



Acta Genet Med Gemellol 34: 15-31 (1985)  
©1985 by The Mendel Institute, Rome

Received 24 July 1984  
Final 6 November 1984

## Multimethod Assessment of Temperament in Twins: Birth to Six Months

M.L. Riese, R.S. Wilson, A.P. Matheny, Jr.  
*University of Louisville School of Medicine, Kentucky*

---

**Abstract.** The predictive relationship between neonatal behavior and temperament at 6 months was assessed in 62 pairs of twins. The neonatal assessment evaluated irritability, resistance to soothing, reactivity, and activity level across various activities. Temperament at 6 months was appraised by a laboratory assessment, ratings on the Infant Behavior Record, and a questionnaire completed by the mother. Examination for temperament consistency from the neonatal period to 6 months indicated that (a) neonates rated by an examiner as irritable and difficult to soothe were reported by the mother to be more negative, irritable, and slower to adapt at 6 months; (b) neonates rated as more irritable and active were more attentive and responsive during mental testing at 6 months; (c) there was no relationship between the neonatal measures and the laboratory ratings, possibly reflecting methodological problems influenced by maturation and vegetative processes at 6 months. Analyses for patterns of concordance within the various data sets for MZ and DZ twins indicated that there was no pattern of differential concordance favoring MZ twins for the temperament measures at birth or at 6 months. It was concluded that a modest predictive relationship between temperament variables had been demonstrated from birth to 6 months, but no specific genetic effects were detected at these early ages.

**Key words:** Temperament, Neonatal Behavior, Twin concordance, Infant twin assessment

---

### INTRODUCTION

The Louisville Twin Study has for some years been engaged in longitudinal research with young twins, principally in the area of mental development and physical growth [eg 19,22]. Beginning in 1976, however, the principal focus of the research program turned to the assessment of temperament in infant twins.

The transition was predicated on two major factors: (a) in our twin sample, an appreciation of the central role played by the twins' temperament in shaping the parent's initial reaction, especially where differences in temperament were evident and (b) a detailed and provocative body of research on temperament from the New York Longitudinal Study [16,17]. The focus of the latter program was perhaps best stated in the authors' words:

"We have focused ... on behavioral differences found among infants during the first three months of life. By tracing the fate of these attributes ... we have sought to determine which of the initial patterns proved to be persistent as the infants grew older, and to explore the ways in which they were significant in influencing later behavioral development" [17: p 4].

This premise has served as the basic rationale for many of the current research studies in infant temperament. The studies have ranged from neonatal assessments to follow-up studies at later ages, often using temperament questionnaires filled out by the parents. The results have sometimes been disappointing, however; and two recent reviews have pointed to certain limitations in the questionnaire studies, while urging that controlled laboratory studies be performed [5,6].

### **Research Rationale**

Within this context, the research program of the Louisville Twin Study was designed to get at the earliest manifestations of temperament in the infant. In broad view, the goal was to make an appraisal in the neonatal period, before the infants were exposed to the socializing experiences of the home; and to create a laboratory assessment of temperament at 6 months. If the infant brings the rudiments of an intrinsic response style with him/her at birth, that pattern might be manifested in the first days of life, and then be sustained over 6 months. The task was to construct an assessment protocol that would adequately measure these aspects of temperament, without being overwhelmed by the vegetative-state processes that play such a dominant role in the early months of life.

### **Related Background Studies**

In a prior longitudinal study of infant twins, a battery of standardized laboratory procedures was developed to serve as probes for temperament [10,23]. The infant was presented with some age-appropriate challenges, the most notable being separation from the mother, and then a structured series of play activities and soothing techniques were employed to alleviate distress and upset. The infant's behavior was rated from videotapes by trained observers, to yield a composite measure of each infant's temperament profile.

In addition, a detailed neonatal assessment had been developed [13], which focused on behaviors expressive of temperament (eg, irritability, resistance to soothing, activity, and reactivity). The procedure searched for the predominant characteristics of each neonate as displayed during a three-hour appraisal of feeding, sleeping, orienting behaviors, and responsiveness to noxious and soothing stimuli. If the rudiments of temperament have a constitutional basis and are evident shortly after birth, they should be detected by this assessment. And if these rudiments continue to shape the infant's behavior, then there should be some stability between neonatal reactions and later temperament.

While the primary measure of temperament at 6 months would be obtained from the laboratory appraisal, it also seemed desirable to capitalize upon the extended observations

of the mother at home. For this purpose, the mother completed a temperament questionnaire for each twin after the 6-months visit, synthesizing her impressions of the infant's temperament as revealed in daily activities. The questionnaire results would supplement the laboratory appraisal to form a composite measure of infant temperament.

The present report focuses upon the predictive relationships between the neonatal assessment and the collective measures of temperament made at 6 months. It also considers the role of zygosity in early temperament – are MZ twins more concordant than DZ twins in patterns of temperament displayed? The latter comparison should help identify any genetic influences on early temperament.

