

Developmental Follow-up: Outcomes, Controversies, and Suggestions

Glen P. Aylward, Ph.D., ABPP
Southern Illinois University School of Medicine
Developmental-Behavioral Pediatrics/Psychology
Springfield, IL
gaylward@siumed.edu

Disclosure

Dr. Aylward is the author of the upcoming Bayley-4, and will have a financial relationship with the Publisher, Pearson. This version will not be discussed in this presentation.

He receives no compensation from the Publisher for the earlier Bayley-II or Bayley-III which will be discussed.

Topics to be Addressed

- General issues in developmental testing
- Factors that affect outcome/prediction
- Outcomes (IQ, Academic, VMI, Language, EF, ADHD, Emotional/Behavioral, ASD)
- Levels of Follow-up
- Controversies (correction, NDI, comparison groups, who to follow, etc.)
- Suggestions on how to address controversies

General Issues in Developmental Testing

- There is no 'gold standard'; there are 'reference standards'
- Flynn effect: .3-.5 pt. increase per year
- How much is too much? Balance between the conceptual and the pragmatic
- Neurologic→motor→sensorimotor→cognitive
- Canalized behavior

Caveats Regarding Outcomes

- Sample selection procedures/ cohort/exclusion/controls
- Age at assessment/outcome measures employed
- Social backgrounds
- Emphasis on birth weight vs. gestational age
- Factors: social risk, family capital, SES, gender, prenatal issues
- Loss to follow-up
- Changes in neonatal care
- BPD, NEC, IVH, PVL, ROP, SGA
- Changing 'yardsticks' with regard to outcome assessments

Gestational Age Groupings

- Extremely preterm (EPT) <28 weeks
- Very preterm (VPT) 28-31 weeks
- Moderate preterm (MPT) 32-33 weeks
- Late preterm (LPT) 34-36 weeks
- Early term (ET) 37-38 weeks
- Term 39-41 weeks
- Late term 42-44 weeks

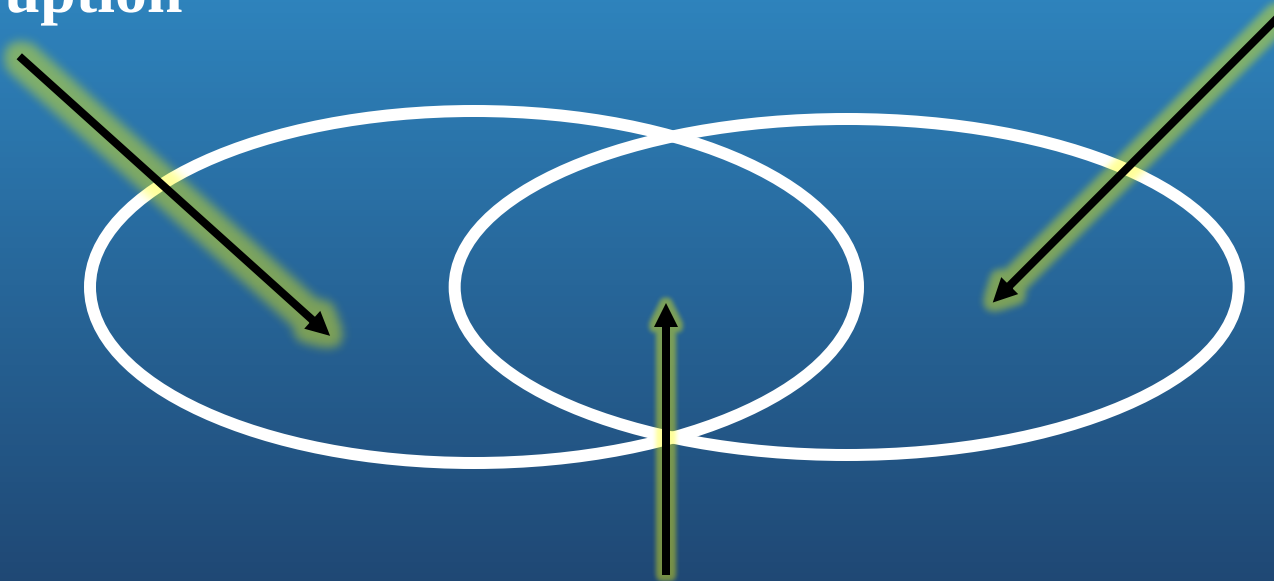
Prediction

- Disruption/Insult—reorganization/recovery
- Tests used
- Ages at testing
- Areas assessed
- Child's level of functioning
- Environment
- Effects of intervention

