Recent Supplements to Acta Paediatrica

436 *UK Hot Topics in Neonatology.*
Edited by **A Greenough.** 2001

437 *UK Hot Topics in Neonatology.*
Edited by **A Greenough.** 2002

438 *Neonatal Hematology and Immunology. International Workshop and Conference, Orlando, Florida, November 14–16, 2002.*
Edited by **RD Christensen.** 2002

439 *Lysosomal Storage Diseases, Fabry disease: clinical heterogeneity and management challenges.*
Proceedings of the 2nd International Symposium, Cannes, April 2002.
Edited by **M Beck, TM Cox and MT Vanier.** 2002

440 *CPICS Child and Parents' Interaction Coding System in Dyads and Triads.*
Edited by **M Hedenbro and A Lidén.** 2002

441 *Aspects on Infant Nutrition. Proceedings of the Giovinazzo Symposium 2001.*
Edited by **S Vigi and A Marini.** 2003

442 *Nutrition and Brain Development of the Infant.*
Edited by **PR Guesry, C Garcia-Rodenas and J Rey.** 2003

443 *Lysosomal Diseases: Pathophysiology and Therapy. Proceedings of the 3rd International Symposium, Santiago de Compostela, May 2003.*
Edited by **M Beck, TM Cox and R Ricci.** 2003

444 *UK Hot Topics in Neonatology*
Edited by **A Greenough.** 2004

445 *Cutting Edge Information in Pediatrics: The Ospedale Pediatricio Bambino Gesú (OPBG)/Mayo Clinic Experience*
Edited by **G Franco and RM Jacobson.** 2004

446 *Coronary Arteries in Children. Anatomy, Flow and Function*
Edited by **E Pesonen and L Holmberg.** 2004

447 *Lysosomal diseases: natural course, pathology and therapy*
Edited by **M Beck, TM Cox, AB Mehta and U Widmer.** 2005

448 *1st Scandinavian Pediatric Obesity Conference 2004*
Edited by **Carl Erik Flodmark, Inge Lissau and Angelo Pictrobelli.** 2005

449 *Current Issues on Infant Nutrition*
Edited by **Fabio Mosca, Silvia Fanaro and Vittorio Vigi.** 2005

ACTA PÆDIATRICA
INTERNATIONAL JOURNAL OF PÆDIATRICS

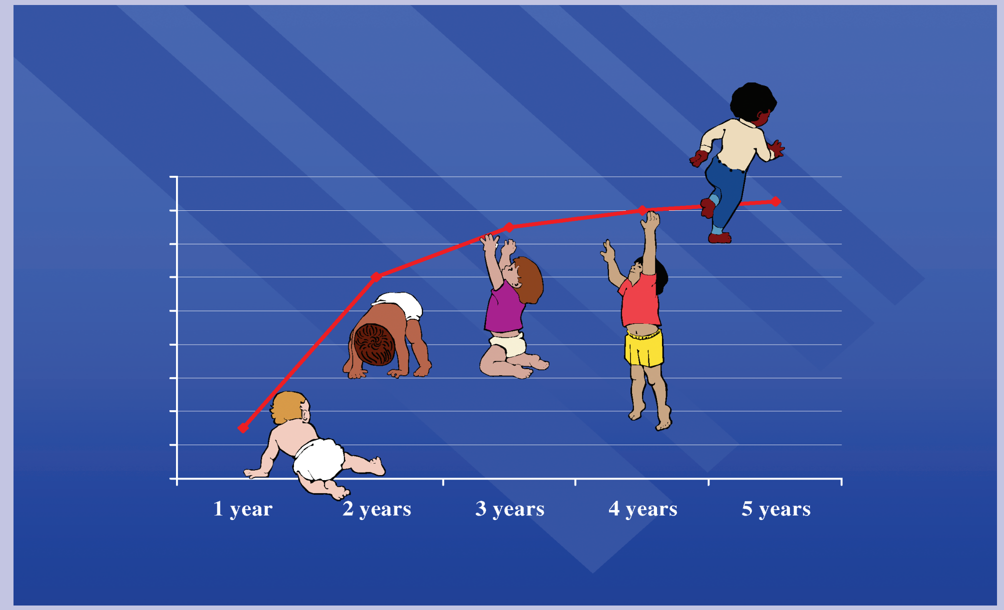
ACTA PÆDIATRICA VOLUME 95 APRIL 2006 SUPPLEMENT 450 PAGES 1–104

ISSN 0803-5326
VOLUME 95 APRIL 2006 SUPPLEMENT 450

ACTA PÆDIATRICA

INTERNATIONAL JOURNAL OF PÆDIATRICS

WHO Child Growth Standards



Guest Editors

Mercedes de Onis
Cutberto Garza
Adelheid W. Onyango
Reynaldo Martorell

Taylor & Francis
Taylor & Francis Group
www.tandf.no/paed

ACTA PAEDIATRICA publishes papers in English covering both clinical and experimental research, in all fields of paediatrics including developmental physiology. *Acta Paediatrica* (formerly *Acta Paediatrica Scandinavica*) will consider contributions from all countries. Articles are accepted for publication on the condition that they have not been submitted to any other journal. *Acta Paediatrica* accepts review articles, original articles, short communications, therapeutic notes and case reports. Review articles should give the present state-of-the-art on topics of clinical importance and include an internationally relevant bibliography. Case reports are accepted only if they provide new data that will improve the understanding, diagnosis, treatment and prevention of a particular disease. Short communications and therapeutic notes are intended as preliminary reports.
A Correspondence section constitutes a forum for comments and short discussions.

Correspondence concerning manuscripts and editorial matters should be addressed to: Acta Paediatrica, International Journal of Paediatrics, Building X5:01, (Borgmästarvillan), Karolinska University Hospital, Karolinska vägen 29, SE-171 76 Stockholm, Sweden. Tel: +46 8 517 724 87; Fax: +46 8 517 740 34; E-mail: mail@actapaediatrica.se. Assistant Editors: Cathrin Andersson, Ann-Christin Lundgren.

Correspondence concerning copyright, requests for permissions, should be addressed to: Taylor & Francis, Marie Larsson, Box 3255, SE-103 65 Stockholm, Sweden.
Tel: +46 8 440 80 57; Fax: +46 8 440 80 50;
E-mail: marie.larsson@se.tandf.no

Correspondence concerning commercial reprints and permissions should be addressed to: Taylor & Francis, Sales Dept., Att: Johanna Rydhem, PO Box 12 Posthuset, NO-0051 Oslo, Norway, Tel: +47 2310 3460; Fax: +47 2310 3461;
E-mail: johanna.rydhem@tandf.no

Correspondence concerning advertising should be addressed to: sales@tandf.no

Correspondence concerning subscriptions, distribution, and back issues should be addressed to: Taylor & Francis, an Informa Business, Customer Services, Sheepen Place, Colchester, Essex, CO3 3LP, UK. Tel: +44 (0) 20 7017 5544; Fax: +44 (0) 20 7017 5198; E-mail: tf.enquiries@tfinforma.com
Customers in the US and Canada: ***U.S. Address:*** Taylor & Francis Group, Journals Customer Service, 325 Chestnut Street, 8th Floor, Philadelphia, PA 19106, Tel: +1 (800) 354-1420 or +1 (215) 625-8900; Fax: +1 (215) 625-2940;
E-mail: info@taylorandfrancis.com

Acta Paediatrica, ISSN 0803-5253, is published monthly by Taylor & Francis, an Informa Business, *Acta Paediatrica, ISSN 0803-5253*.

Subscription rates, Vol. 95, 2006: *Institutions*: USD694. *Individual*: USD300. Prices include air speed delivery.

ACTA PAEDIATRICA (USPS permit number 007937) is published monthly. The 2006 US institutional subscription price is $694. Periodicals postage paid at Jamaica, NY by US Mailing Agent Air Business Ltd. C/O Priority Airfreight NY Ltd, 147–29 182nd Street, Jamaica, NY 11413. US Postmaster: Please send address changes to Air Business Ltd, C/O Priority Airfreight NY Ltd, 147–29 182nd Street, Jamaica, NY 11413.

U.S. Address: Taylor & Francis Group, Journals Customer Service, 325 Chestnut Street, 8th Floor, Philadelphia, PA 19106,
Tel: +1 (800) 354-1420 or +1 (215) 625-8900;
Fax: +1 (215) 625-2940; E-mail: info@taylorandfrancis.com

Subscriptions in Japan should be ordered through the Maruzen Co. Ltd., 3–10 Nihonbashi 2-chome, Chuo-ku, Tokyo 103, Japan.

Articles in *Acta Paediatrica* are covered by the following indexing and abstracting services: Biological Abstracts; BIOSIS; Chemical Abstracts; Current Clinical Cancer; Current Contents/Clinical Medicine; Elsevier BIOBASE/Current Awareness in Biological Sciences; EMBASE/Excerpta Medica; FaxonFinder, Index Medicus/MEDLINE; Nutrition Abstracts; Science Citation Index; Sci Search; Automatic Subject Citation Index; Bibliography of Developmental Medicine and Child Neurology; Current Advances in Genetics and Molecular Biology; Current Advances in Ecological Sciences; CIS Abstracts; Dokumentation Arbeitsmedizin; Index to Dental Literature; Helminthological Abstracts; Medical Documentation Service; Nutrition Research Newsletters; Protozoological Abstracts; Reference Update Research Alert; Risk Abstracts; SIIC UnCover.

Acta Paediatrica is included in the ADONIS service. Accordingly, copies of individual articles are available on compact disks (CD-ROM) and can be printed out on demand. An explanatory leaflet is available from the publisher.

PRINTERS
Typeset by Datapage Intl, Dublin and printed by Henry Ling.

Instructions for Authors: www.tandf.no/paed

PREPARING FOR SUBMISSION
Submitted manuscripts should be arranged according to the rules stated in "Uniform requirements for manuscripts submitted to biomedical journals" Ann Intern Med 1997; 126: 36–47, or JAMA 1997; 277: 927–34. The full document is available at www.icmje.org

When submitting a paper, the author should always make a full statement to the editor about all submissions and previous reports that might be regarded as redundant or duplicate publication of the same or very similar work.

However, publication of abstracts and publication in a minority language is not considered to be a duplicate publication. But authors are requested to report if any such publication has occurred. Submit approval of the paper for publication, signed by all authors, to the Editorial Office. In case research has been supported by pharmaceutical or other industries this should be stated. An author must have made significant contributions to the design and execution and analysis and writing of the study, and he or she must share responsibility for what is published. We ask authors to specify their individual contributions (Contributors' List) as concise as possible and, if appropriate, we publish this information. Regular papers exceeding six printed pages (including illustrations, tables and references) will incur a page charge, USD 200, for each exceeding page. Short Communications may not exceed two printed pages and Clinical Observations (Case Reports) three printed pages. As to Clinical Observations, we only accept reports containing new data that will improve the understanding, diagnosis, treatment and prevention of a particular disease.

Reports on randomised trials must conform to Consort guidelines and should be submitted with their protocols.

Conflict of interest and funding: Authors are responsible for recognising and disclosing financial and other conflicts of interest that might bias their work. They should acknowledge in the manuscript all financial support for the work and other financial or personal connections to the work.

Statistic validity: If statistical data are provided the authors may be requested to submit an official statement issued by a certified statistician (with a proper affiliation) regarding the validity of methods used.

Ethics and consent: When reporting experiments on human subjects, indicate whether the procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation and with the Helsinki Declaration of 1975, as revised in 1983. Do not use patients' names, initials, or hospital numbers, especially in illustrative material. Papers including animal experiments or clinical trials must be approved by the institutional ethics committee.

Identifying information should not be published in written descriptions, photographs, and pedigrees unless the information is essential for scientific purposes and the patient (or parent or guardian) gives written informed consent for publication. Informed consent for this purpose requires that the patient be shown the manuscript to be published.

Copyright: It is a condition of publication that authors assign copyright or licence the publication rights in their articles, including abstracts, to Taylor & Francis. This enables us to ensure full copyright protection and to disseminate the article, and of course the Journal, to the widest possible readership in print and electronic formats as appropriate. Authors may, of course, use the article elsewhere *after* publication without prior permission from Taylor & Francis, provided that acknowledgement is given to the Journal as the original source of publication, and that Taylor & Francis is notified so that our records show that its use is properly authorised. Authors retain a number of other rights under the Taylor & Francis rights policies documents. These policies are referred to at http://www.tandf.co.uk/journals/authorrights.pdf for full details. Authors are themselves responsible for obtaining permission to reproduce copyright material from other sources.

SUBMISSION
Electronic Manuscripts
All submissions should be made online at **Acta Paediatrica's Manuscript Central site http://mc.manuscriptcentral.com/spae** to facilitate rapid accessibility of your work to the readers. New users should first create an account. Once a user is logged onto the site submissions should be made via the Author Centre. For assistance with any aspect of the site, please refer to the User Guide which is accessed via the 'Get Help Now' button at the top right of every screen.

Manuscripts Layout
Please use these simple guidelines when preparing your electronic manuscript.

Guidelines: (i) Key elements consistently throughout. (ii) Do not break words at the ends of lines. Use a hyphen only to hyphenate compound words. (iii) One word space only at the ends of sentences. (iv) Do not use underlining; use the italics feature instead. (v) Leave the right-hand margin unjustified. (vi) Use a double hyphen to indicate a dash. (vii) Do not use the lower case "ell" for 1 (one) or the upper case O for 0 (zero). (viii) When indenting paragraphs or separating columns in tables, use the TAB key, not the spacebar.

Authors should note that where submissions exceed the permitted number of pages (see table at the end of this document) each exceeding page incurs a charge of USD 200.

Double space the entire manuscript and use the SI system of notation. Prepare the manuscript with each of the following parts starting on a new page: (1) The title, with authors' names and affiliations (as a rule the number of authors should be limited to six. The names of others who contributed to the article in varying degree should be mentioned under the heading "Acknowledgements"), the address of the corresponding author and a short running title; (2) The abstract ending with one or two sentences of conclusion, summarising the message of the article including keywords; (3) The text; (4) The references; (5) The tables; (6) The figure legends.

LANGUAGE Manuscripts must be in English. Authors from non-English speaking countries are requested to have their text thoroughly checked by a competent person whose native language is English. Manuscripts may be rejected on the grounds of poor English. Revision of the language is the responsibility of the author.

NOTES/FOOTNOTES Incorporate notes/footnotes in the text, within parentheses, rather than in their usual place at the foot of the page.

ABBREVIATIONS Do not use abbreviations in the title or Abstract, and in the text use only standard abbreviations, i.e. those listed in the latest editions of any of the recognized medical dictionaries (e.g. Dorland's, Butterworth's). The full term for which an abbreviation stands has to precede its first use in the text, unless it is a standard unit of measurement. Redefine abbreviations used in the figure legends.

ILLUSTRATIONS Graphic elements should be kept as separate files in EPS-, PDF- or TIFF-format. These formats guarantee that the quality of the graphics is good throughout the publishing process if provided within a sufficient resolution. Photo illustrations should have at least 300DPI and please use CMYK colour conversion. Halftones and colour photos should be enclosed separately. Please state in the page charge agreement which figures you wish to print in colour. Colour printing incurs a charge, USD 865 per page. If you want to print in black and white, please provide black and white originals, if possible. Glossy photocopies or good quality hard copies are to be preferred rather than low-resolution electronic files. In case a manuscript contains photographs of patients, we require a certificate by the author that consent to publish such a photograph has been given by the child's parent or caretaker. Please submit four originals. Illustrations will only be returned to the author(s) on request.

TITLE PAGE Example of a title page manuscript showing content, underlining (for italics) and spacing. Avoid subtitles. (Leave 7–8 cm space at the top of the page):

Mechanics of breathing in the newborn (title)

L Andersson and K Pettersson (authors)

Department of Paediatrics, University Hospital, Lund, Sweden

Short title: Studies in neonatal hypoglycaemia

Corresponding author: K. Pettersson, Department of Paediatrics, University Hospital, S-221 85 Lund, Sweden. Tel: +00 0 000 00 00; Fax: +00 0 000 00 00; E-mail: K.Pettersson@lu.se

ABSTRACT The abstract of a regular article should not exceed 200 words and be structured with following headings: Aim, Methods, Results and end with one or two sentences of Conclusion summarising the message of the article, including max. 5 keywords listed alphabetically. Type as illustrated below: More detailed information can be found at www.tandf.no/paed

Abstract

Huppke P, Roth C, Christen HJ, Brockmann K, Hanefeld F. Endocrinological study on growth retardation in Rett syndrome. Acta Paediatr 2001;90:1257–1261. Stockholm. ISSN 0803-5253

Aim: To determine whether primary or secondary growth hormone … (text) *Methods:* In 38 patients with Rett syndrome … Results: … Conclusion: … Keywords: *Endocrinology, growth hormone, growth retardation …*

TEXT PAGES Leave a left-hand margin of about 4 cm. Number the pages in the top right-hand corner, beginning with the title page. Headings (left-hand margin): Patients and Methods, Results, Discussion, Acknowledgements, References.

REFERENCES Number the references consecutively in the order in which they are first mentioned in the text. Identify references in the text, tables and legends by arabic numerals (in parentheses). Type list of references as illustrated. Observe the punctuation carefully. The number of references should not exceed 30 in regular articles. (When more than six authors, list first six and add et al).
Abbreviations of journal titles; please consult the *List of Journals Indexed in Index Medicus*, published annually as a list in the January issue of *Index Medicus*, also available at www.nlm.nih.gov. More detailed information can be found at www.tandf.no/paed

For journal article in *electronic* format use the following style: Morse SS. Factors in the emergence of infectious diseases. Emerg Infect Dis [serial online] 1995 Jan-Mar [cited 1996 Jun 5]; 1(1): [24 screens]. Available from: URL: http://www.cdc.gov\ncidod\EID\eid.htm

PROOFS AND REPRINTS Page proofs will be sent to the corresponding author. Return the master proof and the offprint order form within three days, by air mail, to Taylor & Francis, P.O. Box 3255, SE-103 65 Stockholm, Sweden

CONTRIBUTORS' LIST (example)
Dr A had primary responsibility for protocol development, patient screening, enrolment, outcome assessment, preliminary data analysis and writing the manuscript.

Drs B and C participated in the development of the protocol and analytic framework for the study, and contributed to the writing of the manuscript.

Dr D contributed as B and C, and was responsible for patient screening.

Dr E supervised the design and execution of the study, performed the final data analyses and contributed to the writing of the manuscript.

For more specific guidelines, information and support visit www.tandf.no/paed or send an e-mail to; post@tandf.no.

WHO Child Growth Standards

Guest Editors

Mercedes de Onis
Cutberto Garza
Adelheid W. Onyango
Reynaldo Martorell

Contents

Acta Pædiatrica, 2006; 450: 5–6

Taylor & Francis
Taylor & Francis Group

FOREWORD

Growth charts are an essential component of the paediatric toolkit. Their value resides in helping to determine the degree to which physiological needs for growth and development are being met during the important childhood period. However, their usefulness goes far beyond assessing children's nutritional status. Many governmental and United Nations agencies rely on growth charts for measuring the general well-being of populations, formulating health and related policies, and planning interventions and monitoring their effectiveness.

The origin of the WHO Child Growth Standards dates back to the early 1990s and the appointment of a group of experts to conduct a meticulous evaluation of the National Center for Health Statistics/World Health Organization (NCHS/WHO) growth reference, which had been recommended for international use since the late 1970s. This initial phase documented the deficiencies of the reference and led to a plan for developing new growth charts that would document how children *should* grow in all countries rather than merely describing *how* they grew at a particular time and place. The experts underscored the importance of ensuring that the new growth charts were consistent with "best" health practices.

A logical outcome of this plan was the WHO Multicentre Growth Reference Study (MGRS), which was implemented between 1997 and 2003 and serves as a model of collaboration for conducting international research. The MGRS is unique in that it was purposely designed to produce a standard rather than a reference. Although standards and references both serve as a basis for comparison, each enables a different interpretation. Since a standard defines how children should grow, deviations from the pattern it describes are evidence of abnormal growth. A reference, on the other hand, does not provide as sound a basis for such value judgements, although in practice references often are mistakenly used as standards.

The MGRS data provide a solid foundation for developing a standard because they are based on healthy children living under conditions likely to favour achievement of their full genetic growth potential. Furthermore, the mothers of the children selected for the construction of the standards engaged in fundamental health-promoting practices, namely breastfeeding and not smoking.

A second feature of the study that makes it attractive as a standard for application everywhere is that it included children from a diverse set of countries: Brazil, Ghana, India, Norway, Oman and the USA. By selecting privileged, healthy populations the study reduced the impact of environmental variation. Nevertheless, the sample had considerable built-in ethnic or genetic variability in addition to cultural variation in how children are nurtured, which further strengthens the standard's universal applicability.

A key characteristic of the new standards is that they explicitly identify breastfeeding as the biological norm and establish the breastfed child as the normative model for growth and development. Another distinguishing feature of the new standards is that they include windows of achievement for six gross motor development milestones. In the past, although WHO issued recommendations concerning attained physical growth, it had not previously made recommendations for assessing motor development.

This supplement, which presents the first set of the new WHO Child Growth Standards and related data, is divided into five sections. The first three papers provide an overview of the MGRS sample statistics and baseline characteristics, document compliance with the study's feeding criteria, and describe the sample's breastfeeding and complementary feeding practices. The following two papers describe the methods used to standardize the assessment of anthropometric measurements and motor development assessments, and present estimates of the assessments' reliability. The sixth and seventh papers examine differences in linear growth and motor milestone achievement among populations and between sexes, and evaluate the appropriateness of pooling data for the purpose of constructing a single international standard. Next is an overview of the methods used to construct the growth standards based on length/height, weight and age, followed by the windows of achievement for the six gross motor development milestones, and the resulting growth curves and actual windows of achievement. The tenth and final paper examines the relationship between physical growth indicators and ages of achievement of gross motor milestones in the sample population used to construct the standards.

ISSN 0803-5326 print/ISSN 1651-2227 online
DOI: 10.1080/08035320500495373

The WHO Child Growth Standards provide a technically robust tool for assessing the well-being of infants and young children. By replacing the NCHS/WHO growth reference, which is based on children from a single country, with one based on an international group of children, we recognize that children the world over grow similarly when their health and care needs are met. In the same way, by linking physical growth to motor development, we underscore the crucial point that although normal physical growth is a necessary enabler of human development, it is insufficient on its own. Together, three new elements—a *prescriptive* approach that moves beyond the development of growth references towards a standard, inclusion of children from around the world, and links to motor development—provide a solid instrument for helping to meet the health and nutritional needs of all the world's children.

Mercedes de Onis, World Health Organization
Cutberto Garza, United Nations University

Acta Pædiatrica, 2006; Suppl 450: 7–15

Enrolment and baseline characteristics in the WHO Multicentre Growth Reference Study

WHO MULTICENTRE GROWTH REFERENCE STUDY GROUP[1,2]

[1]*Department of Nutrition, World Health Organization, Geneva, Switzerland, and* [2]*Members of the WHO Multicentre Growth Reference Study Group are listed at the end of this paper*

Abstract

Aim: To describe the WHO Multicentre Growth Reference Study (MGRS) sample with regard to screening, recruitment, compliance, sample retention and baseline characteristics. *Methods:* A multi-country community-based study combining a longitudinal follow-up from birth to 24 mo with a cross-sectional survey of children aged 18 to 71 mo. Study subpopulations had to have socio-economic conditions favourable to growth, low mobility and $\geq$ 20% of mothers practising breastfeeding. Individual inclusion criteria were no known environmental constraints on growth, adherence to MGRS feeding recommendations, no maternal smoking, single term birth and no significant morbidity. For the longitudinal sample, mothers and newborns were screened and enrolled at birth and visited 21 times at home until age 24 mo. *Results:* About 83% of 13 741 subjects screened for the longitudinal component were ineligible and 5% refused to participate. Low socio-economic status was the predominant reason for ineligibility in Brazil, Ghana, India and Oman, while parental refusal was the main reason for non-participation in Norway and USA. Overall, 88.5% of enrolled subjects completed the 24-mo follow-up, and 51% (888) complied with the MGRS feeding and no-smoking criteria. For the cross-sectional component, 69% of 21 510 subjects screened were excluded for similar reasons as for the longitudinal component. Although low birthweight was not an exclusion criterion, its prevalence was low (2.1% and 3.2% in the longitudinal and cross-sectional samples, respectively). Parental education was high, between 14 and 15 y of education on average.

Conclusion: The MGRS criteria were effective in selecting healthy children with comparable affluent backgrounds across sites and similar characteristics between longitudinal and cross-sectional samples within sites.

Key Words: *Child nutrition, growth standards, longitudinal study, socio-economic status, survey methodology*

Introduction

The origin of the WHO Multicentre Growth Reference Study (MGRS) [1] dates back to the early 1990s when the World Health Organization (WHO) initiated a comprehensive review of the uses and interpretation of anthropometric references and conducted an in-depth analysis of growth data from breastfed infants [2,3]. This analysis showed that the growth pattern of healthy breastfed infants deviated to a significant extent from the National Center for Health Statistics (NCHS)/WHO international reference [2,3]. The review group concluded from these and other related findings that the NCHS/WHO reference did not adequately describe the physiological growth of children and that its use to monitor the health and nutrition of individual children, or to derive estimates of child malnutrition in populations, was flawed. Moreover, the review group recommended that a standard rather than a reference be constructed, adopting a novel approach that would describe how children *should* grow when free of disease and when their care follows healthy practices such as breastfeeding and non-smoking [1]. The MGRS was launched in 1997 [4] and drew the participation of children from six sites around the world: Brazil (South America), Ghana (Africa), India (Asia), Norway (Europe), Oman (Middle East) and the USA (North America). The growth charts that have been constructed based on the MGRS data are presented in a companion paper in this supplement [5]. The objective of this paper is to provide an overview of the MGRS sample with regard to screening, recruitment, sample attrition, and compliance with the study's feeding and no-smoking criteria. We also provide a description of the baseline characteristics of the study sample.

Correspondence: Mercedes de Onis, Study Coordinator, Department of Nutrition, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland. Tel: +41 22 791 3320. Fax: +41 22 791 4156. E-mail: deonism@who.int

ISSN 0803-5326 print/ISSN 1651-2227 online
DOI: 10.1080/08035320500495407

Methods

The MGRS (1997–2003) had two components: a longitudinal follow-up in which children were recruited at birth and followed up at home until they were 24 mo of age, and a cross-sectional survey involving children aged 18 to 71 mo. The study populations lived under socio-economic conditions favourable to growth, with low mobility and ≥20% of mothers practising breastfeeding [4]. As part of the site selection process in Ghana, India and Oman, surveys were conducted to identify socio-economic characteristics that could be used to select groups whose growth was not environmentally constrained [6–8]. Local criteria for screening newborns, based on parental education and/or income levels, were developed from those surveys. Pre-existing survey data were available from Brazil, Norway and the USA for this purpose [9–11]. A detailed description of the MGRS protocol and its implementation in the six sites has been provided elsewhere [4,9–14].

Longitudinal sample

Infants for the longitudinal component were recruited from selected hospitals and clinics where at least 80% of the subpopulations of interest delivered. Within 24 h of birth, mother–infant pairs were screened for participation in the study. Subjects were enrolled if they met the study's eligibility criteria: specifically, no environmental or economic constraints on growth, mothers' willingness to follow the study's feeding recommendations (i.e. exclusive or predominant breastfeeding for at least 4 mo, introduction of complementary foods by the age of 6 mo, and partial breastfeeding continued to age ≥12 mo), gestational age ≥37 completed weeks and <42 wk, single birth, non-smoking mother, and the absence of significant morbidity in the newborn [4]. Due to large numbers of maternity facilities used by the subpopulations targeted for the MGRS in Ghana and India, these sites implemented a two-stage screening procedure. First, newly delivered mothers in Ghana were pre-screened on area of residence and socio-economic status [12], while in India pre-screening took place during pregnancy [13]. The second and final screening stage at both sites was completed within 24 h of birth. Following screening, children were classified as eligible if all criteria had been met or ineligible if one or more eligibility criteria had not been met. The former were invited to participate in the study.

At the first follow-up home visit (2 wk after delivery) mothers were re-screened to confirm eligibility. This enabled study teams to identify "hidden refusals" (those who repealed their decision to participate) and "hidden ineligibles" (e.g. mothers who had not complied with the feeding recommendations). These infants were dropped from the study and replaced in the sample. Thus, at 2 wk, all children screened for the longitudinal follow-up fell into one of three categories: 1) enrolled subjects; 2) ineligible (including ineligibles identified at first contact and hidden ineligibles); and 3) refusals (those who refused at first contact and hidden refusals). Those who left the study after this point were considered dropouts and were not replaced in the sample [4]. Only children of mothers who complied with the MGRS feeding and no-smoking criteria have been included in the growth standards' sample [5]. However, regardless of compliance status, the entire cohort was followed up.