## Method and Procedure

**Subjects.** The twins have been recruited as part of an ongoing longitudinal study described in more detail elsewhere [22]. Recruits were drawn from the entire twin birth registry in the Louisville area, with a special effort made to enroll and retain families of low socioeconomic status (SES). The full SES distribution was represented in the sample, with 27% of the families in the lowest two deciles of occupational rating scales [11] and the remaining families distributed in roughly equal proportions through the other eight deciles (exception: 4% in highest decile).

Recruited twins began making quarterly visits to the research center at 3 months, but this first visit was largely preoccupied with becoming acquainted with the family, and handling the necessary caretaking activities with the twins. By 6 months the twins were on a more definite schedule; they remained awake for longer periods, and there were more activities in which they could become engaged. The lab assessment at 6 months thus furnished the first reasonably complete appraisal of temperament for these infants.

The data to be reported were obtained from 124 infants who had completed the 6-months visit, with virtually complete data for all programmed episodes. The 124 infants were drawn from 24 male-male pairs, 26 female-female pairs, and 12 opposite-sex pairs. For technical and psychological reasons, the twins are not bloodtyped until they are three years old [20], and only 24 same-sex pairs have been diagnosed by this method to date (14 MZ pairs and 10 DZ pairs). Another 23 pairs, however, have been diagnosed by staff consensus, which has turned out to be very reliable (98%) when compared with bloodtyping for twins observed repeatedly throughout the first year. The remaining three same-sex pairs were of questionable zygosity and have been omitted from the twin analyses.

Neonatal data were available for a somewhat smaller number of twins in the longitudinal study, since the assessments did not begin until after several pairs had been recruited. Similarly, the infant temperament questionnaire did not become available until 1978, and the early pairs missed the questionnaire at 6 months. The appropriate sample size is reported separately for each analysis.

## Neonatal Assessments

The neonatal assessments were designed and conducted by MLR, and a complete description of the test protocol and its derivation may be found in Riese [12]. A full report of results from the standardization sample, including a comparison of full-term and preterm neonates, has been presented elsewhere [13,14].

By way of brief description, infants were examined on the second day of life whenever possible. When twins were born prematurely with medical complications, they were examined when they reached a point of medical stability to allow for the necessary handling (usually just before leaving the hospital).

Neonates were examined during an entire metabolic cycle; that is, from one feeding period to the next. This allowed for standardization of the entire sequence of appraisals, and for observation of the neonate's responsivity during several behavioral states. The following assessments were made:

- A) Behavioral state and irritability were noted before, during and after feeding, as well as activity and feeding ability.
- B) State organization and state changes were recorded after feeding, as well as spontaneous behaviors and activity during sleep.
- C) Midway between feedings, the infant was awakened so that maturational level, sensorimotor status, and orienting behaviors believed to involve cortical processing could be assessed, using a series of items adapted from the Einstein Neonatal Neurobehavioral Assessment Scale [7]. The measures of reactivity were derived primarily from these items.
- D) Reactivity to a stressor was studied by placing a cold disc on the infant's thigh, then noting initial responsivity and subsequent calming after soothing stimuli were applied [2].
- E) Throughout the assessment period, an evaluation was made of spontaneous irritability and consolability. Various types of soothing techniques were applied, including acceptance of a pacifier, vocal stimulation, manual stimulation, placement in the prone position, lifting to shoulder, cradling in arm, and swaddling in blanket. Individual responsivity to the various types of soothing, and degree of intervention necessary for soothing, were assessed.

**Neonatal Scales.** The scores and ratings obtained from the assessment have been combined into four composite scales – irritability, resistance to soothing, reactivity, and reinforcement value. In addition, two measures of activity were included in the behavioral profile, one during sleep and one while awake. These six composite scores served as a condensed behavioral profile for each neonate, based on most of the responses elicited during the neonatal exam.<sup>1</sup>

### Laboratory Assessment of Temperament

There were two primary considerations in designing the laboratory assessment of temperament. The first requirement was to depend solely upon observable behaviors elicited in a standardized setting, and which would be rated by trained observers. The possibility of maternal bias would thus be excluded from the primary data.

Second, the behavior of the infants would be recorded on videotape, and the ratings of temperament would be made from the videotape. The latter confers an enormous advantage in terms of preserving the exact behavioral style of the infant, and of making possible more accurate ratings, from repeated viewings if necessary.

In designing the laboratory protocol, the infant would be confronted with some typical age-related challenges, and the staff would employ a graded series of soothing techniques and diversionary play activities, as required. The principal questions would concern the infant's preponderant reactions to the challenges – whether frequently distres-

<sup>1</sup> Interrater reliabilities, based on 13 infants scored by two separate examiners, were as follows: irritability,  $r = 0.94$ ; resistance to soothing,  $r = 0.99$ ; reactivity,  $r = 0.94$ ; reinforcement value,  $r = 0.90$ ; activity awake,  $r = 0.79$ ; and activity asleep,  $r = 0.92$ .

sed or much more adaptable – and what level of soothing and/or diversion was required to placate the infant.