“Two-Hit” Hypothesis: CNS of Premature Infants

**Developmental
disruption**

Insult



Both

OUTCOMES

Major Handicaps/Severe Disability

Intellectual Disability

(moderate/severe)

Sensory deficits

(vision/hearing)

Cerebral Palsy

High Prevalence/Low Severity Dysfunctions

Learning Disorders

Low average/ Borderline IQ

ADHD

Neuropsychological deficits
(EF, visual-spatial)

Behavior Disorders/
Internalizing

IQ

- ELBW/VLBW (VPT/EPT) children have mean group IQs in borderline to average range, with low average being the mode (~ 10 pts; .66 SD)
- Smaller or younger the infant, lower the mean *group IQ—gradient*
- *However, .5-1.0 SD below normal wt/FT peers; higher rates of cognitive impairment*
- Effect exacerbated w/ addition of biomedical risks (BPD, abn'l ultrasound, NEC, severe ROP, sepsis)
- Does not occur in isolation; subtle, synergistic fashion to produce functional difficulties.
- Males < females (10 IQ pts)
- Each week g.a. <33; average decline in IQ 1.5-2.5 pts

Academic Achievement

- Direct relationship with gestational age
- Problems: math > written expression > spelling > reading (decoding, comprehension)
- 7.2-11.4 points < controls
- Grade retention: VPT-25-40%; MPT-20-30%; LPT-17%; FT-2.3-8%
- Many have non-verbal learning disabilities (NVLD)
- Heredity, gender, environment
- Despite broadly normal IQ/controlling for IQ → multiple LDs (1/3 > 1 LD)
- A positive environment (social advantage) can improve AA (and cognitive scores) over time.

Visual-Motor Outcomes

- Majority manifest some type of visual motor/visual perceptual problem
- Copying, perceptual matching/planning, spatial processing, finger tapping, pegboards, visual memory, spatial organization, visual-sequential memory
- Mean VMI scores 1.6 SD below classmates; 30% < 15th percentile
- FM problems 70+%; more overflow movements
- Higher percentage are left-handed (22-28% vs. 10%-12%)
- May contribute to problems with written expression

Visual-Motor

- Probability of glasses is 3x greater than FT's; ~50% ELBW/EPT
- Poorer legibility; slower handwriting speed
- Deficits in spatial judgment, concepts of orientation/directionality, perceptual deficits
- Impact on academics, particularly elementary grades
- Visual-motor problems typically do not occur in isolation
- Developmental coordination disorder (DCD); FM/GM 31-34% VPT & 50% EPT
- Transient dystonia of prematurity

Language Outcomes

- Many language functions are reasonably intact (e.g., vocabulary, receptive language, memory for prose, naming vocabulary)
- More complex verbal processes are deficient: understanding of syntax, semantics, abstract verbal skills, verb production, VWM, auditory discrimination, lang. processing, reasoning, understanding complex instructions, fluency
- articulation
- Deficits are subtle, but are critical in communication, social and academic endeavors
- More problems in those born <32 weeks (VPT), male

Executive Function

- EF: Coordination of interrelated processes; planning, goal directed behavior, self-regulation
- Instrumental in cognitive, behavioral, emotional, & social function
- Metacognition: initiate, working memory, shift/switching, planning/organizing, monitor
- VPT/EPT 2-3x more likely to have trouble: starting activities, problem-solving flexibility, STM, planning sequences of actions in advance, organizing information
- Preschool data: decreased WM capacity, cognitive inflexibility, maintenance of info on-line

Executive Function

- 2.1-3.5x more likely to have working memory impairment (rate 19%-41%) than FTs
- Impact on IQ, academics, fluid intelligence, social competence
- Related to white matter damage
- Processing speed: difficulty maintaining high level of efficiency when faced with increasingly complex tasks

ADHD

- Relative risk 2.7 < 34 weeks
- 2-3 fold increase in VPT/VLBW
- 4-fold increase in EPT/ELBW
- *NO* male predominance
- *NO* strong association with ODD/CD
- Weaker association with sociodemographic/ family risk in EPT/ELBW
- Inattention vs. hyperactivity/impulsivity
- Related to deficits in working memory/processing speed (SCT); “concentration deficit disorder” (Barkley,2016)