Cross-sectional sample

Children aged 18 to 71 mo were targeted for the cross-sectional component, with recruitment strategies varying by site. In Brazil, India and the USA, children were recruited through a door-to-door survey of selected study areas. In Norway and Oman, children were identified through a national or health registry, and in Ghana, from crèches and nursery schools. Details of the sampling procedures employed at each site are provided elsewhere [9–14]. The cross-sectional survey sampling strategy aimed at recruiting children with backgrounds similar to those in the longitudinal sample. Thus, the same exclusion criteria and site-specific socio-economic criteria were applied, with the exception of infant feeding practices, where a minimum duration of 3 mo of any breastfeeding was required for inclusion in the cross-sectional sample [4].

Results

Longitudinal sample

Tables I and II show the enrolment statistics and reasons for ineligibility by study site for the longitudinal component. Out of 13 741 mother–infant pairs screened, 1743 (12.7%) were enrolled in the longitudinal sample (Table I). Overall, about 83% of the subjects screened were ineligible (ranging between 30.9% in the USA and 91.8% in Brazil) and about 5% refused to participate (mainly in the USA, Norway and India). Inability to meet the study's socio-economic criteria was the main reason for ineligibility in Brazil (54.3%), Ghana (74.2%), India (24.4%) and Oman (47.3%) (Table II). Smoking accounted for 19% and 9.2% of the total ineligibility in Brazil and Norway, respectively. The two main reasons for ineligibility, i.e. residence out of study area and low socio-economic status, together accounted for 71.2% of the exclusions (Table II).

Table I. Enrolment statistics for the longitudinal sample by site.

	Brazil	Ghana	India	Norway	Oman	USA	All
Pre-screened[a], *n*		1519	259				1778
Screened, *n*	4801	538	433	836	4957	398	11963
Ineligibles[b], *n* (%)	4407 (91.8)	1681 (81.7)	310 (44.8)	402 (48.1)	4428 (89.3)	123 (30.9)	11351 (82.6)
Refusals[c], *n* (%)	84 (1.7)	47 (2.3)	81 (11.7)	134 (16.0)	234 (4.7)	67 (16.8)	647 (4.7)
Enrolled at 2 wk, *n* (%)	310 (6.5)	329 (16.0)	301 (43.5)	300 (35.9)	295 (6.0)	208 (52.3)	1743 (12.7)

[a] The total number of pre-screened subjects in Ghana and India are 2057 and 692, respectively, including 538 (Ghana) and 433 (India) that completed screening at birth.
[b] Ineligibles: ineligibles at first hospital contact plus hidden ineligibles at 2 wk.
[c] Refusals: refusals at first hospital contact plus hidden refusals at 2 wk.

Overall, 888 (50.9%) mother–child pairs complied with the study's feeding and no-smoking criteria and completed the 2-y follow-up, ranging across sites from 21.6% in Brazil to 69.3% in Ghana (Table III). The great majority of compliant children (96%) completed the study. Attrition (dropout) rates and reasons for discontinuing participation are summarized in Table IV. Only 11.5% of the enrolled subjects failed to complete the 24-mo follow-up. The main reasons across sites for dropping out were the family moving out of the study area (57.7%) and the parents' request (33.8%).

The characteristics of the families enrolled in the longitudinal component are shown in Table V. The majority of the families had fewer than three children, the median number of children being two for the entire sample. Parental educational attainment was generally high across sites. About 59% of mothers and 63% of fathers had completed at least 15 y of education, 89% of both parents having completed at least 10 y of education. Mean maternal age for the entire sample was 29.4 y. As expected, fathers were taller (175.1 cm) than mothers (161.6 cm), Norwegian parents being the tallest and Omanis the shortest among the sites. Household monthly income was standardized by converting to US dollar equivalents based on the exchange rate prevailing at the beginning of the study in each site. In Ghana the exchange rates were different for the longitudinal and cross-sectional components because of local currency devaluation between the starting points of the two components. Over 99% of families in the longitudinal sample had access to piped water, a flush toilet, refrigerator and a gas or electric cooker, and over 93% and 86% had telephones and cars, respectively.

The characteristics of the children in the longitudinal sample (Table VI) indicate about 73% vaginal deliveries, high Apgar scores at 1 and 5 min, and a low prevalence of low birthweight across sites (overall 2.1%). The all-site mean birthweight, length and head circumference were 3.3 kg, 49.6 cm and 34.2 cm, respectively.

Table II. Reasons for ineligibility for the longitudinal sample by site.

	Brazil	Ghana	India	Norway	Oman	USA	All
Total screened (*n*)	4801	2057	692	836	4957	398	13741
Total ineligible (*n*)	4407	1681	310	402	4428	123	11351
Reasons for ineligibility[a] *(%)*							
Resides out of study area	24.9	11.4	6.2	14.2	31.2	0.0	22.8
Multiple birth	2.2	0.8	0.0	2.9	1.3	0.8	1.5
Perinatal morbidity[b]	6.1	1.3	1.7	12.2	5.0	5.8	5.1
Gestational age outside range	8.7	1.5	4.5	6.2	6.5	3.3	6.3
Breastfeeding non-compliance	1.0	0.2	6.1	1.2	6.7	14.1	3.6
Mother is a smoker	19.0	0.1	0.4	9.2	0.6	1.5	7.5
Low socio-economic status	54.3	74.2	24.4	0.0	47.3	0.8	48.4
Language difficulty	0.0	0.0	0.0	6.8	14.0	4.3	5.6
Late notification of birth	0.0	1.2	1.2	0.0	0.0	1.8	0.3
Incomplete screening	1.9	0.0	0.0	0.0	0.2	0.0	0.7
Child illness/death	0.0	0.1	0.0	0.5	0.2	1.0	0.2
Moving away	0.1	0.6	0.6	0.1	0.5	0.5	0.4
Other reasons	0.1	0.0	0.0	0.0	0.1	0.0	0.1

[a] The ineligibility tally may exceed 100% because of subjects excluded for multiple reasons.
[b] High perinatal morbidity in Norway is due to breech births.

Table III. Compliance with feeding and no-smoking criteria in the longitudinal sample by site.

	Brazil	Ghana	India	Norway	Oman	USA	All
Enrolled at 2 wk	310	329	301	300	295	208	1743
Compliant, study completed, *n* (%)	67 (21.6)	228 (69.3)	173 (57.5)	148 (49.4)	153 (51.8)	119 (57.2)	888 (50.9)
Compliant, study not completed, *n* (%)	3 (1.0)	6 (1.8)	8 (2.6)	7 (2.3)	4 (1.4)	10 (4.8)	38 (2.2)
Not compliant, study completed, *n* (%)	220 (71.0)	64 (19.5)	96 (31.9)	114 (38.0)	107 (36.3)	53 (25.5)	654 (37.5)
Not compliant, study not completed, *n* (%)	20 (6.4)	31 (9.4)	24 (8.0)	31 (10.3)	31 (10.5)	26 (12.5)	163 (9.4)

Cross-sectional sample

Table VII presents enrolment statistics for the cross-sectional component. A total of 21 510 children were screened in the six countries, ranging from 837 in the USA to 5185 in Norway. Of these, 6697 (31.1%) were enrolled in the study. The common reasons for exclusion were low socio-economic status (ranging from nil in Norway to 64.1% in Oman), maternal smoking (0.1% in Ghana to 28.5% in Brazil), gestational age outside range (2.8% in Oman to 16.3% in Norway), child breastfed for less than 3 mo (1.4% in Oman to 28.7% in Brazil) and residence outside the study area (nil in Norway and USA to 23.3% in India). Refusal to participate in the study was lowest in Brazil (0.1%) and highest in Norway (11.8%). The "other exclusions" in Ghana (25.9%) and Norway (18.9%) were for varied reasons, including inability to contact the family, and children who had travelled out of the area or had outgrown the maximum age limit for the study.

Average years of schooling for fathers ranged from about 11 in Brazil to 19 in Ghana, and for mothers from 11 y in Brazil to 17 y in India (Table VIII). For a median number of two children per family (range 1 to 15), the average maternal age of 33 y was high. Average maternal weights were between 62.6 kg in India and 74.5 kg in Ghana. Mothers in Norway were the tallest (167.7 cm) and those in Oman the shortest (156.6 cm), as was the case in the longitudinal sample. Although incomes expressed in US dollars varied widely among sites (lowest in Ghana and highest in Norway), the populations selected for the study in the developing country sites belonged to the upper socio-economic strata, while in Norway and the USA all socio-economic groups were included. Other socio-economic status markers, as assessed by ownership of material goods, ranged from 91.1% for cars overall to 99.8% for gas/electric cookers and refrigerators (Table VIII).

With regard to the baseline characteristics of enrolled children (Table IX), as was the case in the longitudinal sample, there was a slight predominance of males (51.7%) in the total sample, primarily due to the higher percentage of male children (56.5%) in the Indian sample. Overall, a quarter of deliveries were by caesarean section, with the highest rates in Brazil (55.6%) and India (36.2%) and the lowest rates in Oman, Norway and the USA (12–14%). The average birthweight was 3338 g, infants from Norway being the heaviest at birth (3636 g). The average duration of breastfeeding ranged from 12 mo in Brazil to 17 mo in Oman. Infant formula or other milks were introduced at mean ages ranging from 5.2 mo in Oman to 12.4 mo in the USA, and solids/semi-solids between 4.1 mo in Oman and 5.8 mo in Ghana (Table IX).

Discussion

The MGRS was designed to describe how children should grow under optimal conditions in any setting. To achieve this aim, a prescriptive approach was adopted for the study [4]. This paper summarizes the characteristics of children who were enrolled in the MGRS after application of selection criteria aimed at accessing children with unconstrained growth. Not surprisingly, high rates of ineligibility due to low socio-economic status were reported in Brazil, Ghana, India and Oman. On the other hand, parental refusal to participate in the study was the main reason for

Table IV. Follow-up rate and reasons for dropout in the longitudinal sample by site.

	Brazil	Ghana	India	Norway	Oman	USA	All
Enrolled at 2 wk	310	329	301	300	295	208	1743
Completed 2-y follow-up	287 (92.6)	292 (88.8)	269 (89.4)	262 (87.3)	260 (88.1)	172 (82.7)	1542 (88.5)
Dropouts after week 2	23 (7.4)	37 (11.2)	32 (10.6)	38 (12.7)	35 (11.9)	36 (17.3)	201 (11.5)
Reason for dropout							
Child illness	0	1	1	2	4	0	8
Moved away	10	20	27	26	10	23	116
Unknown or other reason	0	1	0	2	2	4	9
Parents' wish	13	15	4	8	19	9	68

Table V. Baseline characteristics of families in the longitudinal sample by site.

	Brazil ($n=310$)	Ghana ($n=329$)	India ($n=301$)	Norway ($n=300$)	Oman ($n=295$)	USA ($n=208$)	All ($n=1743$)
Reproductive history of mother:							
Children born alive; median (range)	2 (1–7)	2 (1–8)	1 (1–3)	1 (1–5)	2 (1–12)	1 (1–5)	2 (1–12)
With < 3 children (%)	81.6	68.7	96.7	87.7	51.4	84.1	78.1
Primiparous (%)	49.0	38.1	53.5	55.0	27.8	53.4	45.7
Parental characteristics:							
Years of education completed							
Mother (mean ± SD)	11.1±3.5	15.1±2.7	17.5±1.5	15.4±2.6	11.9±3.3	16.7±2.1	14.5±3.6
<10 y	33.6	2.7	0	1.3	24.8	0	10.9
10–14 y	41.9	36.9	0.6	31.7	52.2	12.5	30.3
15–19 y	24.5	56.4	90.4	64.0	22.7	75.5	54.5
≥20 y	0	4.0	9.0	3.0	0.3	12.0	4.3
Father (mean ± SD)	10.2±3.6	18.1±3.0	17.4±1.8	15.2±2.8	12.8±3.6	16.9±2.6	15.0±4.1
<10 y	39.4	0.9	0	2.0	19.7	1.5	11.2
10–14 y	44.5	6.5	1	33.3	46.4	14.4	24.9
15–19 y	16.1	65.0	87.0	62.0	33.2	64.9	54.1
≥20 y	0.0	27.6	12.0	2.7	0.7	18.8	9.8
Maternal age (mean±SD)	28.3±6.3	30.8±4.0	28.9±3.5	30.6±4.4	27.5±4.9	30.8±4.8	29.4±4.9
<20 y	11	0	0	0	3.1	1.4	2.6
20–24 y	15.8	4	11.3	9	24.7	7.2	12.1
25–29 y	29.4	36.2	43.5	30.3	44.1	30.3	35.9
30–34 y	27.7	39.8	38.5	42.3	17.6	37.5	33.8
>35y	16.1	20.0	6.7	18.4	10.5	23.6	15.6
Mother's height (cm) (mean±SD)	161.1±6.0	161.9±5.2	157.6±5.4	168.7±6.6	156.6±5.5	164.5±6.9	161.6±7.2
Father's height (cm) (mean±SD)	173.6±6.9	173.0±6.6	172.7±6.3	182.2±6.7	170.4±6.4	178.9±7.4	175.1±7.9
Use of alcoholic beverages by mothers							
Never	85.8	90.2	98.7	68.3	100	78.3	85.0
<1/wk	11.0	9.2	1.3	30.3	–	20.7	13.8
≥1/wk	3.2	0.6	0	1.4	–	1	1.2
Socio-economic factors:							
Family income per month (median)							
Local currency[a]	1100	1 700 000	45 000	48 482	1140	5000	
USD	1019	739	957	6296	2938	5000	
Piped water	100	100	100	100	100	100	100
Flush toilet	100	98.8	100	100	100	100	99.8
Refrigerator	100	98.5	100	100	100	100	99.7
Gas/electric cooker	100	98.2	100	100	100	100	99.7
Telephone	85.2	81.4	99.0	100	98.3	100	93.4
Car	71	81.4	90.4	83.3	97.3	99.0	86.2

Note: All responses are percentages unless otherwise specified.
[a] Local currency (USD equivalent): Real for Brazil (1.08); Cedis for Ghana (2300 in 1999); Rupees for India (47); Kroner for Norway (7.7); Omani Rials for Oman (0.388).

Table VI. Baseline characteristics of children in the longitudinal sample by site.

	Brazil ($n=310$)	Ghana ($n=329$)	India ($n=301$)	Norway ($n=300$)	Oman ($n=295$)	USA ($n=208$)	All ($n=1743$)
Male sex,%	52.3	48.9	54.2	53.3	50.2	50.0	51.5
Apgar score							
1 min	8.7±1.1	7.7±1.2	8.3±1.0	8.7±0.9	8.5±0.9	7.9±1.6	8.3±1.2
5 min	9.7±0.5	9.2±0.9	9.1±0.6	9.4±0.6	9.8±0.6	8.9±0.6	9.4±0.7
Mode of delivery (%)							
Vaginal	46.1	72.9	59.5	90.0	85.8	87	72.6
Caesarean	53.9	27.1	40.5	10.0	14.2	13	27.4
Low birthweight,% (<2500 g)	1.9	1.5	4.7	0.7	2.7	0.5	2.1
Birthweight, kg	3.3±0.4	3.3±0.4	3.1±0.4	3.6±0.5	3.2±0.4	3.6±0.5	3.3±0.5
Birth length, cm	49.6±1.9	49.4±1.9	49.0±1.8	50.4±1.9	49.2±1.7	49.7±2.0	49.6±1.9
Head circumference, cm	34.6±1.1	34.3±1.2	33.8±1.2	34.9±1.2	33.4±1.0	34.2±1.3	34.2±1.3

non-participation in Norway and the USA. The fraction of children excluded from the cross-sectional sample was somewhat lower than for the longitudinal sample, but the main reasons for exclusion were the same, i.e. low socio-economic status in Brazil, Ghana, India and Oman, and refusals in Norway and the USA. Overall, the completion rate in the longitudinal component was very high (88.5%) despite the intense follow-up of 21 home visits over the 2-y period. This was largely due to the interest shown by participating mothers in the growth and development of their children.

There were no notable intra-site differences in parental education and height between the longitudinal and cross-sectional samples. Incomes were also largely comparable within sites, except for Ghana where, in local currency terms, the cross-sectional sample's median income was 1.8 times higher than that of the longitudinal sample. In US dollar equivalents, however, the longitudinal sample's median income was almost twice that of the cross-sectional sample, reflecting a dramatic drop in the foreign exchange value of the Ghanaian cedi between 1999 (start of the longitudinal follow-up) and 2001 when the cross-sectional survey began. The higher proportion across sites of primiparous mothers in the longitudinal compared with the cross-sectional sample is consistent with the fact that mothers in the latter were invariably older—by an average of over 3 y—than their counterparts in the longitudinal sample. The all-site ratio of male/female children in the two study components was the same, with slight variations in individual sites. Although some intra-site variation was seen with respect to rates of low birthweight, average weight at birth measured in the longitudinal sample and based on parental records for the cross-sectional sample was equal in all sites. In overall terms, therefore, the longitudinal and cross-sectional samples were taken from the same subpopulation in each site.

The study selected subjects with overall high parental education across sites, notably in Ghana [12] and India [13] where education was applied as

Table VII. Enrolment statistics for the cross-sectional sample by site.

	Brazil	Ghana	India	Norway	Oman	USA	All
Screened, *n*	2292	4818	3886	5185	4492	837	21510
Enrolled, *n* (%)	487 (21.2)	1406 (29.2)	1490 (38.3)	1387 (26.8)	1447 (32.2)	480 (57.3)	6697 (31.1)
Refusals, *n* (%)	2 (0.1)	60 (1.2)	107 (2.8)	614 (11.8)	57 (1.3)	76 (9.1)	916 (4.3)
Ineligibles, *n* (%)	1803 (78.7)	3352 (69.6)	2289 (58.9)	3184 (61.4)	2988 (66.5)	281 (33.6)	13897 (64.6)
Reasons for ineligibility[a] *(%)*							
Outside study area	5.2	2.3	23.3	0.0	0.2	0.0	5.3
Multiple birth	1.9	1.8	1.8	0.0	0.3	3.0	1.1
Gestational age outside range	9.7	8.3	8.8	16.3	2.8	11.2	9.4
Mother is a smoker	28.5	0.1	1.1	13.6	1.0	4.4	6.9
Low socio-economic status	59.3	36.6	29.6	0.0	64.1	0.5	33.3
Child with disease	2.2	0.9	2.4	5.6	5.3	0.6	3.4
Child breastfed <3 mo	28.7	2.0	12.0	5.3	1.4	15.1	7.8
Language difficulty	0.0	0.1	0.3	8.9	1.1	0.7	2.5
Longitudinal study participant	0.0	2.0	0.0	1.2	0.0	0.0	0.7
Longitudinal study sibling	0.0	0.0	0.0	0.0	0.0	3.9	0.2
Other exclusions	0.6	25.9	0.0	18.9	0.1	0.0	10.4

[a] The ineligibility tally may exceed 100% because of subjects being excluded for multiple reasons.

Table VIII. Baseline characteristics of families in the cross-sectional sample by site.

	Brazil ($n=487$)	Ghana ($n=1406$)	India ($n=1490$)	Norway ($n=1387$)	Oman ($n=1447$)	USA ($n=480$)	All ($n=6697$)
Reproductive history of mother							
Children born alive, median (range)	2 (1–9)	2 (1–10)	2 (1–5)	2 (1–6)	3 (1–15)	2 (1–6)	2 (1–15)
Primiparous (%)	37.6	20.6	41.1	26.3	8.8	23.8	25.2
Years of education, mean ± SD							
Mother	11.2 ± 3.5	15.2 ± 3.1	17.3 ± 1.8	15.4 ± 2.7	11.8 ± 3.5	16.5 ± 2.2	14.8 ± 3.6
Father	10.8 ± 3.7	18.8 ± 3.2	17.4 ± 1.8	15.6 ± 2.8	13.1 ± 3.6	16.8 ± 2.6	15.8 ± 3.8
Maternal age (mean ± SD)	32.0 ± 6.5	34.6 ± 4.7	31.9 ± 4.1	34.9 ± 4.7	30.8 ± 5.1	35.3 ± 5.1	33.1 ± 5.1
Weight in kg, mean ± SD							
Mother	63.5 ± 12.5	74.5 ± 14.3	62.6 ± 10.2	66.2 ± 10.5	66.0 ± 14.3	66.4 ± 13.3	66.9 ± 13.2
Father	79.7 ± 13.3	78.2 ± 13.6	76.2 ± 12.1	83.5 ± 11.9	77.0 ± 13.5	83.9 ± 14.1	78.9 ± 13.2
Height in cm, mean ± SD							
Mother	160.0 ± 6.2	161.9 ± 5.7	157.6 ± 5.7	167.7 ± 6.5	156.6 ± 5.4	164.3 ± 6.7	161.0 ± 7.2
Father	173.2 ± 7.0	172.6 ± 6.6	172.1 ± 6.0	181.2 ± 7.2	169.2 ± 6.4	178.0 ± 7.4	173.8 ± 7.9
Socio-economic factors							
Family income per month (median)							
Local currency [a]	1400	3 300 000	37 250	56 767	1150	5833	
USD	1296	404	793	7372	2964	5833	
Piped water supply	100	99.9	100	100	100	99.6	100
Own flush toilet	100	97.4	100	100	100	99.6	99.4
Own refrigerator	99.6	99.3	100	100	100	99.6	99.8
Own gas/electric cooker	100	99.4	100	99.9	100	99.6	99.8
Own telephone	97.7	95.2	99.5	99.9	99.9	99.6	98.6
Own car	75.8	83.3	92.7	92.1	98.4	99.6	91.1

Note: All responses are percentages unless otherwise specified.
[a] Local currency (USD equivalent): Real for Brazil (1.08); Cedis for Ghana (8172 in 2002); Rupees for India (47); Kroner for Norway (7.7); Omani Rials for Oman (0.388).

a screening criterion. Although family income in US dollar equivalents varied widely across sites, other indicators of socio-economic status, such as availability of basic household amenities, were relatively evenly distributed. The income differences should not be viewed in absolute terms since the cost of living varied from site to site. There were disparities across sites in the weights and heights of mothers and fathers, and in the mode of delivery, reflecting variations in secular trends of physical growth and cultural differences in birthing choices.

Despite the problems of unreliability inherent in recalled information, the early child feeding practices in the cross-sectional sample (reported) tallied well

Table IX. Baseline characteristics of children in the cross-sectional sample by site.

	Brazil ($n=487$)	Ghana ($n=1406$)	India ($n=1490$)	Norway ($n=1387$)	Oman ($n=1447$)	USA ($n=480$)	All ($n=6697$)
Male sex (%)	49.7	48.7	56.5	52.4	49.6	52.3	51.7
Mode of delivery (%)							
Vaginal	44.4	71.6	63.8	86.7	87.7	86.0	75.5
Caesarean	55.6	28.3	36.2	13.3	12.3	14.0	24.5
Low birthweight (%)[a]	1.6	3.6	5.1	0.5	4.8	0	3.2
Birthweight (g)	3423 ± 458	3316 ± 524	3113 ± 448	3636 ± 455	3187 ± 443	3582 ± 457	3338 ± 507
Breastfeeding duration (mo)[b]	12.0 ± 10.9	14.3 ± 5.8	12.6 ± 8.3	13.1 ± 6.3	17.0 ± 7.6	16.8 ± 10.2	14.3 ± 7.9
Age in months other milks or formula introduced	11.1 ± 17.2	5.9 ± 8.8	6.5 ± 12.9	10.0 ± 9.2	5.2 ± 7.7	12.4 ± 16.1	7.6 ± 11.4
Age in months solids/semi-solid food introduced	5.1 ± 2.2	5.8 ± 6.0	4.4 ± 3.5	5.2 ± 1.3	4.1 ± 1.3	5.7 ± 1.8	4.9 ± 3.5

Note: All figures are mean ± SD unless otherwise specified.
[a] Birthweight <2500 g.
[b] Breastfeeding for at least 3 mo was an inclusion criterion in the cross-sectional sample.

with comparable data in the longitudinal sample (observed prospectively). For example, the mean duration of breastfeeding was in between the durations observed in the longitudinal sample's feeding compliant and non-compliant groups in all sites except Oman [15]. This overall pattern is expected given the shorter breastfeeding duration required for inclusion in the cross-sectional sample. The similarity in average age at introduction of complementary feeding was even more striking, being equal in Ghana and within a month of each other in the other sites [16].

The prevalence of low birthweight in the MGRS samples in Brazil, Ghana, India and Oman was much lower than national prevalence rates of 8.5% for Brazil [17], 11% for Ghana, 30% for India and 8% for Oman [18]. This suggests that the selection criteria applied in these sites were effective in excluding most children from low socio-economic status households where the risk of low birthweight is high. The children enrolled in the longitudinal component were quite similar across sites for weight, length and head circumference at birth, and, as described in a companion paper in this supplement, the patterns of linear growth thereafter were strikingly similar among the six sites [19]. Thus, it appears that the selection criteria applied were successful in screening for children who were healthy at birth and with a high probability of experiencing unconstrained growth.

Acknowledgements

This manuscript was prepared by Anna Lartey, Nita Bhandari, Mercedes de Onis, Adelheid W. Onyango, Deena Alasfoor, Roberta J. Cohen, Cora L. Araújo and Anne Baerug on behalf of the WHO Multicentre Growth Reference Study Group. The statistical analysis was conducted by Amani Siyam and Alain Pinol.

Members of the WHO Multicentre Growth Reference Study Group

Coordinating Team

Mercedes de Onis [Study Coordinator], Adelheid Onyango, Elaine Borghi, Amani Siyam, Alain Pinol (Department of Nutrition, World Health Organization)

Executive Committee

Cutberto Garza [Chair], Mercedes de Onis, Jose Martines, Reynaldo Martorell, Cesar G. Victora (up to October 2002), Maharaj K. Bhan (from November 2002)

Steering Committee

Coordinating Centre (WHO, Geneva). Mercedes de Onis, Jose Martines, Adelheid Onyango, Alain Pinol

Investigators (by country). Cesar G. Victora and Cora Luiza Araújo (Brazil), Anna Lartey and William B. Owusu (Ghana), Maharaj K. Bhan and Nita Bhandari (India), Kaare R. Norum and Gunn-Elin Aa. Bjoerneboe (Norway), Ali Jaffer Mohamed (Oman), Kathryn G. Dewey (USA)

Representatives United Nations Agencies. Cutberto Garza (UNU), Krishna Belbase (UNICEF)

Advisory Group

Maureen Black, Wm. Cameron Chumlea, Tim Cole, Edward Frongillo, Laurence Grummer-Strawn, Reynaldo Martorell, Roger Shrimpton, Jan Van den Broeck

Participating countries and investigators

Brazil

Cora Luiza Araújo, Cesar G. Victora, Elaine Albernaz, Elaine Tomasi, Rita de Cássia Fossati da Silveira, Gisele Nader (Departamento de Nutrição and Departamento de Medicina Social, Universidade Federal de Pelotas; and Núcleo de Pediatria and Escola de Psicologia, Universidade Católica de Pelotas)

Ghana

Anna Lartey, William B. Owusu, Isabella Sagoe-Moses, Veronica Gomez, Charles Sagoe-Moses (Department of Nutrition and Food Science, University of Ghana; and Ghana Health Service)

India

Nita Bhandari, Maharaj K. Bhan, Sunita Taneja, Temsunaro Rongsen, Jyotsna Chetia, Pooja Sharma, Rajiv Bahl (All India Institute of Medical Sciences)

Norway

Gunn-Elin Aa. Bjoerneboe, Anne Baerug, Elisabeth Tufte, Kaare R. Norum, Karin Rudvin, Hilde Nysaether (Directorate of Health and Social Affairs; National Breastfeeding Centre, Rikshospitalet University Hospital; and Institute for Nutrition Research, University of Oslo)

Oman

Ali Jaffer Mohamed, Deena Alasfoor, Nitya S. Prakash, Ruth M. Mabry, Hanadi Jamaan Al Rajab, Sahar Abdou Helmi (Ministry of Health)

USA

Kathryn G. Dewey, Laurie A. Nommsen-Rivers, Roberta J. Cohen, M. Jane Heinig (University of California, Davis)

Financial support

The project has received funding from the Bill & Melinda Gates Foundation, the Netherlands Minister for Development Cooperation, the Norwegian Royal Ministry of Foreign Affairs, and the United States Department of Agriculture (USDA). Financial support was also provided by the Ministry of Health of Oman, the United States National Institutes of Health, the Brazilian Ministry of Health and Ministry of Science and Technology, the Canadian International Development Agency, the United Nations University, the Arab Gulf Fund for United Nations Development, the Office of the WHO Representative to India, and the Department of Child and Adolescent Health and Development. The Motor Development Study was partially supported by UNICEF.