**Outline for Each Visit.** The pretesting evolved a basic outline for the temperament assessment, as follows: After a brief warm-up period, the mother left both twins with the staff while she was being interviewed. Subsequently, each twin was left alone with the staff while the alternate twin was taken for mental testing. During these solo episodes, the staff engaged each twin in a prescribed set of activities, or vignettes, for a fixed period of time, so that there would be uniformity of treatment across all participants. The schedule was carefully organized to yield one hour of videotaping for each pair, and in a format that was exactly duplicated for all twins.

For illustration, two of the vignettes employed at 6 months are briefly described below. The complete set of vignettes may be found in Matheny and Wilson (1981), and a fuller description of the background of the laboratory assessment may be found in Wilson and Matheny [23].

### A. SMILE-LAUGH-PLAY

E attempts peek-a-boo, tickle, or vocalizes a noise so as to elicit a smile or laugh (ie, frolic play) from the infant who is seated on E's knees (2 min).

### B. INTERRUPTION OF FEEDING

The purpose of this vignette is to obtain a view of the infant's response to the interruption of feeding, and to consider how easily the infant can be soothed after interruption.

1. The infant is cradled in the feeding position and allowed to feed for 1 minute before the bottle is removed.
2. The bottle is removed and placed out of sight. If the infant does not become upset within 30 seconds, the bottle is returned. If the infant becomes upset, the following soothing procedures are initiated in order: (a) talking, (b) stroking, (c) placing infant prone on E's lap with continued stroking, (d) lifting infant to shoulder, with talking and stroking continued, (e) resumption of feeding.

### Behavioral Rating Scales

These 9-point scales have been designed for ratings of emotional tone, attention, activity, and orientation to staff. They were drawn initially from Bayley's Infant Behavior Record [1], then extensively anchored with behavioral descriptors through pretesting [see 9]. The scales are described below, with illustrations of the anchor points at each extreme.

- A) **Emotional Tone.** Refers to principal emotional state manifested during the rating period: (1) extremely upset, crying vigorously ... (9) excited or animated. This scale serves as the principal measure of emotional activity.
- B) **Attention.** This category refers to the degree to which an infant is alert to and maintains attention on objects and events: (1) unoccupied, non-focused, vacant staring ... (9) continued and persistent attention to the point of being "glued" to object or event.
- C) **Activity.** Refers to body motion with or without locomotion; may involve whole or part body movements: (1) stays quietly in one place, with practically no self-initiated movements ... (9) hyperactive; cannot be quieted for sedentary tests.

- D) **Orientation to Staff.** (1) Actively negativistic, struggling, strongly avoidant ... (9) very strongly oriented, demanding, possessive of interaction.

### Scoring

After the visit was completed, the raters worked from the videotapes and made the appropriate scale ratings for each successive 2-minute period of the visit. No rater scored the episodes for which she was the principal interactionist with the twin. The schedule was organized so that each vignette was presented in a separate 2-minute period. After the scoring was completed, the ratings were then summed across all periods to yield a summary score for each scale. One additional score was computed of the infant's reactions to physical measurements, which was sometimes upsetting; and these scores collectively defined the infant's temperament profile.

### Temperament Questionnaires

From the questionnaires available, the one by Carey and McDevitt [3] was selected because of its wide usage, available norms, and extensive research and clinical applications. The questionnaire included 95 items which were scorable on the nine NYLS temperament categories, as illustrated below.

- 1) **Activity Level** – motor activity during daily routines as well as motility during the sleep-wake cycles.
- 2) **Rhythmicity** – regularity of vegetative functions.
- 3) **Approach or Withdrawal** – initial positive or negative response to a new stimulus.
- 4) **Adaptability** – ease of transition to new or altered situations.
- 5) **Intensity of Reaction** – degree of response.
- 6) **Quality of Mood** – amount of positive or negative emotional behavior.
- 7) **Attention Span and Persistence** – degree to which focal interest is maintained or an activity is pursued in the face of obstacles.
- 8) **Distractibility** – effectiveness of extraneous stimuli to alter ongoing behavior.
- 9) **Threshold of Responsiveness** – level of stimulation required to evoke a response.

The questionnaires, one for each twin, were provided at the 6-month visit with instructions to have them completed at home and mailed back within a week. Experience to date indicated that about 80 to 90 percent of the questionnaires would be returned with no prompting. Each infant was thus represented by nine category scores from the questionnaire which would define the temperament profile, and which could be compared with the lab assessment and the neonatal scores.

### Infant Behavior Record (IBR)

A mental test was administered to each twin while the co-twin was involved in the solo episodes of the temperament assessment. The test setting was a more structured and demanding situation for the infant, and when the test was completed, the examiner made summary ratings of the infant's behavior on the Infant Behavior Record (Bayley 1969).

A recently-performed factor analysis of the IBR ratings [8] extracted three primary factors that appeared as early as six months. These three factors – task orientation, test affect-extraversion, and activity – captured the nuances of behavior in the test setting, and consequently the resultant factor scores could be combined with the other temperament scores to create a multivariate temperament profile for each infant.

## RESULTS

### Neonatal Assessment

The summary scores for the neonatal assessment, along with birth weight, were intercorrelated and the results are shown in Table 1. Also shown are the factor loadings for a principal-components (rotated) analysis performed on the six summary scores.

