Semin Neonatol 2001; **6:** 393–401
doi:10.1053/siny.2001.0060, available online at http://www.idealibrary.com on IDEAL®

Randomized clinical trials in infancy: methodologic issues

Lynn T. Singer

Departments of Pediatrics and Psychiatry, Rainbow Babies and Childrens Hospital, Case Western Reserve University School of Medicine, Cleveland, OH, USA

Randomized clinical trials of early nutritional interventions in infants are necessary to establish safety and efficacy of supplementation of infant formulas with LCPUFAS for term and preterm infants. Such trials pose significant methodologic difficulties when applied to infants because of the rapidly changing development of the infant's central nervous system and its interdependence with multiple environmental factors. Current assessments of infant cognitive development in the first year of life lack stability and predictive relationships to later outcomes. Thus, intervention trials need to extend beyond the first two years of life. Small sample sizes, high attrition rates, and lack of attention to confounding variables known to be related to child outcomes are additional problems in the majority of studies to date, especially for preterm populations. Attention to selection and attrition biases and the roles of mediating and moderating factors in affecting intervention effects are also necessary to determine the benefits vs risks of LCPUFA supplementation of infant formula.

Key words: infancy, development, LCPUFAS, nutrition, randomized clinical trial, methodology, assessment, Bayley Scales

The issue of whether long-chain polyunsaturated fatty acids (LCPUFAS) should be added to infant formula is currently one of the most controversial in infant nutrition. In particular, two of these fatty acids, arachidonic acid (AA) and docasohexaenoic acid (DHA), are incorporated into the central nervous system during development and are thought to be particularly important for brain and retinal development [1]. Because these two essential nutrients are absent from manufactured infant formulas, but abundant in breast milk, it is plausible to consider that formula-fed infants might be cognitively disadvantaged from 'deficiencies' of AA and DHA. Indeed, many studies indicate that breast-milk-fed infants are apparently advantaged in cognitive outcomes in comparison to formula fed infants [2].

Preterm infants have been of particular concern, since the accumulation of DHA in the brain and retina occurs most rapidly in the perinatal period [3] and the preterm infant's supply of AA and DHA from the mother is limited by parturition. Due to these concerns, numerous research studies have been conducted over the past decade to attempt to discern whether early fortification of infant formula with AA and DHA leads to significant, long-term growth, visual and intellectual benefits to preterm and/or term infants. To date, studies have been inconclusive, but because AA and DHA supplementation now occurs routinely in manufacture of European and Asian formulas [4], and because there is also significant pressure to add these nutrients to formula in North American markets, determination of the safety and efficacy of these supplements has acquired some urgency. For a detailed synthesis and review of relevant research, the reader is referred to a volume by Dobbing [1].

Coincident with these controversies, the application of the Randomized Clinical Trial (RCT) to assess the effects of an intervention in infancy has become the standard, and the preferred method of

Correspondence to: Lynn T. Singer, Ph.D., Case Western Reserve University, School of Medicine, The Triangle Building, 11400 Euclid Avenue, Suite 250-A, Cleveland, OH 44106, USA. Tel.: +1 216 844 6212; Fax: +1 216 844 6233; E-mail: LXS5@PO.CWRU.EDU

understanding the impact of a treatment on developmental outcome. Such trials have been widely used in adult studies to establish efficacy of treatment interventions, but have only recently been applied to studies of infants on a large scale, with the multi-site Infant Health and Development Program for low birthweight infants considered seminal [5]. The establishment of the NICHHD Neonatal Intensive Care Unit Network in the 1980s [6] similarly has initiated numerous RCTs to increase evidence-based clinical practice, including such investigations as the use of dexamethasone in ventilator-dependent preterm infants and Vitamin A supplementation for extremely-low-birthweight infants. The value of the RCT in establishing treatment efficacy in infants has thus been clearly established, along with the complexity of methodologic issues inherent in their conduct [7]. Because of an explosion of knowledge recently in scientific areas relevant to infant development, RCTs initiated during the infancy period to assess the effects of adding LCPUFAS to infant formulas can benefit from knowledge obtained from decades of research on infant development. These broad areas which inform current studies include research on continuity and discontinuity of typical infant development [8], early educational and therapeutic interventions for at-risk infants [5], and behavioral teratology studies of fetal alcohol, drug, and environmental contaminant exposure [9].

