level of education of 15.7 (range=12–20) years, ranging in age from 30 to 60 years. We sampled this age range recognising that foster families and grandparents could complete PediaTrac. Interviews were conducted with established cognitive interview methods, using a think-aloud technique with scripted probes. Caregivers were asked to comment on: (1) perceived understanding of each item, (2) whether they were able to

answer the item and (3) if they did not understand the item or were unable to answer it, why. Responses were recorded on a nominal scale (yes, no). The investigators used a summary of the quantitative and qualitative results obtained to provide feedback to the team to revise or remove item content.

## ETHICS

This investigation has single IRB multisite approval from the University of Michigan (IRB HUM00151584). The results of this investigation will be presented at prominent conferences and published in peer-reviewed scientific journals.

The most likely risks to participants include: (1) potential breach of confidential personal and health-care material given the digital nature of the study, (2) potential breach of confidentiality if suspected abuse or neglect needs to be reported to Child Protective Services by research personnel who are mandated reporters, (3) stress and fatigue that new caregivers may encounter participating in the study due to the time involved, (4) stress that caregivers may experience if they become aware or perceive that their child is not developing consistent with expectation (eg, based on test results); and (5) discomfort that the caregiver experiences while answering sensitive questions such as those about their own mental health or parenting stressors.

To address these concerns, each participant's data will be used only for research purposes and will be kept with strict confidence as allowed by law. Caregivers will be informed of the limitations of confidentiality and an adverse event/risk protocol is implemented in the event that abuse or neglect is discovered. Caregivers are informed that they have the option to discontinue the project at any time without penalty. To mitigate fatigue, there is no requirement that caregivers must complete instruments or PediaTrac at one sitting. They will be given a 30-day window to complete and return measures at all but the NB period. As noted, order effects that might occur due to fatigue will be managed at the group level by having six separate surveys for which the order of the domains is rotated. If a developmental delay is detected during the course of the study, the Site PI will send a letter and/or make phone contact with the caregiver to relay the concern and to recommend consultation with their healthcare provider. Caregivers who endorse a significant level of clinical symptoms on one or more of the self-report questionnaires will be contacted by study personnel with a mental health license to further assess distress and provide resources for local mental health services as warranted. If imminent risk is suspected, caregivers will be immediately contacted for further assessment of their safety, referral to mental health professionals or implementation of emergency interventions as appropriate.

## DISSEMINATION

Methods used to communicate our study findings will include: (1) publications in peer-reviewed journals and

scientific presentations at regional, national and international professional conferences, (2) media outreach such as newsletters/newspapers, radio, television, social media and PediaTrac website, (3) personal and professional contacts and (4) key stakeholders such as paediatric clinicians and subspecialties and publishers of assessment tools/methods.

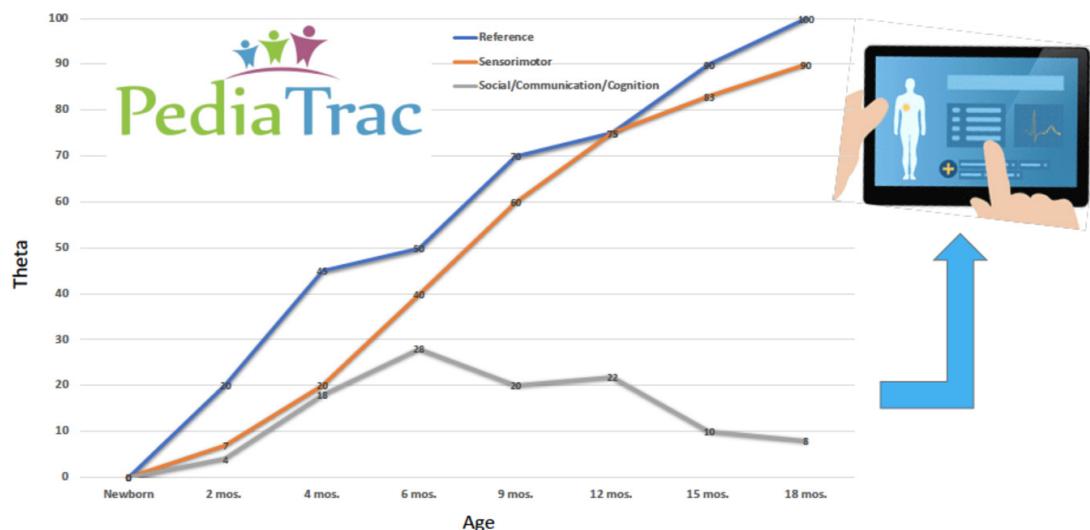
## Data sharing

Deidentified data from several legacy measures will be shared with the National Database for Autism Research (NDAR) data repository. The data will be deposited into NDAR no later than within 5 years of completion of the project.

