Modeling Environmental Influences on Child Growth in the MAL-ED Cohort Study: Opportunities and Challenges

Stephanie A. Richard,[1] Benjamin J. J. McCormick,[1] Mark A. Miller,[1] Laura E. Caulfield,[2] William Checkley,[1,3] and The MAL-ED Network Investigators[a]

[1]Fogarty International Center, National Institutes of Health, Bethesda; and [2]Program in Human Nutrition, and [3]Program in Global Disease Epidemiology and Control, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland

Although genetics, maternal undernutrition and low birth weight status certainly play a role in child growth, dietary insufficiency and infectious diseases are key risk factors for linear growth faltering during early childhood. A primary goal of the Etiology, Risk Factors and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health and Development (MAL-ED) study is to identify specific risk factors associated with growth faltering during the first 2 years of life; however, growth in early childhood is challenging to characterize because growth may be inherently nonlinear with age. In this manuscript, we describe some methods for analyzing longitudinal growth to evaluate both short- and long-term associations between risk factors and growth trajectories over the first 2 years of life across 8 resource-limited settings using harmonized protocols. We expect there will be enough variability within and between sites in the prevalence of risk factors and burden of linear growth faltering to allow us to distinguish some of the key pathways to linear growth faltering in the MAL-ED study.

Keywords. diarrhea; growth; MAL-ED; malnutrition; stunting.

Linear growth in childhood is the manifestation of a complex interaction between genetic and environmental factors [1]. Childhood stunting has historically been associated with increased child morbidity and mortality [2] and impaired cognitive development [3, 4]. In resource-limited settings of low- and middle-income countries (LMICs), poor growth in children is very common, with stunting (length-for-age or height-for-age z score <-2) prevalence of 28% and underweight (weight-for-age z score <-2) prevalence of 17%, on average [5]. In the Etiology, Risk Factors and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health and Development (MAL-ED) cohort study [6], we will describe the interrelations of diet, infection, and illness as they affect child health in 8 LMICs [7]. Children in MAL-ED are followed from birth through 2 years of age, and during follow-up, data are collected regularly on the cohort participants' illnesses, diet, pathogen burden, gut function, micronutrient status, cognitive development, and growth, among others. This manuscript will describe methods used to collect high-quality monthly anthropometric data (eg, length, weight, and head circumference), and will describe some traditional and novel methods that can be used to evaluate child growth and its determinants.

Growth of children is monitored as a sentinel indicator of overall well-being because it reflects the adequacy of the environment as it influences growth and development, which ultimately affects health, performance, and survival. Child growth is not constant; rather, children experience different rates of linear growth over time. During the first months of life, children generally exhibit a fast rate of linear growth, with a deceleration beginning during infancy and continuing as the child ages [8]. The linear growth of individual children follows

this general path, but growth velocity may accelerate or slow for short periods of time. Reasons for these changes are not always understood, but for children in LMIC, weight faltering (as a result of poor infant feeding practices, food insecurity, or infection, for example) can precede linear growth faltering [9]. With appropriate inputs, accelerated growth (ie, catch-up growth) can occur that will return the child to what can be thought of as their original height trajectory [10]. Van Ijzendoorn and colleagues reported in a meta-analysis of international adoption studies the potential for catch-up linear growth in children, particularly those adopted before 18 months of age [11]. In the case of persistent food insecurity or serial infectious diseases, however, opportunities for catch-up linear growth may be inadequate and stunting may result.