The RCT seeks to establish the safety and efficacy of an intervention and must document a reliable (statistically significant) association between the intervention and outcome, with causality inferred through this relationship as well as through the time order, (i.e. the causal factor must precede the effect), and through demonstration that the relationship of the intervention to the outcome is not due to some other factor [10]. To demonstrate such a relationship generally requires appropriately large and representative cohorts for study, accurate information about all factors important to the intervention model, high cohort retention, and measurements of the intervention and outcome variables which are accurate, valid, and consistent with the intervention model. While the application of RCTs to evaluation of infant formula supplementation may appear simple, in fact, achievement of experimental design requirements in infancy studies is particularly difficult. This difficulty stems from a number of issues inherent in infant developmental research, including the lack of predictive relationships between infant behaviors and later cognitive functioning; the absence of valid, reliable measurements of infant cognitive processes; the rapid, discontinuous developmental shifts occurring in infancy; and the powerful influence of numerous extraneous competing factors on developmental outcome which must be controlled. These difficulties are exacerbated in studies of preterm infants, for whom lipid supplementation is of special concern. In this paper, I will outline selected methodologic issues important to consider in conducting and evaluating studies of dietary supplementation on developmental outcomes in infants.

Neurobiological advances have informed our understanding of the long, continuous, and orderly sequence of brain development, a process which is highly dependent on specific genetic factors and environmental supports for normal functioning [11]. Nutritional factors have been demonstrated to be among the most important factors, as considered by the increased occurrence of neural tube disorders with folic acid deficiency in the early period of prenatal development [12]. The sequence of central nervous system development is precise, complex, and extensive, occurring from days after conception until years after birth. During the third trimester, cell differentiation predominates, including cell death, synaptogenesis and myelination, with significant cortical development and pruning continuing postnatally, all dependent on environmental, including nutritional factors [13]. Thus, deficiencies in essential fatty acids may detrimentally affect important neurobiologic processes, especially for preterm infants, impacting on central nervous system development. Such deficiencies may have variable effects on later intellectual development dependent on their timing, severity, and duration as well as their interaction with other environmental and genetic factors [14,15]. This rapidly changing brain organization, strongly interdependent with the infant's caregiving environment as predicted by transactional models makes it difficult to infer effects from an early nutritional intervention, given the known plasticity of brain organization in infancy.

Measurement instruments

The use of infant intellectual outcomes as dependent measures in evaluating the efficacy of early nutritional interventions is a relatively new

phenomenon which recognizes the relationship between fetal and postnatal brain organization and later cognitive functioning. Prior nutritional studies focused primarily on physical parameters, such as mortality, growth, and stool consistency. In such studies, physical properties which are easily observable, well-defined, and operationally precise, can be measured reliably. Because psychological research measures implied, less clear, hypothetical constructs (such as intelligence) the choice and timing of the assessment are critical and pose significant methodologic problems in infancy studies. To meet standards for clinical trials, an assessment instrument must be standardized, reliable, and valid. Standardization requires the development and publishing of uniform procedures in administering and scoring a test, including comparable testing conditions, detailed directions, scoring procedures, and examiner training requirements. Normative samples should have representative racial/ethnic, gender, socioeconomic, and geographic diversity [16]. The establishment of a normative group is particularly important since test performance is judged on empirical data to assess normal vs deviant performance. Since norms change over time, they must be periodically updated or the assessment instrument will yield invalid information on the rate of deviant performance. For example, an upward drift in Bayley Mental Development scores became apparent in population based studies in the 1980s, leading to restandardization of the scale in 1993 [17]. The restandardized scale resulted in mean scores 12 to 15 points, (or >0.75 of a standard deviation) lower than the old scale. These differences likely led to significant underestimations of developmental delay in very low birthweight outcome studies prior to 1992 [6].

Reliability refers to the consistency of a measurement. The range of fluctuation in an infant's developmental test scores limits the extent to which performance can be considered reflective of intellectual ability vs error variance. Validity refers to whether or not the assessment measures the construct it is supposed to measure, and is highly dependent on reliability. Although later childhood IQ scores are very stable, all infant assessments prior to 18–24 months of age have relatively low stability, with correlations declining with younger ages [18,19], and would be considered inadequate for adult clinical trials [20]. Predictive validity refers to the test's relationship to later outcomes, important because skill measurement in infancy would be irrelevant if it were unrelated to child functional outcomes.

Because visual and cognitive functions have been primary outcomes of interest in nutritional intervention studies, the majority of studies to date have relied on a small number of infant measures, each of which have specific limitations in interpreting findings. The Bayley Scales of Infant Development [21], useful for 1 to 42 month old infants, have been the primary outcomes because of ubiquitous use, extensive standardization and documentation of psychometric properties. The Mental Scale yields a global standard score, representing the cumulative achievement of the infant across a number of diverse skills, including attention, problem solving, language, and social abilities. This conglomerate score may be insensitive, however, to even major deficits in a particular function, since higher functioning in one domain may mask deficits in another. Despite their widespread use, the Bayley Scales have been particularly criticized because of their low predictive validity to later IQ prior to 2 years of age, except for the very low scores achieved by the most compromised infants [22]. Moreover, the mental scale is unacceptable for use with infants who have physical handicaps because of its reliance on motor skills for task performance, a difficulty which may affect significant numbers of infants in preterm studies. Whether such infants are excluded from cognitive outcome data, tested with alternative measures, or assigned the lowest scores may have variable consequences for interpretation of study results.

