Małgorzata Augustyniak[1] , Dorota Mrożek-Budzyn[1,2], Agnieszka Kiełtyka[1], Renata Majewska[1]

STABILITY OF THE MENTAL AND MOTOR BAYLEY SCALES OF INFANT DEVELOPMENT (2ND ED.) IN INFANTS OVER FIRST THREE YEARS OF LIFE

[1]Chair of Epidemiology and Preventive Medicine, Jagiellonian University Medical College
[2]Regional Sanitary-Epidemiological Station in Myślenice

ABSTRACT

Despite the Bayley Scales of Infant Development second edition (BSID-II) are wide used both clinically and in research settings, only a few published studies have been examined their stability over time.
AIM OF STUDY. The aim of this study was to examine the stability of mental and motor BSID-II scores over the first three years of life.
MATERIAL AND METHODS. All children included in this study were a sample followed up in a study on the susceptibility of the fetus and child to environmental factors. The cohort recruited prenatally in Krakow Poland, included 408 children. The mental and motor scales of BSID-II were administered to infants at the end of 12th, 24th and 36th month of life. Stability of the test scores from first to second and third assessment was evaluated using the Pearson's correlation coefficient calculated for the entire group of infants, and for the each gender separately.
RESULTS. The older infants obtained the better outcomes in BSID-II. The correlation between the first and second assessments for the Psychomotor Developmental Index (PDI) was r = 0.30, for the Mental Developmental Index (MDI) was r = 0.33, and between the second and third assessments the correlation for the PDI was r = 0.40, for the MDI was r = 0.59. That results suggest a low or moderate degree of relationship between the tests outcomes. Only 8,8% to 34,8% of variance in the infants later BSID-II scores could be explained by their earlier scores. The correlation between the tests scores was higher between outcomes obtained by girls than boys.
CONCLUSION. The BSID-II should not be treated as a useful measure for predictive purposes of infants development.

Key words: *child development, Bayley Scales of Infant Development, outcomes stability*

ABBREVIATIONS: BSID-II – Bayley Scales of Infant Development 2nd ed.
MDI - Mental Developmental Index, PDI - Psychomotor Developmental Index

INTRODUCTION

Bayley Scales of Infant Development second edition (BSID-II) have been considered the criterion standard for the developmental assessment of infants and subsequent diagnosis of cognitive or motor delays. BSID-II have been used as a gold standard to evaluate the other tools of assessment of infants development (1). Bayley evaluating BSID-II has established test-retest stability in the standardization sample: 1, 12, 24, and 36 months. The interval between the two tests ranged from 1 to 16 days. The stability coefficients for both tests were high and confirmed the high reliability of BSID-II (2). Despite its wide use both clinically and in research settings, only a few studies were published that have examined the stability of BSID-II outcomes over time. These studies which have been mostly related to high risk infants with multiple medical conditions have revealed low or moderate stability of BSID-II scores over time. The evaluation a test-retest stability of BSID-II using longer intervals measured the predictive value of the early tests for the future outcomes (3). Ideally, studies that examine stability of test scores typically should use heterogeneous groups. There is important to establish utility of the single BSID-II outcomes obtained in certain time of infant life for the development outcomes in future, both in the high risk and the low risk infants (4).

The aim of this study was to examine the stability of BSID-II scores during the first three years of life in infants from general population, included mostly the

low risk infants, and to establish clinical utility of the single BSID-II outcomes for prediction of children development in future.

MATERIAL AND METHODS

All children included in this study were a sample followed up in a collaborative study with Columbia University in New York on the susceptibility of the fetus and child to environmental factors. The cohort recruited prenatally in Krakow Poland, included the children of nonsmoking women aged 18 to 35 years, with singleton pregnancies, who had lived in Krakow for at least one year prior pregnancy. Mothers were free from chronic diseases, HIV infection and illicit drug use. The study population included 408 children. Despite BSID-II comprises three scales, only the Mental and Motor Scales were administered in the present study in 12th, 24th and 36th month of life (within 4 weeks of the target age). The Psychomotor Scale assesses control of gross and fine muscle groups (rolling, crawling, creeping, sitting, standing, walking, running, and jumping). The Mental Scale includes items that assess memory, habituation, problem solving, early number concepts, generalization, classification, vocalization, language, and social skills. Test scores are adjusted for the age of the child to obtain the Psychomotor Development Index (PDI) and the Mental Development Index (MDI). Test results are in one of four categories: 1) accelerated performance (score > 115), 2) within normal limits (score, 85 to 114), 3) mildly delayed performance (score, 70 to 84), and 4) significantly delayed (score < 69) (2).