Irritability and resistance to soothing were moderately intercorrelated, and both entered strongly into Factor I. Both were negatively related to reinforcement value as perceived by the examiner, and the latter variable entered Factor I with a negative weight. Activity while awake was positively correlated with irritability and resistance to soothing, and became the fourth defining variable to load on Factor I.

In aggregate, these variables defined a fussy, active, difficult-to-soothe neonate. The correlations with birth weight and gestational age (not shown) also showed that these fussy neonates tended to be heavier and more mature at birth, while the premature infants were more passive and tractable. These weight relationships, however, were clearly superimposed upon a basic dimension of irritability/activity/fussiness in these neonates, and individual differences were evident in the earliest days of life.

Factor II was defined primarily by reactivity, which summarized the neonate's orienting behaviors and alertness. High-scoring neonates were more reactive to visual and auditory stimuli, and were rated as more reinforcing by the examiner. This factor appeared to relate more to a dimension of developmental status than to temperament, and may be associated with early mental development scores [15].

### Laboratory Ratings of Temperament

The intercorrelations among the laboratory ratings are shown in Table 2, along with factor loadings from a rotated principal-components analysis.

The results showed a strong relationship between emotional tone and attentiveness – the better-humored infant was more attentive to the stimuli presented in the vignettes – and these two variables defined Factor I.

Activity level, and period-to-period change in activity and emotional tone, made up the second factor, while reaction to physical measurement was the sole variable for Factor III. In the latter case, those infants that were upset by the restraints of physical measurement were not necessarily the ones that had been upset during the lab vignettes; and in general, the physical measurements were much less provocative for 6-month infants than for older infants. The main factor to emerge in the lab ratings represented a dimension of positive/attentive vs distressed/inattentive reactions, as revealed under the successive challenges of the lab vignettes.

### Temperament Questionnaire

The intercorrelations among the nine categories of temperament are shown in Table 3, along with the factor loadings from a rotated principal-components analysis.

The core variables for Factor I involved mood, approach/withdrawal, adaptability and distractibility; and recalling that high scores on these categories reflected more problematic behaviors, this factor detected those infants described by their mothers as irritable and negative in mood, avoidant in the face of new stimuli, slow to adapt, and easily distracted.

The second factor was anchored chiefly by activity, intensity, and threshold of

response, all of which were positively intercorrelated, and all of which were negatively correlated with the first-factor categories. The sign inversion is something of an artifact, due to the direction of scoring for the starred categories in Table 3, and in fact high activity and high intensity were associated with avoidant and slow-to-adapt behaviors in these 6-month-old infants.

The three anchor variables for Factor II all had moderate loadings on Factor I before rotation ( $-0.37$  to  $-0.61$ ), so the original behavioral relationships preserved a stronger linkage between the irritability cluster and activity/intensity than implied by the orthogonal rotation in Table 3. Collectively, these characteristics corresponded in broad terms to the description of the difficult infant advanced by Thomas et al [17].

### Infant Behavior Record

The three primary factors extracted from the Infant Behavior Record are displayed in Table 4, along with the highest-loading items for each factor, and their respective weights [from 8]. The factor weights were employed to generate factor scores for each infant in this study, as a measure of the infant's reaction to the demands of testing. The resultant factor scores were then held for coordination with the other measures of temperament.

### Predictive Relationships

There were certain common themes that emerged in the descriptions for all data sets, particularly for the first-factor scores, and the primary question was whether these descriptively similar characteristics were in fact displayed with some consistency across data sets. Specifically, did the neonatal assessment provide any predictive relationships with temperament as later appraised at 6 months?

As a first step, the correlations were computed between the three neonatal measures (including birth weight), and the factor scores from the temperament appraisal at 6 months. The results are presented in Table 5.

Turning first to the laboratory data, the results showed virtually no relationships between the neonatal measures and the laboratory ratings of temperament at 6 months. The anticipated linkage would have been between Neonate I and Lab I with a negative sign, but the obtained correlation was clearly too low to represent any significant association. No other correlation approached significance, with the single exception of birth weight and Lab Factor III.

The latter represented reactions to physical measurements, and the heavier infants at birth were more upset by and resistant to the physical measurements at 6 months. This relationship recalled the earlier link between birth weight and irritability in the neonatal assessment (cf Table 1); the heavier, more mature neonates were somewhat more irritable, active, and nonsoothable than the less mature neonates. Apparently that initial tendency persisted and was later provoked by the restraints necessary for measurement, while by contrast the premature infant remained more passive and compliant when restrained.

The absence of association between the neonatal scores and the lab ratings at 6 months was something of a disappointment, but on reflection it highlighted several methodological issues that are raised in the discussion section. It also prompted an analysis of individual vignettes in place of the overall combined score, as reported later.