Despite poor predictive validity for individual infants, the Bayley Scales have proved sensitive to differences in at-risk infant groups even in the first year of life, discriminating outcomes on the basis of cocaine exposure, alcohol exposure, failure to thrive, very low birthweight, PCB exposure, and iron deficiency [23–27]. The predictive power of the re-normed Bayley Scales in the first year of life awaits documentation, although there is some evidence that sensitive and specific detection of developmental delay remains elusive in the young infant. In our studies of a large cohort of cocaine-exposed vs non-exposed infants, the 1993 Bayley Scales indicated almost no incidence of cognitive delay at 12 months in the sample (2%) [23], but by 2 years the overall rate was 12% and by 4 years, about 20%, suggesting continued insensitivity of the scales in the first year of life to detection of individual deficits. The re-normed scales have also been

criticized as unsuitable for preterm infants because changes in establishing baseline performance may inflate scores of preterm infants [28].

Given current data, it is imperative that longitudinal outcome studies of nutritional interventions extend at least until 18–22 months of age, since findings are not generalizable to later outcomes until then [8,19]. For very low birthweight groups for whom nutritional interventions have been specifically targeted, even older ages may be required before reasonable generalizability of benefit can be inferred, due to the standard practice of 'correcting' for prematurity in infant developmental testing. With 3 to 4 months prematurity, gestational-age-corrected scores can have significant error variance prior to 2 years of age when chronologic age scores are widely used. To date, few LCPUFA intervention studies have follow-up samples to at least 18 months of age, with the work of Lucas and colleagues [29] and Birch and colleagues [30] the exceptions.

Because standardized infant developmental tests, such as the Bayley Scales, have poor predictive validity in the first year of life, other measures have been developed which are conceptualized as tapping into more discrete areas of functioning underlying intellectual abilities, such as attention, visual perception, memory, and information-processing speed [20]. The methodology for establishing infants' visual preferences has produced a number of promising alternative assessments of infant cognition using visual habituation and paired-comparison tasks to measure visual recognition memory [18,31]. Some of these alternative assessments of cognitive processes, such as measurement of visual novelty preference and duration of looking time, have been used in studies of LCPUFA supplementation [32,33,34]. Such tasks have been shown to have better predictive validity to childhood IQ measures than standard sensorimotor assessments in the first year of life [18,22].

Such tasks are plausible candidates to demonstrate potential mechanisms underlying developmental differences related to early fatty acid nutrition. Both animal [33] and human [32] studies have found differences in look duration based on LCPUFA status. Differences in novelty preferences have also been found in human preterm studies [34]. Of these promising experimental tasks, however, only the Fagan Test of Infant Intelligence (FTII) [35] meets minimal clinical trial standards for psychometric adequacy. Even so, there are significant deficiencies in its psychometric properties which preclude firm conclusions from being drawn about intervention effects. These include poor short-term reliabilities, the need for multiple administrations over ages to produce stability and the lack of standard scores. Further, much of the predictive validity data reported is based on varying test items rather than for the test as currently published and constructed [36]. Even these predictive correlations are only modestly higher than those for the Bayley Scales, thus raising questions about clinical utility.

A further issue is the use of derived measures, either as part of the FTII, or in other experimental paradigms, such as the measure of look duration which has been construed to reflect information processing speed, an important component of intelligence measures at other ages. Derived measures are statistically weak because they cannot be assumed to have the psychometric properties of the 'parent' measure [37]. While of heuristic and exploratory interest, derived or experimental measures should not be used as primary outcomes in intervention studies.

A growing number of infant assessments are currently being developed which hold promise as candidates for use in early intervention studies [38]. These tasks vary in age-appropriateness, level of analysis, and technological advancement. Assessments of vagal tone, cortical electrophysiology, cry analysis, learned expectancies, and the A not B task will undoubtedly be considered in further research studies, but they remain experimental in nature [39,40]. Two of the more psychometrically developed assessments potentially useful in identifying specific benefits of an intervention at early ages include the MacArthur Communicative Development Index [41], and the Preschool Language Scale-3 [42].