Psychiatric/Emotional Issues

- “*preterm behavioral phenotype*” (Johnson & Marlow, 2011)
- *Inattention*
- *Anxiety*
- *Social difficulties* (withdrawal, poor peer relationships, problematic social skills)
- Inverse incremental associations with BWT/GA
- 3-4 fold increase in risk for disorders in childhood
- Higher group scores even though often not in abnormal range
- Internalizing > externalizing
- Males > females

Autism Spectrum Disorder

- Inflated false positive rate for EPT/VPT infants who fail ASD screen b/c of other neurodevelopmental disabilities (ND)...not indicative of autistic traits per se;
- Different etiology for PT's than the general population → altered brain development produces behaviors that mimic ASD symptoms
- Nonetheless there appears to be some increase in ASD: **5%-8%** w/o NDI; **16-17%** with NDI
- 3 different ASD screens; 20% $\geq$ 1 screen positive; 1% all 3 (Stephens et al, 2012)
- Moore et al (2012): all w/ hearing/visual impairment had a + M-CHAT screen; 95.5% those unable to walk independently by 2-years;

ASD

- 23 fold increase in + M-CHAT if child could not sit or stand independently (Kuban, 2008)
- 7x increase if child needed assistance to walk
- 13x quad CP
- 8x greater with vision or hearing deficit; Moore et al (2012) *all* with hearing/visual impairment had + M-CHAT
- Not high functioning ASD or Asperger's
- Kaiser Permanente (2014); Hazard Ratios compared to FT's: 24-26 wk = **2.7**; 27-33 wk = **1.4**; 34-36 wk = **1.3**

Summary

- Spectrum of sequelae in non-handicapped children born prematurely does not differ drastically from the array of problems found in the general population
- However, there is a disproportionately greater incidence and complexity of these problems
- While both biomedical and environmental factors affect outcome, impact of biomedical factors increases as BWT/GA decreases

Summary

- Constellations of deficits vs. 1 area, this due to interrelated circuits
- Those born at later ga's (34-36 weeks) still do not do as well as FTs
- Establishing whether a variable is part of a causal chain can be difficult, particularly when the etiology of outcome is multifactorial (e.g., cognitive impairment)
- Outcomes due to interaction of medical/biologic, social, environmental and genetic factors

Bayley II vs. Bayley-III

- **MDI/PDI**→ Cognitive, Language (RC, EC), Motor (FM, GM) Adaptive (ABAS), Social-Emotional (Greenspan)
- **BSID/BSID-II**: MDI decreased by 12 pts; PDI by 7pts
- Bayley II/Bayley III: COG *increased* by 6-10 pts.; Motor *increased* by 8-18 pts.
 - Correlation: COG and MDI .60-.67
 - Correlation: LANG and MDI .71-.87
 - Correlation: Motor and PDI .60-.65
- Bayley III norms included 10% of at-risk infants and toddlers (PTs, DS, asphyxia, etc.); this would inflate norms
- if purpose is to identify children with DD, inclusion of those with risk of DD will affect diagnostic accuracy

Resultant Problems

- Longitudinal studies where the ‘yardstick’ changes: what happens to comparability of data?
- What are differences in scores due to: Intervention, test issues, or both?
- Impact on power issues: use BSID-II for research proposal, but Bayley-III data are obtained.
- Is the Bayley-II too conservative? Does not explain why some control populations using the Bayley-III score > 1 SD above the mean
- Underdiagnosis
- ‘Reverse Flynn effect’

Attempts at Solutions

- Combine Bayley-III cognitive and language and average the scores (Moore et. al, 2012). Still 7 pts higher than MDI, 8% 1-2 SD below average vs. 15% using MDI.
- Algorithm using cognitive & language scores. Slight improvement
- Least squares regression (Lowe et al, 2012); differences of 27 pts at low end, 7 pts at upper end
- Use of a developmental quotient score (Milne, 2012) Adds additional layers of imprecision.

