Phenoscreening: a developmental approach to research domain criteria-motivated sampling

Colleen M. Doyle
University of Minnesota

Carolyn Lasch
University of Minnesota

Elayne P. Vollman
University of Chicago

Christopher D. Desjardins
University of Minnesota

Nathaniel E. Helwig
University of Minnesota

Suma Jacob
University of Minnesota

Jason J. Wolff
University of Minnesota

Jed T. Elison
University of Minnesota

October 2020

Division of the Social Sciences, Committee on Education, University of Chicago.
The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B140048

The Journal of Child Psychology and Psychiatry

Journal of Child Psychology and Psychiatry 62:7 (2021), pp 884–894 doi:10.1111/jcpp.13341

Phenoscreening: a developmental approach to research domain criteria-motivated sampling

Colleen M. Doyle,[1] Carolyn Lasch,[1] Elayne P. Vollman,[2] Christopher D. Desjardins,[1] Nathaniel E. Helwig,[3,4] Suma Jacob,[5,6] Jason J. Wolff,[7] and Jed T. Elison[1,6]

[1]Institute of Child Development, University of Minnesota, Minneapolis, MN,USA; [2]Department of Comparative Human Development, University of Chicago, Chicago, IL, USA; [3]Department of Psychology, University of Minnesota, Minneapolis, MN, USA; [4]Department of Statistics, University of Minnesota, Minneapolis, MN, USA; [5]Department of Psychiatry, University of Minnesota, Minneapolis, MN, USA; [6]Department of Pediatrics, University of Minnesota, Minneapolis, MN, USA; [7]Department of Educational Psychology, University of Minnesota, Minneapolis, MN, USA

Background: To advance early identification efforts, we must detect and characterize neurodevelopmental sequelae of risk among population-based samples early in development. However, variability across the typical-to-atypical continuum and heterogeneity within and across early emerging psychiatric/neurodevelopmental disorders represent fundamental challenges to overcome. Identifying multidimensionally determined profiles of risk, agnostic to DSM categories, via data-driven computational approaches represents an avenue to improve early identification of risk. **Methods:** Factor mixture modeling (FMM) was used to identify subgroups and characterize phenotypic risk profiles, derived from multiple parent-report measures of typical and atypical behaviors common to autism spectrum disorder, in a community-based sample of 17- to 25-month-old toddlers (n = 1,570). To examine the utility of risk profile classification, a subsample of toddlers (n = 107) was assessed on a distal, independent outcome examining internalizing, externalizing, and dysregulation at approximately 30 months. **Results:** FMM results identified five asymmetrically sized subgroups. The putative high- and moderate-risk groups comprised 6% of the sample. Follow-up analyses corroborated the utility of the risk profile classification; the high-, moderate-, and low-risk groups were differentially stratified (i.e., HR > moderate-risk > LR) on outcome measures and comparison of high- and low-risk groups revealed large effect sizes for internalizing (d = 0.83), externalizing (d = 1.39), and dysregulation (d = 1.19). **Conclusions:** This data-driven approach yielded five subgroups of toddlers, the utility of which was corroborated by later outcomes. Data-driven approaches, leveraging multiple developmentally appropriate dimensional RDoC constructs, hold promise for future efforts aimed toward early identification of at-risk-phenotypes for a variety of early emerging neurodevelopmental disorders. **Keywords:** Development; infancy; social behavior; communication; autism spectrum disorder.

Introduction

Chronic mental health cases account for a third of morbidity and mortality globally (Vigo, Thornicroft, & Atun, 2016). In the United States, mental illness is the most common childhood health issue, affecting 20%–25% of all school-aged youth (Kessler et al., 2005; Merikangas et al., 2010). Similar estimates of the prevalence of DSM disorders (19.5%) have been observed and reported in preschool-aged children (Egger & Angold, 2006). Notably, it is estimated that once impairing symptoms are consolidated in adulthood, <30% of the burden of chronic mental illness can be averted, even with optimal care and access to services (Andrews, Issakidis, Sanderson, Corry, & Lapsley, 2004). These findings highlight the urgency of initiating a focused effort on early identification and early intervention, and if possible, strategic prevention (Insel & Scholnick, 2006). However, identifying incipient risk markers in early childhood of later emerging psychiatric sequelae requires novel methodological approaches.

Operating within a traditional DSM-defined nosological framework, several recent longitudinal studies of samples selected based on familial risk illustrate the potential of presymptomatic prediction of a later emerging diagnostic profile (Hafeman et al., 2017; Hazlett et al., 2017; Rice et al., 2017). These studies forecast a future of prevention trials for high-familial-risk children. Extending these types of results to samples not selected on familial risk remains a fundamental challenge. Yet, promising examples include the following: (a) selecting samples based on prodromal profiles as determined by structured clinical interviews and direct behavioral assessments (e.g., Cannon et al., 2016) and (b) identifying risk as determined by exceeding a specified threshold on targeted parent-report questionnaires (e.g., Wakschlag et al., 2015), a strategy amendable to assessment of large samples sizes. Indeed, large samples are required to address the phenotypic heterogeneity observed within and across DSM-defined disorders, as well as the variability observed around diagnostic thresholds, referred to here as the typical-to-atypical continuum (Cicchetti, 1993).

The Research Domain Criteria (RDoC) initiative (Cuthbert & Insel, 2013) has formalized research strategies for parsing the heterogeneity/variability inherent to the etiology, phenotypic presentation,

Conflict of interest statement: No conflicts declared.

and treatment response of major psychiatric disorders. Because DSM categories are not 'natural kinds', greater measurement precision is required to identify more homogenous subgroups. Examples of data-driven subtyping of DSM-defined disorders are rapidly accumulating (Bebko et al., 2014; Georgiades et al., 2013; Karalunas et al., 2014), with a large corpus of evidence emerging from the B-SNIP study (e.g., Clementz et al., 2016). Yet, RDoC-motivated studies based on samples defined by DSM categories may refine classification or treatment strategies within a category, but cannot inform the biological validity of the category itself (Hyman, 2010). Van Dam et al. (2017) implemented data-driven computational strategies, leveraging a large number of clinical dimensions measured in a community sample (n = 347) with minimal exclusionary criteria, to reveal clinically meaningful subgroups that cross the typical-to-atypical or adaptive-to-maladaptive continuum. This type of approach, if applied to early development with a long-term longitudinal design, could reveal distinct risk profiles that could be addressed with targeted preventive interventions. However, despite conceptual commentary (e.g., Casey, Oliveri, & Insel, 2014; Franklin, Jamieson, Glenn, & Nock, 2015; Mittal & Wakschlag, 2017; Sonuga-Barke, 2014), empirical reports that infuse developmental considerations into an RDoC framework are sparse (but see Fair et al., 2012; Wakschlag et al., 2015).