The Study Group is indebted to many individuals and institutions that contributed to the study in different ways. These are listed in the supplement describing the MGRS protocol [1].

References

[1] de Onis M, Garza C, Victora CG, Bhan MK, Norum KR, editors. WHO Multicentre Growth Reference Study (MGRS): Rationale, planning and implementation. Food Nutr Bull 2004;25 Suppl 1:S1–89.

[2] WHO Working Group on Infant Growth. An evaluation of infant growth. Geneva: World Health Organization; 1994.

[3] Working Group on Infant Growth. WHO. An evaluation of infant growth: the use and interpretation of anthropometry in infants. Bull World Health Organ 1995;73:165–74.

[4] de Onis M, Garza C, Victora CG, Onyango AW, Frongillo EA, Martines J, for the WHO Multicentre Growth Reference Study Group. The WHO Multicentre Growth Reference Study: Planning, study design, and methodology. Food Nutr Bull 2004;25 Suppl 1:S15–26.

[5] WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards based on length/height, weight and age. Acta Paediatr Suppl 2006;450:76–85.

[6] Bhandari N, Bahl R, Taneja S, de Onis M, Bhan MK. Growth performance of affluent Indian children is similar to that in developed countries. Bull World Health Organ 2002;80:189–95.

[7] Owusu WB, Lartey A, de Onis M, Onyango AW, Frongillo EA. Factors associated with unconstrained growth among affluent Ghanaian children. Acta Paediatr 2004;93:1115–9.

[8] Mohamed AJ, Onyango AW, de Onis M, Prakash N, Mabry RM, Alasfoor DH. Socioeconomic predictors of unconstrained child growth in Muscat, Oman. East Mediterr Health J 2004;10:295–302.

[9] Araujo CL, Albernaz E, Tomasi E, Victora CG, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in Brazil. Food Nutr Bull 2004;25 Suppl 1:S53–9.

[10] Baerug A, Bjoerneboe GE, Tufte E, Norum KR, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in Norway. Food Nutr Bull 2004;25 Suppl 1:S72–7.

[11] Dewey KG, Cohen RJ, Nommsen-Rivers LA, Heinig MJ, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in the United States. Food Nutr Bull 2004;25 Suppl 1:S84–9.

[12] Lartey A, Owusu WB, Sagoe-Moses I, Gomez V, Sagoe-Moses C, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in Ghana. Food Nutr Bull 2004;25 Suppl 1:S60–5.

[13] Bhandari N, Taneja S, Rongsen T, Chetia J, Sharma P, Bahl R, et al., for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in India. Food Nutr Bull 2004;25 Suppl 1:S66–71.

[14] Prakash NS, Mabry RM, Mohamed AJ, Alasfoor D, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in Oman. Food Nutr Bull 2004;25 Suppl 1:S78–83.

[15] WHO Multicentre Growth Reference Study Group. Breastfeeding in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450:16–26.

[16] WHO Multicentre Growth Reference Study Group. Complementary feeding in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450:27–37.

[17] Victora CG, Barros FC. Infant mortality due to prenatal causes in Brazil: trends, regional patterns and possible interventions. Sao Paulo Med J 2001;119:33–42.

[18] UNICEF. The state of the world's children 2004. New York: The United Nations Children's Fund (UNICEF); 2003.

[19] WHO Multicentre Growth Reference Study Group. Assessment of differences in linear growth among populations in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450:56–65.

Acta Pædiatrica, 2006; Suppl 450: 16–26

Breastfeeding in the WHO Multicentre Growth Reference Study

WHO MULTICENTRE GROWTH REFERENCE STUDY GROUP[1,2]

[1]*Department of Nutrition, World Health Organization, Geneva, Switzerland, and* [2]*Members of the WHO Multicentre Growth Reference Study Group (listed at the end of the first paper in this supplement)*

Abstract

Aim: To document how children in the WHO Multicentre Growth Reference Study (MGRS) complied with feeding criteria and describe the breastfeeding practices of the compliant group. *Methods:* The MGRS longitudinal component followed 1743 mother–infant pairs from birth to 24 mo in six countries (Brazil, Ghana, India, Norway, Oman and the USA). The study included three criteria for compliance with recommended feeding practices that were monitored at each follow-up visit through food frequency reports and 24-h dietary recalls. Trained lactation counsellors visited participating mothers frequently in the first months after delivery to help with breastfeeding initiation and prevent and resolve lactation problems. *Results:* Of the 1743 enrolled newborns, 903 (51.8%) completed the follow-up and complied with the three feeding criteria. Three quarters (74.7%) of the infants were exclusively/predominantly breastfed for at least 4 mo, 99.5% were started on complementary foods by 6 mo of age, and 68.3% were partially breastfed until at least age 12 mo. Compliance varied across sites (lowest in Brazil and highest in Ghana) based on their initial baseline breastfeeding levels and sociocultural characteristics. Median breastfeeding frequency among compliant infants was 10, 9, 7 and 5 feeds per day at 3, 6, 9 and 12 mo, respectively. Compliant mothers were less likely to be employed, more likely to have had a vaginal delivery, and fewer of them were primiparous. Pacifier use was more prevalent in the non-compliant group.

Conclusion: The MGRS lactation support teams were successful in enhancing breastfeeding practices and achieving high rates of compliance with the feeding criteria required for the construction of the new growth standards.

Key Words: *Breastfeeding, child nutrition, growth curves, growth standards, infant feeding practices*

Introduction

Growth charts are essential instruments in the paediatric toolkit. Their value resides in helping determine the degree to which physiological needs for growth and development are being met during the important childhood period. However, interpretation of the adequacy of growth is highly dependent on the reference data used and may be erroneous if the reference used does not adequately represent physiological growth.

The growth reference recommended for international use since the late 1970s—the National Center for Health Statistics/World Health Organization (NCHS/WHO) reference—has been shown to have a number of drawbacks that make it inappropriate for assessing infant growth [1–3]. One of its most important limitations is that it is based on a sample of predominantly formula-fed infants whose pattern of growth has been demonstrated to deviate substantially from that of healthy breastfed infants [4,5]. The divergence between the growth pattern of healthy breastfed infants and other national growth references that are likewise largely based on formula-fed infants has also been documented [6,7].

Recognizing the shortcomings of the NCHS/WHO international growth reference, in 1994 WHO began planning for the development of new standards which, unlike the current reference, would be based on an international sample of healthy breastfed infants and would portray how children should grow in all countries rather than merely describing how they grew at a particular time and place [8,9]. The WHO Multicentre Growth Reference Study (MGRS), undertaken between 1997 and 2003, focused on the collection of growth and related data from 8440 children from widely differing ethnic backgrounds and cultural settings (Brazil, Ghana, India, Norway, Oman and the USA) [10]. As described elsewhere [10], breastfeeding practices were one of the primary criteria used to select study sites. The intention was to choose populations where breastfeeding was commonly practised and provide lactation support to

Correspondence: Mercedes de Onis, Study Coordinator, Department of Nutrition, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland. Tel: +41 22 791 3320. Fax: +41 22 791 4156. E-mail: deonism@who.int

ISSN 0803-5326 print/ISSN 1651-2227 online
DOI: 10.1080/08035320500495423

mothers enrolled in the study to help them comply with the feeding criteria required to construct the new standards. This paper documents how the children in the MGRS sample complied with the study's feeding criteria in infancy and describes in detail the breastfeeding practices of the feeding-compliant group.

Methods

The MGRS was a population-based study undertaken in the cities of Davis, California, USA; Muscat, Oman; Oslo, Norway; Pelotas, Brazil; and selected affluent neighbourhoods of Accra, Ghana, and South Delhi, India. The MGRS protocol and its implementation at the six sites are described in detail elsewhere [11]. The MGRS combined a longitudinal component from birth to 24 mo of age with a cross-sectional component of children aged 18 to 71 mo. In the longitudinal component, mothers and newborns were screened and enrolled at birth and visited at home at weeks 1, 2, 4 and 6; monthly from 2–12 mo; and bimonthly in the second year. This paper describes infant feeding practices in the longitudinal sample.

The MGRS included three compliance criteria regarding feeding for children to be included in the growth standards sample: 1) exclusive or predominant breastfeeding for at least 4 mo (120 d); 2) introduction of complementary foods between 4 and 6 mo (120 to 180 d); and 3) partial breastfeeding to be continued up to at least 12 mo (365 d). Concerning the first criterion, it is important to note that the MGRS was initiated before WHO's policy on the optimal duration of exclusive breastfeeding changed in 2001 from "4 to 6 months" to "6 months" [12]. Nevertheless, the national policies at three study sites (Brazil, Ghana and India) already recommended 6 mo, and participating mothers in all sites were advised to breastfeed their infants exclusively for as close as possible to 6 mo. For children to be included in the growth standards sample, a fourth criterion, maternal non-smoking, was required.

The MGRS study sites were selected on the basis that a minimum of 20% of mothers in the study's subpopulations were willing to follow the feeding compliance criteria [10]. Mothers were screened at the time of enrolment and those not intending to breastfeed were considered ineligible for the study. In Oman and the USA, screening with regard to child feeding intentions was more stringent: only mothers willing to breastfeed exclusively for at least 4 mo, and to continue breastfeeding up to at least 12 mo of age, were enrolled [13,14].

To ensure a high level of compliance with the three feeding criteria among participating mothers, lactation counselling was made an essential part of the MGRS. Lactation counselling, which was provided by trained lactation counsellors at each site, was designed to help with initiating breastfeeding soon after delivery, preventing and resolving lactation problems, and sustaining exclusive/predominant breastfeeding through 4 mo and partial breastfeeding through at least 12 mo. The first visit by a lactation counsellor took place within 24 h of delivery, and subsequent visits occurred at 7, 14 and 30 d, and monthly thereafter until the sixth month. A 24-h hotline was also made available to mothers for emergency support. Additional visits were carried out whenever feeding problems occurred. Compliance with the feeding criteria was monitored centrally and lactation counselling strengthened as required. Local logistics of the breastfeeding support systems and lactation counselling teams in the six sites are described elsewhere [13–18]. Mothers also received advice on complementary feeding according to locally adapted guidelines. Complementary feeding practices of the MGRS sample are described in a companion paper in this supplement [19].

Exclusive breastfeeding was defined as the infant receiving only breast milk from his/her mother or a wet-nurse, or expressed breast milk, and no other liquids or solids with the exception of drops or syrups consisting of vitamins, mineral supplements or medicines [10]. Predominant breastfeeding consisted of breast milk as the infant's predominant source of nourishment, but the infant could also receive water and water-based drinks (e.g. sweetened and flavoured water, teas, infusions), fruit juice, oral rehydration solution and ritual fluids (in limited quantities) [10].

Compliance with exclusive/predominant breastfeeding was assessed from birth to age 4 mo (visits 1–6) using the cumulative frequency of non-compliant days (i.e. the baby received infant formula or other milk than breast milk and/or more than one teaspoon of solid or semi-solid food). As soon as the number of days of such non-compliance exceeded 12, the child was marked as non-compliant for that and subsequent visits. Timely introduction of complementary foods was assessed from 6 to 12 mo (visits 8–14) on the basis of solid/semi-solid food consumption. Continued breastfeeding until at least 12 mo of age was assessed throughout the first year. Children classified as non-compliant were marked as such for the index and subsequent visits.

Data on feeding practices were collected at each of the follow-up visits [10]. Food frequency reports were used to describe the intake of breast milk, other fluids and milks, and solid and semi-solid foods in the intervals between visits. More detailed data on typical daily feeding were collected by 24-h dietary recalls on what the child ate or drank during each of seven time periods throughout the day. In addition to data collected by follow-up teams, lactation counsellors collected in-hospital information on breastfeeding initiation and at-home information on the establishment of lactation, problems experienced in the first 2 wk, and practices with potentially adverse influences on continued lactation (e.g. pacifier use) [10].

Results

Table I describes the MGRS sample according to compliance with feeding recommendations and completion of follow-up. Of the 1743 enrolled newborns, 903 (51.8%) completed the 24-mo follow-up and met the three operational criteria for compliance with feeding recommendations. Fifteen other children whose mothers did not comply with the study's no-smoking criterion and six with morbid conditions known to affect child growth were further excluded to obtain the sample ($n = 882$) from the MGRS longitudinal component that was used to construct the growth standards [20]. Compliance was highest in Ghana (71.1%), followed by the USA (63.0%), India (60.2%), Norway (55.3%), Oman (53.3%) and Brazil (23.3%). Most of the following analyses focus on the children by compliance group who completed the follow-up.

Table II presents maternal characteristics relevant to breastfeeding choices by compliance group and site. Newborns in all sites were term, single births. Maternal age was not different by compliance group in individual sites; however, when the sample was pooled, the compliant group was significantly older by about 1 y. Maternal education in Norway and Oman was significantly different between compliance groups but in opposite directions. For the overall sample, the compliant group had about 1 y more of education, which was a statistically significant difference. Overall, fewer mothers were employed outside the home in the compliant compared to the non-compliant group. Vaginal delivery was significantly higher, and rate of primiparous mothers significantly lower, for compliers for the overall sample, and no differences were noted in either parity or prevalence of maternal smoking (less than 1% smoked in both groups).

Figure 1 presents compliance with each of the MGRS feeding criteria by site and for all sites together. Overall, 74.7% of infants were exclusively or predominantly breastfed for at least 4 mo, almost all of them (99.5%) were started on complementary foods by the age of 6 mo, and 68.3% were partially breastfed to at least 12 mo of age. Compliance with exclusive/predominant breastfeeding for at least 4 mo was lowest in Brazil (48.6%) and highest in Ghana (89.4%). Norway and the USA also had very high compliance rates for this feeding criterion (86.0 and 82.6%, respectively), and the compliance rates for India and Oman were above 65%. Compliance with the criterion for introduction of complementary foods was above 98% in all sites. Compliance with the third

Table I. Sample classification on overall feeding compliance and completion of follow-up by site.

Compliance/follow-up category	Brazil ($n = 310$)		Ghana ($n = 329$)		India ($n = 301$)		Norway ($n = 300$)		Oman ($n = 295$)		USA ($n = 208$)		All ($n = 1743$)	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Complaint, study completed	69	22.3	228	69.3	173	57.5	159	53.0	153	51.9	121	58.2	903	51.8
Compliant, not completed	3	1.0	6	1.8	8	2.7	7	2.3	4	1.4	10	4.8	38	2.2
Not compliant, study completed	218	70.3	64	19.5	96	31.8	103	34.4	107	36.2	51	24.5	639	36.6
Not compliant, not completed	20	6.4	31	9.4	24	8.0	31	10.3	31	10.5	26	12.5	163	9.4

Table II. Maternal characteristics of compliant and non-compliant subjects.

	Brazil	Ghana	India	Norway	Oman	USA	All
Compliant *n*	69	228	173	159	153	121	903
Non-compliant *n*	218	64	96	103	107	51	639
Maternal age (y), mean ± SD							
Compliant	29.1 ± 6.4	30.9 ± 4.0	29.0 ± 3.5	31.2 ± 4.2	28.1 ± 5.4	31.6 ± 4.6	30.1* ± 4.7
Non-compliant	28.1 ± 6.3	30.4 ± 3.6	29.0 ± 3.6	30.2 ± 4.5	27.2 ± 4.5	31.5 ± 4.0	28.9* ± 5.1
Maternal education (y), mean ± SD							
Compliant	11.9 ± 3.6	15.0 ± 2.7	17.5 ± 1.5	15.9* ± 2.6	11.3* ± 3.5	17.0 ± 1.8	15.0* ± 3.5
Non-compliant	10.9 ± 3.5	15.6 ± 2.2	17.5 ± 1.6	14.7* ± 2.6	12.7* ± 3.1	16.8 ± 1.9	13.7* ± 3.8
Vaginal delivery,%							
Compliant	40.6	76.8	60.7	92.5	88.2	85.1	76.7*
Non-compliant	47.7	71.9	57.3	89.3	85.1	86.3	67.6*
Maternal smoking,%							
Compliant	–	–	–	2.5	–	–	0.4
Non-compliant	0.9	–	–	1.9	–	–	0.6
Maternal employment,%							
Compliant	68.1	84.2	28.9*	79.2	30.1*	58.7	58.9*
Non-compliant	70.6	87.5	47.9*	74.8	64.5*	74.5	68.9*
Type of job, full time,%							
Compliant	–	96.9	88.0	88.1	95.7	52.1*	87.0
Non-compliant	–	96.4	87.0	88.3	97.1	84.2*	91.3
Parity, median (min., max.)							
Compliant	2 (1,6)	2 (1,8)	1 (1,3)	1 (1,4)	3* (1,12)	1 (1,5)	2 (1,12)
Non-compliant	1 (1,7)	2 (1,4)	1 (1,3)	1 (1,5)	2* (1,12)	1 (1,4)	2 (1,12)
Primiparous,%							
Compliant	46.4	36.8	51.4	52.2	18.3*	52.1	42.0*
Non-compliant	51.4	37.5	57.3	58.3	36.4*	51.0	49.5*

*Statistically significant difference (p-value <0.05) between the compliant and non-compliant groups.

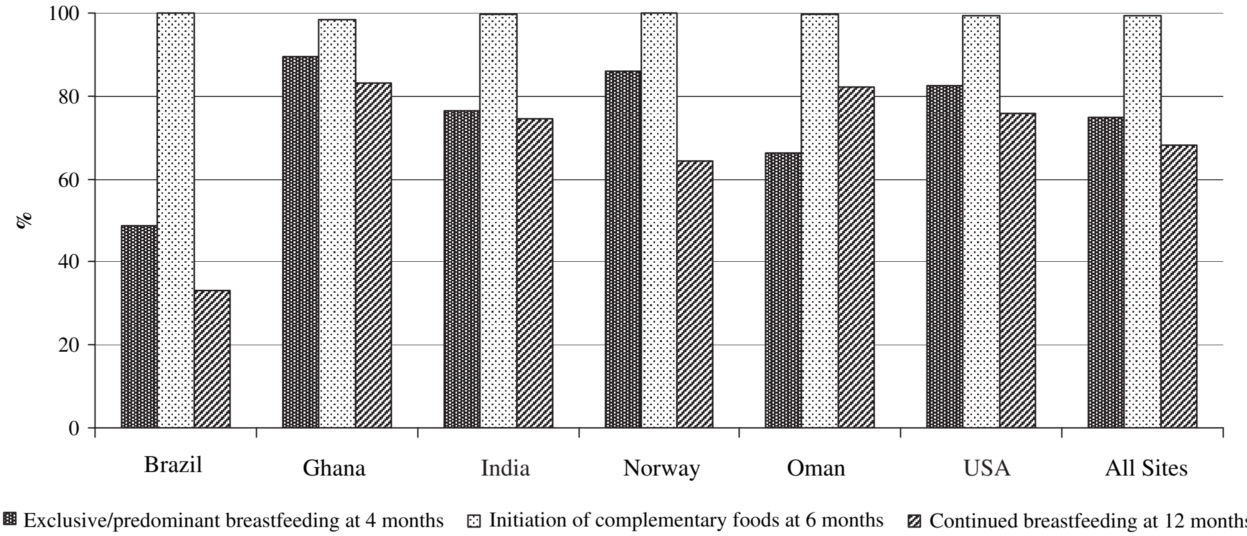


Figure 1. Compliance with MGRS feeding criteria by site and overall.

feeding criterion (i.e. continued breastfeeding up to at least 12 mo of age) was more variable across sites, with Brazil having the lowest compliance rate (33.2%) and Ghana and Oman the highest (83.1 and 82.3%, respectively). Figure 2 shows the percent of overall feeding compliance by site at each follow-up visit up to 12 mo.

Figure 3 displays the prevalence of exclusive, predominant and partial breastfeeding (with and without solids), and the percent of the overall sample not breastfed, from week 2 to 12 mo of age. This figure shows that children classified in the exclusive/predominant category were mainly exclusively breastfed. Moreover, the proportion of infants exclusively breastfed is somewhat underestimated as the data showed that some children moved back and forth between the exclusive and predominant categories between visits. However, for the purpose of constructing the figure, the classification ran only one way; that is, once a child had been classified as predominantly breastfed he/she was not classified back to the exclusively breastfed category even if, at the next visit, the child was being exclusively breastfed. The figure also shows that the overall MGRS sample enjoyed high breastfeeding rates, with 68.3% still being breastfed at 12 mo.

Table III summarizes the frequency and volume of 24-h fluid intake at 6, 9 and 12 mo for compliant children. At 3 mo there was very little consumption of any of these fluids. It is noteworthy that Indian mothers tended to supplement with animal milk, while supplementation with formula seems to have

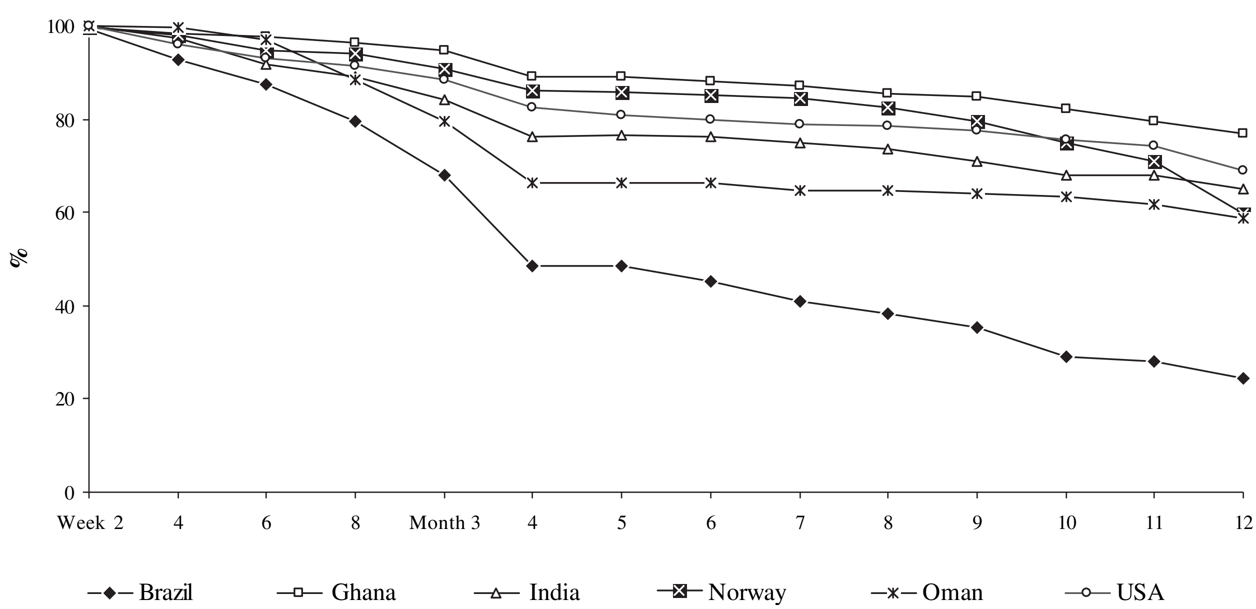


Figure 2. Compliance with MGRS feeding criteria in infancy.

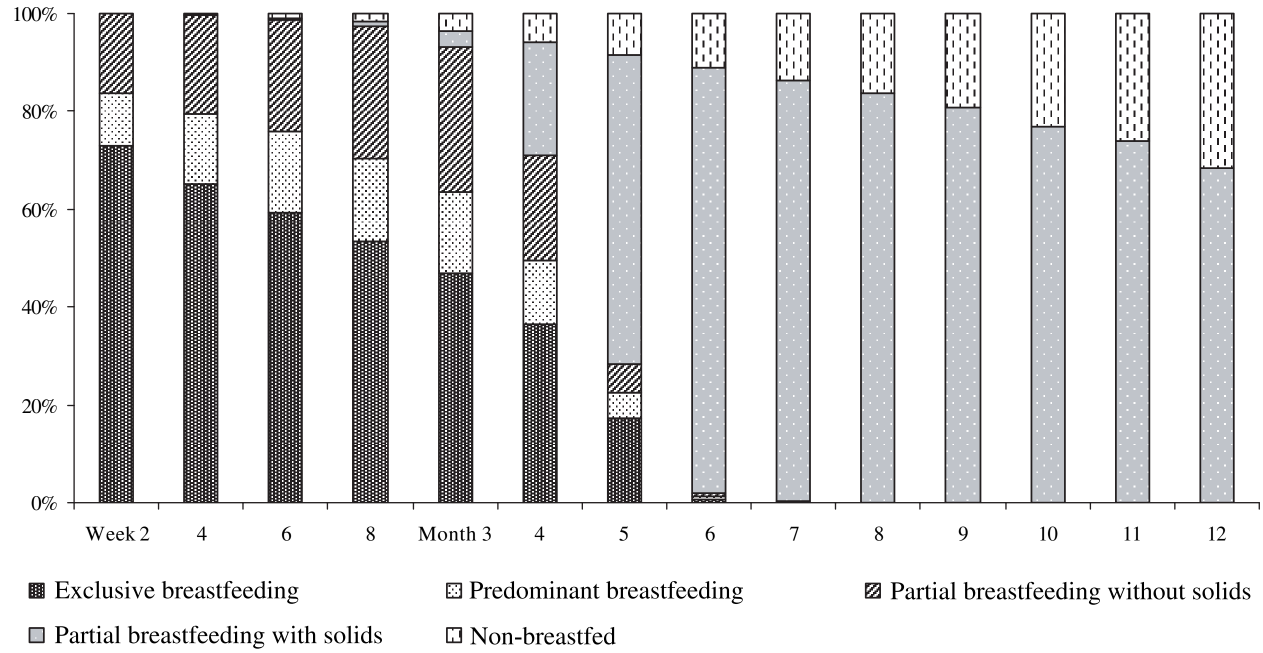


Figure 3. Prevalence of exclusive, predominant and partial breastfeeding, and prevalence of non-breastfed infants for overall sample by age.

been more common in Ghana. Tea was much more common in Brazil, and water supplementation was very common in Ghana, India and Oman. Overall, at 6 mo, supplementation with formula was more common than with animal milk, while at 12 mo the opposite was true. Water was more frequently given to children than juice or tea.

Figure 4 shows the median breastfeeding frequency for each country and all sites at 3, 6, 9 and 12 mo (error bars representing the Q1–Q3 range). At any given time, Ghana and Oman had the highest breastfeeding frequency. The overall median breastfeeding frequency among compliant infants was 10, 9, 7 and 5 feeds per day at 3, 6, 9 and 12 mo, respectively.

Table IV presents the median duration of breastfeeding by compliance group and the percent of children still breastfeeding at 24 mo. The overall median duration in the compliant group was 17.8 mo versus 9.3 mo in the non-compliant group. It should be noted that the median duration in the compliant group is underestimated since 16.2% of the children were still breastfeeding when follow-up was completed. Brazil, India and the USA had the largest proportions of compliant children still breastfeeding at 24 mo. In all sites, both the duration of breastfeeding and the percent of children still breastfeeding at 24 mo were significantly lower statistically in the non-compliant group, with the exception of Ghana and Oman for the percent of children still breastfeeding.

Table V presents, by compliance group, the percentage of newborns breastfed within 1 h of birth; median hours after birth a baby was breastfed for the first time; and pacifier use at 2 wk, and 3 and 6 mo. For the overall sample, the use of pacifiers was significantly higher at 3 and 6 mo, and Norway and the USA had the highest prevalence of use. These data were not available for the Brazilian site.

The most important breastfeeding problems reported among compliant mothers at the week 1 visit (data not shown) were sore nipples (27.9%), engorgement (19%), too much milk (6.3%), mastitis (2.0%) and delayed onset of milk production (2.7%). At the week 2 visit, the prevalence of these problems had decreased substantially: 14.6% sore nipples, 9.9% engorgement, 3.8% too much milk and 2.3% mastitis. Mothers in Norway and the USA most often reported having problems. However, it is important to note that these data were self-reported and their collection was not standardized either across sites or among lactation counsellors within sites. Breastfeeding problems reported by non-compliant mothers did not differ significantly from those of compliant mothers.

Discussion

The results presented here document the success of the MGRS lactation support teams in enhancing breastfeeding practices and achieving high rates of compliance with the study's feeding criteria. Overall, 54% of the sample complied with the three feeding criteria, surpassing the expected compliance rate of 30% used to calculate the study's sample size. This result, coupled with a very low dropout rate (96% of compliant children completed the 24-mo follow-up) yielded a sample for the construction of the standards more than double the size required to ensure stable outer percentiles (i.e. 882 vs 400) [10].

Table III. Twenty-four-hour intake of fluids among compliant infants.