DESCRIPTION OF MAL-ED STUDY GROWTH DATA

Well-trained study staff members measured the length, weight, and head circumference of children upon enrollment into the cohort (within 17 days after birth), and each month thereafter for the first 2 years of life. Mothers or other caregivers were also asked for the birth weight of their child, if available, upon enrollment. There are several quality control measures in place at the 8 study field sites, including the use of standardized techniques and instruments across the sites, same-day review of growth curves to identify unlikely measurements, and duplicate anthropometric measurements taken on a subset of children. Standard recumbent length measuring boards are used to measure length (eg, Shorrboards, seca, UNICEF), infant scales are used to measure weight (eg, Detecto, seca), and nonstretch Teflon synthetic tape is used to measure head circumference (seca). Anthropometric measurements are taken monthly, on the anniversary of the child's birth ±2 days. Once the measurements are collected, the field-workers return to the study office and chart the child's growth on a World Health Organization (WHO) growth chart and review the data with a supervisor. If a highly unusual measurement is observed on the growth chart, the child is remeasured at the earliest opportunity, generally within 2 days. A specific definition for unusual measurements is not used at the study sites; however, supervisors can easily recognize via visual inspection of the growth chart values that are markedly different from the previous growth trajectory. In addition, each month, a supervisor or highly trained study staff member collects a set of duplicate anthropometric measurements for 5% of the children within 24 hours of the original measurement. These values are compared with the original measurement, and if needed, retraining is administered to the study staff. In the central quality control process executed by the MAL-ED Data Coordinating Center (DCC), values that are substantially different from both the previous and following measurements

(weight >1.5 kg different, length >3.5 cm different, and head circumference >2 cm different from both previous and following measurements) are flagged for review by the study site. For values that are flagged by investigators at the DCC, the study site data manager reviews the data reporting form to confirm whether or not the correct value was entered [6].

EVALUATING GROWTH

Although interest in growth has existed since ancient times, it was not until the 19th century that modern epidemiological methods to categorize and quantify growth were developed [12]. The idea of growth distributions began with Quetelet in the 19th century [13], and this concept of growth variability around the mean was further detailed by Roberts [14] and Galton [15] in the United Kingdom, as well as Bowditch [16] in the United States. Advances in biostatistics during the 20th century, as well as the collection of larger longitudinal datasets, furthered the development of epidemiological methods regarding growth. In general, linear growth has historically been considered in 3 ways. First, as a single measurement (attained height or weight) at a certain age; second, compared to a known distribution, by sex, of heights or weights; or third, as a linear function of age. Here we will describe some commonly used methods to describe growth, as well as some novel methods that may enhance the MAL-ED study's ability to link growth impairment with a variety of underlying environmental risk factors across multiple study sites.

Attained Growth Standards

Researchers have commonly generated distributions based on their study populations (by age and sex) and identified children who were growing poorly based on where they fell in that distribution. By convention, attained growth is normally distributed and defined by a mean and standard deviation (SD), with 95% of children falling between ±1.96 SD or z scores. Children falling below −2 SD (ie, in the shortest/lightest 2.5% of the population) are then interpreted to have a low probability of growing to their fullest potential, and this is common in areas with adverse environmental factors.

The culmination of more than a decade of research on international patterns of growth was the WHO Multicenter Growth Reference Study (MGRS), in which anthropometric measures (length, weight, and head circumference) of children in 6 countries were collected to develop a standard of how children could potentially grow with adequate resources across a number of settings [17]. Children in any population can be compared with the WHO MGRS growth standards, allowing for straightforward comparison of growth metrics across populations. The WHO MGRS also demonstrates that ethnicity plays only a small role in attained growth (3% of total variability) compared with

individual environmental factors (70% of total variability) [18]. This point confirms prior research that linear growth is far more dependent on access to food and environmental characteristics than ethnic differences [19, 20]. WHO MGRS z scores calculated using the MAL-ED study data can be used to assess whether attained growth at a certain age, on average, is different based upon the child's experience with certain risk factors.

Growth Velocity Standards

The WHO recently published longitudinal growth velocity standards [21] that can also be used to evaluate child growth in the MAL-ED cohort. The velocity standards describe the rate of length or weight acquisition that is expected between different months of age by sex. Velocity standards exist for 1, 2, 3, 4, and 6-month increments. The WHO suggests that, unlike the tracking that generally occurs with attained growth z scores, growth velocity z scores may naturally vary considerably from one period to the next, which is consistent with the idea of saltatory growth [22]. A single period with a low growth velocity z score may not be indicative of poor growth; however, consecutive periods with very low velocity z scores may have long-term consequences. By using the z scores for growth velocity in conjunction with attained growth, we can describe the relationships among diet, infectious disease, and growth in this multisite cohort study.