Timing of assessments

Numerous considerations affect the timing of data collection points in an RCT of infants. One of the most important is the choice of the time of testing for group differences using instruments based on changes in normal development. Behavioral research is often characterized by measurement scales with unequal intervals, producing 'floor' or 'ceiling' effects [43]. In addition, there is both wide variability in infant skill development within normal ranges, but also high canalization of

developmental skills. The vast majority of infants sit up, smile, talk, and walk, for example, within a relatively short time period. Thus, it is difficult to detect language delay in a 12 month old infant, since most infants at that age will have little variability in their skills. Further, some infant assessments have small ranges of age within which infants can be validly tested. The FTII, for example, has 2-week age windows within which an infant can be assessed for each version of the test with exposure to the familiar stimulus decreasing by seconds at each age, while the Bayley Scales have age-normed scores and can be used from 1–42 months.

Practical considerations of timing infant assessments include the need for comparability to other studies, family burden, research costs, and infant capacity. Infants cannot sustain attention for long periods of time, necessitating that the researcher choose tasks carefully.

As noted above, even minimal confidence in study outcomes is not possible until infants reach 18 months of age, while correlations of about 0.50 are found between infancy tests to later IQ at 2–3 years when symbolic thinking and language skills predominate test content [22]. Assessment to school ages would be necessary to detect more subtle, but functionally important intervention problems or benefits, such as the prevention of specific learning disabilities or attention deficit disorder, which are independent of IQ. Very low birthweight infants of average IQ for example, nevertheless were found to have higher rates of learning problems at 9 years of age in comparison to term children [44]. At school age, a wider range of outcomes can be reliably assessed, thus detecting differences not apparent at earlier ages, sometimes referred to as 'sleeper' effects.

Confounding, mediating and moderating variables

One of the strengths of the RCT is its use of randomization to control extraneous variables which may spuriously affect the relationship between the intervention under study and outcome. Child developmental outcomes are affected by a large number of factors likely to be related to infant nutritional status which need to be considered before attributing benefits or deficits to a specific nutritional intervention [45]. Randomization theoretically controls for these extraneous variables and for selection biases associated with entry into a study [7]. For example, mothers who choose to enroll in a nutritional intervention study may be more motivated, more sociable, psychologically healthier, and more educated than those who do not. In evaluations of AA and DHA effects on infants, it has been difficult to separate the effects of the nutritional differences between breast milk and formula from the effects of socioeconomic and intellectual advantages of mothers who breast feed [46,47].

With randomization, confounding of intervention groups can still occur, especially with smaller sample sizes. However, even with large sample sizes, demonstration of group equivalence is important for all variables which might reasonably relate to the outcome under study. For example, despite a large sample size and randomization, one study [34] found differences in home environment quality confounded with their intervention groups. Identification of potential confounders should be made a priori based on their demonstrated relationships to study variables. Accurate measurement of confounders can aid in identifying the true relationships between the nutritional intervention and outcome, since, if randomization does not 'work', statistical control can be used. Other strategies which address confounders prior to study include exclusion, matching, and stratification.

Numerous factors which relate to child developmental outcome or IQ have been identified as confounding variables which need to be considered in RCT's of infant nutritional interventions. These include socioeconomic status, parental education, race, income, maternal IQ, age, psychological status, parity, marital status, child gender, and number of children in the home. The quality of the home environment, including maternal warmth and responsivity, have also been implicated in child developmental outcomes independent of sociodemographic factors [48,49].

The importance of assessing fetal exposure to a number of potentially teratogenic substances has been demonstrated. Tobacco, alcohol, and marijuana use are not uncommon in pregnant women (with rates of 8–64%) [50] and, dependent on duration and severity of use, may affect infant growth and development [51,52,53]. Confounded with geographic location and age cohorts, substances such as cocaine, heroin, methamphetamines, and 'club' drugs may also be used by pregnant women. Little is also known about the prevalence

of use of prescribed drugs, such as antidepressants and anti-anxiety drugs, during pregnancy or the effects of these drugs on child outcomes [54]. Infants exposed prenatally to fluoxetine were at higher risk for pregnancy complications such as preterm delivery, feeding difficulties, and jitteriness [55]. The depression and anxiety conditions for which such medications are prescribed are common in women of childbearing age and known to exert their own negative effects on fetal development through changes in the hypothalamic–pituitary–adrenal axis, and on postnatal development through less optimal maternal caregiving interactions [56]. Women with postpartum depression (estimated at 10–15% in controlled studies) may be less likely to breast feed [54]. They may be more likely to enter into RCT's because of worry about their infants, or alternatively less likely to enroll because of other depressive symptoms, such as sadness or disinterest. However, to date, no study of LCPUFA supplementation has assessed this factor.