The changing over time the MDI and the PDI scores from first to second and third assessment was evaluated using the Pearson's correlation coefficient calculated for the entire group of infants, and for the each gender separately.

RESULTS

The sample of 408 study participants included mostly the low risk infants with characteristic typical for children recruited from general population (tab. I). Girls had a significantly higher the MDI scores than boys in all age groups. The PDI scores at the age of 24th and 36th months was also higher in girls. The older infants obtained the better outcomes in BSID-II (tab. II). On the other side with increasing age there were the higher number of infants who were no able to perform the test in a reason of a weak cooperation (tab. III). The obtained outcomes divided infants into three groups. The most of infants were included within normal limit group, less to accelerated performance, and a few percent to

Table I. Characteristic of the study group

Characteristic		Number	%
Gender	Boys	206	50.5
	Girls	202	49.5
Parity	1	264	64.7
	2+	144	35.3
Weeks of pregnancy	< 37 weeks	16	3.9
	37-42 weeks	391	95.8
	> 42 weeks	1	0.2
Birth weight	< 2500	11	2.7
	□ 2500	397	97.3
Mothers age	18 - 24	79	19.4
	25 - 29	203	49.8
	30 - 34	126	30.9
Mothers education	primary or vocational school	39	9.6
	technical college	47	11.5
	high school or college	106	26.0
	university	216	52.9

Table II. BSID-II outcomes in 12th, 24th and 36th month of life

Index	Total		Boys		Girls		p
	Mean	SD	Mean	SD	Mean	SD	
MDI12	101.4	10.02	100.3	10.42	102.5	9.48	0.042
PDI12	97.3	11.41	97.2	10.61	97.5	12.19	0.890
MDI24	102.2	12.68	99.0	11.61	105.4	12.93	<0.001
PDI24	99.5	9.59	97.8	9.14	101.3	9.75	<0.001
MDI36	103.6	10.21	101.6	10.01	105.7	10.02	<0.001
PDI36	104.4	10.52	101.9	10.32	107.0	10.11	<0.001

Table III. Category distribution of BSID-II outcomes

The developmental category		Boys		Girls		Total	
		N	%	N	%	N	%
MDI12	accelerated performance	18	8.7%	23	11.4%	41	10.0%
	within normal limits	172	83.5%	168	83.2%	340	83.3%
	mildly delayed	16	7.8%	11	5.4%	27	6.7%
PDI12	accelerated performance	10	4.9%	14	6.9%	24	5.9%
	within normal limits	183	88.8%	164	81.2%	347	85.0%
	mildly delayed	13	6.3%	24	11.9%	37	9.1%
MDI24	accelerated performance	23	11.2%	64	31.7%	87	21.3%
	within normal limits	163	79.1%	125	61.9%	288	70.6%
	mildly delayed	20	9.7%	13	6.4%	33	8.1%
PDI24	accelerated performance	7	3.4%	18	8.9%	25	6.1%
	within normal limits	186	90.3%	175	86.6%	361	88.5%
	mildly delayed	13	6.3%	9	4.5%	22	5.4%
MDI36	accelerated performance	20	9.7%	40	19.8%	60	14.7%
	within normal limits	174	84.5%	158	78.2%	332	81.4%
	mildly delayed	12	5.8%	4	2.0%	16	3.9%
PDI36	accelerated performance	23	11.2%	44	21.8%	67	16.4%
	within normal limits	176	85.4%	155	76.7%	331	81.1%
	mildly delayed	7	3.4%	3	1.5%	10	2.4%

mildly delayed (tab. III). The correlation between the first and second assessments for the PDI was r = 0.30, for the MDI was r = 0.33, and between the second and third assessments the correlation for the PDI was r = 0.40, for the MDI was r = 0.59. The MDI scores had the stronger correlation over time than the PDI score. The stability of the tests scores over time was higher for the outcomes obtained by older infants. That results suggest a low or moderate degree of relationship between the tests outcomes. Only 8,8% to 34,8% of variance in the infants later BSID-II scores could be explained by their earlier scores (tab. IV). The correlation between the tests scores was higher between outcomes obtained by girls than boys (tab. V).