To advance strategic prevention efforts beyond high-familial-risk samples, we propose a version of RDoC-motivated sampling we have termed 'phenoscreening'. Traditional single-threshold criterion screening approaches yield binary outcomes. The primary objective of the phenoscreening approach is to derive multiple 'at-risk' phenotypes by leveraging multiple dimensional constructs selected from the Research Domain Criteria Matrix (Bernard & Mittal, 2015; Garvey, Avenevoli, & Anderson, 2016). This approach represents a synthesis of previous sampling strategies using data collected on multiple dimensional constructs from a large sample and unsupervised data-driven computational strategies (Totah et al., 2016) to derive subgroups defined by homogenous risk profiles based on constellations of complex behavior. The current study functions as a proof of concept for the phenoscreening approach. The constructs we selected and the age range we targeted are most relevant to the psychiatric risk profile for autism spectrum disorders (ASD). However, the general phenoscreening approach is translatable to any complex, multidimensional, heterogeneous psychiatric profile, or risk profiles that may crosscut diagnostic categories.

In the current study, we focus on the characterization of multiple constructs dimensionally distributed in the general population (Ronald, Happé, Price, Baron-Cohen, & Plomin, 2006) related to language, social communication, and repetitive behaviors (including rituals and routines or 'just right' behaviors that may be related to various early emerging anxiety disorders; see Dar, Kahn, & Carmeli, 2012; Evans et al., 1997; Lundström et al., 2011; Pine, Guyer, Goldwin, Towbin, & Leibenluft, 2008). These align with the RDoC constructs of Language (Cognitive Systems domain), Social Communication (Social Processes domain; Production of Facial Communication subconstruct, Production of Non-facial Communication subconstruct), Habit – Sensorimotor (Sensorimotor Systems domain) and Reward Learning (Positive Valence Systems (PVS) domain, Habit – PVS subconstruct). Our selection of these dimensions was informed by growing evidence for their transdiagnostic relevance to early emerging clinical presentations of psychopathology and their relevance to developmental risk profiles that are subclinical or beyond DSM-defined disorders. A large epidemiological study suggests an overlap between autistic-like traits and various DSM-based symptom profiles including ADHD, anxiety, depression, and conduct problems (Lundström et al., 2011). Notably, the Social Responsiveness Scale (SRS), designed to characterize quantitative autistic traits in the general population, differentiates adolescents with varying presentations of psychopathology (high risk for psychosis, psychosis, disruptive behavioral disorders, ADHD and co-occurring motor atypicality, mood, and anxiety disorders) from typically developing controls (Cholemkery, Kitzerow, Rohrmann, & Freitag, 2014; Jalbrzikowski et al., 2013; Pine et al., 2008; Reiersen, Constantino, & Todd, 2008).

The primary objectives of the current study are to: (a) identify homogenous subgroups or behavioral risk profiles in a community sample of toddlers, (b) examine the utility of risk classification through follow-up assessment (see Figure 1 for study design), and (c) examine whether a unidimensional threshold criterion identified risk in a manner equivalent to a multidimensionally derived profile. We hypothesized that we would identify subgroups of children captured by multiple 'risk profiles' that together correspond to epidemiological estimates of risk for ASD and/or learning/developmental delays. We also hypothesized that high-risk profiles would differ from low-risk profiles on a separate measure of developmentally appropriate clinical dimensions, which would provide corroborative evidence of the prognostic utility of our phenoscreening approach. Finally, we hypothesized that a unidimensional threshold criterion approach would not differentiate groups on outcome measures and a factor mixture modeling (FMM) approach.

Methods

Procedure and participants

Parents of 17- to 25-month-old toddlers were recruited to complete three parent-report questionnaires along with a form

characterizing family demographics. This approach was chosen to balance parental burden with acquiring a sufficient amount of data to represent complex behavioral profiles. Participants were drawn from the University of Minnesota Institute of Child Development's Participant Pool, a research registry that includes families from across the state of Minnesota. There were no exclusionary criteria to participate other than toddler age. Data from a final sample of 1,570 toddlers were used in the current analysis (see Table 1 for sample demographics; further details on data collection procedures and a STROBE chart (Figure S1) available online in Supporting Information). There were no differences in demographic indicators between the final sample and excluded participants or between the final sample and the ~65,000 families contained within the registry.

To examine the prognostic utility of the data-derived risk profiles, toddlers from each latent subclass were recruited to complete a follow-up assessment when the children were between 18 and 41 months (n = 107; see Table S1 for follow-up sample demographics). To ensure representation of all subgroups in the follow-up sample, all toddlers classified into the asymmetrically smaller, putative 'moderate- to high-risk' profiles (classes 1, 2, 4; total n = 96) were invited to participate, of whom n = 43 participated. A random subsample of toddlers classified into the relatively larger, putative low-risk profiles (classes 3 and 5; total n = 1,474) was invited to participate, of whom n = 64 participated. See Figure 3 for a breakdown of children who participated in the follow-up assessment by risk profile. Research staff were blind to class assignment during follow-up recruitment and data collection. Follow-up analyses examined data from the Infant Toddler Social Emotional Assessment (ITSEA). To test whether a unidimensional threshold would yield similar information regarding risk as the Factor Mixture Modeling (FMM) approach, those children in the top decile of the Video-Referenced Rating of Reciprocal Social Behavior (vrRSB) were compared to the remaining children on ITSEA outcomes.

Ethical considerations

The University of Minnesota institutional review board approved this study, and parents of each participant provided permission and informed consent.

Measures

Repetitive Behavior Scale for Early Childhood (RBS-EC). This instrument has good to excellent psychometric properties and evidence of convergent/discriminant validity and reliability (Lasch, Wolff, & Elison, 2020; Wolff, Boyd, & Elison, 2016; further details in Appendix S1). Based in large part on the RBS-R (Bodfish et al., 2000), the questionnaire dimensionally quantifies normative variation in discrete forms of repetitive behavior in young children. We examined the frequency of reported behaviors across four subscales: repetitive movement, restricted interests, ritualistic behavior, and self-directed/ self-injurious behavior. While lower-order repetitive behaviors (motor mannerisms/stereotypies and self-injurious behaviors) are associated with intellectual and developmental disabilities, higher-order repetitive behaviors (i.e., rigid adherence to routines, insistent on sameness behavior) are observed across various anxiety disorders.

Video-Referenced Rating of Reciprocal Social Behavior (vrRSB). The vrRSB is a downward extension of the Social Responsiveness Scale (SRS; Constantino et al., 2003; c.f., Lasch et al., 2020) which quantifies trait-like behavior that differentiates ASD from control participants, but also differentiates adolescents with varying psychiatric/ neurodevelopmental disorders from controls. The vrRSB's video-based exemplar affords increased standardization via direct comparison with the child of interest. A total summary score (encompassing 48 items) was used to describe toddlers' reciprocal social communication abilities, with higher scores indicating greater impairment (Marrus et al., 2015).

MacArthur-Bates Communicative Developmental Inventories (MCDI). The MCDI is a well-established standardized measure that characterizes communication, including both gestural and vocal modalities. Two subscales were used to describe toddlers' word production and gesture use (Fenson et al., 2007).