### Questionnaire

When the neonatal measures were correlated with the questionnaire scores, a moderate

Table 1 - Intercorrelations and Factor Loadings for Neonatal Assessment Scores

Variables	Correlations							Factor loadings	
	1	2	3	4	5	6	7	I	II
	IRR	SOO	AWK	ASL	REA	RFV	BWT		
1. Irritability	-	0.67	0.33	0.07	-0.30	-0.65	0.31	0.77	-0.40
2. Resistance to soothing		-	0.49	0.02	-0.19	-0.53	0.31	0.86	
3. Activity awake			-	0.12	-0.06	-0.35	0.17	0.77	
4. Activity sleep				-	-0.08	0.03	-0.09		
5. Reactivity					-	0.39	0.00		0.91
6. Reinforcement value						-	-0.18	-0.66	0.53
7. Birth weight							-	ni	ni
% Variance explained								53.9	19.7

ni = not included in factor analysis. Factor loadings < 0.30 omitted.

Table 2 - Intercorrelations and Factor Loadings for Laboratory Ratings of Temperament

Variables	Correlations						Factor loadings		
	1	2	3	4	5	6	I	II	III
	EMO	EM:V	ACT	AC:V	ATT	RST			
1. Emotional tone	-	-0.21	0.06	0.18	0.71	0.16	0.92		
2. Emotional tone (Var)		-	0.32	0.09	-0.18	-0.02	-0.35	0.66	
3. Activity			-	0.22	0.02	-0.02		0.80	
4. Activity (Var)				-	-0.02	0.04		0.60	
5. Attentiveness					-	0.10	0.90		
6. Reaction to restraint (Physical measures)						-			0.98
% Variance explained							30.8	23.9	16.1

N = 124. Factor loadings < 0.30 omitted. (Var) = variability.

Table 3 - Intercorrelations and Factor Loadings for Temperament Categories from Infant Temperament Questionnaire

Categories	Correlations									Factor loadings			
	1	2	3	4	5	6	7	8	9	I	II	III	IV
	ACT	RHY	APP	ADP	INT	MOD	ATT	DIS	THR				
1. Activity	-	-0.17	-0.42	-0.19	0.51	-0.18	-0.03	-0.26	0.35				
2. Rhythmicity*		-	0.20	0.42	-0.20	0.16	-0.04	0.24	-0.10			0.81	0.75
3. Approach/ withdrawal*			-	0.55	-0.20	0.62	0.04	0.39	-0.14		0.77		
4. Adaptability*				-	-0.12	0.50	-0.09	0.23	0.04		0.56		0.67
5. Intensity					-	0.04	-0.23	-0.20	0.20			0.74	-0.36
6. Mood*						-	0.01	0.47	-0.22		0.91		
7. Attention/ persistence*							-	-0.02	-0.17				0.98
8. Distractibility*								-	-0.16		0.61		
9. Threshold of response									-			0.62	0.36
% Variance explained										32.5	17.3	11.8	10.2

N = 72. Factor loadings < 0.30 omitted.

\* High scores on scales so marked represent presumed problem behaviors, ie, nonrhythmic in vegetative functions, withdrawn and avoidant in new situations, slow to adapt, etc.

Table 4 - Factor Loadings for Selected Items from Bayley's Infant Behavior Record

Items	Factors		
	I	II	III
Object orientation	0.83		
Goal directedness	0.76		
Attention span	0.69		
Reactivity	0.66		
Social orientation		0.70	
Cooperativeness		0.65	
Fearfulness		-0.77	
Emotional tone		0.74	
Endurance		0.62	
Activity			0.82
Body motion			0.78
Energy			0.71
% Variance explained	25.8	11.5	8.5

Factor loadings < 0.30 omitted.  
 Results drawn from Matheny [8] based on N = 415.

Table 5 - Correlations Between Neonatal Predictors and 6-Month Outcome Measures

Neonatal Variables	Outcome measures									
	Lab I	Lab II	Lab III	QST I	QST II	QST III	QST IV	IBR I	IBR II	IBR III
Neonate I	-0.03	-0.14	-0.15	0.32*	0.13	0.02	0.19	0.25*	-0.17	0.10
Neonate II	0.00	0.04	0.10	0.06	0.17	-0.05	0.07	0.07	0.11	0.03
Birth weight	0.00	-0.02	-0.35**	0.16	0.01	0.01	-0.14	0.37**	0.19	-0.02

N = 76 for Lab and IBR correlations; N = 56 for Quest. correlations.

\* P < 0.05

\*\* P < 0.01

correlation was obtained between Neonate I and Questionnaire I ( $r = 0.32$ ,  $P < 0.05$ ). Those newborns who were rated by the examiner as more irritable and difficult to soothe in the earliest days of life were also reported by the mothers to be more negative and irritable and slower to adapt during the first 6 months. Thus, the temperamental characteristics detected in the neonate by a trained examiner were persistent enough to register and be perceived by the mother during early infancy. In spite of differences in observers and rating instruments, and in the face of the hectic circumstances of caring for a new infant at home, there still emerged a convergent view of the infant from these two sources.

It is notable that the mother's report showed a relationship to the neonatal assessment whereas the lab ratings did not, and some reasons for this discrepancy will be considered later. But the main theme was this: The attributes of temperament seen in the neonate persisted in significant fashion and led to corresponding maternal ratings of temperament during the first 6 months.