## DISCUSSION

PediaTrac is designed to meet the critical need for an efficient, low cost, yet comprehensive assessment tool to track infant/toddler development prospectively from birth to 18 months across multiple domains. This manuscript describes the study protocol that will be used to validate PediaTrac V.3.0. The item bank will be revised from its prior version to both improve PediaTrac's psychometric properties and allow for better estimates of development. The previously published pilot investigation of PediaTrac V.2.0 that included expert panel reviews and cognitive interviews with parents revealed that the items and core domains of the PediaTrac instrument had the potential for producing meaningful estimates of infant development, although refinement of the item bank was warranted.<sup>27</sup> The pilot investigation revealed the need for more varied items that could sample the range of ability at each age as well as a method to yolk items across adjoining assessments to ensure that growth could be effectively modelled. Binary response options (eg, yes, no) were also deemed not suitable for modelling growth and developmental trajectories.<sup>39 40</sup> Finally, in order to ensure that the parameter estimates of our items and latent traits (eg, sensorimotor) being estimated by theta provided sufficient 'information', it was necessary to include a more representative sample of infants with a broader range of ability.

The item bank for the current version of PediaTrac (V.3.0), for which the reliability and validity will be evaluated by this investigation, includes some duplicate items at contiguous time periods to ensure a sufficient sampling of the range of abilities across development and to allow for modelling of developmental trajectories. To provide more precise estimates of the items and domains assessed by PediaTrac V.3.0 as well as improved modelling, binary choices will be replaced by ordinal response options (ie, 5-point Likert scales). The current version of PediaTrac will also include 15-month and 18-month items and assessments using a methodology previously reported.<sup>27</sup> In light of literature suggesting that cortical functions are substantially intercorrelated and less differentiated prior to 6 months of age,<sup>49 50</sup> these revisions were intended to



**Figure 2** Hypothetical graphical display of PediaTrac sensorimotor and social/communication/cognition trajectories over infancy/toddlerhood. Reference refers to population estimates that would be derived and continually updated based on national samples.

help clarify associations between developmental domains at different ages and to investigate potential changes in these associations across infancy and toddlerhood. Finally, for this protocol, both term and preterm infants will be included to ensure a more representative sample of the range of developmental abilities and to determine if the items or domains of PediaTrac can reliably predict developmental status.

At the completion of this project, our findings will allow us to: (1) validate and demonstrate the psychometric properties of PediaTrac V.3.0 with a person-centred approach using innovative IRT methodology and modelling, (2) characterise unique developmental subgroups (eg, typical/atypical trajectories) using longitudinal multidimensional IRT modelling in a sample of about 500 demographically diverse caregiver/infant dyads comprised of both term and preterm infants (see figure 2) and (3) examine the validity of PediaTrac in predicting development at 24 months of age.

### Study limitations

The study has several methodological limitations that will affect the generalisability of the findings. The exclusion criteria will preclude participation by the most disadvantaged portion of the population that does not have access to digital devices as well as those who are not fluent in English. It is anticipated that instrument refinement will result in a significantly reduced set of questions that could be presented as short forms in paper–pencil format, if necessary. Separate projects will be necessary to adapt, translate, back-translate and validate PediaTrac into other languages. The PediaTrac instrument assesses infant/toddlers up to 18 months of age, yet there are rapid critical phases of early development that extend well beyond that age. We anticipate the need for future projects to develop modules at a minimum for 24-month, 36-month and 48-month time points.

Designed as a longitudinal birth cohort study,<sup>51</sup> PediaTrac has the potential to capture more comprehensive infant/toddler data than current systems, complemented with a practical and efficient methodology similar to online diagnostic and healthcare management systems.<sup>19 20</sup> While not intended as a replacement for assessments by paediatric care teams, PediaTrac will advance knowledge about how developmental problems emerge during infancy and toddlerhood, the factors related to their emergence, and ways to provide for early identification and treatment of at-risk infants.