Table IV. Correlation coefficients between BSID-II outcomes

Index outcomes	MDI12	MDI24	MDI36
MDI12	1	0.329*	0.241*
MDI24		1	0.590*
MDI36			1
	PDI12	PDI24	PDI36
PDI12	1	0.298*	0.161*
PDI24		1	0.396*
PDI36			1

*Correlation coefficients significant bilaterally p<0.001

Table V. Correlation coefficients between BSID-II outcomes in boys and girls

Index outcomes	Boys			Girls		
	MDI12	MDI24	MDI36	MDI12	MDI24	MDI36
MDI12	1	0.196**	0.132	1	0.434**	0.329**
MDI24		1	0.583**		1	0.557**
	PDI12	PDI24	PDI36	PDI12	PDI24	PDI36
PDI12	1	0.248***	0.136	1	0.344***	0.188*
PDI24		1	0.320**		1	0.417**

**Correlation coefficients significant bilaterally p<0.001
*Correlation coefficients significant bilaterally p<0.01

DISCUSSION

The evaluation of children's developmental progress is an important part of routine pediatric care. The methods recommended as a reference tool in assessing infants development are the BSID-II and BSID-III (5). Despite the high reliability and validity of BSID-II that have been established in the U.S. there is still important to evaluate a stability of the tests outcomes over time in different groups of infants. The low risk infants and the high risk infants with multiple medical conditions characterize different pattern of development. The differential pattern of results over time for BSID-II when used with medically fragile infants could be interpreted as evidence for questioning the validity of tests, but it is more likely a result of group differences. Despite matching samples on primary diagnosis, age of the first assessment, age at the second assessment, gender, and geographic region in which the tests were administered, groups may have differed on other factors, like psychosocial and environmental factors which may be associated with differential developmental outcomes (6). The same factors influence children development in general population from which we derived infants for our study. Our study has confirmed the low or moderate stability of the BSID-II outcomes. The obtained correlation coefficients between the PDI and the MDI scores over time can be compared only to the results of a few previous studies which had the similar study design. There is necessary to perform the BSID-II tests in infants at the same age with the same time interval between re-assessments to have a possibility to compare the results. The most of previous studies started to perform the initial tests in younger infants than in our study but finished the assessment earlier, nearly at the end of the second year of life. The correlation between the first and second assessments for the PDI (r = 0.30), and for the MDI (r = 0.33) in our group of infants was somewhat weaker than in other studies, which assessed more homogenous groups compared to our infants (3,7) In the other studies participants were divided into groups of low risk and high risk infants and among the second group infants were matched to samples with primary diagnosis. In homogenous samples of infants the pattern of development was more stable than in heterogeneous group (8). We had many information about infants giving the possibility of division our group into the low risk and the high risk sample but there were too small number of infants who could be included to the high risk sample, than we assessed the stability of the BSID-II for entire group without any exclusions. The heterogeneity of our group of infants is probably the reason of the lower stability of the BSID-II compared to the results of the other studies (7,8,9). Clearly, the length of the interim period was closely related to the strength of the correlation: the longer the interval, the lower the correlation. Furthermore, independently of the length of the interim period, the correlation coefficients were higher for older infants and for the MDI comparing to the PDI scores. There is no information in previous studies about stability of the BSID-II in gender groups. Our study revealed the higher stability of the BSID-II in girls. It confirmed the differential pattern of development between genders (5). The advantage of performing this study in older group of infants is possibility to evaluate the BSID-II stability in period which was missed in most previous study design (3,7). The greater stability of the BSID-II in older infants, the more difficulties in performing the tests for a reason of infants weak cooperation. The problems with coopera-

tion concerned mostly infants with lower BSID-II scores but happened to infants who performed the earlier tests very well, too. That problem did not occur in the youngest infants. The weak cooperation in a few percent of three years old infants decreases the clinical utility of the BSID-II in that group of age.