## IBR

The first IBR factor titled task orientation was significantly associated with Neonate I, and even more strongly with birth weight. The neonate who was heavier and more irritable/active at birth tended to be more task-oriented in the testing session at 6 months. The latter captured the qualities of attentiveness and responsiveness to the test items, so apparently the more mature and fussy neonate later displayed greater interest and involvement with the test materials.

Prematurity (via birth weight) seemed to be the principal link, and just as the premature infant was more passive and compliant when restrained for physical measurements, so also the less reactive infant at birth was less attentive and responsive to the test materials at 6 months. This effect may be age-limited; the suppressive effects of prematurity are substantially offset by the second year for both growth and mental development [18,19], and a similar relation might be expected for the task-orientation behaviors measured on the IBR.

## Multivariate Analysis

While the univariate correlations in Table 5 are informative, each infant's temperament profile was defined by the collective vector of scores from all data sets. The variations in profile became the distinguishing features of temperament for each infant, and the question may then be posed whether the multivariate profile from the neonatal assessment would predict the multivariate temperament profile at 6 months.

For this analysis, a canonical correlation analysis was performed to isolate the relationships between the neonatal scores and the 6-month temperament scores. The canonical analysis searched for the linear combination of scores in each set that would yield the largest correlation between sets [4]. The canonical variate loadings then identified which variables in each set contributed to the cross-linkage between sets.

The analysis was specified to include the four outcome variables that represented the principal factors from each separate aspect of the assessment: Lab Factor I, Questionnaire Factor I, physical measurements (Lab Factor III), and behavior during testing (IBR Factor I). These four measures furnished a condensed behavioral profile for each infant that accounted for the largest proportion of variance from the original data sets. The question was how successfully these scores would be predicted from the neonatal scores.

**Table 6 - Canonical Correlation Between Neonatal Predictors and Selected Outcome Variables**

Canonical Correlation	First Canonical Variate	
	Predictors	Canonical loadings
R = 0.64**	Neonate I	0.70
	Neonate II	
	Birth weight	0.72
	Outcome	
	Lab I	
	Lab III	-0.45
	Quest I	0.61
	IBR I	0.64

N = 56. Canonical loadings below 0.30 omitted.

\*\* P < 0.01

**Table 7 - Twin Correlations for Temperament Factor Scores**

Variable	Zygoty	
	MZ	DZ
Neonate Factor I	0.16	0.38*
Laboratory Factor I	0.32	0.18
Questionnaire Factor I	0.65***	0.77**
IBR Factor I	0.46*	0.45*

Median N of 20 pairs per group

\* P < 0.05; \*\* P < 0.01

The results of the canonical analysis are shown in Table 6. A canonical correlation of  $R = 0.64$  was obtained ( $P < 0.01$ ); and the canonical loadings identified the principal contributors in each set – Neonate I and birth weight for the predictor set, and Quest I and IBR I for the outcome set. Neonate I was primarily linked to Quest I, while birth weight was linked to IBR I, and these variables made independent contributions to the canonical relationship. In broad terms, neonatal measures did have a significant association with the multivariate temperament profile at 6 months, and the most challenging result was that a neonatal measure of temperament correlated with the mother's rating of temperament at 6 months.

The association between neonatal status and 6-month temperament was not overwhelming – the predictor set accounted for 12.9% of the variance in the outcome set – but in an area where any significant results have been difficult to come by, the evidence for some degree of continuity is encouraging. It might be anticipated that as the measures are further refined and perfected, an increasing degree of continuity may be found.

### Reanalysis of lab ratings

The absence of relationships involving lab factor scores prompted a reanalysis that focused on the individual vignettes. Initially, all ratings had been pooled to give a composite measure of emotional tone, but inevitably, this blurred some rather striking differences between vignettes in emotional tone provoked. Therefore, the emotional tone ratings for 10 vignettes were analyzed, both for possible relationships to the neonatal measures, and to the other temperament measures at 6 months. When the correlations were examined, two distinctive results were evident.

First, emotional tone ratings for the smile/laugh/play vignette were significantly correlated with the first-factor scores from the questionnaire ( $r = -0.37$ ,  $P < 0.01$ ), and the feeding vignettes were correlated at a lower level with Quest I ( $r = -0.16$ ). Taken together, these laboratory vignettes yielded a multiple correlation of  $R = 0.39$  ( $P < 0.01$ ) with Quest I.

The infant who did not respond positively to frolic play with the examiner was also reported by the mother to be irritable and avoidant at home. Apparently this particular vignette, with its strong emphasis on playful interaction between infant and caretaker, came the closest to resembling the daily experiences of the mother with this infant. It is worth noting that it was the absence of a positive response where normally expected that furnished the key link between home and lab – the infant that could not be cajoled into good humor by playful interaction retained this sour disposition in both settings.