Infant Toddler Social Emotional Assessment (ITSEA). The ITSEA is a well-established measure of social–emotional and behavioral development designed to

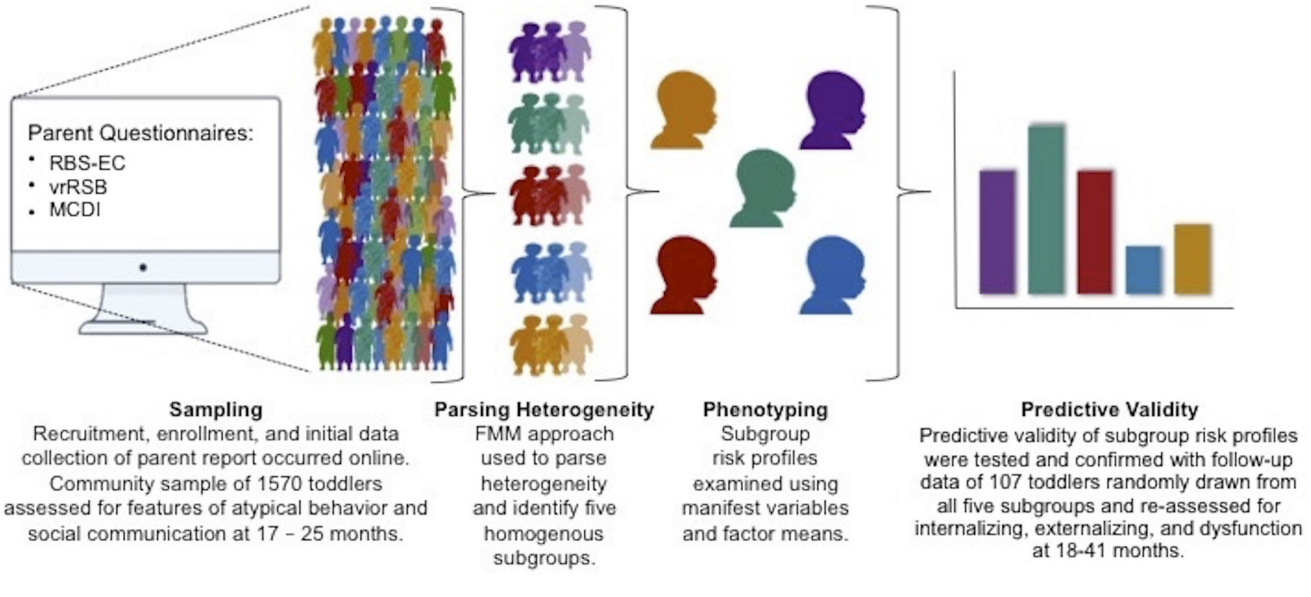

Figure 1 Illustration of study design

identify early deficits or delays (Carter, Briggs-Gowan, Jones, & Little, 2003). Three subscales were used as independent, distal outcomes to examine differences between subgroups: internalizing (e.g., withdrawal, general anxiety, inhibition to novelty), externalizing (e.g., impulsivity, aggression), and dysregulation (e.g., negative emotionality, sleep/eating dysregulation, sensory sensitivity).

Statistical methods

Primary analyses. To identify homogenous developmental profiles, we used factor mixture modeling (FMM), a statistical method for parsing population heterogeneity into homogenous subgroups (i.e., latent classes; c.f., Bolhuis et al., 2017; Lubke & Muthén, 2005; further details available online). FMM combines factor analysis, used to estimate unobserved continuous variables, with latent class analysis, used to identify unobserved categorical groups. Thus, FMM can overcome significant constraints in child psychiatry (Coghill & Sonuga-Barke, 2012; Pickles & Angold, 2003) related to dual goals of modeling behavioral phenomena dimensionally and identifying subgroups with putative clinical value. Additionally, because subgroups are represented through multiple observed manifest variables and factor means, phenotypic risk profiles generated by FMM have the potential to be more informative than a measure of any single domain by itself.

As a prerequisite to fitting a FMM (see Clark et al., 2013), a confirmatory factor analysis (CFA) was conducted to test

Table 1 Descriptive statistics

	N	Mean/%	*SD*	Min	Max
Family variables					
Parent age (years)[1]	1,570	33.43	3.92	22.01	55.37
Parent education level[1]					
Some high school	2	0.13			
High school	16	1.02			
Some college/2 year. degree	142	9.04			
College degree	654	41.66			
Some graduate school	105	6.69			
Graduate degree	651	41.46			
Family income					
<$25,000	16	1.02			
$25,000–$49,999	105	6.69			
$50,000–$74,999	234	14.90			
$75,000–$99,999	324	20.64			
$100,000–$149,999	514	32.74			
$150,000–$199,999	235	14.97			
$200,000 and over	142	9.04			
Child variables					
Age at assessment (months)	1,570	20.06	2.02	17.22	26.46
Birthweight (grams)	1,570	3,508.32	482.86	2,027.00	5,018.00
GA at birth (weeks)	1,570	39.50	1.49	32.57	45.71
Sex					
Female	745	47.45			
Male	825	52.55			
Ethnicity					
Hispanic	53	3.38			
Non-Hispanic	979	62.35			
Unknown/Did not report	538	34.27			
Race					
American Indian/Alaska Native	0	0.00			
Asian	30	1.91			
Black	9	0.57			
Multi-racial	112	7.13			
Native Hawaiian/Pacific Islander	0	0.00			
White	1,371	87.32			
Other	22	1.40			
Unknown/ not reported	26	1.66			
FMM parameters					
RBS-EC subscales					
Repetitive Movement	1,570	14.86	11.13	0.00	36.00
Restricted Interests	1,570	2.77	3.31	0.00	35.00
Ritual and Routine	1,570	4.14	4.24	0.00	28.00
SDSI Behavior	1,570	2.09	2.70	0.00	22.00
vrRSB subscale					
Social communication	1,570	19.23	6.46	3.00	65.00
MCDI subscales					
Total gestures	1,570	46.74	9.19	0.00	63.00
Words produced	1,570	108.20	104.45	0.00	396.00

Abbreviations: [1] = reporting parent; FMM = Factor Mixture Model; GA = gestational age; Max = Maximum; MCDI = MacArthur-Bates Communicative Developmental Inventories; Min = Minimum; RBS-EC = Repetitive Behavior Scale for Early Childhood; *SD* = standard deviation; SDSI = self-directed, self-injurious; vrRSB = Video-Referenced Rating of Reciprocal Social Behavior.