Confounding factors in preterm studies

In addition to the factors noted above, studies of preterm and VLBW infants pose additional methodologic problems because of the constantly changing medical care practices of neonatal intensive care units and the many medical and neurologic conditions associated with prematurity which have relationships to cognitive outcomes. These include gestational age, birthweight, intrauterine growth retardation, multiple birth, the presence of patent ductus arteriosis, septicemia, bilirubinemia, apnea, retinopathy of prematurity, necrotizing enterocolitis (NEC), and bronchopulmonary dysplasia (BPD) or chronic lung disease [25].

Direct effects on cognitive outcomes are found with neurologic complications, such as intraventricular hemorrhage, periventricular leukomalacia, meningitis, seizures, asphyxia, and various forms of hypoxic-ischemic main injuries [6]. Because these numerous perinatal complications may cluster nonrandomly and with other sociodemographic factors, large sample sizes are needed as well as clear documentation of their incidence across study groups. Infants with BPD, for example, tend to be smaller and sicker than their VLBW counterparts [25], making it problematic when BPD has been used as an outcome in LCPUFA interventions [57]. Some confounding factors prevent enteral feeding, effectively excluding infants from the feeding protocol and restricting generalizability. Because of the large sample sizes needed to control for confounding factors, often multiple sites are needed for preterm studies, creating additional size requirements due to practice variations and other differences across sites. One study estimated a needed recruitment sample size of 470 to detect a 0.5 standard deviation difference in Bayley outcomes across 7 sites [34].

The importance of accurate measures of all variations in practice cannot be underestimated. For example, early studies suggested postnatal steroid interventions to be safe for preterm infants, while recent randomized trials have indicated their association with cerebral palsy at $4\frac{1}{2}$ years [58]. Confounding factors may also differ in their effects across various developmental domains as well. We found no effect at 3-years-of-age of patent ductus arteriosis on cognitive outcome of preterm infants [25], but found a significant relationship to language outcome at the same age [58].

Mediating vs moderating variables

A special consideration is whether important variables related to outcome are designated as confounding vs mediating or moderating [60]. Failure to consider these characteristics can lead to over or underestimation of the effects of an intervention. Mediating variables form part of the causal chain between the intervention and the outcome, while moderators are characteristics which change the effect of an intervention. Thus, hypothetically, if early nutritional intervention reduced the incidence of BPD or necrotizing enterocolitis in preterm infants, and these conditions were considered confounding variables, their control in analyses would obscure the effects of the intervention. When considered as mediators, however, it can be demonstrated that the intervention led to a reduced incidence of BPD or NEC which, in turn, led to better developmental outcomes. An example of a moderating variable can be found in the Infant Health and Development study which demonstrated strong intervention effects on IQ except for the subgroup of infants who weighed less than 1500 grams [5]. Such moderating effects should be explored in nutritional intervention studies, since

effects of supplemental LCPUFAS may vary by infant characteristics such as birthweight.

Selection bias and attrition

As noted above, characteristics of women who enter nutritional trials are likely to be different from those who do not enroll. Researchers need to delineate as much as possible the demographic or other characteristics of infants and their parents who selected out of study. This is particularly important in studies in which feeding conditions include women who breast feed given the overriding influences of education, income, intelligence, and caregiving characteristics associated with breast feeding in some countries.

Similarly, attrition is a serious problem for longitudinal studies because it is usually non-random [61]. In fact, a number of studies demonstrate that dropout is selective [62,63]. Estimates of attrition need to be considered in determination of power analyses and sample size. Attrition tends to increase over time, a particular difficulty for early infant interventions which require years of follow-up to determine outcomes. The effects of attrition are of concern when sample sizes are so reduced that adequate power is lost. However, attrition may produce misleading effects even when statistical power is good. For example, in a RCT of term infants supplemented with LCPUFAs, retention rate at 18 months was 71%, with only 18–20 infants per study group [30]. With such small sample sizes, the intervention effects could potentially be influenced by a few outliers retained or lost to follow-up. High retention rates are necessary to have confidence in findings. The NICCHD network requires an 80% retention rate to 18 month follow-up [6]. Even that rate may be too low for firm conclusions about intervention effects to be made. In one study [64] increasing the follow-up retention from 89% to 96% from 18 months to 7.5 years detected a significant neurodevelopmental impairment difference between study groups that could have been found at the earlier age had retention been better.