36 mo. Cog RS=69; Composite 90; 28 mo; $28/36 = 8$ mo. delay; DQ=78

18 mo. Cog RS=48; Composite 90; 16 mo; $16/18 = 2$ mo delay; DQ=89

- Use different cut-offs (Vohr et al 2012): <80 vs <70; some say <85; 70-84 moderate NDI; Motor Composite <73 (Duncan et al. 2015)

Bayley: Going Forward

- Prospectively use normative FT comparison group
- Compare normative group to standardized norms; if different, use normative group scores for comparison
- No mixed norms for standardization; this inflates scores
- Bayley-II and Bayley-III reconciliations should probably stop
- Categorically, <80-85 could be considered significant impairment; separate out CP and neurosensory disabilities

Controversies

- How long to follow-up? Specific ages?
- Correction for prematurity (yes/no; how long)
- What should be used as outcome measures?
- How to handle Flynn effect?
- What should be considered as neurodevelopmental impairment (NDI)?
- Relationship of ASD to prematurity?
- Use test norms, comparison groups, or both?
- Which specific groups should be followed?

How Long to Follow-up? Specific Ages?

- 2-years, minimum
- 5-years, desirable
- 8-years, optimal
- The more subtle a problem is, the older the child must be to detect it.
- The older the child, the harder it is to draw associations between perinatal issues and outcomes because of other variables coming into play

Age at Testing and Prediction to School Age Function

18-24 months:

- Ability to predict school age function improves
- Cognitive and motor functions diverge
- Language and reasoning skills develop
- No IQ tests, only developmental
- Environmental influences; signal to noise ratio

3-4 years

- IQ, EF, pre-academic readiness, VMI, verbal/non-verbal
- Better prediction/closer in time
- Stronger environmental influences
- Some IQ tests have weak test floors at early ages
- Still difficult to predict high prevalence/low severity dysfunctions

To Correct or Not to Correct

- Wilson-Ching, et al (2014); theoretical model using Bayley-III cognitive scores
- Took baseline raw scores that yielded a score of 70, 85, 100; recalculated for 1, 2, 3, & 4 mo. prematurity, allowing comparison of corrected (C) and non-corrected (NC) scores
- e.g., 12 mo. Standard score of 100 (RS 40-41): 13 mo=95, 14 mo=90, 15 mo=85, 16 mo=80
- At 1 & 2 mo PT, difference C/NC was **5** pts at 24 mo and **0** points at 36 mo
- At 3 & 4 mo PT, difference at 24 mo. was **5-10** pts; at 36 mo, **5 pts**

To Correct or Not to Correct

Differences greater for:

- increasing degree of prematurity
- Younger age at assessment
- Higher baseline scores (e.g., 100 vs 85 or 70)
- Van Veen et al (2016): age 5- 0-15 pts:
- Greater differences with 1) lower GA's, 2) higher IQ scores,
- Correct to 40 weeks
- Correct for LPT/ET infants?

Selection of Outcome Measures

- Early:

Cognitive, motor, language

Adaptive

- Later:

Cognitive/Academic Achievement

Executive Function

VMI/Visual perception

Comorbidities

Adaptive (QoL)

Flynn Effect

- Do not use tests normed >15 years ago
- Update norms; minimize changes in test content because this limits comparability
- Use of a comparison group could minimize this issue because groups still could be compared on same measure
- Still has a negative impact on longitudinal data

What Should be Considered NDI?

- Cognitive, CP, sensory should not be combined
- Could have diplegia but normal cognitive, yet still be grouped as NDI—very heterogeneous
- **Moderate NDI:** 70-85 COG, MOT, *or* LANG; **Severe:** <70 on COG, MOT, *or* LANG (Chalak, 2014)
- NRN: **Moderate NDI:** COG/MOT 70-84
- NRN: **Severe COG:** 55-69; **Profound COG:** <55 (Vohr, 2014)
- CP and DCD
- Sensory Hearing deficits (permanent hearing loss, minimal understanding w/o amplification) more prevalent than vision impairment (bilateral blindness, corrected to < 20/200)

Relationship of ASD to Prematurity

- The revised M-CHAT may yield different results:
 - Any 3 positive findings → follow-up interview (M-CHAT/F)
 - Any 7 positive findings → full evaluation (>82% +)
- If M-CHAT & M-CHAT/F are positive
 - 54% had ASD; 98% had clinically significant developmental concerns (Chlebowski et al, 2013)
- Look at different GA groupings and consider other impairments
- < 29 wks. Age 2, 4; M-CHAT FI & ADOS (Pritchard et al, 2016)
 - 22% + screen; 1.8% had ASD
 - M-CHAT FI-> communication impairment

*Use Test Norms, Comparison
Groups or Both?*

BOTH!!