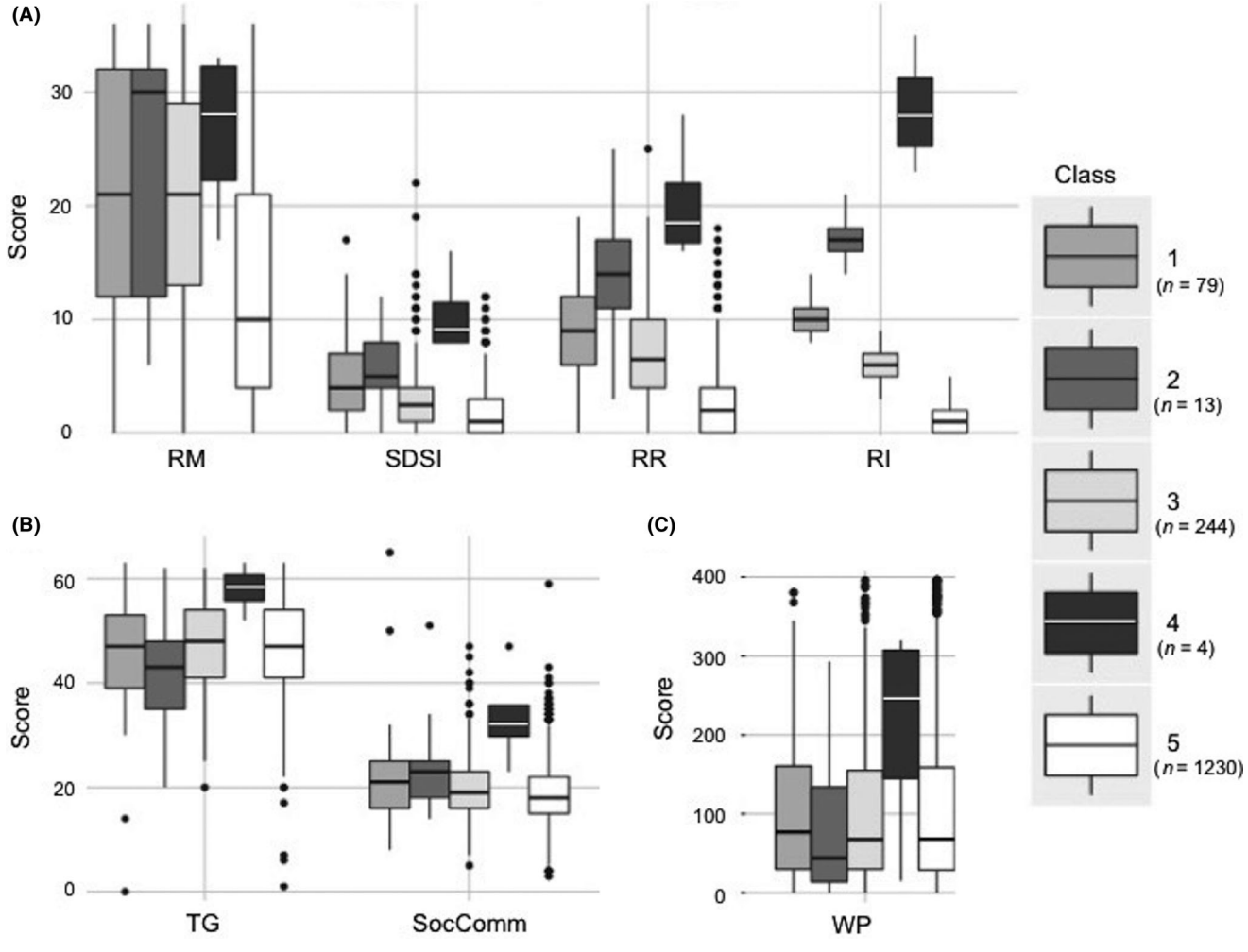

Figure 2 Factor mixture model results by risk subgroups. (A) Boxplots of manifest variables included in factor 1 (atypical behavior) drawn from the Repetitive Behavior Scale for Early Childhood (RBS-EC). Note: RM = repetitive movement, SDSI = self-directed/self-injurious behavior, RR = ritual and routine, RI = restricted interests. (B and C) Boxplots of manifest variables included in factor 2 (social communication) drawn from the MacArthur-Bates Communicative Developmental Inventories (MCDI) and Video-Referenced Rating of Reciprocal Social Behavior (vrRSB). Note: TG = total gestures used, SocComm = social communication impairment, WP = words produced. Variables in B and C are graphed separately due to differences in scale

our hypothesized factor structure (Figure S2). Consistent with previous data from individuals with ASD and a general population twin sample (Georgiades et al., 2013; Ronald et al., 2006), two factors fit the data well (RMSEA = 0.07; CFI = 0.94; TLI = 0.91). Factor mixture modeling was conducted using MPlus 7.4 (Muthén & Muthén, 2012). Detecting subgroups that reflect epidemiological rates of some hypothesized risk phenotypes (e.g., phenotypes reflecting ASD) requires identifying asymmetric subgroup sizes; class (es) totaling approximately 20 children within our sample of 1,570 would reflect ASD prevalence rates of 1 in 68 children or 1.5%. Therefore, to detect subgroups of this size, assessing measurement invariance was not possible and strict factorial invariance was assumed (Lubke & Muthén, 2005).

During model fitting, seven manifest variables loaded onto two factors then used to enumerate latent classes. Specifically, four RBS-EC subscales (repetitive movement, restricted interests, ritualistic behavior, self-directed/self-injurious behavior) loaded onto the first factor, representing atypical behaviors. Three additional indices drawn from the vrRSB and MCDI (social communication, words produced, gestures used), loaded onto the second factor, representing social communication behaviors. Following an iterative process (Clark et al., 2013), factor mixture models were fit with an increasing number of classes, from 2 to 7 classes, and the fit of different models was compared (see Table S2). A model with seven classes did not converge and is not reported. The best fitting model was selected using best practice criteria (Clark et al., 2013; further details in Figure S3). Taken together, the 5-class model was selected as the final model, as converging evidence drawn from fit indices, entropy value, and the results of sensitivity analysis showed this model offered a combination of the relative best fit. Once generated, class profiles were examined by comparing manifest variable and factor means (see Table 2 and Figure 2).

Follow-up analyses. To test the utility of the latent classes identified by our 5-class factor mixture model, follow-up data from a subset of 107 toddlers were used. First, to visually examine group differences, we generated boxplots of ITSEA subscales by risk subgroups (Figure 3A). Using a MANOVA framework, we examined whether putative high-risk (i.e., combined classes 2 and 4) and moderate-risk classes (i.e., class 1) could be differentiated from low-risk classes (i.e., combined classes 3 and 5). Planned Tukey HSD post hoc