Examiner confounds

Few published studies of early nutritional interventions have addressed the effects of the examiner. While such effects are important to consider in all studies, examiner effects are potentially more problematic in infant studies since infants require more personal interactions from the examiner during assessment [16]. Both study participants and examiners must be 'blinded' or 'masked' to intervention status to prevent beliefs about the intervention effects from affecting the examiner's ratings of infant performance [65]. Examiner characteristics such as enthusiasm also can affect infant performance but can be minimized by consistent training and establishment of inter-examiner reliabilities prior to study initiation. Continued maintenance of reliabilities at regular intervals throughout the study is also needed to prevent examiner 'drift' [66]. When multiple sites are needed, this problem is magnified. The use of a standard training procedure for examiners at all sites, with continued monitoring is necessary [67]. The number of examiners in any study should be noted and the effects of examiner differences on outcome should be examined statistically.

Summary

The use of the randomized clinical trial to assess the safety and benefits of LCPUFA supplementation in infant formula on cognitive development poses special methodologic problems when applied to infancy interventions. The randomized clinical trial must document a reliable association between the intervention and outcome with causality inferred through this relationship as well as through the time order. Few infant assessment instruments meet standards for use in clinical trials. Currently available assessments to measure infant cognitive abilities lack adequate stability and predictive validity, requiring longitudinal follow-up of interventions into beyond the second year of life to establish safety and efficacy. Timing of data collection points is also critical because of the rapid growth of infant developmental skills in the first two years of life. Numerous confounding factors known to relate to child developmental outcome must be controlled through measurement exclusion matching or stratification. Medical complications which cluster non-randomly are also an issue in preterm studies which require larger sample sizes to accommodate for control of medical covariates and often must be conducted over multiple sites. Failure to consider whether covariates are mediating or

moderating variables can also lead to over- or underestimation of the effects of an intervention. Selection and attrition biases and examiner effects need also to be considered if valid findings are to be obtained.

Acknowledgements

Supported by Grants MC-00127-04 from the Maternal and Child Health Bureau and RO1-07957-07 from the National Institute on Drug Abuse. With thanks for the assistance of Terri Lotz-Ganley.