MODELING GROWTH

The relationships among diet, infectious disease, and growth may be best explored through modeling child length or weight over time. Historically, scientists have devised multiple parametric models to characterize child growth (Table 1). Longitudinal growth models have been proposed that include both linear and nonlinear combinations of parameters. Polynomial and 3-parameter linear models have been used to generate individual growth curves using random effects [23, 24], as have more computationally intensive nonlinear models [27–29]. Recently, systems biology models have been proposed that offer a more holistic approach to evaluate child growth and its determinants. Each of these approaches is described below.

Models that use linear combinations of parameters encompass the most commonly used approach to modeling child growth in the literature. These models can be relatively straightforward in their calculation, summarizing patterns of linear or ponderal growth into a handful of numbers. Examples of linear growth models are listed in Table 1. Linear models can be made more complex by incorporating flexible piecewise polynomials such as regression splines to provide a semiparametric, data-driven specification of the growth curves [30, 31]. These linear models can use a marginal approach such as generalized estimating equations [32], a method that aims to estimate the average response over the population with a possible unknown

Table 1. Examples of Parametric Growth Models

Model Name	Formula
Linear models	
Count [23]	Height = a + b*age + c*ln(age)
Wingerd [24]	Height = a + b*age + c*age^0.5
Guo [25]	Height = a + b*log(age + 1) + c*(age + 1)^0.5
Quadratic	Height = a + b*age + c*age^2
Reed first order [26]	Height = a + b*age + c*log(age + 1) + d/(age + 1)
Cubic	Height = a + b*age + c*age^2 + d*age^3
Reed second order [26]	Height = a + b*age + c*log(age + 1) + d/(age + 1) + e/(age + 1)^2
Quartic	Height = a + b*age + c*age^2 + d*age^3 + e*age^4
Power	Height = a + b*age + age^c
Nonlinear models	
Jenss & Bayley [27]	Height = a + b*age − exp(c + d*age)
Karlberg [28]	Height[age < d] = a + b*(1-exp(-c*age[1:(height [age > d] = a + b*(1-exp(-c*age[d:end])) + e* (d-age[d:end])d-1)]]

correlation between outcomes. Random-effects models that aim to model heterogeneity in child growth curves and provide an estimate of individual children's trajectory could also be used [33]. Random effects models are also able to adequately capture heteroscedasticity between an individual's growth data within a population over time, and individual growth trajectories can be estimated from these types of models. In the setting of linear mixed effects models, serial correlation between repeated measurements can be modeled. Moreover, we can demonstrate how the growth curve of that child may be affected according to a set of exposures (ie, independent variables).

A drawback of linear models is that overall inferences are predicated on comparing population means between groups. Particular independent variables may not have consistent relationships with growth across children of all growth percentiles or even within the same child over time. Stunted children are more likely to have micronutrient deficiencies that are associated with poorly functioning immune systems (eg, zinc) [34]; therefore, linear growth faltering may be more likely in stunted children as a result of infection. This question can be explored by performing subgroup analyses, such as those based on whether the child was stunted or not at baseline.

Quantile regression is a statistical method that estimates the conditional percentiles of an outcome for a set of covariates [35]. This method allows the estimation of different relationships between risk factors and growth for children in different percentiles. The continuing challenge with this method is the incorporation of random effects, and one method currently in use to do this involves regularization [35]. Because children can cross growth percentiles as they age, quantile regression provides a good estimate of population or quantile-specific

estimates, but not necessarily for individual children. The application of quantile regression is currently limited in comparison to more standard methods for parameter estimation because the statistical software programs for longitudinal analysis using quantile regression are not readily available. Incorporating the inherent complications of time series data (including serial autocorrelation) and the nonindependence of repeatedly sampling individuals into the quantile regression method is not an insignificant challenge.