Table 2 Phenotype risk profiles for five class factor mixture model

	Class 1 (N = 79)				Class 2 (N = 13)				Class 3 (N = 244)				Class 4 (N = 4)				Class 5 (N = 1230)			
	Mean/%	*SD*	Min	Max	Mean/%	*SD*	Min	Max	Mean/%	*SD*	Min	Max	Mean/%	*SD*	Min	Max	Mean/%	*SD*	Min	Max
Sex																				
Female	39.24				30.77				43.85				50.00				51.14			
Male	60.76				69.23				56.15				50.00				48.86			
Age (months)	20.43	2.01	17.60	25.80	20.70	3.07	17.32	25.53	20.28	2.29	17.30	25.93	21.80	3.04	18.72	24.66	19.98	1.94	17.22	26.45
Repetitive movement	21.30	10.72	0	36	24.77	10.76	6	36	20.75	9.67	0	36	26.50	7.51	17	33	13.13	10.82	0	36
Restricted interests	10.49	1.62	8	14	16.85	2.23	14	21	5.84	1.32	3	9	28.50	5.20	23	35	1.43	1.29	0	5
Ritual and routine	9.68	4.39	0	19	14.08	5.48	3	25	7.30	4.73	0	25	20.25	5.44	16	28	3.00	3.14	0	18
SDSI Behavior	4.48	3.71	0	17	6.23	3.39	0	12	3.28	3.43	0	22	10.50	3.79	8	16	1.63	2.17	0	12
Factor 1	8.98	0.97			15.56	1.18			4.32	1.01			27.45	2.22			0.08	0.70		
Social comm.[1]	21.47	7.82	8	65	24.31	10.07	14	51	20.06	6.55	5	47	33.50	9.95	23	47	18.82	6.17	3.00	59
Total gestures	45.69	10.55	0	63	41.00	12.78	20	62	47.32	8.40	20	62	58.00	4.69	52	63	46.72	9.19	1.00	63
Words produced	106.06	98.53	0	381	89.77	104.34	0	293	109.78	109.26	0	396	206.50	140.49	15	320	107.90	103.74	0	396
Factor 2	0.87	3.86			3.06	4.73			0.07	3.25			−1.84	2.68			0.001	3.32		

Abbreviations: [1] = higher scores indicate greater difficulty/risk; Max = maximum; Min = minimum; *SD* = standard deviation; SDSI = self-directed, self-injurious; Social Comm. = Social communication.

comparisons among the three groups corrected for multiple comparisons.

Results

Factor mixture model results

Our final model identified five latent classes (see Figure S4). Classes were not significantly different across any child or family demographic characteristics, including child age, race/ethnicity, family socioeconomic status, and primary language spoken in the home. Sensitivity analysis provided evidence that homogenous, asymmetrically sized subgroups were captured effectively (see Figure S3), as class membership was conserved across model specifications, indicating stable, 'genuine' subgroups were identified. The 5-class model identified two small subgroups (classes 4 and 2; combined n = 17) showing elevated rates of atypical behavior (factor 1) and social communication impairments (factor 2); within our total sample of 1,570, this collapsed high-risk group approximates the phenotype and expected prevalence rate of ASD (17/1,570 = 1%).

Based on evidence gained from examining class profiles (see Figure 2), subgroups were interpreted to represent stratified risk phenotypes. Classes 4 (n = 4) and 2 (n = 13) were interpreted to represent high-risk groups, as their profiles revealed high levels of atypical behaviors. Additionally, class 4 exhibited the highest levels of gesture and word production along with the highest levels of social communication impairment, while class 2 showed the lowest levels of gesture and word production and a relatively increased level of social communication impairment. Altogether, the profile features of these two classes indicate risk phenotypes aligning with characteristics of subgroups of individuals with ASD who may exhibit atypical behaviors coupled with verbal or nonverbal social communication impairments. Class 1 (n = 79) was interpreted to represent a moderate-risk group, as this class profile showed relatively higher atypical behaviors and average social communication development. In contrast to these moderate- to high-risk subgroups, classes 3 (n = 244) and 5 (n = 1230) were interpreted to represent low-risk phenotypes, as their profiles revealed low levels of atypical behaviors, average gesture and word production, and low levels of social communication impairment. Notably, an examination of manifest variables (see Figure 2) revealed that the subgroups were clearly differentiated by factor 1 variables. For example, class profiles show very good stratification among subgroups for self-directed/self-injurious behavior, ritual and routine, and restricted interests.

Follow-up results

A MANOVA with three dependent variables (i.e., externalizing, internalizing, and dysregulation

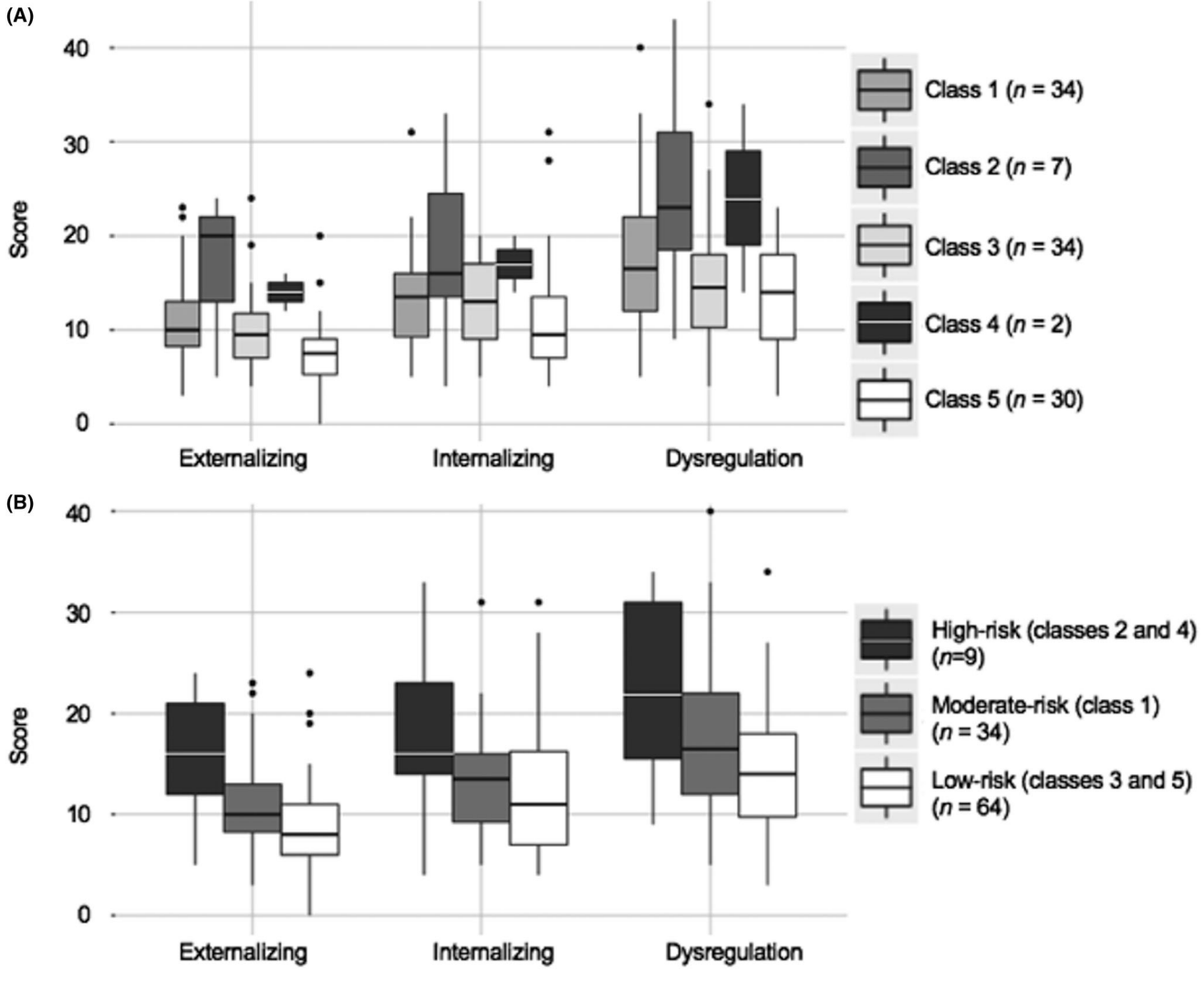

Figure 3 Follow-up results. (A) Boxplots illustrating risk classes by ITSEA subscale medians, 1st and 3rd quartile range, and the 95% confidence interval. (B) Boxplots illustrating the same metrics for high- (classes 2 and 4), moderate- (class 1), and low-risk (classes 3 and 5) groups