	Brazil ($n = 69$)		Ghana ($n = 228$)		India ($n = 173$)		Norway ($n = 159$)		Oman ($n = 153$)		USA ($n = 121$)		All ($n = 903$)	
At 6 mo														
Animal milk, *n* (%)	2	(2.9)	14	(6.1)	90	(52.0)	0	(0.0)	10	(6.5)	0	(0.0)	116	(12.8)
frequency, median (min., max.)	2	(2,2)	2	(1,5)	2	(1,8)			1.5	(1,2)			2	(1,8)
volume, median (min., max.)	180	(160,200)	40	(10,260)	90	(15,900)			60	(10,150)			75	(10,900)
Infant formula, *n* (%)	2	(2.9)	112	(49.1)	14	(8.1)	20	(12.6)	48	(31.4)	11	(9.1)	207	(22.9)
frequency, median (min., max.)	1	(1,1)	2	(1,10)	1	(1,5)	1	(1,6)	1	(1,3)	2	(1,5)	2	(1,10)
volume, median (min., max.)	135	(90,180)	150	(30,1180)	152.5	(20,480)	100	(15,600)	45	(10,390)	135	(15,480)	120	(10,1180)
Tea, *n* (%)	9	(13.0)	1	(0.4)	2	(1.2)	0	(0.0)	6	(3.9)	0	(0.0)	18	(2.0)
frequency, median (min., max.)	2	(1,3)	2	(2,2)	1	(1,1)			1	(1,3)			1	(1,3)
volume, median (min., max.)	50	(18,120)	120	(120,120)	27.5	(10,45)			60	(30,100)			60	(10,120)
Water, *n* (%)	12	(17.4)	153	(67.1)	118	(68.2)	23	(14.5)	142	(92.8)	26	(21.5)	474	(52.5)
frequency, median (min., max.)	1	(1,5)	3	(1,9)	2	(1,10)	1	(1,3)	3	(1,8)	1	(1,3)	3	(1,10)
volume, median (min., max.)	30	(10,280)	70	(10,360)	47.5	(10,300)	50	(10,200)	75	(10,500)	37.5	(10,135)	60	(10,500)
Juice, *n* (%)	19	(27.5)	21	(9.2)	34	(19.7)	1	(0.6)	69	(45.1)	4	(3.3)	148	(16.4)
frequency, median (min., max.)	1	(1,2)	1	(1,3)	1	(1,3)	1	(1,1)	1	(1,5)	1	(1,4)	1	(1,5)
volume, median (min., max.)	80	(3,250)	30	(10,240)	50	(10,150)	120	(120,120)	50	(15,250)	52.5	(30,180)	50	(3,250)
At 9 mo														
Animal milk, *n* (%)	13	(18.8)	64	(28.1)	118	(68.2)	3	(1.9)	29	(19.0)	2	(1.7)	229	(25.4)
frequency, median (min., max.)	2	(1,6)	2	(1,6)	2	(1,20)	1	(1,1)	1	(1,4)	1	(1,1)	2	(1,20)
volume, median (min., max.)	280	(100,960)	50	(10,1000)	150	(10,1075)	50	(10,80)	100	(10,360)	90	(60,120)	100	(10,1075)
Infant formula, *n* (%)	3	(4.3)	94	(41.2)	6	(3.5)	14	(8.8)	61	(39.9)	24	(19.8)	202	(22.4)
frequency, median (min., max.)	2	(1,3)	2	(1,10)	2.5	(1,6)	1.5	(1,5)	1	(1,5)	2	(1,4)	2	(1,10)
volume, median (min., max.)	160	(150,460)	180	(25,1050)	322.5	(25,600)	150	(80,680)	80	(10,500)	120	(15,420)	120	(10,1050)
Tea, *n* (%)	10	(14.5)	6	(2.6)	4	(2.3)	1	(0.6)	6	(3.9)	0	(0.0)	27	(3.0)
frequency, median (min., max.)	1	(1,3)	1	(1,1)	1	(1,1)	1	(1,1)	1	(1,4)			1	(1,4)
volume, median (min., max.)	45	(20,150)	75	(10,150)	22.5	(10,90)	5	(5,5)	60	(25,100)			50	(5,150)
Water, *n* (%)	12	(17.4)	174	(76.3)	157	(90.8)	102	(64.2)	149	(97.4)	56	(46.3)	650	(72.0)
frequency, median (min., max.)	2	(1,3)	4	(1,12)	4	(1,24)	3	(1,10)	4	(1,8)	1	(1,6)	4	(1,24)
volume, median (min., max.)	40	(20,150)	175	(20,640)	120	(10,750)	75	(10,400)	160	(15,1050)	60	(15,240)	120	(10,1050)
Juice, *n* (%)	34	(49.3)	44	(19.3)	27	(15.6)	9	(5.7)	71	(46.4)	25	(20.7)	210	(23.3)
frequency, median (min., max.)	1	(1,3)	1	(1,4)	1	(1,1)	1	(1,4)	1	(1,2)	1	(1,6)	1	(1,6)
volume, median (min., max.)	95	(10,500)	50	(7,240)	50	(15,120)	30	(10,150)	60	(10,200)	60	(10,180)	55	(7,500)

Table III (*Continued*)

	Brazil ($n = 69$)		Ghana ($n = 228$)		India ($n = 173$)		Norway ($n = 159$)		Oman ($n = 153$)		USA ($n = 121$)		All ($n = 903$)	
At 12 mo														
Animal milk, *n* (%)	31	(44.9)	107	(46.9)	139	(80.3)	65	(40.9)	41	(26.8)	39	(32.2)	422	(46.7)
frequency, median (min., max.)	3	(1,8)	2	(1,8)	2	(1,9)	2	(1,5)	1	(1,3)	2	(1,5)	2	(1,9)
volume, median (min., max.)	420	(50,1900)	60	(5,480)	240	(25,1575)	100	(10,500)	100	(20,500)	105	(10,600)	120	(5,1900)
Infant formula, *n* (%)	2	(2.9)	75	(32.9)	5	(2.9)	17	(10.7)	60	(39.2)	21	(17.4)	180	(19.9)
frequency, median (min., max.)	1.5	(1,2)	2	(1,8)	2	(1,3)	1	(1,4)	2	(1,6)	2	(1,5)	2	(1,8)
volume, median (min., max.)	240	(150,330)	180	(30,990)	300	(120,400)	130	(40,600)	105	(10,900)	210	(30,660)	150	(10,990)
Tea, *n* (%)	5	(7.3)	9	(4.0)	11	(6.4)	2	(1.3)	7	(4.6)	1	(0.8)	35	(3.9)
frequency, median (min., max.)	1	(1,4)	1	(1,5)	1	(1,2)	1	(1,1)	1	(1,1)	1	(1,1)	1	(1,5)
volume, median (min., max.)	90	(20,150)	60	(24,300)	30	(15,100)	35	(20,50)	50	(30,60)	15	(15,15)	50	(15,300)
Water, *n* (%)	22	(31.9)	184	(80.7)	160	(92.5)	138	(86.8)	150	(98.0)	84	(69.4)	738	(81.7)
frequency, median (min., max.)	1	(1,3)	5	(1,13)	5	(1,13)	3	(1,20)	4	(1,11)	2	(1,9)	4	(1,20)
volume, median (min., max.)	60	(10,220)	240	(20,1000)	150	(20, 1000)	130	(10,600)	212.5	(20,1050)	120	(15,480)	180	(10, 1050)
Juice, *n* (%)	36	(52.2)	56	(24.6)	35	(20.2)	20	(12.6)	93	(60.8)	52	(43.0)	292	(32.3)
frequency, median (min., max.)	2	(1,3)	1	(1,6)	1	(1,2)	1	(1,5)	1	(1,3)	1	(1,8)	1	(1,8)
volume, median (min., max.)	150	(30,360)	60	(10,720)	50	(20,200)	50	(10,400)	70	(20,300)	90	(7,480)	72.5	(7,720)

Compliance with feeding recommendations varied across sites depending on the initial baseline levels of breastfeeding and the sociocultural characteristics of each of the study subpopulations. Compliance was highest in Ghana and lowest in Brazil. Many Brazilian paediatricians recommended use of water and tea in the early months, prescribed formula when it was not necessary, and recommended complementary foods before children were 4 mo old [21]. Nevertheless, the efforts of the Brazilian lactation team made a substantial difference to the rates of exclusive/predominant breastfeeding and the duration of breastfeeding, resulting in a remarkable improvement compared to national and local rates [21]. In Ghana, breastfeeding is the norm, although exclusive breastfeeding rates in the general population are low. However, the provision of lactation support to the MGRS mothers increased the exclusive breastfeeding well beyond national levels [22].

Mothers who complied with the MGRS feeding criteria were less likely to be employed outside the home and more likely to have had a vaginal delivery, and fewer were primiparous. Similarly, pacifier use was more prevalent in the non-compliant group. Pacifier use has been associated with early weaning [23] and might partly explain the relatively early termination of breastfeeding in the Norwegian site despite long maternity leave (10 mo with 100% salary or 12 mo with 80% salary). Maternal education differed significantly between compliance groups when all sites were considered simultaneously, i.e. more highly educated mothers were more likely to comply with feeding criteria. However, the relationship went in opposite directions in the individual sites (Norway and Oman) where schooling was statistically different by compliance group. This might suggest cultural differences in the influence of education on breastfeeding practices.

Low rates of exclusive breastfeeding worldwide have raised concerns about the practicality of recommending a diet for children that occurs so infrequently [24]. However, recent evidence demonstrates that community-based breastfeeding counselling is a cost-effective way to increase exclusive breastfeeding rates [25–28]. Experience from the MGRS confirms this observation in six very different settings. The breastfeeding support team at each site served a critical role, particularly in providing lactation support during the first week or two after hospital discharge. Mothers were provided with information about avoiding sore nipples through correct breastfeeding technique, early management of nipple trauma when it occurred, prevention and early treatment of breast engorgement, the disadvantages of early introduction of any food besides human milk, and overall

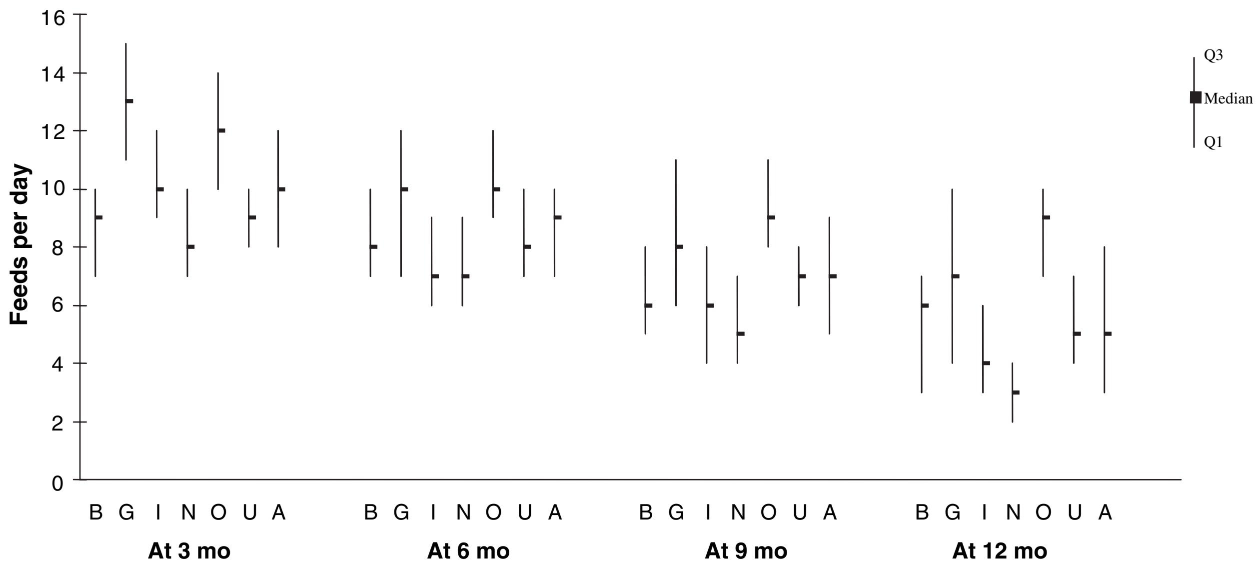


Figure 4. Median breastfeeding frequency among compliant infants by site and overall. B: Brazil; G: Ghana; I: India; N: Norway; O: Oman; U: USA; A: all sites.

Table IV. Median breastfeeding duration and continued breastfeeding at 24 mo by compliance category.

	Brazil	Ghana	India	Norway	Oman	USA	All
Compliant *n*	69	228	173	159	153	121	903
Non-compliant *n*	218	64	96	103	107	51	639
Duration of breastfeeding, median months (min., max.)							
Compliant	19.5* (12,24)	16.1* (12.1,24)	17.8* (12,24)	15.2* (12,24)	23.2* (12.3,24)	18.3* (12,24)	17.8* (12,24)
Non-compliant	6.3* (0.5,24)	10.3* (2,24)	9.4* (2,24)	10.3* (1,24)	17.4* (1.5,24)	10.5* (1.4,24)	9.3* (0.5,24)
Percent still breastfeeding at 24 mo							
Compliant	33.3*	5.7	23.1*	8.2*	11.8	32.2*	16.2*
Non-compliant	4.1*	1.6	4.2*	1.0*	9.3	5.9*	4.4*

*Statistically significant difference (p-value <0.05) between the compliant and non-compliant groups.

Table V. Breastfeeding initiation and pacifier use by compliance category and site.

	Brazil	Ghana	India	Norway	Oman	USA	All
Compliant *n*	69	228	173	159	153	121	903
Non-compliant *n*	218	64	96	103	107	51	639
Baby breastfed within 1 h of birth, %							
Compliant	–	57.1*	23.1	84.9	96.7	77.7	65.7
Non-compliant	–	40.7*	16.7	76.7	95.3	64.7	61.0
Median hours after birth baby breastfed for first time, h (min., max.)							
Compliant	–	5 (1,25)	4 (2,37)	2 (1,21)	2 (2,3)	2 (2,8)	4 (1,37)
Non-compliant	–	6 (2,28)	5 (2,50)	3 (1,20)	2 (1,5)	2 (2,25)	4 (1,50)
Use of pacifier at 2 wk, %							
Compliant	–	3.3	0.6	18.2	0.0	12.4	6.4
Non-compliant	–	1.8	1.0	18.6	0.9	3.9	5.8
Use of pacifier at 3 mo, %							
Compliant	–	3.2	0.6	44.3*	2.0	41.7	16.0*
Non-compliant	–	8.5	1.0	61.0*	2.9	45.1	23.0*
Use of pacifier at 6 mo, %							
Compliant	–	2.4	0.6	47.5	1.3*	41.3	16.3*
Non-compliant	–	1.9	0.0	60.2	5.8*	42.0	22.1*

*Statistically significant difference (p-value <0.05) between the compliant and non-compliant groups.

raised consciousness regarding the importance of breastfeeding for mothers and babies. The challenge is to extend this support, including guidance on breastfeeding techniques and ways to resolve problems, ideally as part of routine health services for the entire population.

The MGRS was designed to construct growth standards based on healthy breastfed infants and thereby establish coherence with national [29] and international [12] infant feeding guidelines that recommend breastfeeding as the optimal source of nutrition during infancy. Recognizing the adequacy of human milk to support not only healthy growth [24,29,30] but also cognitive development [31] and long-term health [32,33], the resulting growth standards [20] are recommended for application to all children independently of type of feeding.

Acknowledgements

This paper was prepared by Mercedes de Onis, Laurie A. Nommsen-Rivers, Anne Baerug, Anna Lartey, Adelheid Onyango, Elaine Albernaz, Nita Bhandari, Jose Martines, Nitya S. Prakash and Isabella Sagoe-Moses on behalf of the WHO Multicentre Growth Reference Study Group. The statistical analysis was conducted by Amani Siyam and Alain Pinol.

References

[1] WHO. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. Technical Report Series No. 854. Geneva: World Health Organization; 1995.

[2] de Onis M, Habicht JP. Anthropometric reference data for international use: recommendations from a World Health Organization Expert Committee. Am J Clin Nutr 1996;64: 650–8.

[3] de Onis M, Yip R. The WHO Growth Chart: historical considerations and current scientific issues. Bibl Nutr Dieta 1996;53:74–89.

[4] WHO Working Group on Infant Growth. An evaluation of infant growth. Geneva: World Health Organization; 1994.

[5] WHO Working Group on Infant Growth. An evaluation of infant growth: the use and interpretation of anthropometry in infants. Bull World Health Organ 1995;73:165–74.

[6] Cole TJ, Paul AA, Whitehead RG. Weight reference charts for British long-term breastfed infants. Acta Paediatr 2002;91: 1296–300.

[7] de Onis M, Onyango AW. The Centers for Disease Control and Prevention 2000 growth charts and the growth of breastfed infants. Acta Paediatr 2003;92:413–9.

[8] de Onis M, Garza C, Habicht JP. Time for a new growth reference. Pediatrics 1997;100:E8.

[9] Garza C, de Onis M, for the WHO Multicentre Growth Reference Study Group. Rationale for developing a new international growth reference. Food Nutr Bull 2004;25 Suppl 1:S5–14.

[10] de Onis M, Garza C, Victora CG, Onyango AW, Frongillo EA, Martines J, for the WHO Multicentre Growth Reference Study Group. The WHO Multicentre Growth Reference Study: planning, study design and methodology. Food Nutr Bull 2004;25 Suppl 1:S15–26.

[11] de Onis M, Garza C, Victora CG, Bhan MK, Norum KR, editors. WHO Multicentre Growth Reference Study (MGRS): Rationale, planning and implementation. Food Nutr Bull 2004;25 Suppl 1:S1–89.

[12] Fifty-fourth World Health Assembly. Resolution WHA54.2, Infant and young child nutrition. Geneva: World Health Organization; 2001.

[13] Prakash NS, Mabry RM, Mohamed AJ, Alasfoor D, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in Oman. Food Nutr Bull 2004;25 Suppl 1:S78–83.

[14] Dewey KG, Cohen RJ, Nommsen-Rivers LA, Heinig MJ, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in the USA. Food Nutr Bull 2004;25 Suppl 1:S84–9.

[15] Araujo CL, Albernaz E, Tomasi E, Victora CG, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in Brazil. Food Nutr Bull 2004;25 Suppl 1:S53–9.

[16] Lartey A, Owusu WB, Sagoe-Moses I, Gomez V, Sagoe-Moses C, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in Ghana. Food Nutr Bull 2004;25 Suppl 1:S60–5.

[17] Bhandari N, Taneja S, Rongsen T, Chetia J, Sharma P, Bahl R, et al., for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in India. Food Nutr Bull 2004;25 Suppl 1:S66–71.

[18] Baerug A, Bjoerneboe GE Aa, Tufte E, Norum KR, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in Norway. Food Nutr Bull 2004;25 Suppl 1:S72–7.

[19] WHO Multicentre Growth Reference Study Group. Complementary feeding in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450:27–37.

[20] WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards based on length/height, weight and age. Acta Paediatr Suppl 2006;450:76–85.

[21] Albernaz E, Giugliani ER, Victora CG. Supporting breastfeeding: a successful experience. J Hum Lact 1998;14:283–5.

[22] Ghana Statistical Service, Noguchi Memorial Institute for Medical Research and ORC Macro. Ghana Demographic and Health Survey 2003. Calverton, Maryland: GSS, NMIMR and ORC Macro; 2004.

[23] Barros FC, Victora CG, Semer TC, Tonioli Filho S, Tomasi E, Weiderpass E. Use of pacifiers is associated with decreased breast-feeding duration. Pediatrics 1995;95:497–9.

[24] WHO. The optimal duration of exclusive breastfeeding. Report of an Expert Consultation. Geneva: World Health Organization; 2002.

[25] Morrow AL, Guerrero ML, Shults J, Calva JJ, Lutter C, Bravo J, et al. Efficacy of home-based peer counselling to promote exclusive breastfeeding: a randomised controlled trial. Lancet 1999;353:1226–31.

[26] Haider R, Ashworth A, Kabir I, Huttly SR. Effect of community-based peer counsellors on exclusive breastfeeding practices in Dhaka, Bangladesh: a randomised controlled trial. Lancet 2000;356:1643–7.

[27] Albernaz E, Victora CG. Impacto do aconselhamento face a face sobre a duracão do aleitamento exclusivo: um estudo de revisão. Pan Am J Public Health 2003;14:17–24.

[28] Aidam BA, Perez-Escamilla R, Lartey A. Lactation counseling increases exclusive breast-feeding rates in Ghana. J Nutr 2005; 135:1691–5.

[29] American Academy of Pediatrics Policy Statement. Breastfeeding and the use of human milk. Pediatrics 2005;115:496–506.

[30] Harder T, Bergmann R, Kallischnigg G, Plagemann A. Duration of breastfeeding and risk of overweight: A meta-analysis. Am J Epidemiol 2005;162:397–403.

[31] Anderson JW, Johnstone BM, Remley DT. Breastfeeding and cognitive development: a meta-analysis. Am J Clin Nutr 1999; 70:525–35.

[32] Owen CG, Whincup PH, Odoki K, Gilg JA, Cook DG. Infant feeding and blood cholesterol: a study in adolescents and a systematic review. Pediatrics 2002;110:597–608.

[33] Singhal A, Cole TJ, Fewtrell M, Lucas A. Breastmilk feeding and lipoprotein profile in adolescents born preterm: follow-up of a prospective randomised study. Lancet 2004;363:1571–8.

Acta Pædiatrica, 2006; Suppl 450: 27–37

Taylor & Francis
Taylor & Francis Group

Complementary feeding in the WHO Multicentre Growth Reference Study

WHO MULTICENTRE GROWTH REFERENCE STUDY GROUP[1,2]

[1]*Department of Nutrition, World Health Organization, Geneva, Switzerland, and* [2]*Members of the WHO Multicentre Growth Reference Study Group (listed at the end of the first paper in this supplement)*

Abstract

Aim: To describe complementary feeding practices in the Multicentre Growth Reference Study (MGRS) sample. *Methods:* Food frequency questionnaires and 24-h dietary recalls were administered to describe child feeding throughout the first 2 y of life. This information was used to determine complementary feeding initiation, meal frequency and use of fortified foods. Descriptions of foods consumed and dietary diversity were derived from the 24-h recalls. Compliance with the feeding recommendations of the MGRS was determined on the basis of the food frequency reports. Descriptive statistics provide a profile of the complementary feeding patterns among the compliant children. *Results:* Complementary feeding in the compliant group began at a mean age of 5.4 mo (range: 4.8 (Oman)–5.8 mo (Ghana)). Complementary food intake rose from 2 meals/d at 6 mo to 4–5 meals in the second year, in a reverse trend to breastfeeding frequency. Total intake from the two sources was 11 meals/d at 6–12 mo, dropping to 7 meals/d at 24 mo. Inter-site differences in total meal frequency were mainly due to variations in breastfeeding frequency. Grains were the most commonly selected food group compared with other food groups that varied more by site due to cultural factors, for example, infrequent consumption of flesh foods in India. The use of fortified foods and nutrient supplements was also influenced by site-variable practices. Dietary diversity varied minimally between compliance groups and sites.

Conclusion: Complementary diets in the MGRS met global recommendations and were adequate to support physiological growth.

Key Words: *Complementary feeding, dietary diversity index, food frequency, infant feeding, 24-hour dietary recall*

Introduction

The WHO Multicentre Growth Reference Study (MGRS) was designed to collect growth data from an international sample of healthy breastfed infants from widely differing ethnic backgrounds and cultural settings (Brazil, Ghana, India, Norway, Oman and the USA) [1]. These data have been used to create the new length/height- and weight-based growth standards presented in this supplement [2]. As described elsewhere [3], complementary feeding practices were one of the secondary criteria used for selection of the study sites for the MGRS. The intention was to select populations in which feeding practices were unlikely to pose any constraints on growth. Thus, it is important to document how the children in the MGRS sample were fed in each of the sites.

The period of complementary feeding, when other foods are added to the diet of breastfed children, is a time of particular vulnerability to nutritional deficiencies. This is because children at this age are growing and developing rapidly, yet do not consume large quantities of food. Thus, the foods they eat must be of high nutrient density to provide adequate amounts of essential nutrients. In recent years increasing attention has been paid to the importance of complementary feeding [4,5]. The key limiting nutrients identified for breastfed children between the ages of 6 and 24 mo are iron, zinc, vitamin B_6 and, in some populations, riboflavin, niacin, thiamin, calcium, vitamin A, folate and vitamin C. Vitamin D is also of concern in populations with low exposure to sunshine or at high latitudes. In 2003, global guidelines for complementary feeding of the breastfed child were published [6]. These included recommendations on 1) introducing complementary foods at 6 mo of age, 2) continued breastfeeding to 2 y of age or beyond, 3) responsive feeding practices, 4) safe, hygienic preparation and feeding of complementary

Correspondence: Mercedes de Onis, Study Coordinator, Department of Nutrition, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland. Tel: +41 22 791 3320. Fax: +41 22 791 4156. E-mail: deonism@who.int

ISSN 0803-5326 print/ISSN 1651-2227 online
DOI: 10.1080/08035320500495456

foods, 5) amounts of complementary foods needed at each age interval, 6) food consistency, 7) meal frequency and energy density, 8) assuring adequate nutrient intake from complementary foods, 9) use of fortified foods or nutrient supplements, and 10) feeding during and after illness. This paper describes the complementary feeding practices of the sample of infants and young children used to construct the WHO Child Growth Standards, and discusses the patterns observed with regard to several of the global guidelines above, such as age of introduction of complementary foods, meal frequency, dietary quality, and use of fortified foods or nutrient supplements.

Methods

Overview of the MGRS

The MGRS was a six-country community-based project designed to develop new growth standards for infants and young children. The design included a longitudinal component that followed children from birth to 24 mo and a cross-sectional component that enrolled children aged 18 to 71 mo. The pooled sample from all six countries included 8440 children. The study subpopulations were selected so that socio-economic conditions would be favourable to growth, and the selection criteria for individuals specified absence of health or environmental constraints on growth, adherence to recommended infant feeding practices, absence of maternal smoking, single term birth, and absence of significant morbidity. This paper describes data from the longitudinal component of the MGRS where mothers and newborns were screened and enrolled at birth and visited in the home at weeks 1, 2, 4 and 6, monthly from 2 to 12 mo, and every 2 mo in the second year of life. Details of the study design and methods can be found elsewhere [3].

Complementary feeding guidelines

As described elsewhere [7–12], mothers in each site were given guidelines on complementary feeding. Mothers were advised to introduce complementary foods at 4–6 mo (the WHO recommendation prior to 2001) in Norway and the USA and, in line with individual national policies, at 5–6 mo in Oman and at 6 mo in Brazil, Ghana and India. In all sites, continued breastfeeding was recommended and the guidelines emphasized use of a variety of nutrient-rich foods. Most of the sites also included guidelines regarding meal frequency, food consistency, use of a separate bowl for the infant, use of iron-rich and vitamin A-rich foods, and responsive feeding practices. Half of the sites included advice on nutrient supplements (India, Norway and the USA), limitations on use of sugary beverages such as juice (Norway, Oman and the USA), and avoidance of certain foods if there was a family history of allergy (Norway, Oman and the USA). India and Oman provided guidelines on the amounts of foods to be fed. Ghana and India included recommendations regarding hygiene when preparing and feeding complementary foods. Norway and the USA included advice to use infant formula if a supplement to breast milk was needed. India and Oman advised using only iodized salt, while Norway advised against adding salt to baby food.

Compliance criteria

As described elsewhere [13], the MGRS included three compliance criteria with regard to infant feeding: 1) exclusive or predominant breastfeeding for at least 4 mo (120 d), 2) introduction of complementary foods between 4 and 6 mo (120 to 180 d), and 3) partial breastfeeding to be continued for at least 12 mo (365 d). The operational definition of compliance with the first criterion was that the infant did not consume formula, other milk, or more than one teaspoon of solid or semi-solid food on more than 10% of days during the first 4 mo (i.e. $\leq$12 d). This paper focuses on the complementary feeding practices of subjects who were "compliant" with all three feeding criteria, with brief reference to whether the results for the "non-compliant" subjects differed substantially from those for the compliant subjects. The final sample used to construct the growth standards also excluded children whose mothers smoked and those experiencing morbidity with adverse effects on growth [2].

Definitions of variables and data analysis

Data on feeding practices, including a 24-h dietary recall, were collected at each of the follow-up visits [3]. Before conducting the 24-h recall, the interviewer asked the mother if the child's diet on the preceding day was typical. If not (e.g. because of illness or travelling), then the recall data were collected for the last day when the diet was typical. The mother was asked what the child ate or drank in each of seven time periods during the day (when the child woke up; morning; lunch; afternoon; dinner; evening; during the night).