### Author affiliations

<sup>1</sup>Psychology, Eastern Michigan University, Ypsilanti, Michigan, USA

<sup>2</sup>Physical Medicine and Rehabilitation, University of Michigan Michigan Medicine, Ann Arbor, Michigan, USA

<sup>3</sup>Pediatrics, University Hospitals Cleveland Medical Center, Cleveland, Ohio, USA

<sup>4</sup>Department of Pediatrics, Case Western Reserve University and Rainbow Babies & Children's Hospital, Cleveland, Ohio, USA

<sup>5</sup>Department of Pediatrics, Nationwide Children's Hospital Research Institute and The Ohio State University, Columbus, Ohio, USA

<sup>6</sup>Dietetics and Human Nutrition, Eastern Michigan University, Ypsilanti, Michigan, USA

<sup>7</sup>Nursing, Eastern Michigan University, Ypsilanti, Michigan, USA

<sup>8</sup>Survey Research Center, Institute for Social Research, University of Michigan Institute for Social Research, Ann Arbor, Michigan, USA

<sup>9</sup>Psychology, University of Windsor, Windsor, Ontario, Canada

**Acknowledgements** The authors acknowledge the invaluable support of Study Coordinators, Jennifer Cano, Shannon Franz, and Lesa Dieter, as well as the tireless effort of Research Assistants, Casey Swick, Samantha Goldstein, Michelle Lobermeier, Amanda Hicks, Najae Bolden, Kirsten Oard, Shay Robinson, Yanisa Robbins, Jazmine Kirkland and Emily Gorjanc.

**Contributors** RL-O initiated and conceived the study and in collaboration with JB and AL developed earlier versions of the research protocol. SW, AH-B and HGT in collaboration with RL-O, JB and AL assisted in critically revising the protocol. RL-O, SW, AH-B, HGT and AL, assisted in the development of the recruitment procedures. RL-O, SW, AH-B, JB, AL, ADS and LE assisted in the development and selection of study measurement tools. RL-O, AS, TR, PB and SS assisted in the development of the data analytic methods. All authors approved the final version of this protocol.

**Funding** Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number R01HD095957.

**Disclaimer** The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Competing interests** None declared.

**Patient consent for publication** Not applicable.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Renee Lajiness-O'Neill <http://orcid.org/0000-0002-2448-6215>