subscales from the ITSEA) and three groups (i.e., high-risk comprising classes 2 and 4; moderate-risk comprising class 1; low-risk comprising classes 3 and 5) was implemented to test the null hypothesis that risk groups are statistically equivalent on ITSEA subscale scores (Figure 3B). There was a statistically significant difference in internalizing, externalizing, and dysregulation scores based on risk group, F (6, 204) = 4.29, $p < .005$; Wilk's $\Lambda = .79$, partial $\eta^2 = .11$. Planned follow-up comparisons revealed that the high-risk group differed from the moderate-risk group (externalizing: $p = .007$, $d = 0.99$; internalizing: $p = .046$, $d = 0.67$; dysregulation: $p = .041$, $d = 0.72$) and the low-risk group (externalizing: $p < .001$, $d = 1.39$; internalizing: $p = .006$ $d = 0.83$; dysregulation: $p < .001$, $d = 1.19$) on all three variables and that the moderate-risk group significantly differed from the low-risk group on the dysregulation subscale: $p = .022$, $d = 0.47$; see Table S3).

To test a 'single-instrument, single-threshold' approach to risk identification, vrRSB total scores were used to identify a high-risk and low-risk group. Participants in the top decile ($n = 16$) were designated 'high-risk' ($M = 30.44$; $SD = 3.03$), and all other participants ($n = 91$) were designated 'low-risk' ($M = 15.81$; $SD = 6.87$). A one-way ANOVA showed a significant difference in vrRSB scores by threshold group, F (2,104) = 41.34, $p < .001$. Results of a MANOVA showed there was no statistically significant difference in internalizing, externalizing, and dysfunction scores from the ITSEA based on threshold group, F (3, 103) = 1.47, $p = .23$; Wilk's $\Lambda = .96$, partial $\eta^2 = .04$ (see Table S4).

Discussion

Factor mixture modeling (FMM) was used to identify latent classes representing risk profiles in a community sample of 17- to 25 month-olds, and these risk profiles were corroborated with follow-up data. From the sample of 1,570 toddlers, 1% ($n = 17$) were captured by two risk profiles. An additional 5%

(n = 79) could be considered a moderate-risk group. As a whole, those at elevated risk showed more externalizing symptoms and dysregulated behaviors than low-risk comparison toddlers when measured an average of 10 months after the initial battery of questionnaires, corroborating the prognostic utility of the phenoscreening approach. The question of whether psychiatric classification is best conceptualized as a latent taxon, latent dimension, or a combination of the two has a distinguished history and continues to be scrutinized (e.g., Kotov et al., 2017). FMM incorporates dimensional and categorical properties into a single analytic framework to provide one avenue for exploring the nature of emerging psychopathology.

Realizing the goal of strategic prevention requires methods for detecting incipient profiles or patterns of behavior in population-based samples prior to the consolidation of psychiatric symptomatology. We selected multiple dimensions of behavior representing meaningful variability across the typical-to-atypical continuum. Notably, restricted patterns of play behavior were completely stratified across the five latent classes, as were 'just right' behaviors (Evans et al., 1997) to a slightly lesser degree, as captured by the rituals and routines subscale. Whether improved conceptualization and characterization of repetitive behaviors during toddlerhood will improve characterization of early emerging risk warrants further study.

This study has several important strengths. Notably, we verify that parents will complete ~45 min of questionnaires about their child's behavior if incentivized. Single instrument, single-threshold population-based screening approaches that yield binary outcomes may not provide clinicians with sufficient descriptive specificity to make precise clinical recommendations. The phenoscreening strategy has the potential to advance a more personalized approach to clinical recommendations based on early screening by providing distinct, multidimensional profiles of risks and relative strengths for subgroups to guide recommendations for subsequent stage 2 screening or direct developmental assessments. For example, although classes 4 and 2 are both putative high-risk groups, their distinct profiles warrant tailored recommendations. Toddlers in class 2 may benefit from assessments and interventions targeting language delays or deficits, toddlers in class 4 may benefit from interventions to strengthen reciprocal social communication skills, and toddlers in both groups may benefit from evidence-based treatments targeting specific restricted, repetitive behaviors (see Boyd, McDonough, & Bodfish, 2012). Moreover, deriving data-driven subgroups through the phenoscreening strategy may minimize or reduce the phenotypic heterogeneity observed in DSM-defined samples, a feature that has hindered basic scientists' search for pathophysiological mechanisms. Lastly, we captured complex behavioral profiles in toddlers, prior to the preschool-aged years when a substantial proportion of maladaptive behavior patterns begin to consolidate.

A primary limitation of the current study is the sample size. A larger starting sample would likely identify a greater number of children in high and moderate-risk profiles, affording the opportunity to examine each profile independently at outcome. An additional limitation includes sample homogeneity in regard to SES and race/ethnicity, which prevents our study from addressing important environmental risk factors, and limits the generalizability of results. Follow-up assessment with a larger and more diverse sample, using a broader suite of direct clinical assessments, is warranted to further investigate the clinical utility of our approach. Acquiring additional biological measures, for example, structural and functional connectivity data and/or polygenic risk scores, could verify the biological substrates of risk profiles. Finally, our analyses were exclusive to behavioral features at one time point. Future studies using FMM or other approaches (Feczko et al., 2019) to identify latent subgroups may benefit from combining behavioral and biological factors, or change over time in these factors, to identify and inform risk profiling. Given the range of potential outcomes for infants/toddlers at risk, parsing the heterogeneity of developmental trajectories over more than one time point may provide insight into why a child follows one developmental trajectory and not another.

Conclusions

There is an urgent need to identify incipient signatures of later emerging mental illness in order to implement preventive interventions. Identifying instantiating pathophysiological markers, beyond high-familial-risk cases, requires an approach to population-based sampling that yields clinically meaningful risk profiles at sensitive periods in development. Phenoscreening, which leverages data-driven computational approaches to minimize phenotypic heterogeneity across the typical-to-atypical continuum during toddlerhood, is one such approach.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix S1. Additional methodological description.
Table S1. Demographics of follow-up sample (N=107).
Table S2. Class enumeration model fit statistics.
Table S3. MANOVA with planned Tukey HSD post-hoc tests.
Table S4. Single-instrument single-threshold results.
Figure S1. STROBE chart illustrating data collection and exclusion prior to analysis.