Nonlinear mixed effects models provide an alternative for the analysis of longitudinal data. In their simplest form, they do not assume an additive combination of parameters and might include power functions or multiplicative relationships (Table 1). Similar to linear models, nonlinear models are predicated on fitting mean population effects and cannot make inferences of the effects of exposures on growth at different percentiles. Many more parameters might be invoked in a nonlinear model than a linear one to reflect the more varied probability distribution(s) assumed to underlie the observed data. However, parameters come at the expense of tractability and reduce the overall statistical power of a model.

A systems biology approach, increasingly common in ecology [36], may prove to be a novel way to explore the variety of risk factors and outcomes collected in the MAL-ED study. By mathematically representing the mechanisms of growth inputs and processes, patterns are generated and compared to observed data. This is in contrast to the more traditional approach of constructing a model based on identifying statistical matches between putative risk factors and outcome variables of interest [37]. In the latter, the aim is to identify correlates from empirical evidence, whereas the former makes assumptions about the relationships between the outcome and processes that generate it that are embedded in equations (ie, the mechanisms are assumed). For MAL-ED, systems modeling approaches allow for quantification of the variability in growth that can be explained by mechanistic descriptions of putative risk factors. The strengths of this approach include the identification of knowledge gaps (eg, identifying particular mechanisms that result in considerable growth variability) and describing growth based on explicit mechanisms rather than statistical associations (which may be borne out of an unobserved chain of intermediate factors between a factor and a health outcome).

Anticipated methodological challenges, regardless of the modeling approach, include (1) efficient management of high-dimensional large-scale data, (2) evaluation of drivers of short- and long-term growth outcomes, (3) time delays in growth faltering (eg, lagged time covariate effects, lagged time to detection, lagged time to response), (4) effect correlations and accumulations (eg, correlations and accumulations in frequency and severity of illness), and (5) nonlinear causal relationships (eg, threshold effects, saturation, and recovery effects).

MODELING THE ASSOCIATION BETWEEN DIARRHEA AND GROWTH

There are several analytical challenges when studying the association between diarrhea and growth in early childhood. The first is proper development of a longitudinal model to capture child growth, as discussed above, because an inadequate goodness-of-fit in the growth model may lead to a gross over- or underestimation of effects. The biostatistical models discussed above account for the serial dependence of repeated measurements in the same child and heterogeneity in growth across children. In the past, modeling correlation from repeated measurements and random effects were commonly ignored and may lead to an overstatement of statistical significance.

Second is the need to account for the longitudinal nature of diarrheal episodes [30, 31], other infectious diseases, and dietary intake. For example, in longitudinal field studies of long duration, children may experience multiple episodes of diarrhea. Studies that use short time intervals to examine the effects of diarrhea on growth may either overestimate growth deficits, because short intervals do not allow time to detect possible catch-up growth, or underestimate growth deficits, because they do not allow time to detect possible delayed effects. Third is the need to account for the interaction effects of multiple episodes of diarrhea on growth, which includes simultaneous modeling of both lagged effects [30, 31] and potential catch-up growth between diarrhea-free periods. An important contribution of the MAL-ED study will be the exploration of modern developments in biostatistics for the analysis of longitudinal data in which we simultaneously model lagged effects and catch-up growth.

DISCUSSION

Growth in early childhood can be challenging to model because growth is inherently nonlinear with age. Children grow at different rates as they age, and may alternate between growth spurts and periods of slower growth due to either intrinsic or environmental factors. This saltatory process is thought to be normal; however, slower growth that occurs as a result of dietary insufficiency and infectious disease may prevent or limit the frequency of faster growth periods and thereby result in linear growth faltering. By considering the growth process in a longitudinal manner, one can better describe the synergistic relationships among diet, infection, and growth, both over the short and long term. In the MAL-ED study, data are collected involving a variety of different illnesses and exposures that may be linked to poor growth. We do not, however, collect information on tuberculosis or human immunodeficiency virus infection in a standardized way. Although both of those illnesses are expected to be relatively rare in the cohort based on local knowledge of

the populations at the inception of the study (with the exception of the South Africa site), information on those diseases may be gathered at the site through other methods, and subanalyses can be performed.