References

1 Dobbing J. *Developing Brain and Behavior: The Role of Lipids in Infant Formula*. New York: Academic Press, 1977.
2 Lucas A, Morley R, Cole T, Lister G, Leesom-Payne C. Breast milk and subsequent intelligence quotient in children born preterm. *Lancet* 1992; **339:** 261–264.
3 Clandinin MR, Chappell JE, Leong S, Heim J, Swyer PR, Chance GW. Intrauterine fatty acid accretion rates in human brain. *Early Hum Dev* 1980; **4:** 121–129.
4 Food and Agricultural Organization/WHO. *Fats and Oils in Human Nutrition*. Rome: FAO/WHO, 1994; 6–7, 49–55.
5 Brooks-Gunn J, McCarton C, Van Rossem R, *et al*. Early intervention of low birthweight premature infants. *JAMA* 1994; **272:** 1257–1262.
6 Vohr BR, Wright LL, Dusick AM, *et al*. Neurodevelopmental and functional outcomes of extremely low birthweight infants in the NICHHD Neonatal Research Network, 1993–1994. *Pediatrics* 2000; **105:** 1216–1226.
7 Kerlinger F. *Foundations of Behavioral Research*, 3rd ed. New York: Holt, Rinehart, and Winston, 1986.
8 McCall RB. Nature, nurture and the two realms of development: A proposed integration with respect to mental development. *Child Dev* 1981; **55:** 1–12.
9 Fein GG, Schwartz PM, Jacobson SW, Jacobson JL. Environmental toxins and behavioral development. A new role for psychological research. *American Psychologist* 1983; **38(11):** 1188–1197.
10 Gordon R. The design and conduct of randomized clinical trials. *IRB – A Review of Human Subjects Research* 1985; **7:** 1–3, 12.
11 Katz LC, Schatz CJ. Synaptic activity and the construction of cortical circuits. *Science* 1996; **274:** 1133–1138.
12 Wald NJ. Folic acid and neural tube defects. In: Bock G, Marsh J (eds) *Neural Tube Defects, Ciba Foundation Symposium (8)*. London, England: John Wiley, 70–89.
13 Huttenlocher PR. Synaptogenesis in human cerebral cortex. In: Davison G, Fischer KA (eds) *Human Behavior and the Developing Brain*. New York: Guilford Press, 2000; 137–152.
14 Nelson CA. Change and continuity in neurobehavioral development: Lessons from the study of neurobiology and neural plasticity. *Infant Behavior and Development* 1999; **22:** 415–429.
15 Vorhees CV. Principles of behavioral teratology. In: Riley EP, Vorhees CV (eds) *Handbook of Behavioral Teratology*. New York: Plenum Press, 1986; 23–48.
16 Anastasi A. *Psychological Testing*. New York: Macmillan, 1996.
17 Campbell SK, Siegel E, Parr CA, Ramey CT. Evidence for the need to renorm the Bayley Scales of Infant Development based on the performance of a population-based sample of 12 month old infants. *Topics in Early Childhood Special Education* 1986; **6:** 83–96.
18 Bornstein M, Sigman M. Continuity in mental development from infancy. *Child Dev* 1986; **57:** 251–274.
19 Fagan JF, Singer LT. Infant recognition memory as a measure of intelligence. In: Lipsitt L (ed.) *Advances in Infancy Research*. Norwood, NJ: Ablex, 1983; **2:** 31–78.
20 Colombo J. *Infant Cognition: Predicting Later Intellectual Functioning*. Sage Publications: Newbury Park, 1993.
21 Bayley N. *The Bayley Scales of Infant Development*. San Antonio, TX: Psychological Corporation, 1993.
22 McCall R, Carriger M. A meta analysis of infant habituation and recognition memory performance as predictors of later IQ. *Child Dev* 1993; **64:** 57–59.
23 Singer LT, Arendt R, Minnes S, Salvator A, Siegel AC, Lewis BA. Developing language skills of cocaine-exposed infants. *Pediatrics* 2001; **107:** 1057–1064.
24 Singer LT, Yamashita TS, Lilien L, Collin M, Baley J. A longitudinal study of infants with bronchopulmonary dysplasia and very low birthweight. *Pediatrics* 1997; **100:** 987–993.
25 Singer LT, Fagan JF. The cognitive development of the failure-to-thrive infant: A three year longitudinal study. *J Pediatr Psychol* 1984; **9:** 363–384.
26 Jacobson JL, Jacobson SW. Intellectual impairment in children exposed to polychlorinated biphenyls in utero. *N Engl J Med* 1996; **335:** 783–789.
27 Lozoff B. Has iron deficiency been shown to cause altered behavior in infants? In: Dobbing J (ed.) *Brain Behavior and Iron in the Infant Diet*. London: Springer-Verlag, 1990.
28 Ross G, Lawson K. Using the Bayley II: Unresolved issues in assessing the development of prematurely born children. *Developmental and Behavioral Pediatrics* 1997; **18:** 109–111.
29 Lucas A, Morley R, Cole TJ, Gore SM. A randomized multicenter study of human milk versus formula and later development in preterm infants. *Archives of Diseases of Childhood* 1994; **70:** 141–146.
30 Birch EE, Garfield S, Hoffman DR, Uauy R, Birch DG. A randomized controlled trial of early dietary supply of long-chain polyunsaturated fatty acids and mental development in term infants. *Dev Med Child Neurol* 2000; **42:** 174–181.
31 Rose S, Wallace I. Visual recognition memory: A predictor of later cognitive development in preterms. *Child Dev* 1985; **56:** 843–852.
32 Werkman SH, Carlson S. A randomized trial of visual attention of preterm infants fed docosohexaenoic acid until nine months. *Lipids* 1996; **31:** 91–97.
33 Reisbick S, Neuringer M, Gohl E, Wald R, Anderson GJ. Visual attention in infant monkeys: Effects of dietary fatty acids and age. *Developmental Psychology* 1997; **33:** 387–395.

34 O'Connor DL, Hall R, Adamkin D, Auestad N, *et al*. Growth and development in preterm infants fed long-chain polyunsaturated fatty acids: A prospective randomized control trial. *Pediatrics* in press.

35 Fagan JF, Shepherd P. *The Fagan Test of Infant Intelligence Training Manual*. Cleveland, OH: Infantest Corp, 1989.

36 Benasich AA, Bejar I. The Fagan Test of Infant Intelligence: A critical review. *Journal of Applied Developmental Psychology* 1992; **13:** 153–171.

37 Cronbach L. *The Essentials of Psychological Testing*, 4th ed. New York: Harper and Row, 1984.

38 Singer LT. General issues in infant assessment and development. In: Singer LT, Zeskind PS (eds) *Biobehavioral Assessment of the Infant*. New York: Guilford Publications, 2001.

39 Singer LT, Zeskind SP (eds). *Biobehavioral Assessment of the Infant*. New York: Guilford Publications, 2001.

40 O'Connor MJ, Cohen SL, Parmelee AH. Infant auditory discrimination in preterm and full-term infants as a predictor of 5-year intelligence. *Developmental Psychology* 1984; **20:** 159–165.

41 Fenson L, Dale PS, Reznick JS, *et al*. *MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. San Diego: Singular Publishing Group, 1993.

42 Zimmerman IL, Steiner VG, Pond RE. *Preschool Language Skills 3rd Edition*. San Antonio, TX: The Psychological Corp., 1992.

43 Kenny T, Holden EW, Santilli L. The meaning of measures: Pitfalls in behavioral and developmental research. *J Dev Behav Pediatr* 1991; **12:** 355–360.