The results presented here come from several questions in the follow-up questionnaire. Age of introduction of solid or semi-solid foods was derived from a question about whether the child had received certain fluids or foods since the previous visit (which at this age was an interval of 1 mo). If the answer was yes for either of the two non-fluid choices ("fruit" or "solid or semi-solid foods"), the age of the child at the current visit was taken as the age of introduction. Meal frequency was derived from the 24-h recall. If a child ate twice within 45 min, it was considered a single meal. Water, tea, juice or other beverages consumed on their own were not considered as meals, nor were small snacks (e.g. a small cookie or a spoonful of mashed fruit). The total number of meals included both solid/semi-solid foods and milk-only meals (including breast milk). Milk-only meals included breastfeeds and feedings of formula, milk or yogurt.

Data on the types of foods consumed and dietary diversity were also derived from the 24-h recall. Foods were grouped into 12 categories based on type and nutrient content: grain products, legumes/nuts, tubers, milk products, flesh foods (meat, poultry and fish), eggs, vitamin A-rich fruits and vegetables, other fruits and vegetables, juices, sweetened beverages, soups, and fats/oils (in Brazil, 11 categories were used, without separating vitamin A-rich fruits and vegetables from other fruits and vegetables). To assess dietary diversity, an index developed by other investigators [14] was used, based on the following eight food groups: 1) grain products and tubers, 2) legumes/nuts, 3) milk products, 4) flesh foods, 5) eggs, 6) vitamin-A rich fruits and vegetables, 7) other fruits and vegetables and juices, and 8) fats/oils. This categorization of foods was chosen so that a higher total score would be likely to reflect greater consumption of foods of higher nutrient density, such as animal-source foods (three categories) and fruits and vegetables (two categories). The number of food groups represented in the child's diet (range 0–8, except in Brazil where it was 0–7), regardless of the amount consumed from each food group, was calculated as a measure of dietary diversity. Thus, for example, if the 24-h recall showed that six out of the eight food groups were represented in the diet, then the dietary diversity for that day was 6.

Use of fortified foods was derived from a question that followed the 24-h recall: "Were any of the foods fortified with any of the following nutrients: a) iron, b) vitamin A, c) vitamin C, d) vitamin D, e) other (specify)?" The site-specified fortificants in the "other" category were calcium and zinc (India, Norway, the USA), vitamin E and vitamin B-complex (Ghana), and folic acid (Oman). Use of salt was determined by asking: "Do you add salt to his/her food?" If the mother answered "yes", this was followed by: "Please show me the type of salt you put in your baby's food. I would like to check if it contains iodine, which is important for the baby."

Use of nutrient supplements was determined by asking: "Since the last visit, has your baby received any vitamins or minerals?" If the response was yes, data were recorded on the brand name of the supplement, the dose given and the frequency of supplementation (per day, week or month). The nutrient contents of all supplements used at each site were recorded in order to determine which specific nutrients were taken by each child.

Basic summary statistics such as means, standard deviations, medians, summary ranges and frequency distributions were used in these analyses.

Results

Table I shows the mean and median age of introduction of solid or semi-solid foods for the compliant subjects. The overall mean was 5.4 mo, ranging from 4.8 mo in Oman to 5.8 mo in Ghana. For non-compliant subjects (data not shown), the overall mean was somewhat lower (4.8 mo).

Figure 1 shows the mean number of non-milk meals between 4 and 24 mo for the compliant subjects at each site. Values were close to zero at 4 mo, increasing to an overall average (meals/d) of about 2 at 6 mo, 4 at 9 mo and 4–5 at 12–24 mo. During the first year of life, meal frequency was generally similar across sites, but in the second year the children in Ghana tended to eat somewhat more often and the children in Brazil less often than the children in the other sites. For non-compliant subjects, non-milk meal frequency was slightly higher at 6 mo (~3 meals/d) compared to compliant subjects, but at the older ages the mean values were similar. Figure 2 shows the mean number of all meals (including milk-only meals) for the compliant subjects. The average decreased with age, from an overall mean of ~11 meals/d at 6–12 mo to ~9 at 18 mo and ~7 at 24 mo. Because of differences in breastfeeding frequency across sites [13], total meal frequency at 4–12 mo tended to be higher in Ghana and Oman and lower in Norway than in the other sites. After 12 mo, total meal frequency remained high in Oman, was lowest in Norway and dropped steadily throughout the second year in

Table I. Mean and median age (in months) of the introduction of solid or semi-solid foods for compliant children.

Site	*n*	Mean (SD)	Median (min., max.)[a]
Brazil[b]	68	5.5 (0.7)	6.0 (4.0, 7.1)
Ghana	228	5.8 (0.6)	6.0 (1.4, 7.5)
India	173	5.0 (0.6)	5.0 (3.0, 7.0)
Norway	159	5.5 (0.8)	5.1 (3.0, 7.1)
Oman	153	4.8 (0.6)	5.0 (3.2, 7.0)
USA	121	5.4 (0.7)	5.1 (3.9, 7.2)
All	902	5.4 (0.7)	5.1 (1.4, 7.5)

[a] The minimum age of introduction is less than 4 mo in several sites because the operational definition for compliance with exclusive or predominant breastfeeding for at least 4 mo allowed for occasional consumption of solid or semi-solid foods, as long as the number of days on which this occurred did not exceed 12.

[b] Excludes one child with missed visits at ages 6 and 7 mo.

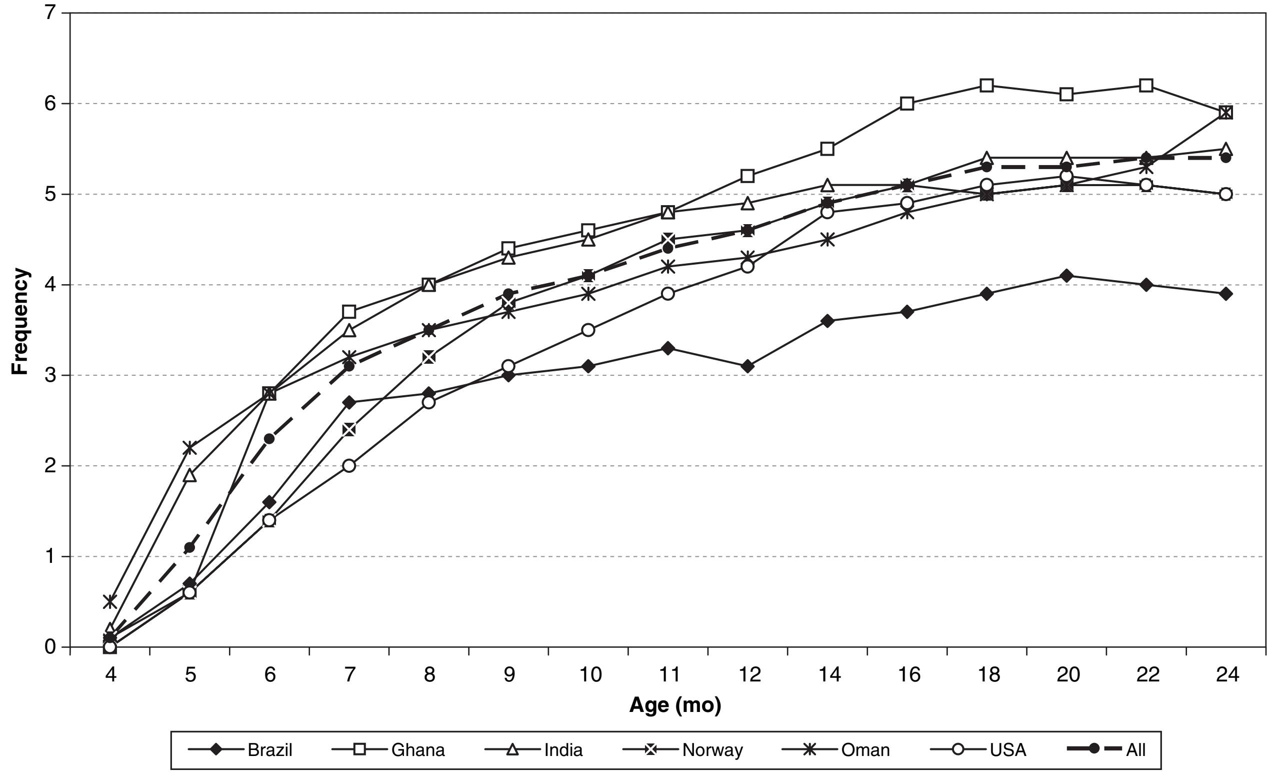


Figure 1. Mean number of non-milk meals (age 4 to 24 mo) per day for compliant children.

Ghana. For non-compliant subjects (data not shown), total meal frequency was lower by about 1–2 meals/d at 6–18 mo in comparison with the values shown for compliant subjects.

Subjects' food consumption patterns were evaluated by categorizing the foods reported in the 24-h recall into 12 food groups. The percentages of compliant children fed foods from these food groups

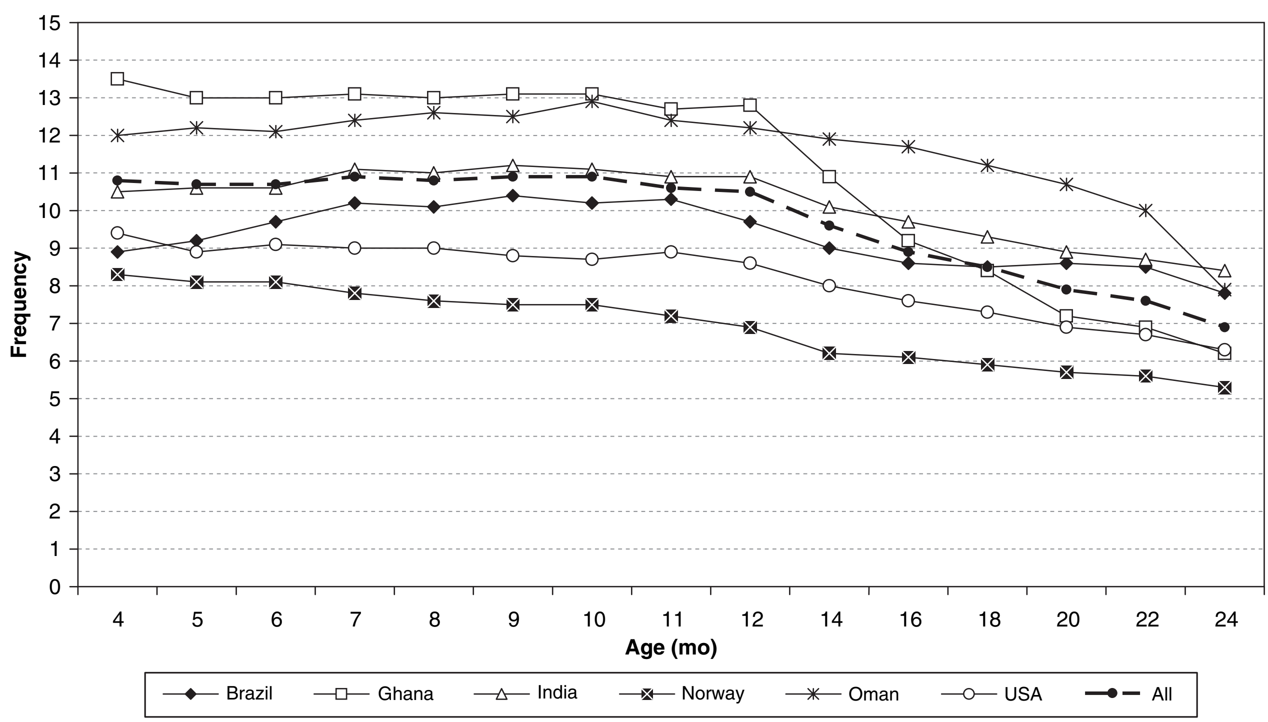


Figure 2. Mean number of all meals (age 4 to 24 mo) per day for compliant children.

by the stated ages are reported in Table II. Grain products were consumed by the vast majority of subjects at all ages (except in Brazil at 6 mo). There was wide variability in the percentage of children who consumed legumes or nuts after 6 mo: <6% in Norway, 12–21% in Oman, 9–43% in the USA, 36–47% in Ghana, 39–60% in Brazil, and 71–91% in India. Consumption of tubers was uncommon at 6 mo (except in Oman), but increased thereafter to 33–51% overall, with the highest percentages in Ghana, India and Oman. Consumption of milk products varied by site at 6–9 mo (high in Ghana, India and Oman, lower in Norway and the USA), but at 12–24 mo >75% of children in all sites consumed milk products. Flesh foods were rarely consumed at 6 mo (except in Oman), but intake rose thereafter. In all sites except India, the percentage of children consuming flesh foods on the day of the recall was >50% at 12 mo, >66% at 18 mo, and >75% at 24 mo; in India ≤11% of children consumed flesh foods on the day of the recall. Egg consumption varied by age and by site, with the overall percentage being 3–10% at 6–9 mo and ~20–30% thereafter. Eggs were rarely consumed in Norway (at all ages) and in the USA at 6–9 mo, whereas they were consumed by almost half the children in Oman at 24 mo. Consumption of vitamin A-rich fruits and vegetables was relatively low at 6 mo (except in Oman and the USA), but increased thereafter to 43–48% overall, with the highest percentages reported for Ghana and the USA. Other fruits and vegetables were consumed by 35% of children at 6 mo, with intake rising thereafter to 70–87% overall. At 6 mo, juice was infrequently consumed by infants in Ghana, Norway and the USA, but consumed by 20–45% of infants in Brazil, India and Oman; thereafter, juice consumption rose in all sites, with the highest percentages at 24 mo reported for Oman and the USA. Sweet beverages were rarely consumed at 6–9 mo (except in Brazil), with intake rising to 15–34% overall from 12 to 24 mo. Consumption of soup was highly variable across sites: it was common in Brazil, Oman and Ghana but uncommon in India, Norway and the USA. Consumption of fats and oils after 6 mo was also highly variable across sites, being very common in Ghana and Norway, less common in India and the USA, and rare in Brazil and Oman.

Food consumption patterns of the non-compliant subjects (data not shown) did not differ dramatically from those of compliant subjects with the following exceptions. Because they were less likely to be breastfeeding, non-compliant children were more likely to consume milk products. Compared to compliant subjects, they tended to have a lower consumption of vitamin A-rich fruits and vegetables and fats and oils (at all ages), and a higher consumption of soups at 6–9 mo.

Mean and median dietary diversity are shown in Table III for compliant subjects. The median number of food groups consumed, out of a maximum of eight (seven in Brazil), was two at 6 mo, four at 9 mo, and five at 12–24 mo. Values for Brazil were lower than for the other sites because the dietary diversity index included seven rather than eight food groups. For the other sites, dietary diversity at 6 mo was lower in Norway and higher in India and Oman compared to the overall median; by 18–24 mo dietary diversity was similar among sites except for Ghana, where it was higher. Dietary diversity was similar between compliant and non-compliant subjects (data not shown).

Use of fortified foods varied by age and by site. For simplicity, only data for foods fortified with iron or vitamin A are shown. Figure 3 shows the percentage of compliant subjects consuming iron-fortified foods at each age. The majority of infants consumed such foods at 6 mo in all sites, ranging from ~60% in Oman to 75–90% in Ghana and the USA. Thereafter, the percentage remained very high in the USA and rose from ~60% to ~85% in Oman, but declined in the other sites to ~40% in Ghana and ~20% in India and Norway by 24 mo (data were unavailable for Brazil). Figure 4 shows the percentage of compliant subjects consuming vitamin A-fortified foods. In Norway and Ghana, the percentage was ~80–100% at all ages, whereas in Oman it was 50–60%, in India it decreased from ~65% at 6 mo to ~20% at 24 mo, and in the USA it increased from ~20% at 6 mo to >90% at 24 mo. Use of iron- or vitamin A-fortified foods was generally somewhat lower among compliant than among non-compliant subjects (data not shown).

Salt was commonly used in the foods provided to the children, particularly after 6–9 mo. The percentage using salt between 6 and 24 mo increased from 71% to 99% in Brazil, 48% to 100% in Ghana, 82% to 100% in India, 2% to 80% in Norway, 40% to 100% in Oman, and 0% to 36% in the USA. Of those using salt in food, over 93% used iodized salt, except in Norway where 8–17% used non-iodized salt at 12–24 mo.

Use of nutrient supplements varied greatly by site. Table IV shows the percentage of children in the compliant group who received supplements that contained one or more of the specified nutrients. The fat-soluble vitamins A, D and E are often combined in one supplement for infants, and this combination was commonly used in Norway throughout the age range 6–24 mo (73–80% of children). Vitamins A and D were taken by 30–35% of children in Ghana and 12–40% of children in India. Between 12 and 44% of children in Norway, Ghana and India also used supplements containing vitamins C, B_1, B_2 and B_6. In Norway folate was taken by 15–22% of children, in Ghana niacin was taken by 23–29% of

Table II. Twenty-four-hour dietary intake (prevalence and median) from selected food subgroups by compliant children at 6–24 mo.

		Sites						
Food subgroups	Age (mo)	Brazil ($n=69$)	Ghana ($n=228$)	India ($n=173$)	Norway ($n=159$)	Oman ($n=153$)	USA ($n=121$)	All ($n=903$)
Grains	6	24.6 (1)	88.6 (3)	86.1 (2)	79.2 (1)	85.6 (2)	73.6 (1)	79.1 (2)
	9	59.4 (2)	94.3 (4)	97.7 (3)	97.5 (3)	94.8 (2)	93.4 (2)	92.8 (3)
	12	85.5 (3)	97.4 (4)	100.0 (4)	100.0 (3)	97.4 (3)	97.5 (3)	97.5 (3)
	18	95.7 (4)	98.7 (4)	97.1 (4)	98.1 (4)	98.7 (3)	99.2 (4)	98.1 (4)
	24	97.1 (3)	97.8 (4)	99.4 (4)	98.7 (3)	100.0 (4)	97.5 (4)	98.6 (4)
Legumes & nuts	6	5.8 (1)	20.2 (2)	35.3 (1)	0.0 (0)	10.5 (1)	0.8 (1)	14.2 (1)
	9	39.1 (1)	36.4 (2)	70.5 (1)	1.3 (1)	11.8 (1)	9.1 (1)	29.1 (1)
	12	43.5 (2)	38.6 (1)	87.3 (1)	3.8 (1)	15.7 (1)	33.9 (1)	37.7 (1)
	18	60.0 (2)	45.2 (1)	87.9 (2)	2.5 (1)	17.6 (1)	43.0 (1)	42.0 (1)
	24	56.5 (1)	47.4 (1)	90.8 (2)	5.7 (1)	20.9 (1)	43.0 (1)	44.0 (1)
Tubers	6	5.8 (1)	9.2 (1)	12.1 (1)	10.7 (1)	47.7 (1)	3.3 (1)	15.5 (1)
	9	18.8 (1)	39.0 (1)	38.2 (1)	30.2 (1)	45.8 (1)	11.6 (1)	33.2 (1)
	12	23.2 (1)	49.6 (1)	50.9 (1)	33.3 (1)	45.8 (1)	21.5 (1)	40.5 (1)
	18	38.6 (1)	55.7 (1)	59.0 (1)	44.0 (1)	51.0 (1)	21.5 (1)	47.6 (1)
	24	37.7 (1)	64.5 (1)	59.5 (1)	46.5 (1)	55.6 (1)	22.3 (1)	51.2 (1)
Milk (dairy) products	6	20.3 (1)	59.2 (2)	64.2 (2)	13.2 (1)	49.7 (1)	9.1 (2)	40.8 (2)
	9	75.4 (1)	73.2 (2)	82.7 (2)	33.3 (1)	73.2 (2)	43.0 (2)	64.1 (2)
	12	75.4 (2)	76.3 (2)	89.6 (3)	82.4 (2)	83.0 (2)	86.0 (2)	82.3 (2)
	18	91.4 (4)	89.9 (2)	92.5 (4)	95.6 (4)	94.1 (3)	97.5 (4)	93.3 (3)
	24	88.4 (4)	93.0 (2)	96.0 (4)	95.6 (4)	94.1 (4)	97.5 (3)	94.5 (3)
Flesh foods	6	2.9 (1)	11.4 (1)	0.0 (0)	3.1 (1)	26.8 (1)	1.7 (1)	8.4 (1)
	9	10.1 (1)	70.6 (2)	2.3 (1)	42.1 (1)	63.4 (1)	26.4 (1)	40.8 (1)
	12	50.7 (2)	81.1 (2)	6.4 (1)	66.7 (1)	77.8 (1)	59.5 (1)	58.5 (1)
	18	65.7 (2)	91.2 (2)	9.2 (1)	79.9 (2)	84.3 (1)	69.4 (1)	67.5 (2)
	24	81.2 (1)	93.9 (2)	11.0 (1)	80.5 (2)	77.8 (1)	76.9 (1)	69.7 (2)
Eggs	6	2.9 (1)	4.4 (1)	3.5 (1)	0.0 (0)	7.8 (1)	0.0 (0)	3.3 (1)
	9	7.2 (1)	13.2 (1)	8.1 (1)	1.9 (1)	22.2 (1)	5.8 (1)	10.3 (1)
	12	8.7 (1)	27.2 (1)	13.9 (1)	3.1 (1)	31.4 (1)	18.2 (1)	18.5 (1)
	18	27.1 (1)	35.5 (1)	26.0 (1)	5.7 (1)	42.5 (1)	21.5 (1)	27.1 (1)
	24	26.1 (1)	39.0 (1)	33.5 (1)	8.8 (1)	47.1 (1)	19.0 (1)	30.3 (1)
Vitamin A-rich fruits and vegetables[a]	6	–	7.5 (1)	15.0 (1)	7.5 (1)	38.6 (1)	34.7 (1)	17.3 (1)
	9	–	46.9 (1)	32.4 (1)	32.7 (1)	56.9 (1)	69.4 (1)	42.7 (1)
	12	–	53.9 (1)	31.8 (1)	29.6 (1)	46.4 (1)	78.5 (2)	43.3 (1)
	18	–	69.7 (2)	38.7 (1)	35.8 (1)	36.6 (1)	79.3 (2)	48.1 (1)
	24	–	73.2 (2)	38.7 (1)	34.6 (1)	37.3 (1)	75.2 (1)	48.4 (1)
Other fruits and vegetables	6	73.9 (1)	11.4 (1)	52.6 (1)	23.9 (1)	39.9 (1)	37.2 (1)	34.6 (1)
	9	71.0 (1)	68.9 (1)	73.4 (1)	67.9 (1)	62.1 (1)	84.3 (2)	70.7 (1)
	12	76.8 (2)	78.9 (2)	85.0 (2)	79.9 (2)	69.9 (1)	89.3 (2)	80.0 (2)
	18	75.7 (2)	94.3 (2)	87.9 (2)	84.3 (2)	81.7 (1)	90.9 (2)	87.3 (2)
	24	72.5 (2)	94.7 (2)	91.9 (2)	78.0 (2)	83.7 (2)	89.3 (2)	86.9 (2)
Juice	6	27.5 (1)	9.2 (1)	19.7 (1)	0.6 (1)	45.1 (1)	3.3 (1)	16.4 (1)
	9	49.3 (1)	19.3 (1)	15.6 (1)	5.7 (1)	46.4 (1)	20.7 (1)	23.3 (1)
	12	52.2 (2)	24.6 (1)	20.2 (1)	12.6 (1)	60.8 (1)	43.0 (1)	32.3 (1)
	18	55.7 (1)	29.8 (1)	17.3 (1)	27.7 (1)	56.9 (1)	50.4 (2)	36.4 (1)
	24	40.6 (1)	44.7 (1)	29.5 (1)	39.0 (1)	63.4 (1)	65.3 (1)	46.4 (1)
Sweet beverages	6	23.2 (1)	1.8 (1.5)	1.7 (1)	6.3 (1)	4.6 (1)	0.0 (0)	4.4 (1)
	9	14.5 (1)	6.1 (1)	6.4 (1)	11.3 (1)	5.2 (1)	0.0 (0)	6.8 (1)
	12	34.8 (1.5)	11.8 (1)	9.8 (1)	24.5 (1)	11.8 (1)	8.3 (1)	15.0 (1)
	18	48.6 (2)	17.5 (1)	16.2 (1)	44.0 (1)	25.5 (1)	33.9 (1)	27.9 (1)
	24	1.4 (3)	25.4 (1)	23.1 (1)	59.7 (1)	35.9 (1)	51.2 (1)	34.4 (1)
Soup	6	66.7 (1)	4.4 (1)	12.1 (1)	0.0 (0)	38.6 (1)	0.0 (0)	15.1 (1)
	9	63.8 (2)	26.3 (1)	10.4 (1)	1.9 (1)	42.5 (1)	0.8 (1)	21.2 (1)
	12	47.8 (2)	34.2 (1)	8.7 (1)	2.5 (1)	40.5 (1)	3.3 (1)	21.7 (1)
	18	37.1 (1)	39.0 (1)	5.2 (1)	6.9 (1)	30.1 (1)	5.0 (1)	20.7 (1)
	24	21.7 (1)	34.2 (1)	10.4 (1)	9.4 (1)	26.1 (1)	2.5 (1)	18.7 (1)

Table II (*Continued*)

Food subgroups	Age (mo)	Brazil ($n=69$)	Ghana ($n=228$)	India ($n=173$)	Norway ($n=159$)	Oman ($n=153$)	USA ($n=121$)	All ($n=903$)
		Sites						
Fats & oils	6	0.0 (0)	6.6 (1)	15.6 (1)	2.5 (1)	1.3 (1)	0.0 (0)	5.3 (1)
	9	0.0 (0)	59.2 (1)	32.9 (1)	36.5 (1)	3.9 (2)	0.8 (2)	28.5 (1)
	12	0.0 (0)	72.8 (2)	45.1 (1)	64.2 (2)	2.0 (1)	10.7 (1)	40.1 (1)
	18	0.0 (0)	89.5 (2)	45.1 (1)	81.8 (2)	3.9 (1)	22.3 (1)	49.2 (2)
	24	2.9 (1)	94.3 (2)	43.4 (1)	83.6 (2)	4.6 (1)	19.8 (1)	50.5 (2)

[a] In Brazil, vitamin A-rich fruits and vegetables were not separated from other types of fruits and vegetables.

children, and in India vitamin B_{12} was taken by 10–17% of children. In Brazil, Oman and the USA, use of vitamin supplements was rare (generally <10%). Use of mineral supplements was rare except for iron in Brazil (7–19% of children) and iron and zinc in India (9–13% of children). Among non-compliant subjects, use of nutrient supplements was generally similar to the patterns observed for compliant subjects.

Discussion

These results document that the complementary feeding practices for the subjects included in the "compliant" group for the MGRS were generally consistent with the recently published *Guiding Principles for Complementary Feeding of the Breastfed Child* [6].

The overall mean age of introduction of solid or semi-solid foods was 5.4 mo, with relatively little variability across sites. The MGRS was initiated before the WHO policy on the optimal duration of exclusive breastfeeding was changed in 2001 from "4–6 months" to "6 months" [15,16], although in three of the six sites (Brazil, Ghana and India) national policy recommended 6 mo. The two highest mean values for age of introduction of complementary foods were in two of these three sites (5.8 mo in Ghana and 5.6 mo in Brazil), though the means in Norway (5.5 mo) and the USA (5.4 mo) were not much lower. The lowest mean value was in Oman (4.8 mo), where the policy at the time was to recommend introduction at 5 mo. It should be noted that these mean values are biased towards older ages because the actual age of introduction of solid or semi-solid foods could have occurred up to a month prior to the date of the interview.

Solids or semi-solids were fed on average about twice per day at 6 mo, three times per day at 9 mo, four times per day at 12 mo and 4–5 times per day at 14–24 mo. These means are consistent with the recommendations in the *Guiding Principles*, which state that breastfed infants should be given meals of complementary foods 2–3 times per day at 6–8 mo and 3–4 times per day at 9–11 and 12–24 mo, with additional nutritious snacks offered 1–2 times per day as desired [6].

There was considerable variability in the types of food consumed by children in each of the sites, which

Table III. Mean and median dietary diversity index[a] at selected ages.