Seth Warschausky <http://orcid.org/0000-0003-2850-2144>

Angela D Staples <http://orcid.org/0000-0002-7678-5794>

Laszlo Erdodi <http://orcid.org/0000-0003-0575-9991>

#### REFERENCES

- Guttmacher A. *NICHD 'S scientific vision : the NICHD mission is to ensure...*, 2012.
- NICHD Child Development and Behavior Branch (CDBB). Eunice Kennedy Shriver National Institute of child health and human development. Available: <https://www.nichd.nih.gov/about/org/der/branches/cdbb/Pages/overview.aspx>
- Centers for Disease Control and Prevention (CDC.gov). National center for health statistics. Available: <https://www.cdc.gov/growthcharts/>
- Visser SN, Danielson ML, Bitsko RH, *et al.* Trends in the parent-report of health care provider-diagnosed and medicated attention-deficit/hyperactivity disorder: United States, 2003-2011. *J Am Acad Child Adolesc Psychiatry* 2014;53:34-46.
- Dosman CF, Andrews D, Goulden KJ. Evidence-Based milestone ages as a framework for developmental surveillance. *Paediatr Child Health* 2012;17:561-8.
- Buescher AVS, Cidav Z, Knapp M, *et al.* Costs of autism spectrum disorders in the United Kingdom and the United States. *JAMA Pediatr* 2014;168:721-8.
- U.S. Department of Health and Human Services C for DC and P. Prevalence of autism spectrum disorder among children aged 8 years — autism and developmental disabilities monitoring network, 11 sites, United States, 2014 2018;67.
- Centers for Disease Control and Prevention (CDC). Autism spectrum disorder, 2021. Available: <https://www.cdc.gov/ncbddd/autism/facts.html>
- Kessler RC, Berglund P, Demler O, *et al.* Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National comorbidity survey replication. *Arch Gen Psychiatry* 2005;62:593. doi:10.1001/archpsyc.62.6.593
- Kessler RC, Chiu WT, Demler O, *et al.* Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National comorbidity survey replication. *Arch Gen Psychiatry* 2005;62:617-27.
- Rapin I, Dunn M. Update on the language disorders of individuals on the autistic spectrum. *Brain Dev* 2003;25:166-72.
- National healthcare quality report, 2006. Available: <http://archive.ahrq.gov/qual/nhdr06/nhdr06report.pdf>
- Centers for Disease Control and Prevention (CDC.gov). Cerebral palsy. Available: <https://www.cdc.gov/ncbddd/cp/index.html>
- Ashwal S, Russman BS, Blasco PA, *et al.* Practice parameter: diagnostic assessment of the child with cerebral palsy: report of the Quality Standards Subcommittee of the American Academy of Neurology and the practice Committee of the child Neurology Society. *Neurology* 2004;62:851-63.
- Bayley N, Aylward GP. *Bayley scales of infant and toddler development*. 4 edn. Bloomington, MN: NCS Pearson, Inc, 2019.
- Mullen EM. *Mullen scales of early learning*, 1995.
- Shipman SA, Lan J, Chang C-H, *et al.* Geographic maldistribution of primary care for children. *Pediatrics* 2011;127:19-27.
- Hus V, Taylor A, Lord C. Telescoping of caregiver report on the Autism Diagnostic Interview--Revised. *J Child Psychol Psychiatry* 2011;52:753-60.
- Child Health & Development Interactive System (CHADIS). Available: <http://www.chadis.com/>
- Cella D, Yount S, Rothrock N, *et al.* The patient-reported outcomes measurement information system (PROMIS). *Med Care* 2007;45:S3-11. doi:10.1097/01.mlr.0000258615.42478.55
- Squires J, Bricker D, Twombly E. *Ages & stages questionnaires: A parent-completed child monitoring system-3*, 2009.
- Bayley N. *Bayley scales of infant development*. 3 edn, 2006.
- Baumhauer JF. Patient-Reported Outcomes - Are They Living Up to Their Potential? *N Engl J Med* 2017;377:6-9.
- Planalp EM, Braungart-Rieker JM. Trajectories of regulatory behaviors in early infancy: determinants of infant self-distraction and self-comforting. *Infancy* 2015;20:129-59.
- Robins DL, Casagrande K, Barton M, *et al.* Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics* 2014;133:37-45.
- Øien RA, Schjølberg S, Volkmar FR, *et al.* Clinical features of children with autism who passed 18-month screening. *Pediatrics* 2018;141. doi:10.1542/peds.2017-3596. [Epub ahead of print: 21 May 2018].
- Lajiness-O'Neill R, Brooks J, Lukomski A, *et al.* Development and validation of PediaTrac™: a web-based tool to track developing infants. *Infant Behav Dev* 2018;50:224-237.
- Hambleton RK, Swaminathan H, Jane Rogers H. *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications, Inc, 1991.
- Harris PA, Taylor R, Minor BL, *et al.* The REDCap Consortium: building an international community of software platform partners. *J Biomed Inform* 2019;95:103208.
- Harrison P, Oakland T. *Adaptive behavior assessment system*. 2 edn. San Antonio, TX: The Psychological Corporation, 2003.
- Bush SS, Ruff RM, Tröster AI, *et al.* Symptom validity assessment: practice issues and medical necessity NAN policy & planning committee. *Arch Clin Neuropsychol* 2005;20:419-26.
- Bush SS, Heilbronner RL, Ruff RM. Psychological assessment of symptom and performance validity, response bias, and malingering: official position of the association for scientific advancement in psychological injury and law. *Psychol Inj Law* 2014;73:197-205.
- Chafetz MD, Williams MA, Ben-Porath YS, *et al.* Official position of the American Academy of clinical neuropsychology social security administration policy on validity testing: guidance and recommendations for change. *Clin Neuropsychol* 2015;29:723-40.
- Heilbronner RL, Sweet JJ, Morgan JE, *et al.* American Academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *Clin Neuropsychol* 2009;23:1093-129.
- Dandachi-FitzGerald B, Ponds RWHM, Merten T. Symptom validity and neuropsychological assessment: a survey of practices and beliefs of neuropsychologists in six European countries. *Arch Clin Neuropsychol* 2013;28:771-873.
- Martin PK, Schroeder RW, Odland AP. [Formula: see text] Neuropsychologists' Validity Testing Beliefs and Practices: A Survey of North American Professionals. *Clin Neuropsychol* 2015;29:741-6.
- Merten T, Dandachi-FitzGerald B, Hall V, *et al.* Symptom validity assessment in European countries: development and state of the art. *Clínica y Salud* 2013;24:129-38.
- Viglione DJ, Giromini L, Landis P. The Development of the Inventory of Problems - 29 : A Brief Self-Administered Measure for Discriminating Bona Fide From Feigned Psychiatric and Cognitive Complaints for Discriminating Bona Fide From Feigned Psychiatric and Cognitive Complaints. *J Pers Assess* 2017;99:534-44.
- Samejima F. Graded Response Model. In: van der Linden WJ, Hambleton RK, eds. *Handbook of modern item response theory*. New York, N.Y.: Springer, 1997.
- de ARJ. The theory and practice of item response theory. *Ref Res B News* 2009;24.
- Team RC. A language and environment for statistical computing. R foundation for statistical computing, 2020. Available: <https://www.r-project.org/>