44 Klein NK, Hack M, Breslau N. Children who were very low birthweight: Developmental and academic achievement at nine years of age. *J Dev Behav Pediatr* 1989; **10:** 32–37.

45 Sameroff AJ, Chandler MJ. Reproductive risk and the continuum of caretaking casualty. In: Horowitz FD, Hetherington M, Scarr-Salapatek S, Siegel G (eds) *Review of Child Development Research* (Vol. 4). Chicago: University of Chicago Press, 1975; 187–244.

46 Jacobson SW, Jacobson JL. Breast feeding and intelligence (letter to the editor). *Lancet* 1992; **339:** 926.

47 Lucas A, Cole TJ, Morley, R, *et al*. Factors associated with a mother's choice to provide breast milk for a pre-term infant. *Arch Dis Child* 1988; **63:** 48–52.

48 Caldwell B, Bradley R. *Home Observation for Measurement of the Environment*. Little Rock: University of Arkansas, 1984.

49 Bradley R, Caldwell B, Rock S. Home environment and school performance: A ten year follow-up and examination of three models of environmental action. *Child Dev* 1988; **58:** 852–867.

50 US Department of Health and Human Services. Preliminary estimates from the 1998 National Household Survey on Drug Abuse (Report #21), and Mental Health Services Administration. Rockville, MD., 1998.

51 Coles CD, Platzman KA, Smith I, James ME, Falek A. Effects of cocaine and alcohol use in pregnancy on neonatal growth and neurobehavioral status. *Neurotoxicol Teratol* 1992; **14:** 23–33.

52 Fried PA, Watkinson B. 12 and 24 month neurobehavioral follow-up of children prenatally exposed to marijuana, cigarettes, alcohol. *Neurotoxicol Teratol* 1990; **11:** 49–58.

53 Streissguth AP, Barr HM, Martin DC, Herman CS. Effects of maternal alcohol, nicotine, and caffeine use during pregnancy on infant mental and motor development at 8 months. *Alcohol Clin Exp Res* 1980; **4:** 152–164.

54 Wisner KL, Gelenberg AJ, Leonard H, Zarkin D, Frank E. Pharmacologic treatment of depression during pregnancy. *JAMA* 1999; **282(13):** 1264–1269.

55 Chambers C, Johnson K, Dick LM, Felix R, Jones KL. *N Eng J Med* 1996; **335:** 1010–1015.

56 Singer LT, Yamashita T, Davillier M, Preuss L, Hawkins S, Yamashita T, Baley J. Feeding interactions in infants with very low birthweight and BPD. *J Dev Behav Pediatr* 1996; **17:** 69–76.

57 Carlson SE. Long-chain polyunsaturated fatty acid supplementation of preterm infants. In: Dobbing J (ed.) *Developing Brain and Behavior: The Role of Lipids in Infant Formula*. San Diego: Academic Press, Ltd., 1997; 41–102.

58 Shinwell ES, Karplus M, Recih D, Weintraub Z, Blazer S, Bader D, *et al*. Early postnatal dexamethasone treatment and increased incidence of cerebral palsy. *Arch Dis Child* Fetal Neonatal Ed 2000; **83(3):** F177–181.

59 Singer LT, Siegel AC, Lewis B, Hawkins S, Yamashita T. Preschool language outcomes of children with history of bronchopulmonary dysplasia and very low birth weight. *J Dev Behav Pediatr* 2001; **22:** 19–25.

60 Barron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986; **51:** 1173–1182.

61 Flick SN. Managing attrition in clinical research. *Clinical Psychology Review* 1988; **8:** 499–515.

62 Kalverboer A. Follow-up of biological high risk groups. In: Rutter M (ed.) *Studies of Psychosocial Risk*. New York: Cambridge University Press, 1988; 114–137.

63 Tin W, Fritz S, Wariyar U, Hey E. Outcome of very preterm birth: Children reviewed with ease at 2 years differ from those followed up with difficulty. *Arch Dis Child* Fetal Neonatology (ed.) 1998; **79:** F83–F87.

64 Lucas A. In: Dobbing J (ed.) *Developing Brain and Behavior: The Role of Lipids in Infant Formula*. New York: Academic Press, 1997; 237.

65 Stern M, Hildebrandt KA. Prematurity stereotyping: Effects on mother infant interaction. *Child Dev* 1986; **57:** 308–315.

66 Morley R. Data collection and management for research studies. *Arch Dis Child* 1995; **73:** 364–367.

67 Berk RA. Generalizability of behavioral observations: A clarification of interobserver agreement and interobserver reliability. *American Journal of Mental Deficiency* 1979; **83:** 460–463.