Age (mo)		Brazil ($n=69$)	Ghana ($n=228$)	India ($n=173$)	Norway ($n=159$)	Oman ($n=153$)	USA ($n=121$)	All ($n=903$)
		Sites						
6	Mean (SD)	1.4 (0.8)	2.2 (1.3)	2.8 (1.3)	1.3 (0.9)	2.9 (1.3)	1.6 (1.0)	2.1 (1.3)
	Median (min., max.)	1.0 (0,4)	2.0 (0,7)	3.0 (0,7)	1.0 (0,4)	3.0 (0,7)	2.0 (0,4)	2.0 (0,7)
9	Mean (SD)	2.8 (1.0)	4.7 (1.8)	4.1 (1.2)	3.2 (1.5)	4.1 (1.2)	3.3 (1.3)	3.9 (1.5)
	Median (min., max.)	3.0 (0,5)	5.0 (0,8)	4.0 (0,7)	3.0 (0,7)	4.0 (0,7)	3.0 (0,6)	4.0 (0,8)
12	Mean (SD)	3.5 (1.3)	5.3 (1.6)	4.6 (1.1)	4.3 (1.2)	4.4 (1.2)	4.8 (1.2)	4.6 (1.4)
	Median (min., max.)	4.0 (0,6)	6.0 (0,8)	5.0 (2,7)	4.0 (2,7)	4.0 (0,7)	5.0 (0,8)	5.0 (0,8)
18	Mean (SD)	4.3 (1.1)	6.2 (1.2)	4.9 (1.3)	4.9 (1.1)	4.7 (1.0)	5.3 (1.0)	5.2 (1.3)
	Median (min., max.)	4.0 (0,6)	6.0 (0,8)	5.0 (0,8)	5.0 (0,7)	5.0 (0,7)	5.0 (3,8)	5.0 (0,8)
24	Mean (SD)	4.3 (1.2)	6.3 (1.3)	5.1 (1.0)	4.9 (1.2)	4.8 (1.0)	5.3 (1.2)	5.3 (1.3)
	Median (min., max.)	4.0 (0,6)	6.0 (0,8)	5.0 (2,8)	5.0 (0,7)	5.0 (2,8)	5.0 (0,8)	5.0 (0,8)

[a] Dietary diversity index: the sum (1 = yes, 0 = no) of eight food groups (seven food groups for Brazil): 1) grains and tubers; 2) legumes and nuts; 3) milk products; 4) flesh foods; 5) eggs; 6) vitamin-A rich fruits and vegetables; 7) other fruits and vegetables and juices; 8) fats and oils.

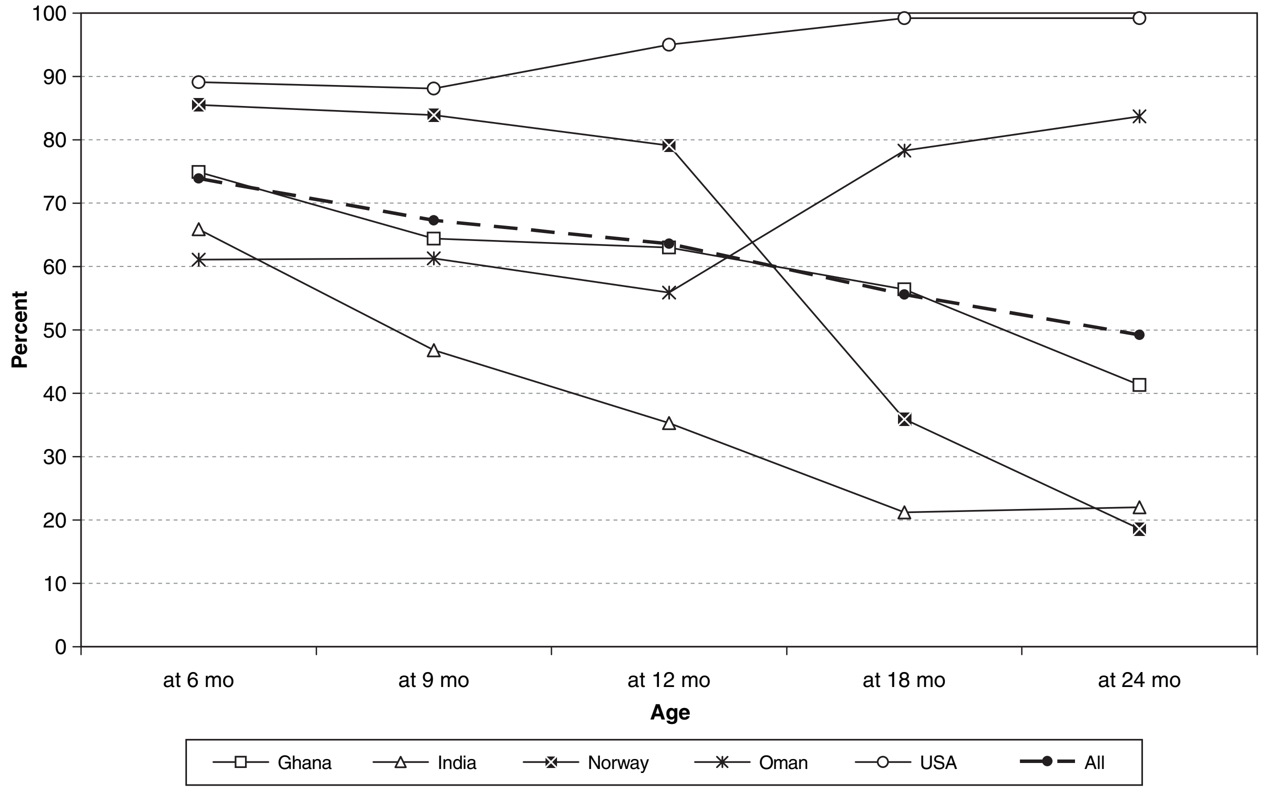


Figure 3. Percentage of compliant children consuming iron-fortified foods at selected ages.

is not surprising given the cultural differences in food habits across countries. Nonetheless, there were certain commonalities that indicate that the diets were generally of high nutritional quality in all sites. For example, in the second year of life, >75% of children in each site consumed milk products and fruits/vegetables, and 50–95% consumed meat, poultry or fish (except in India) on the day of the recall. These dietary characteristics reflect the high socioeconomic status of the subjects included in the MGRS. Some of the differences across sites may be due to variability in the complementary feeding guidelines that parents were given, either from the MGRS staff or from healthcare providers. For example, advice to avoid potentially allergenic foods such as eggs and nuts (in families with a history of allergy, though this caveat is not always added by healthcare providers) was given in Norway, Oman and the USA, which may explain the lower percentage of children with intake from the egg (except Oman) and legumes/nuts food groups in these sites, at least during the first year of life. The guidelines in these three sites also advised limiting the intake of juice, which may account for the low frequency of juice consumption at 6–9 mo in Norway and the USA (though this was not evident in Oman). In addition, the guidelines in Norway advised against adding salt to foods for infants, and the rates of salt usage during the first year of life were correspondingly low in that site (though they were also low in the USA, which may reflect general public concern about excessive salt intake).

Median dietary diversity on the day of the recall increased from two food groups at 6 mo to five food groups (out of a maximum of eight) at 12–24 mo. Using the same dietary diversity indicator [14], the values at 9–12 mo (generally 4–5 food groups) are higher than the averages observed for low-income populations in Peru (3.7 food groups), Ghana (3.4 food groups) and Bangladesh (2.1 food groups). This indicates that MGRS subjects generally consumed a varied diet, which on any given day typically included fruits and/or vegetables and at least one type of animal-source food, in addition to the usual staple foods. Dietary diversity is correlated with nutritional adequacy of the complementary food diet at this age ($r = 0.4–0.7$) [14].

Use of fortified foods and nutrient supplements varied greatly across sites. Most infants received iron-fortified foods at 6 mo, but the percentage continuing to receive such foods through the first and second years of life was not consistently high. This probably reflects the lack of uniform policies about the recommended duration of use of such products for infants and toddlers. Vitamin supplements (which included vitamin D) were commonly given in Norway, presumably because of recommendations that breastfed infants in populations at high latitudes receive an

Table IV. Percentages of compliant children who received supplements at selected ages.

Supplement	Age (mo)	Brazil ($n = 69$)	Ghana ($n = 228$)	India ($n = 173$)	Norway ($n = 159$)	Oman ($n = 153$)	USA ($n = 121$)	All ($n = 903$)
Vitamin A	6	8.7	29.8	39.3	78.6	2.6	0.0	30.0
	9	7.2	35.1	28.9	80.5	11.8	1.7	31.3
	12	13.0	36.4	24.3	78.0	2.6	2.5	29.3
	18	2.9	30.3	18.5	74.2	1.3	5.0	25.3
	24	1.4	33.8	12.1	74.8	0.7	14.0	26.1
Vitamin D	6	8.7	29.8	39.9	78.6	2.0	0.0	30.0
	9	7.2	35.1	30.6	80.5	0.7	1.7	29.8
	12	13.0	36.4	26.0	80.5	2.0	2.5	30.0
	18	4.3	32.0	25.4	76.7	0.7	5.0	27.5
	24	1.4	34.2	16.8	77.4	0.7	14.0	27.6
Vitamin E	6	4.3	7.0	25.4	77.4	0.0	0.0	20.6
	9	1.4	9.2	19.1	78.6	0.7	0.0	20.0
	12	4.3	7.9	15.6	77.4	0.0	1.7	19.2
	18	2.9	7.0	8.1	73.0	0.0	3.3	16.8
	24	0.0	7.5	5.2	76.1	0.0	11.6	17.8
Vitamin C	6	2.9	26.3	34.1	20.8	3.3	0.0	17.6
	9	5.8	31.1	23.1	30.2	1.3	1.7	18.5
	12	8.7	31.1	19.7	28.9	3.3	2.5	18.3
	18	5.7	27.6	18.5	29.6	0.7	6.6	17.1
	24	1.4	25.0	12.1	32.7	2.0	15.7	16.9
Vitamin B_1	6	2.9	26.3	43.9	21.4	3.3	0.0	19.6
	9	5.8	36.4	31.8	30.2	1.3	0.0	21.3
	12	7.2	33.3	27.7	28.9	3.3	1.7	20.2
	18	7.1	39.9	22.5	29.6	0.7	2.5	20.6
	24	0.0	36.0	15.6	32.7	2.0	9.9	19.5
Vitamin B_2	6	2.9	25.9	43.9	21.4	3.3	0.0	19.5
	9	5.8	35.5	31.2	30.2	1.3	0.0	20.9
	12	7.2	32.5	27.2	28.9	3.3	1.7	19.8
	18	7.1	38.6	22.0	29.6	0.7	2.5	20.1
	24	0.0	32.9	16.2	32.7	2.0	10.7	18.9
Vitamin B_6	6	2.9	25.9	43.9	20.8	2.6	0.0	19.3
	9	1.4	36.0	30.6	30.2	1.3	0.0	20.6
	12	4.3	32.9	26.0	28.9	2.6	1.7	19.4
	18	4.3	38.6	22.5	29.6	0.7	2.5	20.0
	24	1.4	36.0	16.2	32.7	2.0	11.6	19.9
Vitamin B_{12}	6	2.9	2.6	12.1	5.7	0.0	0.0	4.2
	9	5.8	7.9	9.8	8.2	0.7	0.0	5.9
	12	7.2	9.2	13.3	8.8	0.0	0.8	7.1
	18	7.1	17.1	16.8	11.3	0.0	3.3	10.5
	24	0.0	19.3	15.0	14.5	0.7	10.7	11.8
Folate	6	2.9	0.0	8.7	15.7	0.0	0.0	4.7
	9	1.4	0.9	8.7	22.0	0.0	0.0	5.9
	12	4.3	1.3	6.9	20.1	0.0	0.0	5.5
	18	2.9	3.1	4.0	19.5	0.0	2.5	5.5
	24	1.4	3.1	5.2	20.1	0.7	9.9	6.9
Niacin	6	2.9	22.8	0.0	6.3	3.3	0.0	7.6
	9	5.8	29.4	0.0	8.2	1.3	0.0	9.5
	12	7.2	28.1	0.0	8.8	3.3	1.7	10.0
	18	7.1	28.5	0.0	11.3	0.7	2.5	10.2
	24	1.4	25.9	0.0	14.5	2.0	10.7	11.0
Iron	6	7.2	1.8	13.9	0.6	0.7	0.0	3.9
	9	17.4	4.4	9.2	0.0	0.7	0.8	4.4
	12	18.8	3.9	13.3	0.0	1.3	4.1	5.8
	18	10.0	6.1	9.8	1.3	1.3	2.5	5.0
	24	8.7	7.5	8.7	0.6	0.7	9.9	5.8

Table IV (*Continued*)

Supplement	Age (mo)	Brazil ($n = 69$)	Ghana ($n = 228$)	India ($n = 173$)	Norway ($n = 159$)	Oman ($n = 153$)	USA ($n = 121$)	All ($n = 903$)
		Sites						
Zinc	6	2.9	0.4	12.1	0.6	0.0	0.0	2.8
	9	1.4	0.9	12.1	0.0	0.0	0.0	2.7
	12	4.3	1.8	11.0	0.0	0.0	0.0	2.9
	18	2.9	4.8	10.4	1.3	0.0	2.5	4.0
	24	0.0	5.3	8.7	1.3	0.7	7.4	4.3
Iodine	6	0.0	0.0	1.2	0.6	0.0	0.0	0.3
	9	0.0	0.0	0.6	0.0	0.0	0.0	0.1
	12	1.4	0.0	0.6	0.0	0.0	0.0	0.2
	18	0.0	0.0	1.7	1.3	0.0	2.5	0.9
	24	0.0	0.0	0.0	1.3	0.0	6.6	1.1
Calcium	6	2.9	0.4	2.9	0.6	0.0	0.0	1.0
	9	1.4	0.9	2.9	0.6	0.0	0.0	1.0
	12	4.3	0.4	2.9	0.6	0.0	0.8	1.2
	18	5.7	0.4	6.9	0.6	0.0	2.5	2.3
	24	0.0	1.8	5.2	1.3	0.7	6.6	2.7

external source of vitamin D. Vitamin supplements were given to up to 40% of children in Ghana and India but were rarely used in Brazil, Oman and the USA. Mineral supplements were not commonly used in any of the sites.

In general, except for practices that were related to the reasons for non-compliance—introduction of solid or semi-solid foods at an earlier age, fewer "milk-only" meals because of a lower frequency of breastfeeding, and greater consumption of milk products other than breast milk—there were few substantive differences in complementary feeding practices between the compliant and non-compliant subjects of the MGRS. This indicates that the compliant group was not an "atypical" subset of the overall MGRS sample with respect to most complementary feeding practices among the relatively economically well-off groups that we studied.

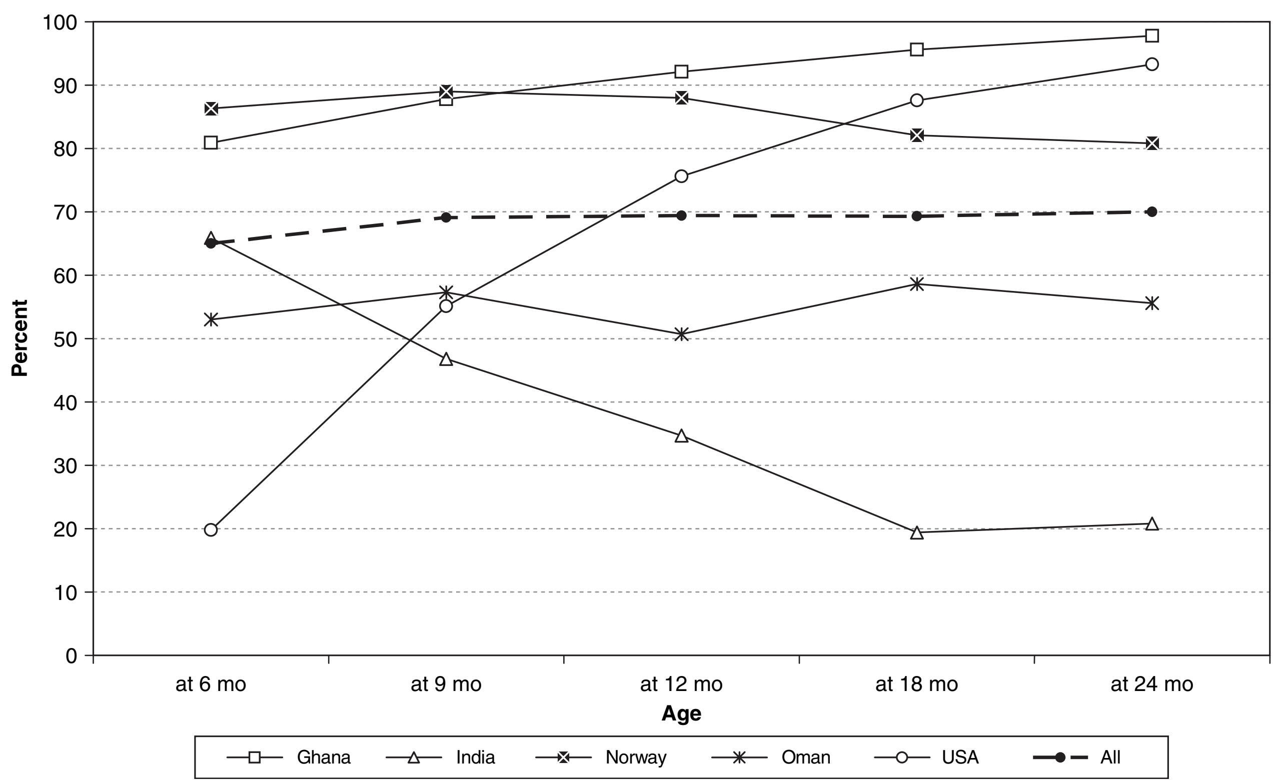


Figure 4. Percentage of compliant children consuming vitamin A-fortified foods at selected ages.

To summarize, these results indicate that the complementary food diets of children in the MGRS were generally of high quality. Global recommendations for complementary feeding stress the need for frequent intake of animal-source foods as well as fruits and vegetables [6]. After the initial period of ~6–9 mo, when new foods were still being introduced, the majority of children consumed animal-source foods and fruits and vegetables on the day of each dietary recall in all of the MGRS sites. Dietary diversity was relatively high and meal frequency was in accord with global guidelines. The majority of children received iron-fortified complementary foods during the first year of life, and many continued to receive them during the second year of life. Thus, the risk of nutritional deficiencies was low. We conclude that the complementary food patterns of MGRS subjects were adequate to support physiological growth.

Acknowledgements

This paper was prepared by Kathryn G. Dewey, Adelheid W. Onyango, Cutberto Garza, Mercedes de Onis, Deena Alasfoor, Elaine Albernaz, Nita Bhandari, Gunn-Elin A. Bjoerneboe and Anna Lartey on behalf of the WHO Multicentre Growth Reference Study Group. The statistical analysis was conducted by Amani Siyam and Alain Pinol.

References

[1] de Onis M, Garza C, Victora CG, Bhan MK, Norum KR, editors. WHO Multicentre Growth Reference Study (MGRS): Rationale, planning and implementation. Food Nutr Bull 2004;25 Suppl 1:S1–89.

[2] WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards based on length/height, weight and age. Acta Paediatr Suppl 2006;450:76–85.

[3] de Onis M, Garza C, Victora CG, Onyango AW, Frongillo EA, Martines J, for the WHO Multicentre Growth Reference Study Group. The WHO Multicentre Growth Reference Study: Planning, study design, and methodology. Food Nutr Bull 2004;25 Suppl 1:S15–26.

[4] WHO/UNICEF. Complementary feeding of young children in developing countries: a review of current scientific knowledge. WHO/NUT/98.1. Geneva: World Health Organization; 1998.

[5] Dewey KG, Brown KH. Update on technical issues concerning complementary feeding of young children in developing countries and implications for intervention programs. Food Nutr Bull 2003;24:5–28.

[6] Pan American Health Organization/World Health Organization. Guiding principles for complementary feeding of the breastfed child. Washington, DC: Pan American Health Organization; 2003.

[7] Araujo CL, Albernaz E, Tomasi E, Victora CG, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in Brazil. Food Nutr Bull 2004;25 Suppl 1:S53–9.

[8] Lartey A, Owusu WB, Sagoe-Moses I, Gomez V, Sagoe-Moses C, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in Ghana. Food Nutr Bull 2004;25 Suppl 1:S60–5.

[9] Bhandari N, Taneja S, Rongsen T, Chetia J, Sharma P, Bahl R, et al., for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in India. Food Nutr Bull 2004;25 Suppl 1:S66–71.

[10] Baerug A, Bjoerneboe GE, Tufte E, Norum KR, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in Norway. Food Nutr Bull 2004;25 Suppl 1:S72–7.

[11] Prakash NS, Mabry RM, Mohamed AJ, Alasfoor D, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in Oman. Food Nutr Bull 2004;25 Suppl 1:S78–83.

[12] Dewey KG, Cohen RJ, Nommsen-Rivers LA, Heinig MJ, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in the United States. Food Nutr Bull 2004;25 Suppl 1:S84–9.

[13] WHO Multicentre Growth Reference Study Group. Breastfeeding in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450:16–26.

[14] Dewey KG, Cohen RJ, Arimond M, Ruel MT. Developing and validating simple indicators of complementary food intake and nutrient density for breastfed children in developing countries. Report to the Food and Nutrition Technical Assistance (FANTA) Project/Academy for Educational Development (AED). Washington DC: FANTA/AED; September 2005.

[15] World Health Assembly. Resolution WHA54.2. Infant and young child nutrition. Geneva: World Health Organization; 2001.

[16] World Health Organization. The optimal duration of exclusive breastfeeding. Report of an expert consultation. Geneva: World Health Organization; 2002.

Acta Pædiatrica, 2006; Suppl 450: 38–46

Taylor & Francis
Taylor & Francis Group

Reliability of anthropometric measurements in the WHO Multicentre Growth Reference Study

WHO MULTICENTRE GROWTH REFERENCE STUDY GROUP[1,2]

[1]*Department of Nutrition, World Health Organization, Geneva, Switzerland, and* [2]*Members of the WHO Multicentre Growth Reference Study Group (listed at the end of the first paper in this supplement)*

Abstract

Aim: To describe how reliability assessment data in the WHO Multicentre Growth Reference Study (MGRS) were collected and analysed, and to present the results thereof. *Methods:* There were two sources of anthropometric data (length, head and arm circumferences, triceps and subscapular skinfolds, and height) for these analyses. Data for constructing the WHO Child Growth Standards, collected in duplicate by observer pairs, were used to calculate inter-observer technical error of measurement (TEM) and the coefficient of reliability. The second source was the anthropometry standardization sessions conducted throughout the data collection period with the aim of identifying and correcting measurement problems. An anthropometry expert visited each site annually to participate in standardization sessions and provide remedial training as required. Inter- and intra-observer TEM, and average bias relative to the expert, were calculated for the standardization data. *Results:* TEM estimates for teams compared well with the anthropometry expert. Overall, average bias was within acceptable limits of deviation from the expert, with head circumference having both lowest bias and lowest TEM. Teams tended to underestimate length, height and arm circumference, and to overestimate skinfold measurements. This was likely due to difficulties associated with keeping children fully stretched out and still for length/height measurements and in manipulating soft tissues for the other measurements. Intra- and inter-observer TEMs were comparable, and newborns, infants and older children were measured with equal reliability. The coefficient of reliability was above 95% for all measurements except skinfolds whose R coefficient was 75–93%.

Conclusion: Reliability of the MGRS teams compared well with the study's anthropometry expert and published reliability statistics.

Key Words: *Anthropometry, bias, measurement error, measurement reliability, precision*

Introduction

Measurement reliability is a direct indicator of data quality. Reducing errors in measurement will increase the probability that any relationships among variables in a study are uncovered. Adherence to recommended procedures will reduce bias in measurement and increase the certainty of inferences about similarities/differences with respect to other populations. For these and other reasons, it is generally cost effective to reduce measurement error to recommended minima. Standardized data collection methodology, rigorous training and monitoring of data collection personnel, frequent and effective equipment calibration and maintenance, and periodic assessment of anthropometric measurement reliability were among the quality assurance measures included in the World Health Organization's (WHO) Multicentre Growth Reference Study (MGRS) of infants and children [1]. Anthropometry standardization sessions were conducted with the goal of monitoring anthropometric measurement techniques, identifying sources of error or bias and retraining teams or individuals as necessary.

Only a few growth studies and surveys [2–11] provide detailed descriptions of anthropometric standardization and measurement reliability assessments. The standardization of measurement techniques in anthropometry by Lohman and colleagues in the late 1980s has been a useful guide and reference for the collection of reliable anthropometric measurements [12]. However, there is a lack of uniformity in the methods employed in collecting reliability data and in reporting the statistics and terminology used in reliability assessment [6,7,9,11,13–16].

The objectives of this article are to describe the approach used in the MGRS to collect and analyse

Correspondence: Mercedes de Onis, Study Coordinator, Department of Nutrition, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland. Tel: +41 22 791 3320. Fax: +41 22 791 4156. E-mail: deonism@who.int

ISSN 0803-5326 print/ISSN 1651-2227 online © 2006 Taylor & Francis
DOI: 10.1080/08035320500494464

reliability information, to present key results about measurement reliability, and to assess the implications of these results for the MGRS. The analyses are based on data collected during anthropometric standardization sessions held regularly over the duration of data collection in each of the six MGRS sites and on duplicate measurements taken during routine data collection.

Methods

Data collection teams and procedures

Data in the MGRS were collected between 1997 and 2003 in Brazil, Ghana, India, Norway, Oman and the USA [17]. Data collection teams were trained in each site during the study's preparatory phase, at which time measurement techniques were standardized against one of two MGRS lead anthropometrists. During the study, one of these experts visited each site annually to participate in standardization sessions [1]. For the longitudinal component of the study, screening teams measured newborns within 24 h of delivery, and follow-up teams conducted home visits until the children reached 24 mo of age. The follow-up teams also carried out measurements in the cross-sectional component of the MGRS involving children aged 18–71 mo.

The anthropometric variables measured were weight and head circumference at all ages, recumbent length in the longitudinal study, height in the cross-sectional study, and arm circumference, triceps and subscapular skinfolds in all children aged $\geq$3 mo. The methodology and equipment used in taking these measurements have been described in detail elsewhere [1]. Briefly, anthropometric data were collected by observers working in pairs. Each observer independently measured and recorded a complete set of measurements, and the two then compared their readings. If any pair of readings exceeded the maximum allowable difference for a given variable (weight 100 g; circumferences 5 mm; length/height 7 mm; skinfolds 2 mm), the observers again independently measured and recorded a second and, if necessary, a third set of readings for the affected variable(s). The availability of duplicate measurements by two observers allows for the estimation of inter-observer reliability statistics under routine data collection conditions. Since weight was measured with near-perfect precision on digital scales, it was not included in the standardization sessions.

During the standardization sessions, screening teams measured newborns while follow-up teams measured older infants. The children involved in the standardization sessions were not part of the MGRS cohort. During these sessions, the observers measured independently but did not compare values with other observers, as was done during routine data collection. No inter-site statistical comparisons are presented because no common set of children was measured by observers from different sites. At each site, the screening teams' standardization sessions stopped when the enrolment of newborns ended (duration 12–14 mo), while the follow-up team sessions continued for the entire 3–3½ y of MGRS data collection. Because the USA site did not have access to newborns for the screening team's standardization exercises, the team measured older infants.

Data management

The MGRS data management protocol, which has been described in detail elsewhere [18], highlights the specific measures applied in detecting errors and cleaning the MGRS anthropometry data. For the standardization sessions, study supervisors in each site were responsible for checking the data collected for any recording errors prior to on-site analysis of measurement error. The data were then sent to the study coordinating centre in Geneva, Switzerland, for further quality control checks and monitoring of the performance of observers and site teams. These data were merged within site to create the standardization master files used in the present analyses. Recorded values that varied by more than 4 standard deviations from a given child's mean (estimated from all values recorded by the observers in the session) were considered errors of transcription or the result of causes unrelated to measurement reliability and were reset as missing [8]. For the purpose of this report, data were analysed only from observers who participated in two or more standardization sessions.

Statistical analysis

Reliability statistics reported for the standardization sessions were intra-observer technical error of measurement (TEM), inter-observer TEM and average bias. Inter-observer TEM achieved in routine data collection was also estimated and used to calculate the coefficient of reliability, R, for six anthropometric variables (excluding weight) measured in the MGRS. The key statistics are defined as follows.

Technical error of measurement (TEM) is a measure of error variability that carries the same measurement units as the variable measured, e.g. centimetres of head circumference. Its interpretation is that differences between replicate measurements will be within $\pm$the value of TEM two-thirds of the time [14]. Similarly, 95% of the differences between replicate measurements are expected to be within $\pm 2 \times$TEM [9], which is referred to as the 95% precision margin elsewhere in this paper. Intra-observer TEM is estimated from differences between replicate measurements taken by one observer, while inter-observer TEM is estimated from single measurements taken by two or more observers. The formulae (1)–(4) for these statistics are given below.