Figure S2. Figure illustrating results of a confirmatory factor analysis conducted to test our a priori hypothesized factor structure including an atypical behavior factor (Factor 1) and a social communication factor (Factor 2).
Figure S3. Table illustrating sensitivity analyses results.
Figure S4. Final factor mixture model with seven manifest variables, two factors, and five latent classes.

Acknowledgements
C.D. and C.L. were funded by NSF-GSF awards. J.J.W. was funded by NIMH K01 MH101653. This study was made possible by an NIMH BRAINS award (R01 MH104324) to J.T.E. The authors thank the families who participated, Dustin Gibson for his consultation, and our research team including Kristen Gault and Emma Shankland. The authors have declared that they have no competing or potential conflicts of interest.

Correspondence
Jed T. Elison, Institute of Child Development and Department of Pediatrics, University of Minnesota, 51 E. River Parkway, Minneapolis, MN 55455, USA; Email: jtelison@umn.edu

Key points

- Early identification and intervention efforts require the ability to detect risk profiles for psychiatric/neurodevelopmental disorders in early childhood. However, variability across the typical-to-atypical continuum and heterogeneity within and across early emerging disorders represent fundamental challenges.
- We used a data-driven approach (factor mixture modeling) to overcome these challenges. We identified and characterized phenotypic profiles for five risk groups in a large community-based sample of toddlers. Independent, distal outcomes demonstrated predictive utility of the risk classification with large effect sizes between high-, moderate-, and low-risk groups.
- Data-driven approaches, leveraging multiple developmentally appropriate dimensional RDoC constructs, holds promise for future efforts aimed toward early identification of at-risk-phenotypes for a variety of early emerging neurodevelopmental disorders.

References
Andrews, G., Issakidis, C., Sanderson, K., Corry, J., & Lapsley, H. (2004). Utilising survey data to inform public policy: comparison of the cost-effectiveness of treatment of ten mental disorders. *The British Journal of Psychiatry*, *184*, 526–533.

Bebko, G., Bertocci, M.A., Fourneir, J.C., Hinze, A.K., Bonar, L., Almeida, J.R., … & Phillips, M.L. (2014). Parsing dimensional vs diagnostic category-related patterns of reward circuitry function in behaviorally and emotionally dysregulated youth in the Longitudinal Assessment of Manic Symptoms study. *JAMA Psychiatry*, *71*, 71–80.

Bernard, J.A., & Mittal, V.A. (2015). Updating the research domain criteria: the utility of a motor dimension. *Psychological Medicine*, *45*, 2685–2689.

Bodfish, J.W., Symons, F.J., Parker, D.E., & Lewis, M.H. (2000). Varieties of repetitive behavior in autism: Comparisons to mental retardation. *Journal of Autism and Developmental Disorders*, *30*, 237–243.

Bolhuis, K., Lubke, G.H., van der Ende, J., Bartels, M., van Beijsterveldt, C.E., Lichtenstein, P., … & Boomsma, D.I. (2017). Disentangling heterogeneity of childhood disruptive behavior problems into dimensions and subgroups. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*, 678–686.

Boyd, B.A., McDonough, S.G., & Bodfish, J.W. (2012). Evidence-based behavioral interventions for repetitive behaviors in autism. *Journal of Autism and Developmental Disorders*, *42*, 1236–1248.

Cannon, T.D., Yu, C., Addington, J., Bearden, C.E., Cadenhead, K.S., Cornblatt, B.A., … & Perkins, D.O. (2016). An individualized risk calculator for research in prodromal psychosis. *American Journal of Psychiatry*, *173*, 980–988.

Carter, A.S., Briggs-Gowan, M.J., Jones, S.M., & Little, T.D. (2003). The infant–toddler social and emotional assessment (ITSEA): Factor structure, reliability, and validity. *Journal of Abnormal Child Psychology*, *31*, 495–514.

Casey, B.J., Oliveri, M.E., & Insel, T. (2014). A neurodevelopmental perspective on the research domain criteria (RDoC) framework. *Biological Psychiatry*, *76*, 350–353.

Cholemkery, H., Kitzerow, J., Rohrmann, S., & Freitag, C.M. (2014). Validity of the social responsiveness scale to differentiate between autism spectrum disorders and disruptive behaviour disorders. *European Child & Adolescent Psychiatry*, *23*, 81–93.

Cicchetti, D. (1993). Developmental Psychopathology: Reactions, reflections, projections. *Developmental Review*, *13*, 471–502.

Clark, S.L., Muthén, B., Kaprio, J., D'Onofrio, B.M., Viken, R., & Rose, R.J. (2013). Models and strategies for factor mixture analysis: An example concerning the structure underlying psychological disorders. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(4), 681–703.

Clementz, B.A., Sweeney, J.A., Hamm, J.P., Ivleva, E.I., Ethridge, L.E., Pearlson, G.D., … & Tamminga, C.A. (2016). Identification of distinct psychosis biotypes using brain-based biomarkers. *American Journal of Psychiatry*, *173*, 373–384.

Coghill, D., & Sonuga-Barke, E.J. (2012). Annual research review: categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders–implications of recent empirical study. *Journal of Child Psychology & Psychiatry*, *53*, 469–489.

Constantino, J.N., Davis, S.A., Todd, R.D., Schindler, M.K., Gross, M.M., Brophy, S.L., … & Reich, W. (2003). Validation of a brief quantitative measure of autistic traits: Comparison

of the social responsiveness scale with the autism diagnostic interview-revised. *Journal of Autism & Developmental Disorders*, *33*, 427–433.

Cuthbert, B.N., & Insel, T.R. (2013). Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Medicine*, *11*, 126.

Dar, R., Kahn, D.T., & Carmeli, R. (2012). The relationship between sensory processing, childhood rituals and obsessive–compulsive symptoms. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*, 679–684.

Egger, H.L., & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: Presentation, nosology, and epidemiology. *Journal of Child Psychology & Psychiatry*, *47*, 313–337.

Evans, D.W., Leckman, J.F., Carter, A., Reznick, J.S., Henshaw, D., King, R.A., & Pauls, D. (1997). Ritual, habit, and perfectionism: The prevalence and development of compulsive-like behavior in normal young children. *Child Development*, *68*, 58–68.

Fair, D.A., Bathula, D., Nikolas, M.A., & Nigg, J.T. (2012). Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 6769–6774.

Feczko, E., Miranda-Dominguez, O., Marr, M., Graham, A.M., Nigg, J.T., & Fair, D.A. (2019). The heterogeneity problem: Approaches to identify psychiatric subtypes. *Trends in Cognitive Sciences*, *23*, 584–601.

Fenson, L., Bates, E., Dale, P.S., Marchman, V.A., Reznick, J.S., & Thal, D. (2007). *MacArthur-Bates communicative development inventories user's guide and technical manual* (2nd edn). Baltimore, MD: Paul H. Brookes Publishing Company.