Second, the smile/laugh/play vignette at 6 months was correlated with the first-factor score from the neonatal assessment ( $r = 0.19$ ,  $P < 0.10$ ). Again, the more irritable and difficult-to-soothe neonate tended to be more unresponsive or upset by playful interaction at 6 months. While the relationship was marginal, it did suggest an emerging link between neonatal behavior and some aspects of the lab assessment.

### Twin Concordance

As a final analysis, the patterns of concordance were examined for MZ twins and DZ twins on the primary measures of temperament. Intraclass correlations were computed for the first-factor scores in each data set, to determine if MZ twins were more concordant than DZ twins. The results are shown in Table 7.

The concordance values were relatively modest for the neonatal scores and the lab scores, with MZ twins being less concordant at birth, but then becoming somewhat more concordant than DZ twins for behavior in the lab. The concordance levels increased markedly for both groups on the questionnaire, and to a somewhat lesser extent for behaviors rated on the IBR.

The high correlations for the questionnaire are consistent with an earlier preliminary report of  $R_{MZ} = 0.74$  for 12 MZ pairs; however, no DZ pairs were included in that report [21]. The present high DZ correlation makes it evident that both twins in a pair were perceived as moderately similar in temperament by the mother, regardless of zygosity. Some of the similarity may reflect continuing perinatal influences that affected irritability and soothability for both twins, and some may reflect the mother's tendency to rate both twins as alike. However, the latter clearly was not the whole story – the IBR correlations were virtually identical for both zygosity groups, but in this case each twin was rated by a separate examiner, so there was no confounding effect. It may also be noted that Matheny [8] reported significant MZ/DZ differences on IBR Factor I for a much larger sample, so the present nonsignificant results may represent a Type II error arising from sampling variation.

Overall, there was no pattern of differential concordance favoring MZ twins for the temperament measures at this early age. Evidently the impact of perinatal and early-infancy experiences obscured any distinctive genetic effects that might have been detected by within-pair comparisons. Recent results have shown that genetic effects did become strongly evident in the second year [24]; and it is also a fact that early mental development showed no MZ/DZ differences in the first year, but then became powerfully differentiated by zygosity in later years [18,22]. The next step is to investigate possible genetic effects on continuity and change in temperament as the key to understanding when MZ/DZ differences should become apparent.

## DISCUSSION

In broad view, these results furnished a glimpse of significant continuity in the rudiments of temperament that spanned the turbulent phases of early infancy. They hinted at an emergent behavioral core that was gradually pulled into focus as development proceeded.

The linkage from newborn to 6 months was modest and the unexplained variations were still large, for two principal reasons – first, maturation and vegetative processes had a profound impact on behavior in this early period; and second, the test procedures and the key behaviors to be measured are still being refined. Ultimately, when the entire sample has reached two years of age and the temperament patterns have stabilized, the two-year ratings can serve as the criterion against which the neonatal measures can be sharpened.

In considering the low-order relationship between the lab ratings and questionnaires at 6 months, which was markedly at variance with the convergent results obtained at 12 through 24 months [10], the videotapes and staff notes were reviewed for possible clues to this discrepancy. It became apparent that one hindrance to the laboratory assessment at 6 months was the pronounced effect of vegetative state.

Some infants at this age were still markedly affected by disruptions in schedule for feeding and sleep, and when they were brought in for a morning visit, the necessary travel time plus the duration of the lab assessment inevitably upset the schedule. These

schedule-bound infants often lapsed into bouts of fussiness arising from fatigue or temporary delays in feeding, although they were not necessarily fussy under the more ordered schedules of home. Teething and occasional bouts of colic also took their toll.

The end result was that vegetative state sometimes played a prominent role in the infant's reaction to the lab vignettes, to the point of obscuring the more typical reactions evident in daily routines. Many parents reported that the limits of home life were set by the infant's schedule for feeding and sleeping, and the infant remained benign as long as the limits were observed. But if the schedule was transgressed, there was a period of upset and distress before calm was restored.

This factor is the likely reason that the questionnaire results correlated with the neonatal assessment, whereas the lab ratings did not. The neonatal assessment was specifically tailored to the infant's schedule, extending from one feeding period to the next. Aspects of irritability, activity, and responsiveness to soothing were evaluated at fixed points in the metabolic cycle, so that vegetative state did not intrude as an uncontrolled influence. The neonate's temperament as thus revealed coincided more closely to the aspects of temperament observed by the mother at home.

### Temperament and Development

The results may be addressed to several questions of general interest. The first was whether irritability and resistance to soothing in the neonate would correlate with later measures of temperament; and the answer is a guarded yes, recognizing that it was the questionnaire which carried the relationship. Follow-up data will reveal whether there are latent relationships that may become evident as the temperament patterns become more fully stabilized in the second year.

*Does prematurity mask individual differences in temperament?* It certainly has a muting effect on emotional tone – the more premature infants were more passive and less expressive of irritability, both as neonates and at 6 months. One incidental observation was that the premature infants, when seen at 6 months, were often closer to 3-month infants in their response to the vignettes. They were less engaged or captured by the activities presented in the vignettes, and in that sense less responsive.