The advantage of our study comparing to the previous studies is a large group of infants derived not from specific population but from general population who were assessed in relatively older age. The most important advantage in comparison with previous studies is the using the same time interval between reassessments what allowed to establish a precise difference between stability of BSID-II in different age. In our study we evaluated the stability of BSID-II for boys and girls separately and we revealed a significant difference of the BSID-II stability depended on gender. Furthermore we have many information on additional maternal and environmental risk factors which may affect the BSID-II scores. For example in our study we established the parents educational level as a most significant factor that influenced the infants development.

Despite some limitations, our study confirmed the previous results that for infants who had completed the BSID-II the correlations between scores in first and second or third year of life indicated a moderate level of systematic change in children development. The developmental delays identified in the end of the first year with BSID-II may indicate a moderate probability of continuing delays, the same relationship concerns infants within normal limits who may change their classification group of the developmental level. Infants from general population characterize a trend of acceleration the development in the second and the third year of life.

While the BSID-II as a measure of infant development may be a valid and useful indicator of current functioning, scores for some infants may not be stable from first to second and third year of life. The instability in scores are due to the nature of infants development rather than deficiencies in the test (10). These findings have clinical implications and are directly relevant to assessment policies and practices in infant development programs. They confirm cautious interpretation of assessments conducted in the early infancy. The results from BSID-II assessments of infants in the first year of life should not be used for predictive purposes, and must be interpreted differently for individual infants, considering specific medical conditions and the other factors which can influence infants development (5).

CONCLUSION

The BSID-II should not be treated as a useful measure for predictive purposes of infants development.

Acknowledgements

This is a part of ongoing comparative longitudinal study on the health impact of prenatal exposure of infants and children to outdoor/indoor air pollution, which is being conducted in the New York City and Krakow. The study received funding from an RO1 grant entitled "Vulnerability of the Fetus/Infants to PAH, $PM_{2.5}$ and ETS" (RO1 ES10165 NIEHS; 02/01/00 - 01/31/04) and from NIEHS (5 RO1 ES010165-0451), the Lundin Foundation, and the Gladys T. and Roland Harriman Foundation. Principal Investigator: Prof. F.P. Perera

REFERENCES

1. Dietrich KN, Eskenazi B, Schantz S, Yolton K, Rauh VA, Johnson CB, Alkon A, Canfield RL, Pessah IN, Berman RF. Principles and practices of neurodevelopmental assessment in children: lessons learned from the Centers for Children's Environmental Health and Disease Prevention Research. Environ Health Perspect 2005;113(10):1437-46.
2. Bayley N. Bayley Scales of Infant Development, 2nd edition. Psychological Corporation, San Antonio, TX, 1993.
3. Niccols A, Latchman A. Stability of Bayley Mental Scale of infant development with high-risk infants. Br J Dev Disabil 2002;48:3-13.
4. Oberklaid F, Efron D. Developmental delay--identification and management. Aust Fam Physician 2005;34(9):739-42.
5. Pinto-Martin JA, Dunkle M, Earls M, Fliedner D, Landes C. Developmental stages of developmental screening: steps to implementation of a successful program. Am J Public Health 2005;95(11):1928-32.
6. Aylward GP. Methodological issues in outcome studies of at-risk infants. J Pediatr Psychol 2002;27(1):37-45.
7. Evensen KA, Skranes J, Brubakk AM, Vik T. Predictive value of early motor evaluation in preterm very low birth weight and term small for gestational age children. Early Hum Dev 2009;85(8):511-8.
8. Harris SR, Megens AM, Backman CL, Hayes VE. Stability of the Bayley II Scales of Infant Development in a sample of low-risk and high-risk infants. Dev Med Child Neurol 2005;47(12):820-3.
9. Ment LR, Vohr B, Allan W, Katz KH, Schneider KC, Westerveld M, Duncan CC, Makuch RW. Change in cognitive function over time in very low-birth-weight infants. JAMA 2003;289(6):705-11.
10. Aylward GP. Prediction of function from infancy to early childhood: implications for pediatric psychology. J Pediatr Psychol 2004;29(7):555-64.

Received: 12.02.2013
Accepted for publication: 7.05.2013

Address to correspondence:
Dorota Mrożek-Budzyn
Chair of Epidemiology and Preventive Medicine
Jagiellonian University Medical College
ul. Kopernika7a, 31-034 Krakow
tel/fax: 12 4231003/12 422 87 95
e-mail: dorota.mrozek-budzyn@uj.edu.pl