Franklin, J.C., Jamieson, J.P., Glenn, C.R., & Nock, M.K. (2015). How developmental psychopathology theory and research can inform the research domain criteria (RDoC) project. *Journal of Clinical Child & Adolescent Psychology*, *44*, 280–290.

Garvey, M., Avenevoli, S., & Anderson, K. (2016). The National Institute of Mental Health Research Domain Criteria and clinical research in child and adolescent psychiatry. *Journal of American Academy of Child & Adolescent Psychiatry*, *55*, 93–98.

Georgiades, S., Szatmari, P., Boyle, M., Hanna, S., Duku, E., Zwaigenbaum, L., … & Thompson, A. (2013). Investigating phenotypic heterogeneity in children with autism spectrum disorder: a factor mixture modeling approach. *Journal of Child Psychology & Psychiatry*, *54*, 206–215.

Hafeman, D.M., Merranko, J., Goldstein, T.R., Axelson, D., Goldstein, B.I., Monk, K., … & Birmaher, B. (2017). Assessment of a person-level risk calculator to predict new-onset bipolar spectrum disorder in youth at familial risk. *JAMA Psychiatry*, *74*, 841–847.

Hazlett, H.C., Gu, H., Munsell, B.C., Kim, S.H., Styner, M., Wolff, J.J., … & Piven, J. (2017). Early brain development in infants at high risk for autism spectrum disorder. *Nature*, *542*, 348–351.

Hyman, S.E. (2010). The diagnosis of mental disorders: The problem of reification. *Annual Review of Clinical Psychology*, *6*, 155–179.

Insel, T.R., & Scolnick, E.M. (2006). Cure therapeutics and strategic prevention: Raising the bar for mental health research. *Molecular Psychiatry*, *11*, 11–17.

Jalbrzikowski, M., Krasileva, K.E., Marvin, S., Zinberg, J., Andaya, A., Bachman, P., … & Bearden, C.E. (2013). Reciprocal social behavior in youths with psychotic illness and those at clinical high risk. *Development and Psychopathology*, *25*, 1187–1197.

Karalunas, S.L., Fair, D., Musser, E.D., Aykes, K., Iyer, S.P., & Nigg, J.T. (2014). Subtyping attention-deficit/hyperactivity disorder using temperament dimensions: Toward biologically based nosologic criteria. *JAMA Psychiatry*, *71*, 1015–1024.

Kessler, R.C., Berglund, P., Demler, O., Jin, R., Merikangas, K.R., & Walters, E.E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 593–602.

Kotov, R., Krueger, R.F., Watson, D., Achenbach, T.M., Althoff, R.R., Bagby, R.M., … & Eaton, N.R. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): a dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*, 454.

Lasch, C., Wolff, J.J., & Elison, J.T. (2020). Examining criterion-oriented validity of the Repetitive Behavior Scales for Early Childhood (RBS-EC) and the Video-Referenced Rating of Reciprocal Social Behavior (vrRSB). *Development & Psychopathology*, *32*, 779–789.

Lubke, G.H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, *10*, 21.

Lundström, S., Chang, Z., Kerekes, N., Gumpert, C.H., Råstam, M., Gillberg, C., … & Anckarsäter, H. (2011). Autistic-like traits and their association with mental health problems in two nationwide twin cohorts of children and adults. *Psychological Med*, *41*, 2423–2433.

Marrus, N., Glowinski, A.L., Jacob, T., Klin, A., Jones, W., Drain, C.E., … & Sant, S.M. (2015). Rapid video-referenced ratings of reciprocal social behavior in toddlers: a twin study. *Journal of Child Psychology & Psychiatry*, *56*, 1338–1346.

Merikangas, K., He, J.P., Burstein, M., Swanson, S.A., Avenevoli, S., Cui, L., … & Swendsen, J. (2010). Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication-Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*, 980–989.

Mittal, V.A., & Wakschlag, L.S. (2017). Research domain criteria (RDoC) grows up: Strengthening neurodevelopmental investigation within the RDoC framework. *Journal of Affective Disorders*, *216*, 30.

Muthén, L.K., & Muthén, B.O. (2012). *Mplus user's guide* (7th edn). Los Angeles: Muthén & Muthén.

Pickles, A., & Angold, A. (2003). Natural categories or fundamental dimensions: On carving nature at the joints and the rearticulation of psychopathology. *Development & Psychopathology*, *15*, 529–551.

Pine, D.S., Guyer, A.E., Goldwin, M., Towbin, K.A., & Leibenluft, E. (2008). Autism spectrum disorder scale scores in pediatric mood and anxiety disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*, 652–661.

Reiersen, A.M., Constantino, J.N., & Todd, R.D. (2008). Co-occurrence of motor problems and autistic symptoms in attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*, 662–672.

Rice, F., Sellers, R., Hammerton, G., Eyre, O., Bevan-Jones, R., Thapar, A.K., … & Thapar, A. (2017). Antecedents of new-onset major depressive disorder in children and adolescents at high familial risk. *JAMA Psychiatry*, *74*, 153–160.

Ronald, A., Happé, F., Price, T.S., Baron-Cohen, S., & Plomin, R. (2006). Phenotypic and genetic overlap between autistic traits at the extremes of the general population. *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*, 1206–1214.

Sonuga-Barke, E.J. (2014). 'What's up,(R) DoC?'–can identifying core dimensions of early functioning help us understand, and then reduce, developmental risk for mental disorders? *Journal of Child Psychology & Psychiatry*, *55*, 849–851.

Totah, N., Akil, H., Huys, Q.J., Krystal, J.H., MacDonald, A.W., III, Maia, T.V., … & Pauli, W.M. (2016). Complexity and heterogeneity in psychiatric disorders. In A.D. Redish & J.A.

Gordon (Eds.). *Computational psychiatry: New perspectives on mental illness* (pp. 33–60). Cambridge, MA: MIT Press.

Van Dam, N.T., O'Connor, D., Marcelle, E.T., Ho, E.J., Craddock, R.C., Tobe, R.H., ... & Milham, M.P. (2017). Data-driven phenotypic categorization for neurobiological analyses: beyond DSM-5 labels. *Biological Psychiatry*, *81*, 484–494.

Vigo, D., Thornicroft, G., & Atun, R. (2016). Estimating the true global burden of mental illness. *Lancet Psychiatry*, *3*, 171–178.

Wakschlag, L.S., Estabrook, R., Peticlerc, A., Henry, D., Burns, J.L., Perlman, S.B., Voss, J.L., ... & Briggs-Gowen, M.L. (2015). Clinical implications of a dimensional approach: the normal:abnormal spectrum or early irritability. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*, 626–634.

Wolff, J.J., Boyd, B.A., & Elison, J.T. (2016). A quantitative measure of restricted and repetitive behaviors for early childhood. *Journal of Neurodevelopmental Disorders*, *8*, 27.

Accepted for publication: 5 October 2020