Intra-observer TEM for one observer is calculated by:

$$\sqrt{\frac{\sum_{i=1}^{N}(M_{i1}-M_{i2})^2}{2^*N}}, \tag{1}$$

where M_{i1} and M_{i2} are the duplicate measurements recorded by a given observer for the ith child, and N is the number of children measured. It can be generalized to k observers as in (2):

$$\sqrt{\frac{\sum_{j=1}^{K}\sum_{i=1}^{N_j}(M_{ij1}-M_{ij2})^2}{2^*\sum_{j=1}^{K}N_j}}, \tag{2}$$

where M_{ij1} and M_{ij2} are the duplicate readings recorded by observer j for the i^{th} child, N_j is the number of children measured by observer j, and K is the number of observers taking the measurements.

The inter-observer TEM in standardization data is calculated by:

$$\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{(K_i-1)}\left[\sum_{j=1}^{K_i}Y_{ij}^2-\frac{\left(\sum_{j=1}^{K_i}Y_{ij}\right)^2}{K_i}\right]\right\}^{1/2}, \tag{3}$$

where Y_{ij} is one of the duplicate measurements taken by observer j for child i (for simplicity in programming the present analyses, the first recorded measurement was selected), K_i is the number of observers that measured child i (this takes care of missing values), and N is the number of children involved. In the routine MGRS data (calculated separately for screening, longitudinal follow-up and cross-sectional survey data), only two observers took measurements, so formula (3) simplifies to:

$$\left\{\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{j=1}^{2}Y_{ij}^2-\left(\sum_{j=1}^{2}Y_{ij}\right)^2/2\right]\right\}^{1/2}, \tag{4}$$

where N is the total number of children measured in respective master files for each anthropometric variable.

Average bias is estimated as the average difference between measurements taken by an expert and those taken by an observer or observers of the same subjects. A negative-signed average bias estimate indicates that the test group underestimates the measurement, while the opposite indicates overestimation. It is calculated by:

$$\frac{\sum_{i=1}^{N_G}\left[\sum_{j=1}^{K}(M_{ij1}+M_{ij2})/(2^*K)-(M_{iG1}+M_{iG2})/2\right]}{N_G}, \tag{5}$$

where M_{ij1}, M_{ij2} and M_{iG1}, M_{iG2} are the duplicate readings recorded by observer j and the expert, respectively, for the i^{th} child, N_G is the set of children measured by the expert, and K is the number of observers measuring the same children.

Coefficient of reliability, R, estimates the proportion of the inter-subject variance (total measurement variance) that is not due to measurement error. A reliability coefficient $R=0.8$ means that 80% of the total variability is true variation, while the remaining proportion (20%) is attributable to measurement error, described by Marks and colleagues [8] as imprecision and unreliability. For the MGRS data, R was calculated using the formula:

$$R=1-\frac{(TEM(Inter))^2}{SD^2}, \tag{6}$$

where *TEM(Inter)* refers to the MGRS data TEM as calculated in formula (4), and SD values for each anthropometric variable are taken from the MGRS population at specified ages. For newborns: head circumference 1.27 cm and length 1.91 cm; and for older children: head circumference 1.40 cm, length 2.60 cm, arm circumference 1.30 cm, triceps skinfold 1.80 mm, subscapular skinfold 1.40 mm (12 mo), and height 4.07 cm (48 mo).

In the MGRS, intra-observer TEM could be calculated for the standardization but not the routine study data, while inter-observer TEM was calculated for both data sets. Intra-observer TEM for each team was calculated using data from all the standardization sessions conducted in a given site. The MGRS anthropometry experts' measurements from all sites were combined to calculate the "gold standard" intra-observer TEM. The assessment of bias was restricted to the data collected during the standardization sessions in which an international lead anthropometrist participated.

Several approaches were used to judge the adequacy of measurement in the MGRS, consistent with guidelines suggested in the literature:

a. TEM values for observers were considered adequate if they were within ± 2 times the expert's TEM, i.e. the expert's 95% precision margin [19].
b. We assessed average bias in terms of magnitude and whether or not site teams systematically over- or underestimated measurements. To be consistent with the criterion used to set the maximum allowable differences between paired observer measurements in the MGRS, bias was

considered to be large if it exceeded the expert's intra-observer TEM $\times 2.8$ [1]. This is equivalent to the limits that were considered to indicate significant deviations from likely "true" values while accommodating the unavoidable imprecision of anthropometric measurements.

c. Our main criterion for judging adequacy of measurement was the coefficient of reliability, R, because it considers the measurement variance in relation to variability in the measurement. As is the case for other related measures of agreement, e.g. kappa, values of 0.8 and greater may be taken to represent "excellent" agreement and those between 0.61 and 0.8 "substantial" agreement [20].
d. Finally, we compared the TEMs obtained by the MGRS observers to those reported in the literature.

Results

The number of standardization sessions at each site ranged from five to nine for the screening teams and 14 to 21 for the follow-up teams (Table I). There was also inter-site variation in the number of observers, which was a function of staff turnover (Ghana had the highest turnover and Oman the lowest). The MGRS anthropometry experts participated in 17 of the standardization sessions.

Screening team

Intra-observer TEMs ranged among sites from 0.16 to 0.28 cm for newborn head circumference and from 0.22 to 0.48 cm for length measurements (Table II). In all cases, observer TEMs were within twice the gold standard TEM, that is, within the 95% precision margin. While there was no evidence of bias in the teams' head circumference measurements compared with the expert's, all four sites for which bias was calculated tended to underestimate length, by −0.21 to −0.37 cm.

Inter-observer TEMs for both the standardization and the routine data collected by the screening teams are presented in Table III. TEMs were very similar for the two data sources. Reliability coefficients, estimated for routine data collection, were greater than 95% in all cases. Inter-observer TEMs were not substantially larger than intra-observer TEMs (Table III versus Table II).

Follow-up team

In almost all cases, the follow-up teams' intra-observer TEMs were less than twice the gold standard TEM (Table IV). Only the Norwegian and Omani teams' TEMs exceeded the expert's 95% precision margin (0.24 cm) for head circumference. All bias estimates but one (Brazil, subscapular skinfold) were within the allowable limits of 2.8 times the gold standard TEM for each measurement. However, the sign of the teams' bias estimates showed that they tended to underestimate arm circumference, length and height, and to overestimate skinfold measurements. Estimates of bias in head circumference had a fair balance of positive and negative signs, and were of the lowest overall magnitude.

The three sets of data (standardization, longitudinal and cross-sectional) represented in Table V had similar inter-observer TEMs within each variable and site. The largest disparity in this regard was for triceps skinfold in India with 0.49 mm for the standardization and 0.71 mm for the longitudinal data. The coefficient of reliability was above 0.95 for all variables except the skinfolds for which R ranged from 0.75 to 0.93. A comparison of inter- and intra-observer TEM based on the standardization data revealed very few substantial differences. The expected pattern (inter-observer TEM larger than intra-observer TEM) was systematic for two measurements (the skinfolds) in all sites, and for all measurements in two sites (Brazil and the USA).

The reliability of both newborn and older-child measurements for the MGRS teams was as good as,

Table I. Standardization sessions and observer participation by team and site.

	Newborn screening team			Follow-up team		
Sites	Sessions[a]	Observers	Expert[b]	Sessions[a]	Observers	Expert[b]
Brazil	6	6	0	20	9	1
Ghana	8	9	2	21	15	4
India	9	9	2	19	10	3
Norway	5	5	1	14	9	3
Oman	9	6	3	19	11	4
USA	0[c]	–	–	17	9	2

[a] The screening team sessions are fewer than the follow-up team sessions because newborn screening for the longitudinal study lasted 12–14 mos while the follow-up team worked through the entire 3–3½ y of data collection.
[b] These are the sessions in which one of the MGRS international lead anthropometrists participated.
[c] The USA did not have access to newborns for the standardization sessions, so the screening team measured older infants.

Table II. Screening team intra-observer technical error of measurement (TEM)[a] and bias[b] relative to MGRS anthropometry expert in the standardization sessions.

		Expert	Site [c]				
			Brazil (n = 20, 60)[d]	Ghana (n = 95)	India (n = 99)	Norway (n = 60)	Oman (n = 102)
Head circumference (cm)	TEM	0.16	0.24	0.25	0.16	0.28	0.27
	Average bias		–	0.00	−0.09	0.08	0.03
Length (cm)	TEM	0.29	0.22	0.29	0.33	0.48	0.37
	Average bias		–	−0.29	−0.21	−0.37	−0.26

[a] The expert's TEM is based on the sum of measurements taken in all sites by the MGRS lead anthropometrists participating in standardization sessions. Site teams' intra-observer TEM is calculated using data from all standardization sessions (initial and bimonthly) conducted in respective sites, average of all observers taking part in ≥2 bimonthly sessions.
[b] Average bias relative to the expert is calculated from the subset of measurements taken in the standardization sessions in which the MGRS lead anthropometrist participated, and thus includes only subjects measured by both the expert and each site's team (n per site: Ghana 31; India 30; Norway 20; Oman 42; Brazil did not hold a separate session for the newborn screening team at the initial standardization where the lead anthropometrist participated).
[c] The USA was excluded from this analysis because the screening team did not measure newborns in the standardization sessions.
[d] Sample size: n = 20 infants for head circumference and n = 60 for length. The earliest enrolled newborns in Brazil had their first head circumference measurement taken at 7 d. The MGRS protocol was amended, and only then did the screening team begin to take head circumference measurements at birth.

or better than, intra-observer TEM estimates reported in other published studies involving children (Table VI).

Discussion

The measurement and standardization protocols of the MGRS provided a mechanism for continuous monitoring of measurement reliability. This helped to identify and resolve problems by retraining individual observers (during or immediately after each standardization session) or site teams, as happened on specific occasions in Norway and the USA. The sources of error in the MGRS were identified with the express intention of correcting them, going beyond what has been implemented in other studies that documented measurement reliability [5,9,11]. A further unique feature of the MGRS is the documentation of measurement reliability in the very data that have been used to construct the WHO Child Growth Standards [21].

The standardization sessions and routine data collection settings are difficult to compare. In the former, workers had to collect duplicate measurements on 10 to 20 children in one session and were not allowed to compare and take new measurements when differences were large. In routine data collection, fieldworkers were dealing with just one child at a time and were allowed to compare their values and re-measure if disparities exceeded preset limits. Despite these differences, measurement error was similar in both settings.

A comparison of reliability statistics between the screening and follow-up teams, and between the longitudinal and cross-sectional samples, shows that newborn and older infants were measured as reliably as were older children. Judging by the site teams' intra-observer TEM relative to the expert's 95%

Table III. Inter-observer technical error of measurement (TEM) for the newborn screening teams in the standardization sessions and routine MGRS data collection.

	Data source and R coefficient [a]	Site						
		Brazil[b] (n)	Ghana (n)	India (n)	Norway (n)	Oman (n)	USA (n)	All (n)
Head circumference (cm)	Standardization	0.42 (20)	0.27 (95)	0.20 (99)	0.25 (60)	0.26 (102)	–	–
	MGRS data	–	0.25 (329)	0.18 (301)	0.24 (300)	0.25 (295)	0.27 (208)	0.22 (1433)
	R coefficient	–	0.96	0.98	0.96	0.96	0.95	0.97
Length (cm)	Standardization	0.32 (60)	0.35 (95)	0.42 (99)	0.48 (60)	0.40 (102)	–	–
	MGRS data	–	0.30 (329)	0.35 (301)	0.39 (300)	0.39 (295)	0.40 (208)	0.34 (1433)
	R coefficient	–	0.98	0.97	0.96	0.96	0.96	0.97

[a] Inter-observer TEM was calculated separately for the standardization and routine screening data of the MGRS longitudinal component. The R coefficient is calculated for the latter data set only.
[b] Inter-observer TEM and R were not calculated for the Brazilian newborn screening data because the site began to duplicate measurements halfway into recruitment. The early data were thus inappropriate for this analysis.

Table IV. Follow-up team intra-observer technical error of measurement (TEM)[a] and bias[b] relative to MGRS anthropometry expert in the standardization sessions.

			Site[c]					
		Expert	Brazil (n =210, 0)	Ghana (n =234, 138)	India (n =200, 160)	Norway (n =162, 80)	Oman (n =200, 90)	USA (n =179, 69)
Head circumference (cm)	TEM	0.12	0.13	0.23	0.19	0.25	0.29	0.19
	Average bias		0.01	−0.01	−0.16	0.04	-0.18	-0.14
Length (cm)	TEM	0.33	0.23	0.37	0.33	0.58	0.43	0.21
	Average bias		0.01	−0.18	−0.15	−0.35	−0.24	−0.70
Arm circumference (cm)	TEM	0.17	0.17	0.20	0.20	0.26	0.27	0.15
	Average bias		−0.10	−0.30	−0.24	−0.31	−0.26	−0.37
Triceps skinfold (mm)	TEM	0.40	0.42	0.39	0.46	0.61	0.49	0.45
	Average bias		−0.81	0.21	0.45	0.11	0.25	0.11
Subscapular skinfold (mm)	TEM	0.30	0.38	0.31	0.32	0.29	0.35	0.41
	Average bias		−1.05	0.28	0.28	0.11	0.03	0.79
Height (cm)	TEM	0.23	–	0.26	0.27	0.29	0.27	0.16
	Average bias		–	−0.30	−0.21	−0.20	−0.22	−0.06

[a] The expert's TEM is based on the sum of measurements taken in all sites by the MGRS lead anthropometrists participating in standardization sessions. Site teams' intra-observer TEM is calculated using data from all standardization sessions (initial and bimonthly) conducted in respective sites, average of all observers taking part in ≥2 bimonthly sessions.
[b] Average bias relative to the expert is calculated from the subset of measurements taken in the standardization sessions in which the MGRS lead anthropometrist participated, and thus includes only subjects measured by both the expert and each site's team (n per site (n height): Brazil 19 (0); Ghana 60 (40); India 40 (30); Norway 41 (30); Oman 50 (30); USA 19 (9)).
[c] The second sample size figure is the number of subjects involved in height standardization. Sites normally began to take this measurement at the inception of the cross-sectional study.

Table V. Inter-observer technical error of measurement (TEM) for the follow-up teams in the standardization sessions and the routine MGRS data.

	Data source and R coefficient[a]	Site						
		Brazil (*n*)	Ghana (*n*)	India (*n*)	Norway (*n*)	Oman (*n*)	USA (*n*)	All (*n*)
Head circumference (cm)	Standardization	0.25	0.24	0.18	0.23	0.29	0.28	–
	Longitudinal	0.23 (5849)	0.23 (6069)	0.23 (5633)	0.25 (5460)	0.26 (5425)	0.23 (3834)	0.24 (32270)
	Cross-sectional	0.25 (1342)	0.23 (1406)	0.21 (1455)	0.24 (1376)	0.29 (1445)	0.28 (1339)	0.25 (8363)
	R coefficient	0.97/0.97	0.97/0.97	0.97/0.98	0.97/0.97	0.97/0.96	0.97/0.96	0.97/0.97
Length (cm)	Standardization	0.40	0.44	0.32	0.48	0.42	0.37	–
	Longitudinal	0.33 (5836)	0.41 (6067)	0.36 (5630)	0.37 (5470)	0.38 (5420)	0.41 (3827)	0.38 (32250)
	Cross-sectional	0.23 (250)	0.34 (286)	0.27 (327)	0.27 (371)	0.35 (356)	0.29 (164)	0.30 (1754)
	R coefficient	0.98/0.99	0.98/0.98	0.98/0.99	0.98/0.99	0.98/0.98	0.98/0.99	0.98/0.99
Arm circumference (cm)	Standardization	0.28	0.27	0.19	0.29	0.26	0.26	–
	Longitudinal	0.26 (4545)	0.25 (4791)	0.26 (4461)	0.29 (4267)	0.23 (4550)	0.26 (3002)	0.26 (25616)
	Cross-sectional	0.22 (1333)	0.20 (1406)	0.18 (1448)	0.22 (1354)	0.21 (1444)	0.23 (1339)	0.21 (8324)
	R coefficient	0.96/0.97	0.96/0.98	0.96/0.98	0.95/0.97	0.97/0.97	0.96/0.97	0.96/0.97
Triceps skinfold (mm)	Standardization	0.67	0.51	0.49	0.83	0.60	0.87	–
	Longitudinal	0.66 (4638)	0.50 (4791)	0.71 (4464)	0.75 (4259)	0.63 (4551)	0.83 (3002)	0.67 (25705)
	Cross-sectional	0.85 (1325)	0.46 (1406)	0.67 (1440)	0.76 (1328)	0.59 (1444)	0.84 (1335)	0.70 (8278)
	R coefficient	0.87/0.78	0.92/0.93	0.85/0.86	0.83/0.82	0.88/0.89	0.79/0.78	0.86/0.85
Subscapular skinfold (mm)	Standardization	0.48	0.42	0.36	0.42	0.41	0.67	–
	Longitudinal	0.47 (4639)	0.42 (4791)	0.45 (4466)	0.43 (4273)	0.46 (4551)	0.69 (3003)	0.48 (25723)
	Cross-sectional	0.59 (1324)	0.44 (1406)	0.44 (1434)	0.39 (1339)	0.49 (1444)	0.62 (1335)	0.50 (8282)
	R coefficient	0.89/0.82	0.91/0.90	0.89/0.90	0.91/0.92	0.89/0.88	0.75/0.80	0.88/0.87
Height (cm)	Standardization	–	0.27	0.23	0.34	0.35	0.33	–
	Cross-sectional	0.15 (1328)	0.39 (1404)	0.23 (1449)	0.34 (1358)	0.26 (1443)	0.32 (1348)	0.29 (8330)
	R coefficient	1.00	0.99	1.00	0.99	1.00	0.99	0.99

[a]"Longitudinal" are the data measured by the follow-up team in the MGRS longitudinal component, and "cross-sectional" are data from the MGRS cross-sectional component. The reliability coefficient R was based on the routine MGRS (not standardization) data: the first figure belongs to the longitudinal measurements and the second to the cross-sectional measurements, and the single figure for height refers to the cross-sectional component.

precision margins, the teams' precision compared favourably with the expert's for all measurements. There was no consistent pattern in the relationship between intra- and inter-observer variability.

Although the magnitude of bias in the teams' measurements was overall within allowable limits compared with the expert, distinct negative and positive tendencies were noticeable for all measurements except head circumference. The "problem" measurements were those that involve manipulation of soft tissues (arm circumference and skinfolds) and those that require careful positioning to ensure that the child is fully stretched out for the measurement (length and height). It is worth noting that the same pattern was observed in the Rotterdam standardization session [1] where, compared with the expert, the session's participants all had negative-signed bias for length, height and arm circumference, and positive-signed bias for the skinfold measurements. In general, the standardization sessions were stressful as the observers had to repeat measurements on often crying and struggling children. Under those conditions, the expert could, with greater self-assurance than the fieldworkers, position the child to full length/height, pause to let the callipers close in on skinfolds before taking the reading, and retain better control of the circumference tape around the child's arm to avoid compressing the soft tissues. The average bias estimate for subscapular skinfold in Brazil was larger than the limits set by the expert's TEM $\times 2.8$ and also in the opposite direction from the other sites. The data used to calculate this estimate were collected at the site's initial standardization, and the team thereafter received remedial training in the measurement of skinfolds.

Considering our main criterion for assessing measurement reliability in the MGRS data, overall R coefficients were higher than the 90% reliability threshold that Marks and colleagues [8] suggest as adequate for the presentation of growth standards. However, Ulijaszek and Lourie [22], while endorsing that cut-off, recognized the characteristic low reliability of skinfold measurements in young children. Indeed, the MGRS skinfold measurements had R coefficients below 90% but mostly above the threshold of 80% applied to other measures of agreement such as the kappa coefficient cut-off for "excellent" agreement [20]. As others have noted, larger inter-observer reliability is expected in measurements that have characteristically low precision [8]. This is illustrated by the lower intra- than inter-observer TEM for the two skinfold measurements in the MGRS. One suggested approach to improving precision for such measurements is to measure twice and report the average of the two values [5,8]. This is what we did in the MGRS, for all the anthropometric measurements used to construct the WHO Child Growth Standards, with the added assurance that the

Table VI. Comparison of intra-observer TEM between the MGRS and other estimates in the literature (child populations).

Age group and variables	MGRS teams	Published estimates	Source (number in ref. list)
Newborn			
Length (cm)	0.22–0.48	0.79, 1.22	Johnson et al., 1997 [23]
Head circumference (cm)	0.16–0.28	0.28, 0.30	Johnson et al., 1997 [23]
Older children			
Length (cm)	0.23–0.58	0.4, 0.8	Ulijaszek and Lourie, 1994, literature review [22]
Height (cm)	0.16–0.29	0.34	Martorell et al., 1975 [6]
		0.49	Malina et al., 1973, NHES III [5]
Head circumference (cm)	0.13–0.29	0.14	Martorell et al., 1975 [6]
MUAC (cm)	0.15–0.27	0.35	Malina et al., 1973, NHES III [5]
		0.18	Martorell et al., 1975 [6]
Triceps skinfold (mm)	0.39–0.61	0.47	Martorell et al., 1975 [6]
		0.80	Johnston et al., 1972, NHES III [24]
Subscapular skinfold (mm)	0.29–0.41	0.27	Martorell et al., 1975 [6]
		1.83	Johnston et al., 1972, NHES III [24]

NHES III: cycle III of the National Health Examination Survey (USA); MUAC: mid-upper arm circumference.

two measurements were within preset margins of difference [1].

Several published studies and reviews of the anthropometry literature provided intra-observer TEM estimates, and these were compared with the MGRS teams' performance [5,6,22–24]. The MGRS teams' reliability was generally better than the published ranges. However, these comparisons should be viewed with the understanding that the numbers of observers and subjects involved, and the measurement protocols and equipment employed, vary widely among studies. For example, the number of subjects measured in the MGRS standardization sessions is larger than has been reported in most other published studies.

The MGRS presents a number of innovations with regard to reliability assessment in anthropometry. These include the use of standardized measurement protocols and equipment at six country sites, the evaluation of the different site teams' reliability using a common gold standard, and the estimation of measurement reliability in the data that have been used to construct growth standards. Ulijaszek and Kerr [15] proposed using "criterion anthropometrist(s)" for the purpose of overseeing and assuring the standard application of measurement procedures, and to set targets for the level of accuracy that fieldworkers in anthropometry could aim to achieve. The use in the MGRS of the international lead anthropometrist's intra-observer TEM to set cut-offs for precision (the expert's 95% precision margin) and the limits of acceptable bias (2.8 times the expert's TEM) is a significant step in this direction, and one that could be applied in other studies to standardize reliability assessment when a gold standard is available. In the absence of a designated individual to serve as gold standard, the average intra-observer TEM of a well-trained group could be used to set both precision and accuracy targets.

Acknowledgements

This paper was prepared by Adelheid W. Onyango, Reynaldo Martorell, Wm Cameron Chumlea, Jan Van den Broeck, Cora L. Araújo, Anne Baerug, William B. Owusu and Roberta J. Cohen on behalf of the WHO Multicentre Growth Reference Study Group. The statistical analysis was conducted by Alain Pinol and Elaine Borghi.

References

[1] de Onis M, Onyango AW, Van den Broeck J, Cameron WC, Martorell R, for the WHO Multicentre Growth Reference Study Group. Measurement and standardization protocols for anthropometry used in the construction of a new international growth reference. Food Nutr Bull 2004;25 Suppl 1:S27–36.

[2] Garn S, Shamir Z. Methods for research in human growth. Springfield: Charles C. Thomas; 1958.

[3] Damon A, Stoudt H, McFarland R. The human body in equipment design. Cambridge: Harvard University Press; 1966.

[4] McGammon R. Human growth and development. Springfield: Charles C. Thomas; 1970.

[5] Malina RM, Hamill PV, Lemeshow S. Selected body measurements of children 6–11 years: United States.Vital Health Stat Series 11 No. 123, 1973:38–48.

[6] Martorell R, Habicht JP, Yarbrough C, Guzman G, Klein RE. The identification and evaluation of measurement variability in the anthropometry of preschool children. Am J Phys Anthropol 1975;43:347–52.

[7] Foster TA, Berenson GS. Measurement error and reliability in four pediatric cross-sectional surveys of cardiovascular disease risk factor variables—the Bogalusa Heart Study. J Chronic Dis 1987;40:13–21.

[8] Marks GC, Habicht JP, Mueller WH. Reliability, dependability, and precision of anthropometric measurements. The Second National Health and Nutrition Examination Survey 1976–1980. Am J Epidemiol 1989;130:578–87.

[9] Chumlea WC, Guo S, Kuczmarski RJ, Johnson CL, Leahy CK. Reliability of anthropometric measurements in the Hispanic Health and Nutrition Examination Survey (HHANES 1982–1984). Am J Clin Nutr 1990;51:902S–7S.

[10] Roche AF. Growth, maturation and body composition: The Fels longitudinal study 1929–1991. New York: Cambridge University Press; 1992.
[11] Moreno LA, Joyanes M, Mesana MI, González-Gross M, Gil CM, Sarría A, et al. Harmonization of anthropometric measurements for a multicenter nutrition survey in Spanish adolescents. Nutrition 2003;19:481–6.
[12] Lohman TG, Roche AF, Martorell R, editors. Anthropometric standardization reference manual. Champaign: Human Kinetics Books; 1988.
[13] Habicht JP, Yarbrough C, Martorell R. Anthropometric field methods: criteria for selection. In: Jelliffe DB, Jelliffe EFP, editors. Nutrition and growth. New York: Plenum Press; 1979. p. 365–87.
[14] Mueller WH, Martorell R. Reliability and accuracy of measurement. In: Lohman TG, Roche AF, Martorell R, editors. Anthropometric standardization reference manual. Champaign: Human Kinetics Books; 1988. p. 83–6.
[15] Ulijaszek SJ, Kerr DA. Anthropometric measurement error and the assessment of nutritional status. Br J Nutr 1999;82: 165–77.
[16] Johnson TS, Engstrom JE. State of the science in measurement of infant size at birth. Newborn Infant Nurs Rev 2002;2: 150–8.
[17] de Onis M, Garza C, Victora CG, Onyango AW, Frongillo EA, Martines J, for the WHO Multicentre Growth Reference Study Group. The WHO Multicentre Growth Reference Study: Planning, study design and methodology. Food Nutr Bull 2004;25 Suppl 1:S15–26.
[18] Onyango AW, Pinol AJ, de Onis M, for the WHO Multicentre Growth Reference Study Group. Managing data for a multi-country longitudinal study: Experience from the WHO Multicentre Growth Reference Study. Food Nutr Bull 2004;25 Suppl 1:S46–52.
[19] WHO. Measuring change in Nutritional Status. Geneva: World Health Organization; 1983.
[20] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
[21] WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards based on length/height, weight and age. Acta Paediatr Suppl 2006;450:76–85.
[22] Ulijaszek SJ, Lourie JA. Intra- and inter-observer error in anthropometric measurement. In: Ulijaszek SJ, Mascie-Taylor CGN, editors. Anthropometry: the individual and the population. Cambridge: Cambridge University Press; 1994. p. 30–55.
[23] Johnson TS, Engstrom JL, Gelhar DK. Intra- and interexaminer reliability of anthropometric measurements of term infants. J Pediatr Gastroenterol Nutr 1997;24:497–505.
[24] Johnston FE, Hamill PVV, Lemeshow S. Skinfold thickness of children 6–11 years: United States. Vital Health Stat Series 11 No. 120, 1972:50–60.

Acta Pædiatrica, 2006; Suppl 450: 47–55

Taylor & Francis
Taylor & Francis Group

Reliability of motor development data in the WHO Multicentre Growth Reference Study

WHO MULTICENTRE GROWTH REFERENCE STUDY GROUP[1,2]

[1]*Department of Nutrition, World Health Organization, Geneva, Switzerland, and,* [2]*Members of the WHO Multicentre Growth Reference Study Group (listed at the end of the first paper in this supplement)*

Abstract

Aim: To describe the methods used to standardize the assessment of motor milestones in the WHO Multicentre Growth Reference Study (MGRS) and to present estimates of the reliability of the assessments. *Methods:* As part of the MGRS, longitudinal data were collected on the acquisition of six motor milestones by children aged 4 to 24 mo in Ghana, India, Norway, Oman and the USA. To ensure standardized data collection, the sites conducted regular standardization sessions during which fieldworkers took turns to examine and score about 10 children for the six milestones. Assessments of the children were videotaped, and later the other fieldworkers in the same site watched the videotaped sessions and independently rated performances. The assessments were also viewed and rated by the study coordinator. The coordinator's ratings were considered the reference (true) scores. In addition, one cross-site standardization exercise took place using videotapes of 288 motor assessments. The degree of concordance between fieldworkers and the coordinator was analysed using the Kappa coefficient and the percentage of agreement. *Results:* Overall, high percentages of agreement (81–100%) between fieldworkers and the coordinator and "substantial" (0.61–0.80) to "almost perfect" (>0.80) Kappa coefficients were obtained for all fieldworkers, milestones and sites. Homogeneity tests confirm that the Kappas are homogeneous across sites, across milestones, and across fieldworkers. Concordance was slightly higher in the cross-site session than in the site standardization sessions. There were no systematic differences in assessing children by direct examination or through videotapes.