The WHO growth standards are useful; however, it should also be noted that the longitudinal data collected during the WHO MGRS were treated as if they were cross-sectional for the purposes of creating the attained growth standard for children 0–59 months of age. Thus, although we call them "growth standards," they do not necessarily describe the longitudinal experience of growth over time. Comparing the longitudinal growth of an individual child to the cross-sectional WHO growth standard does not necessarily assess whether the child is growing in a healthy manner in the context of their growth potential, which is unknown, or their previous growth experience. In LMIC, the entire distribution of attained lengths is shifted downward, such that even children who are not technically stunted (length-for-age z score <-2) are likely not growing to their fullest potential. The addition of the WHO velocity standards is potentially useful but, for a more nuanced assessment of these relationships, we will explore other options for modeling growth and assessing both the long- and short-term impact of environmental insults.

CONCLUSIONS

In applying these different modeling approaches to the MAL-ED study data, we seek to develop novel hypotheses on potential environmental factors common in children who experience atypical growth. We hope to bring into focus the biological processes that are associated with reduced growth potential, and to quantify their impact on measured outcomes. We envision that rank ordering of environmental factors with greatest impact on growth may lead to novel intervention hypotheses. Using the MAL-ED cohort data, we will determine the best way to quantify the relationships between key risk factors and growth among this collection of cross-sectional and longitudinal growth modeling techniques. Finally, we will identify key ages at which growth trajectories are most affected, and provide policy and program suggestions for the prevention of growth faltering.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online (http://cid.oxfordjournals.org). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

References

1. Martorell R, Mendoza F, Castillo R. Poverty and stature in children. In: Waterlow JC. Linear growth retardation in less developed countries Nestle nutrition workshop series Nestle Nutr Workshop Ser. Vol 14. New York: Raven Press, **1988**:57–73.
2. Caulfield LE, de Onis M, Blossner M, Black RE. Undernutrition as an underlying cause of child deaths associated with diarrhea, pneumonia, malaria, and measles. Am J Clin Nutr **2004**; 80:193–8.
3. Grantham-McGregor SM, Fernald LC, Sethuraman K. Effects of health and nutrition on cognitive and behavioural development in children in the first three years of life. Part 1: Low birthweight, breast-feeding, and protein-energy malnutrition. Food Nutr Bull **1999**; 20:53–75.
4. Martorell R, Horta BL, Adair LS, et al. Weight gain in the first two years of life is an important predictor of schooling outcomes in pooled analyses from five birth cohorts from low- and middle-income countries. J Nutr **2009**; 140:348–54.
5. Black RE, Victora CG, Walker SP, et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet **2013**; 382:427–51.
6. The MAL-ED Network Investigators. The MAL-ED study: a multinational and multidisciplinary approach to understand the relationship between enteric pathogens, malnutrition, gut physiology, physical growth, cognitive development and immune responses in infants and children up to two years of age in resource poor environments. Clin Infect Dis **2014**; 59(suppl 4):S193–206.
7. Checkley W, Buckley G, Gilman RH, et al. Multi-country analysis of the effects of diarrhoea on childhood stunting. Int J Epidemiol **2008**; 37:816–30.
8. Cameron N. The biology of growth. Nestle Nutr Workshop Ser Pediatr Program **2008**; 61:1–19.
9. Richard SA, Black RE, Gilman RH, et al. Wasting is associated with stunting in early childhood. J Nutr **2012**; 142:1291–6.
10. Schorling JB, Guerrant RL. Diarrhoea and catch-up growth. Lancet **1990**; 335:599–600.
11. Van Ijzendoorn MH, Bakermans-Kranenburg MJ, Juffer F. Plasticity of growth in height, weight, and head circumference: meta-analytic evidence of massive catch-up after international adoption. J Dev Behav Pediatr **2007**; 28:334–43.
12. Tanner JM. A history of the study of human growth. Cambridge: Cambridge University Press, **1981**.
13. Quetelet A. Sur l'homme et le développement de ses facultés, ou Essai de physique sociale. Paris: Bachelier, imprimeur-libraire, quai des Augustins, no. 55, **1835**.
14. Roberts C. The physical development and proportions of the human body. **1874**.