- 42 Nguyen TH, Han H-R, Kim MT, *et al*. An introduction to item response theory for patient-reported outcome measurement. *Patient* 2014;7:23–35.
- 43 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 1995;57:289–300.
- 44 Bradley RH, Corwyn RF. Socioeconomic status and child development. *Annu Rev Psychol* 2002;53:371–99.
- 45 Reeve BB, Fayers P. Applying item response theory modeling for evaluating questionnaire item and scale properties. In: Fayers P, Hays RD, eds. *Assessing quality of life in clinical trials: methods of practice*. 2 edn. New York: Oxford University Press, 2005: 55–73.
- 46 Embretson SE, Reise SP. *Item response theory for psychologists*. Mahwah, NJ: Erlbaum, 2000.
- 47 Kim S-Y. Sample size requirements in single- and multiphase growth mixture models: a Monte Carlo simulation study. *Struct Equ Model A Multidiscip J* 2012;19:457–76.
- 48 Hanna KM, Scott LL, Schmidt KK. Retention strategies in longitudinal studies with emerging adults. *Clin Nurse* 2014;28:41–5.
- 49 Brown TT, Lugar HM, Coalson RS, *et al*. Developmental changes in human cerebral functional organization for word generation. *Cereb Cortex* 2005;15:275–90.
- 50 Luna B, Thulborn KR, Munoz DP, *et al*. Maturation of widely distributed brain function Suberves cognitive development. *Neuroimage* 2001;13:786–93.
- 51 Landrigan PJ, Trasande L, Thorpe LE, *et al*. The National children's study: a 21-year prospective study of 100,000 American children. *Pediatrics* 2006;118:2173–86.
- 52 a GM, Rothbart MK. Studying infant temperament via the revised infant behavior questionnaire. *Infant Behav Dev* 2003;26:64–86.
- 53 Sadeh A. A brief screening questionnaire for infant sleep problems: validation and findings for an Internet sample. *Pediatrics* 2004;113:e570–7.
- 54 Briggs-Gowan MJ, Carter AS, Irwin JR, *et al*. The brief Infant-Toddler social and emotional assessment: screening for social-emotional problems and delays in competence. *J Pediatr Psychol* 2004;29:143–55.
- 55 Reynolds C, Carey W, Mcdevitt S, *et al*. *Overview of the ASQ : SE, 2000*.
- 56 Wetherby A, Prizant B. *Communication and symbolic behavior scales developmental profile Infant/Toddler checklist, 2001*.
- 57 Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. development of the 10-item Edinburgh postnatal depression scale. *Br J Psychiatry* 1987;150:782–6.
- 58 Derogatis LR, Melisaratos N. The brief symptom inventory: an introductory report. *Psychol Med* 1983;13:595–605.
- 59 Brockington IF, Fraser C, Wilson D. The postpartum bonding questionnaire: a validation. *Arch Womens Ment Health* 2006;9:233–42.
- 60 Haskett ME, Ahern LS, Ward CS, *et al*. Factor structure and validity of the parenting stress index-short form. *J Clin Child Adolesc Psychol* 2006;35:302–12.
- 61 Shea E, Tronick E. The maternal self-report inventory. In: Fitzgerald H, ed. *Theory and research in behavioral pediatrics*. New York: Plenum Press, 1988.
- 62 Morey LC, Inventory PA, Manual P. *Odessa, FL: Psychological Assessment Resources, 1991*.