*How early is a distinctive temperament profile evident for each infant?* In the narrow sense, the vector of scores at each age defined the infant's temperament profile; but the broader sweep of the question is whether the profile was maintained consistently from age to age. The answer at this point is that there was some evidence of continuity from birth to 6 months, but it was considerably weaker than the continuity during the second year [10].

Some of the reasons have been itemized above, and the next step is to try to identify those changes that represent systematic transitions in temperament vs those changes that represent error variance. The powerful maturational processes of infancy propel each infant along a distinctive developmental pathway, and while some degree of behavioral continuity must be expected, it is not the sole criterion of what is systematic in the behavioral-guidance systems. Ultimately, when the full longitudinal data are available for these infant twins, it will be possible to verify whether the behavioral pathways are traced in parallel by identical twins.

**Acknowledgments:** This research was supported in part by research grants from the Office of Human Development (90-C-922), the National Science Foundation (BNS 76-17315), and the John D. and Catherine T. MacArthur Foundation. We are grateful to the many coworkers who have contributed so

much to the program, including S. Nuss, P. Gefert, R. Arbegust, J. Lechleiter, K. Adkins, M. Hinkle, M. Riedesel, P. Murphy, and B. Moss; and to the participating hospitals that provided facilities for the neonatal assessments.

## REFERENCES

1. Bayley N (1969): Bayley Scales of Infant Development. New York: Psychological Corporation.
2. Birns B (1965): Individual differences in human neonates' responses to stimulation. *Child Dev* 36: 249-256.
3. Carey WB, McDevitt SC (1978): Revision of the Infant Temperament Questionnaire. *Pediatrics* 61: 735-739.
4. Cooley WW, Lohnes PR (1971): *Multivariate Data Analysis*. New York: John Wiley.
5. Goldsmith HH, Campos JJ (1982): Toward a theory of infant temperament. In RN Emde, RJ Harmon (eds.), *The Development of Attachment and Affiliative Systems*. New York: Plenum.
6. Humbert NC, Wachs TD, Peters-Martin P, Gandour MJ (1982): The study of early temperament: Measurement and conceptual issues. *Child Dev* 53: 571-600.
7. Kurtzberg D, Vaughan HG, Daum C, Grellong BA, Albin S, Rotkin L (1979): Neurobehavioral performance of low-birthweight infants at 40 weeks conceptional age: Comparison with normal full term infants. *Dev Med Child Neurol* 21: 590-607.
8. Matheny AP, Jr (1980): Bayley's Infant Behavior Record: Behavioral components and twin analyses. *Child Dev* 51: 1157-1167.
9. Matheny AP, Wilson RS (1981): Developmental tasks and rating scales for the laboratory assessment of infant temperament. *JSAS Catalog of Selected Documents in Psychology* 11: 81-82 (Ms. No. 2367).
10. Matheny AP, Jr, Wilson RS, Nuss SM (1984): Toddler temperament: Stability across settings and over ages. *Child Dev* 55:1200-1211.
11. Reiss AJ (1961): *Occupations and Social Status*. New York: Free Press of Glencoe.
12. Riese ML (1982): Procedures and norms for assessing behavioral patterns in full-term and stable pre-term neonates. *JSAS Catalog of Selected Documents in Psychology*, 12: 6 (Ms. No. 2415).
13. Riese ML (1983): Assessment of behavioral patterns in neonates. *Infant Behav Dev* 6: 241-246.
14. Riese ML (1983): Behavioral patterns in full-term and preterm twins. *Acta Genet Med Gemellol* 32: 209-220.
15. Riese ML, Matheny AP (1982): Behaviors of the neonate and cognitive performance of the infant. *Infant Behav Dev* 5: 193 (Abstract).
16. Thomas A, Chess S (1977): *Temperament and Development*. New York: Brunner/Mazel.
17. Thomas A, Chess S, Birch HG, Hertzog ME, Korn S (1963): *Behavioral Individuality in Early Childhood*. New York: New York University Press.
18. Wilson RS (1978): Synchronies in mental development: An epigenetic perspective. *Science*, 202: 939-948.
19. Wilson RS (1979): Twin growth: Initial deficit, recovery, and trends in concordance from birth to nine years. *Ann Hum Biol* 6: 205-220.
20. Wilson RS (1980): Bloodtyping and twin zygosity: Reassessment and extension. *Acta Genet Med Gemellol* 29, 103-120.
21. Wilson RS (1982): Intrinsic determinants of temperament. In R Porter, GM Collins (eds): *Temperamental Differences in Infants and Young Children*. Ciba Foundation Symposium 89: Pitman (London).
22. Wilson RS (1983): The Louisville Twin Study: Developmental synchronies in behavior. *Child Dev* 54: 298-316.
23. Wilson RS, Matheny AP, Jr (1983): Assessment of temperament in infant twins. *Dev Psychol* 19: 172-183.
24. Wilson RS, Matheny AP, Jr (1984): Behavior genetics research in infant temperament. Paper presented at Fifth Occasional Temperament Conference: Keystone, Colorado.

**Correspondence:** M.L. Riese, Ph.D., Child Development Unit, Health Sciences Center, University of Louisville, KY 40292, USA.