15. Galton F. Natural inheritance. London: Macmillan, **1889**.
16. Bowditch HP. The growth of children studied by Galton's percentile grades. 22nd annual report of the state Board of Health of Massachusetts. Boston: Wright and Potter, **1891**:479–525.
17. World Health Organization. WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development. Geneva, Switzerland: WHO, **2006**.
18. Assessment of differences in linear growth among populations in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl **2006**; 450:56–65.
19. Habicht JP, Martorell R, Yarbrough C, Malina RM, Klein RE. Height and weight standards for preschool children. How relevant are ethnic differences in growth potential? Lancet **1974**; 1:611–4.
20. Ulijaszek SJ. Ethnic differences in patterns of human growth in stature. In: Martorell R, Haschke F. Nutrition and growth. Philadelphia: Lippincott Williams and Wilkins, **2001**:1–20.
21. World Health Organization. WHO child growth standards: growth velocity based on weight, length and head circumference. Geneva, Switzerland: WHO, **2009**.
22. Lampl M, Veldhuis JD, Johnson ML. Saltation and stasis: a model of human growth. Science **1992**; 258:801–3.
23. Count EW. Growth patterns of the human physique: an approach to kinetic anthropometry, Part 1. Hum Biol **1943**; 15:1–32.
24. Wingerd J. The relation of growth from birth to 2 years to sex, parental size and other factors, using Rao's method of the transformed time scale. Hum Biol **1970**; 42:105–31.
25. Guo SM, Roche AF, Fomon SJ, et al. Reference data on gains in weight and length during the first two years of life. J Pediatr **1991**; 119: 355–62.
26. Berkey CS, Reed RB. A model for describing normal and abnormal growth in early childhood. Hum Biol **1987**; 59:973–87.
27. Jenss RM, Bayley N. A mathematical method for studying the growth of a child. Hum Biol **1937**; 9:556–63.
28. Karlberg J, Jalil F, Lam B, Low L, Yeung CY. Linear growth retardation in relation to the three phases of growth. Eur J Clin Nutr **1994**; 48(suppl 1): S25–43.
29. Peerson JM, Heinig MJ, Nommsen LA, Lonnerdal B, Dewey KG. Use of growth models to describe patterns of length, weight, and head circumference among breast-fed and formula-fed infants: the DARLING Study. Hum Biol **1993**; 65:611–26.
30. Checkley W, Epstein LD, Gilman RH, Cabrera L, Black RE. Effects of acute diarrhea on linear growth in Peruvian children. Am J Epidemiol **2003**; 157:166–75.
31. Richard SA, Black RE, Gilman RH, et al. Diarrhea in early childhood: short-term association with weight and long-term association with length. Am J Epidemiol **2013**; 178:1129–38.
32. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. Biometrics **1988**; 44:1049–60.
33. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics **1982**; 38:963–74.
34. Shankar AH, Prasad AS. Zinc and immune function: the biological basis of altered resistance to infection. Am J Clin Nutr **1998**; 68(2 suppl): 447S–63S.
35. Wei Y, Pere A, Koenker R, He X. Quantile regression methods for reference growth charts. Stat Med **2006**; 25:1369–82.
36. Odum HT. Systems ecology: an introduction. New York, NY: John Wiley and Sons, **1983**.
37. West GB, Brown JH, Enquist BJ. A general model for ontogenetic growth. Nature **2001**; 413:628–31.