Who Should Be Followed?

- All VPT and EPT (ELBW)
- ? MPT (14.5% twins; 35.5% triplets)
- ? LPT (49.8% twins; 43.6% triplets)
- What grade IVH?
- Solution: different levels (comprehensiveness) of follow-up and different *frequencies* (how many times) of follow-up visits, depending on medical/biologic risk
- Purpose: research, clinical, both

Considerations: Who Should Be Followed

- Probability of sequelae; greater the likelihood of problems, follow-up more frequent, more detailed
- Need for control group
- Resources available
- Ability to switch tracks: depending on findings

Levels: 1. **Screen: telephone, electronic, mailing**

2. **Hands-on screen;(abbreviated Bayley COG/GMFCs; Bayley Screener)**

3. **Comprehensive developmental assessment (e.g. Bayley-III; Battelle)**

4. **Multidisciplinary team (D/B peds, psych., OT/PT, S/L)**

5. **Comprehensive cognitive/neuropsych. evaluation (DAS-II, KABC-2, WPPSI-IV, NEPSY-II)**

Examples

	6 mo.	12 mo.	24 mo.	36 mo.	48 mo.	60 mo.
EPT:22-23 wks.	2	2	3	4	5	5
26-27 wks.	2	2	3	4	-*	5
MPT 32-33 wks.	1	2	3	3/4*	5*	
LPT	1	1	2	3*		5*
HIE/cooling	1	2	3	3*		5*

* =testing depends on previous results

Tests

Developmental:

- **Griffiths** (1996;2016); 0-2yrs, 2-8 yrs: 5 scales
- **Mullen Scales** (1995); 0-68 mo., visual/language receptive/expressive
- **Battelle Dev. Inventory-2** (2005); 5 domains; B-7 y.
- **Bayley Scales of Infant and Toddler Development-III** (2006) 16d-42 mo**

Intelligence:

- **Differential Ability Scales-II** (2007) 2.6-3.5; 4 subtests**
- **Stanford-Binet 5 for Early Childhood** (2003) 2.0-5; 5 scales
- **WPPSI-IV** (2012) 2.6-7.7; FSIQ, PIQ, VIQ
- **K-ABC-2** (2004) 3-18; Sequential/Simultaneous, MPI**
- **NEPSY-II** (2007) 3-4; 5-16 (neuropsych)**
- **CANTAB**

References

- Aylward, G.P. Continuing issues with the Bayley-III: Where to go from here. *J Dev Behav Pediatr.* 2013;34: 697-701.
- Aylward, G.P., Aylward, B.S. The changing yardstick in measurement of cognitive abilities in infancy. *J Dev Behav Pediatr.* 2011;32:465-468.
- Aylward, G.P. Developmental screening and assessment: What are we thinking? *J Dev Behav Pediatr.* 2009; 30:169-173.
- Vohr B, Wright LL, Hack M, Aylward G, Hirtz D. Follow-up care of high-risk infants. *Pediatrics.* 2004; 114:1377-1397 (Supp).
- Vohr, B.R. Neurodevelopmental outcomes of extremely preterm infants. *Clin Perinatol.* 2014; 41:241-255
- Vohr, B.R. Long-term outcomes of moderately preterm, late preterm, and early term infants. *Clin Perinatol.* 2013, 40: 739-751.

References

- Anderson, P.J. Neuropsychological outcomes of children born very preterm. *Sem Fetal Neonat Med.* 2014;19: 90-96.
- Wilson-Ching, M., Pascoe, L., Doyle, L. Anderson, P.J. Effects of correcting for prematurity on cognitive test scores in childhood. *J Paediatr Child Health.* 2014; 50:182-188.
- van Veen S., Aarnoudse-Moens, C.S.H., et al. Consequences of correcting intelligence quotient for prematurity at age 5 years. *J. Pediatr.* 2016;173:90-95.
- Aylward, G.P. Neurodevelopmental outcomes of infants born prematurely. *J.Dev Behav Pediatr.* 2005;26:427-440.
- Joseph R.M., O'Shea, T.M. et al. Neurocognitive and academic outcomes at age 10 years of extremely preterm newborns. *Pediatrics.* 2016;137:1-9.
- Hintz, S.R., Newman J.E., Vohr. B.R.C. Changing definition of long-term follow-up: Should “long-term” be even longer? *Sem Perinatol.* 2016;40:398-409.