Conclusion: These results show that the criteria used to define performance of the milestones were similar and applied with equally high levels of reliability among fieldworkers within a site, among milestones within a site, and among sites across milestones.

Key Words: *Agreement, children, inter-rater reliability, motor development, motor skills*

Introduction

The World Health Organization (WHO), in collaboration with partner institutions worldwide, conducted the WHO Multicentre Growth Reference Study (MGRS) to generate new growth curves for assessing the growth and development of infants and young children [1]. As part of the longitudinal component of the MGRS, the Motor Development Study (MDS) was carried out to assess the acquisition of six distinct key motor milestones by affluent children growing up in different cultures. The assessments were done from 4 mo of age until the children were able to walk independently, or reached 24 mo, in Ghana, India, Norway, Oman and the USA. The details of the MDS's study design and methodology have been described elsewhere [2]. To our knowledge, only two other multi-country studies of motor development have used a longitudinal design [3,4].

Rigorous data collection procedures and quality-control measures were applied in all sites to minimize measurement error when assessing motor milestone achievement and to avoid bias among sites. Variability in methods of measurement can occur for several reasons [5–7]:

1. *The setting in which the assessments are carried out.* Data collection took place at the children's homes and thus the assessment environment was somewhat variable except for what we could control. Where possible, the number of persons present during assessments was limited to three (fieldworker, caretaker and child); also, the surface of the floor where the assessments took place was kept clean and free of objects that

Correspondence: Mercedes de Onis, Study Coordinator, Department of Nutrition, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland. Tel: +41 22 791 3320. Fax: +41 22 791 4156. E-mail: deonism@who.int

ISSN 0803-5326 print/ISSN 1651-2227 online
DOI: 10.1080/08035320500495480

might interfere with locomotion, and a maximum of three toys or objects with which the child liked to play were available [2].

2. *The child's mood.* Children vary in their emotional state during assessments for a variety of reasons, and this cannot be controlled. Care was taken, however, to reassure and calm the children and to record their overall emotional state according to two scales described by Brazelton [8].
3. *The examiner's mood.* Examiners also vary among themselves, and over time, in mood, level of energy and motivation. Efforts were made to keep fieldworkers motivated, to impress upon them the importance of the study, and to repeatedly emphasize the need to adhere to the standardized protocol. In addition, appropriate training, site visits by the MDS coordinator and monitoring of data quality were essential to control for this third possible source of variability and to minimize bias across sites.
4. *Methodological differences among fieldworkers.* Observational assessment tools such as the assessment of motor milestones are particularly prone to error due to differences among fieldworkers in judging when a particular behaviour has been exhibited [9]. Therefore, considerable effort was made to standardize the criteria for assessing when certain motor skills were demonstrated, such as clear instructions and drawings in the procedures manual, periodic standardization sessions in all sites, and the use of videotapes to standardize criteria across sites.

The purpose of this paper is to describe the methods used to standardize the assessment of motor milestones in the MGRS and to present estimates of the reliability of these assessments.

Methods

Periodic site standardization sessions

Standardization sessions were conducted on a regular basis (at 1-mo or 2-mo intervals) during data collection in Ghana, India, Norway and Oman. The North American site did so only once because data collection was nearly completed by the time the decision was taken to conduct regular standardization sessions; also, and for the same reason, this site did not participate in the cross-site standardization exercise. Due to limited data availability, the North American site was thus not included in the analyses for this paper. Brazil, which was the earliest MGRS site, did not assess motor milestones.

During each session, 10 apparently healthy children, aged 6 to 12 mo, were recruited for participation through day-care and health centres. At every session, one of the fieldworkers examined and scored the children for each of the six gross motor milestones: sitting without support, hands-and-knees crawling, standing with assistance, walking with assistance, standing alone and walking alone. A different fieldworker was selected for each session to give everyone a turn. The performance of each milestone was recorded as follows: "inability"—the child tried but failed to perform the test item; "ability"—the child performed the test item according to the specified criteria; "refusal"—the child was calm and alert but uncooperative; and "unable to test"—the child could not be examined because his or her emotional state (drowsiness, fussiness or crying) interfered with the examination or the child's caretaker was distraught. In practice, it proved difficult to distinguish between "refusal" and "unable to test", and these categories were therefore combined. The child's caregiver was present during all assessments but was requested not to interfere with the examination. However, when needed, the examiner asked for the caregiver's assistance, for instance in placing the child into the correct position or in encouraging the child to crawl or walk. The examiner recorded the results discretely, taking care not to disclose the child's rating. Since it was not always possible to get the child to cooperate immediately, the examiner was allowed three tries to assess each milestone.

Assessments of the children were videotaped, and later the other fieldworkers in the same site watched the videotaped sessions and independently rated performances. The videotape of the session and the fieldworkers' ratings were then sent to the MGRS Coordinating Centre at WHO in Geneva where the MDS coordinator viewed the tape and rated the children's performance. The ratings given by the coordinator were considered to be the reference (true) scores.

Cross-site standardization session

The MDS coordinator visited Ghana, India, Norway and Oman to carry out standardization exercises using videotapes of 288 motor assessments made in 51 children. Care was taken to select the best demonstrations of the milestones. The fieldworkers in all four countries viewed the videotapes and independently rated the children's performance.

Statistical analysis

Three outcome categories were examined: 1) observed inability; 2) refusal and/or unable to test; and 3) observed ability.

The degree of concordance between fieldworkers and the MDS coordinator was analysed using the

Kappa (κ) coefficient, a measure of association for categorical variables [10]. Kappa compares the observed agreement between pairs of raters to the agreement expected by chance when judgements are statistically independent [11]. Kappa coefficients vary between 0 and 1. A Kappa coefficient of ≤ 0.20 indicates slight agreement, $\kappa = 0.21 - 0.40$ indicates fair agreement, $\kappa = 0.41 - 0.60$ indicates moderate agreement, $\kappa = 0.61 - 0.80$ indicates substantial agreement and $\kappa > 0.80$ means almost perfect agreement [12].

The percentage of agreement was also estimated because this value can be calculated in all instances [13], whereas Kappa coefficients cannot be calculated if all children are rated similarly by both fieldworkers. The percentage of agreement was calculated by dividing the number of agreements between a fieldworker's rating and the MDS coordinator by the total number of paired observations [13]. Agreement of 90% or more was considered high [2].

Further analysis was based on the methodology suggested by Reed [14] that allows one to judge whether the Kappa coefficients from several studies or clinical centres "belong together" as a set. In the MDS, a key question is whether Kappa coefficients across participating sites pass the homogeneity test. The null hypothesis is that the Kappas of all sites are equal for each of the milestones (H_0: $\kappa_{Ghana} = \kappa_{India} = \kappa_{Norway} = \kappa_{Oman}$). For this purpose, summary Kappa coefficients were calculated for all fieldworkers within a site and for each milestone. The goodness-of-fit test of the null hypothesis H_0 was obtained by using a statistic that is assumed to be χ^2 distributed with n (= number of sites − 1) degrees of freedom. Homogeneity was also assessed for Kappa coefficients across fieldworkers within sites and for each milestone (i.e. do all fieldworkers within a site have similar Kappas for each milestone?) and across milestones within sites (i.e. are the Kappas similar within sites for all six milestones?).

Two sources of information are available about concordance in the ratings of motor milestones between fieldworkers and the MDS coordinator: the site-specific exercises and the cross-site session. Should similar Kappa coefficients be expected? To answer this question, differences in approaches must be considered. All assessments by all fieldworkers in all sites used the same set of videotapes in the cross-site standardization session, whereas the site standardization sessions included local children and assessments by fieldworkers were done either by direct examination of the child or through videotapes. The MDS coordinator assessed video recordings in both types of exercises, although she was present in the sites during the cross-site standardization session. Because the videos were selected for teaching purposes, including clarity in filming and in the demonstration of motor behaviours, better concordance between fieldworkers and the MDS coordinator might be expected in the cross-site session.

Finally, we examined the level of concordance with the MDS coordinator in the rating of motor milestones when fieldworkers assessed children by direct examination or through videotapes by randomly selecting three fieldworkers per site and comparing their Kappa coefficients and percentage of agreement in each site.

All statistical analyses were performed using Stata 8.0 [15].

Results

Periodic site standardization sessions

Kappa coefficients and percent agreement with the MDS coordinator are given in Table I for all fieldworkers, by site, across all standardization sessions. The number of sessions varied by site: Ghana 8, India 11, Norway 2 and Oman 11. The number of children assessed per fieldworker and milestone varied as well because some fieldworkers did not complete the standardization sessions or because some milestone assessments were omitted due to poor filming. In general, there were "substantial" to "almost perfect" levels of agreement between fieldworkers and the MDS coordinator across all milestones and sites. Exceptions were the Kappa coefficients for the milestone "sitting without support" for fieldworker no. 4 in Ghana ($\kappa = 0.585$) and for the milestones "standing alone" and "walking alone" for fieldworker no. 6 in Norway ($\kappa = 0.422$ and 0.345, respectively). The percentage of agreement ranged between 81.0% (Norway, standing with assistance) and 100.0%.

Cross-site standardization session

Table II presents similar data to that in Table I but for the cross-site standardization session, where the MDS coordinator travelled to the sites and showed the same videotapes of 288 motor assessments. The Kappa coefficients indicate "substantial" to "almost perfect" levels of agreement between fieldworkers and the MDS coordinator. The percentage of agreement ranged between 80.9% (Ghana, walking alone) and 100.0%.

Concordance was rated "substantial" to "almost perfect" in both the periodic site and the cross-site standardization sessions but was often slightly higher in the cross-site session for all milestones except "walking alone" (values in Table II tend to be greater than values in Table I).

Table I. Kappa coefficients and % of agreement with the MDS coordinator for all fieldworkers, by site, for the periodic site standardization sessions[a].

		Ghana			India			Norway			Oman		
	Fieldworker	*n*	Kappa	% agree	*n*	Kappa	% agree	*n*	Kappa	% agree	*n*	Kappa	% agree
	1	83	0.851	98.8	107	0.904	98.1	20	0.857	95.0	103	0.923	97.1
	2	63	1.000	100.0	39	0.898	97.4	20	0.857	95.0	103	0.949	98.1
	3	53	0.660	98.1	107	0.900	98.1	20	0.857	95.0	103	0.925	97.1
Sitting without support	4	83	0.585	95.2	107	1.000	100.0	20	0.771	90.0	103	0.950	98.1
	5	53	0.658	98.1	107	0.952	99.1	20	1.000	100.0	103	1.000	100.0
	6	83	0.851	98.8	107	0.952	99.1	20	0.857	95.0	103	0.951	98.1
	7	83	0.851	98.8	39	0.898	97.4	20	1.000	100.0			
	8				77	0.892	97.4						
	9				107	0.908	98.1						
	Overall	501	0.761	98.2	797	0.927	98.5	140	0.884	95.7	618	0.950	98.1
	1	84	0.960	97.6	105	0.949	97.1	22	0.919	95.5	106	0.939	96.2
	2	65	0.949	96.9	35	0.849	91.4	22	0.919	95.5	106	0.939	96.2
	3	55	0.820	89.1	105	0.880	93.3	22	0.833	90.9	106	0.970	98.1
Hands-and-knees crawling	4	84	0.800	88.1	105	0.966	98.1	22	1.000	100.0	106	0.954	97.2
	5	55	0.938	96.4	105	0.931	96.2	22	1.000	100.0	106	0.955	97.2
	6	84	0.960	97.6	105	0.883	93.3	22	0.919	95.5	106	0.924	95.3
	7	84	0.942	96.4	35	0.952	97.1	22	1.000	100.0			
	8				75	0.861	92.0						
	9				105	0.897	94.3						
	Overall	511	0.912	94.7	775	0.911	95.0	154	0.943	96.8	636	0.947	96.7
	1	74	0.808	89.2	97	0.830	91.8	21	0.837	90.5	100	0.857	92.0
	2	57	0.727	86.0	38	0.867	92.1	21	1.000	100.0	100	0.875	93.0
	3	48	0.826	91.7	97	0.831	91.8	21	0.837	90.5	100	0.893	94.0
Standing with assistance	4	74	0.738	86.5	97	0.809	90.7	21	0.837	90.5	100	0.892	94.0
	5	48	0.767	87.5	97	0.785	89.7	21	1.000	100.0	100	0.911	95.0
	6	74	0.813	90.5	97	0.894	94.8	21	0.653	81.0	100	0.892	94.0
	7	74	0.760	87.8	38	0.869	92.1	21	0.755	85.7			
	8				71	0.893	94.4						
	9				97	0.804	90.7						
	Overall	449	0.777	88.4	729	0.839	91.9	147	0.847	91.2	600	0.887	93.7
	1	76	0.905	94.7	104	0.792	87.5	20	1.000	100.0	104	0.793	86.5
	2	60	0.822	90.0	37	0.903	94.6	20	0.917	95.0	104	0.777	85.6
	3	50	0.891	94.0	104	0.839	90.4	20	1.000	100.0	104	0.912	94.2
Walking with assistance	4	76	0.902	94.7	104	0.869	92.3	20	0.917	95.0	104	0.729	82.7
	5	50	0.854	92.0	104	0.836	90.4	20	1.000	100.0	104	0.807	87.5
	6	76	0.856	92.1	104	0.889	93.3	20	0.817	90.0	104	0.808	87.5
	7	76	0.882	93.4	37	0.808	89.2	20	1.000	100.0			
	8				74	0.841	90.5						
	9				104	0.773	86.5						
	Overall	464	0.875	93.1	772	0.838	90.3	140	0.951	97.1	624	0.805	87.3

Table I (*Continued*)

		Ghana			India			Norway			Oman		
	Fieldworker	*n*	Kappa	% agree	*n*	Kappa	% agree	*n*	Kappa	% agree	*n*	Kappa	% agree
	1	72	0.926	95.8	108	0.736	86.1	20	1.000	100.0	105	0.919	95.2
	2	57	0.836	91.2	39	0.875	94.9	20	0.683	85.0	105	0.902	94.3
	3	47	0.800	89.4	108	0.845	91.7	20	0.897	95.0	105	0.968	98.1
Standing alone	4	72	0.897	94.4	108	0.863	92.6	20	0.797	90.0	105	0.798	88.6
	5	47	0.783	89.4	108	0.768	88.0	20	0.797	90.0	105	0.902	94.3
	6	72	0.850	91.7	108	0.884	93.5	20	0.422	75.0	105	0.936	96.2
	7	72	0.873	93.1	39	0.939	97.4	20	0.785	90.0			
	8				78	0.725	87.2						
	9				108	0.823	90.7						
	Overall	439	0.861	92.5	804	0.820	90.7	140	0.776	89.3	630	0.905	94.4
	1	60	0.902	95.0	109	0.732	88.1	19	0.835	94.7	106	0.851	91.5
	2	45	0.773	88.9	40	0.804	92.5	19	0.835	94.7	106	0.834	90.6
	3	35	0.743	88.6	109	0.838	92.7	19	0.835	94.7	106	0.950	97.2
Walking alone	4	60	0.867	93.3	109	0.895	95.4	19	0.835	94.7	106	0.741	85.8
	5	35	0.743	88.6	109	0.820	92.7	19	1.000	100.0	106	0.896	94.3
	6	72	0.827	91.7	109	0.867	93.6	19	0.345	84.2	106	0.821	89.6
	7	72	0.880	94.4	40	0.939	97.5	19	1.000	100.0			
	8				79	0.806	92.4						
	9				109	0.849	93.6						
	Overall	379	0.835	92.1	813	0.835	92.9	133	0.822	94.7	636	0.849	91.5

[a] Analyses combine all standardization sessions per site (8 in Ghana, 11 in India, 2 in Norway and 11 in Oman).

Table II. Kappa coefficients and % of agreement with the MDS coordinator for all fieldworkers, by site, for the cross-site standardization session using videotapes of 288 motor assessments.

		Ghana		India		Norway		Oman	
	Fieldworker	Kappa	% agree	Kappa	%agree	Kappa	% agree	Kappa	% agree
Sitting without support ($n = 49$)	1	1.000	100.0	1.000	100.0	0.866	95.9	1.000	100.0
	2	0.930	98.0	0.936	98.0	0.930	98.0	1.000	100.0
	3	0.930	98.0	0.826	93.9	0.867	95.9	1.000	100.0
	4	0.871	95.9	0.936	98.0	0.879	95.9	0.930	98.0
	5	0.854	95.9	0.936	98.0	0.877	95.9	0.657	87.8
	6	1.000	100.0	1.000	100.0	0.936	98.0		
	7	0.868	95.9	1.000	100.0	0.867	95.9		
	8			1.000	100.0				
	Overall	0.923	97.7	0.952	98.5	0.889	96.5	0.909	97.1
Hands-and-knees crawling ($n = 47$)	1	0.894	93.6	0.964	97.9	0.887	93.6	0.887	93.6
	2	1.000	100.0	0.963	97.9	0.887	93.6	0.887	93.6
	3	0.893	93.6	0.964	97.9	0.735	85.1	0.926	95.7
	4	0.812	89.4	0.927	95.7	0.928	95.7	0.926	95.7
	5	0.963	97.9	0.859	91.5	0.926	95.7	0.776	87.2
	6	0.963	97.9	0.891	93.6	0.852	91.5		
	7	0.928	95.7	0.890	93.6	0.854	91.5		
	8			0.964	97.9				
	Overall	0.922	95.4	0.924	95.5	0.867	92.4	0.880	93.2
Standing with assistance ($n = 51$)	1	0.837	90.2	0.896	94.1	0.746	86.3	0.931	96.1
	2	0.864	92.2	0.828	90.2	0.896	94.1	0.860	92.2
	3	0.896	94.1	0.932	96.1	0.859	92.2	0.896	94.1
	4	0.827	90.2	0.824	90.2	0.863	92.2	0.895	94.1
	5	0.901	94.1	0.863	92.2	0.899	94.1	0.861	92.2
	6	0.897	94.1	0.933	96.1	0.720	84.3		
	7	0.862	92.2	0.898	94.1	0.862	92.2		
	8			0.896	94.1				
	Overall	0.869	92.4	0.888	93.6	0.836	90.8	0.889	93.7
Walking with assistance ($n = 48$)	1	0.962	97.9	0.818	89.6	0.889	93.8	1.000	100.0
	2	0.927	95.8	0.814	89.6	0.890	93.8	0.925	95.8
	3	0.924	95.8	0.890	93.8	0.769	87.5	0.963	97.9
	4	0.962	97.9	0.887	93.8	0.852	91.7	0.887	93.8
	5	0.888	93.8	0.846	91.7	0.887	93.8	0.962	97.9
	6	0.962	97.9	0.925	95.8	0.888	93.8		
	7	0.925	95.8	0.890	93.8	0.927	95.8		
	8			0.753	85.4				
	Overall	0.935	96.4	0.848	91.4	0.872	92.9	0.947	97.1
Standing alone ($n = 46$)	1	0.952	97.8	0.901	95.7	0.819	91.3	1.000	100.0
	2	0.902	95.7	1.000	100.0	0.949	97.8	0.901	95.7
	3	0.857	93.5	0.907	95.7	0.896	95.7	0.951	97.8
	4	0.648	84.8	1.000	100.0	1.000	100.0	0.952	97.8
	5	0.949	97.8	0.952	97.8	0.902	95.7	0.851	93.5
	6	0.949	97.8	0.952	97.8	0.848	93.5		
	7	0.951	97.8	1.000	100.0	0.763	89.1		
	8			0.951	97.8				
	Overall	0.888	95.0	0.964	98.4	0.881	94.7	0.931	97.0
Walking alone ($n = 47$)	1	0.801	93.6	0.678	89.4	0.702	89.4	0.803	93.6
	2	0.780	93.6	0.931	97.9	0.927	97.9	0.803	93.6
	3	0.721	91.5	0.702	89.4	0.780	93.6	0.861	95.7
	4	0.722	91.5	1.000	100.0	0.927	97.9	0.801	93.6
	5	0.780	93.6	1.000	100.0	0.794	93.6	0.813	93.6
	6	0.780	93.6	0.771	93.6	0.781	93.6		
	7	0.861	95.7	0.862	95.7	0.658	87.2		
	8			0.861	95.7				
	Overall	0.778	93.3	0.838	95.0	0.788	93.3	0.816	94.0

Table III. Tests of homogeneity of Kappa coefficients in the MDS: p-values for the periodic site standardization sessions (SSS) and for the cross-site standardization session (CSS).

	Ghana		India		Norway		Oman		Across sites, within milestones
	SSS	CSS	SSS	CSS	SSS	CSS	SSS	CSS	CSS
Sitting without support	NA[a]	0.619	0.925	0.246	0.789	0.848	0.580	NA[b]	0.414
Hands-and-knees crawling	0.198	0.265	0.497	0.646	0.602	0.550	0.903	0.477	0.274
Standing with assistance	0.942	0.983	0.926	0.900	0.355	0.510	0.989	0.916	0.463
Walking with assistance	0.923	0.912	0.772	0.665	0.420	0.790	0.519	0.418	0.082
Standing alone	0.857	0.050	0.613	0.629	0.619	0.318	0.127	0.501	0.084
Walking alone	0.753	0.305	0.656	0.102	0.768	0.452	0.116	0.955	0.890
Across milestones, within sites	0.199	0.546	0.438	0.668	0.384	0.772	0.265	0.662	

[a] Test of homogeneity among Kappas can not be performed because the number of concordant negative ratings (i.e. fieldworker and MDS coordinator recording that the child was unable to perform the milestone) was zero for all fieldworkers for milestone sitting without support.
[b] Test of homogeneity among Kappas can not be performed because the number of discordant (i.e. fieldworker and MDS coordinator recording different ratings for the same child) was zero for three out of five fieldworkers for milestone sitting without support.

Homogeneity

Table III presents results assessing the homogeneity of Kappa coefficients in the site standardization sessions and the cross-site session. *P*-values inside the table (all values but those given in the bottom row and right-hand column) answer the question: Are the fieldworkers homogeneous in assessing motor milestones within a site? *P*-values in the right-hand column answer the question: Are the fieldworkers homogeneous in assessing motor milestones across sites when viewing the same videotapes? *P*-values on the bottom row answer the question: Are the fieldworkers homogeneous in their assessments across milestones within a site? None of the *P*-values were statistically significant ($p < 0.05$), although one value (Ghana, standing alone, CSS) had a p-value of 0.05. These results indicate that the Kappas are homogeneous across sites, across milestones, and across fieldworkers.

Concordance in assessment by direct examination versus videotape

Table IV presents, for 12 randomly selected fieldworkers (three per site), the Kappa coefficients and percentage of agreement with the MDS coordinator when fieldworkers tested children by direct examination or using videotapes. Overall, there were no systematic differences to indicate that one way of conducting the assessment is more concordant with the MDS coordinator than the other.

Discussion

This is the first longitudinal study to use a standardized protocol to describe gross motor development among healthy children from different countries and to carry out standardization sessions on a regular basis. Kappa coefficients were used to estimate the concordance of independent pairs of raters, specifically one of several fieldworkers and always the MDS coordinator. These values estimate the quality of the MDS testing procedures [2] and the fieldworkers' ability to apply the rating criteria consistently.

Overall, high percentages of agreement between fieldworkers and the MDS coordinator, and "substantial" to "almost perfect" Kappa coefficients, were obtained for all fieldworkers, milestones and sites. Homogeneity tests confirm that the Kappa coeffi-

Table IV. Comparison of Kappa coefficients and percentage agreement when three randomly selected fieldworkers per site assessed children by direct examination or through videotapes.

Site	Assessment	Milestone[a]	Kappa	% agreement
Ghana	Direct	2	1.000	100.0
	Video		0.945	96.9
Ghana	Direct	2	0.808	90.0
	Video		0.796	87.8
Ghana	Direct	5	0.912	94.1
	Video		0.929	96.4
India	Direct	1	1.000	100.0
	Video		0.948	99.0
India	Direct	2	0.805	87.5
	Video		0.887	93.8
India	Direct	4	0.821	90.0
	Video		0.839	90.4
Norway	Direct	2	1.000	100.0
	Video		0.896	94.1
Norway	Direct	4	1.000	100.0
	Video		0.902	93.8
Norway	Direct	5	0.556	75.0
	Video		0.360	75.0
Oman	Direct	3	0.841	90.0
	Video		0.896	94.4
Oman	Direct	4	0.628	75.0
	Video		0.755	84.5
Oman	Direct	6	0.814	88.9
	Video		0.834	90.9

[a] Milestone: 1 = sitting without support; 2 = hands-and-knees crawling; 3 = standing with assistance; 4 = walking with assistance; 5 = standing alone; 6 = walking alone.

cients are a homogeneous set across sites, across milestones, and across fieldworkers. Concordance was slightly higher in the cross-site session (i.e. when fieldworkers rated the same set of videotapes) than in the periodic site standardization sessions where different sets of local children were assessed.

The forgoing analyses show that the standardization of milestone assessments made in any one site were consistently high among fieldworkers within a site, among milestones within a site, and among sites across all six milestones. Also, the cross-site exercise indicates that the fieldworkers could reliably rate motor milestones of children both in their own and in the other sites.

There are few reports of inter-rater agreement [16–19] in motor milestones assessments, and what information is available suggests that the MDS concordance is very good relative to other studies. For example, the mean percentage of agreement between four examiners during the standardization of the Denver Developmental Screening Test was 90%, with a range of 80–95% [17]. Using the Movement Assessment of Infants, Haley et al. [16] reported only 2% of the items demonstrated excellent ($\kappa > 0.75$) inter-rater reliability beyond chance, with 58% in the fair-to-good ($0.40 < \kappa < 0.75$) range.

The six milestones were selected for the study because they were considered to be both fundamental to the acquisition of self-sufficient erect locomotion and simple to administer and evaluate. They should measure observable behaviour with a clear pass or fail score. The high degree of inter-rater reliability confirms that these milestones were simple to administer and feasible to standardize. These results were probably attributable to the clarity of the instructions for administering and rating the performance of the milestones, and to the fact that fieldworkers were well trained. As observed in other studies [18,19], the multiple standardization sessions no doubt added to the fieldworkers' skills and confidence in conducting motor development assessments.

The organization of reliability sessions is often logistically demanding and places considerable stress on both researchers and family members. An attractive alternative is to estimate inter-rater reliability coefficients with the aid of videotapes instead of having several examiners test a group of children more than once. Stuberg et al. [20] found that minimizing the handling of children and relying on observation help achieve more accurate test results. Children can behave differently from one time to the next [17], and these differences may influence the reliability coefficients. By using videotapes, these results reflected the fieldworker's ability to rate the test items under controlled conditions, that is without having to deal with children's moods and behaviours. On the other hand, Gowland et al. [21] concluded that observing task performances from a videotape appeared to be a major source of variability because taping frequently did not capture the full performance, or part of the body to be observed was not filmed fully or from an appropriate angle. Our study excluded milestone assessments that could not be rated for these reasons, and we found no systematic difference in the Kappa coefficients and percentage of agreement when fieldworkers rated children by direct examination or through videotapes.

We found several advantages, which were also common to other studies [6,22,23], in using video recordings to evaluate rating performances. Videotapes helped to alleviate problems with recruiting children and scheduling sessions. Fieldworkers were able to rate the motor development assessments when convenient to them. The MDS coordinator could examine the tape with the fieldworkers to explore possible reasons for disagreement. Most importantly, children did not have to endure repeated assessments by numerous fieldworkers. Russell et al. [6] cited as a main disadvantage that this method tests only the participant's ability to rate the videotaped assessments but provides no indication of the participant's ability to administer and score them in a clinical or study situation. This is a fair criticism, and for this reason studies should assess the quality of assessments in both direct examination and video settings. This is what we did, but in our case we did not find systematic differences between these settings.

The MDS protocol was designed to provide a simple method of evaluating six gross motor milestones in young children. The WHO MGRS, in implementing this protocol, provided the opportunity to evaluate these milestones in multiple countries and, for the first time, to use the data collected to construct an international standard for the achievement of six universal gross motor development milestones [24,25]. Assessing children's behaviour, including gross motor milestones, is demanding for both fieldworkers and children. The results of this study demonstrate that, with careful attention to protocol and training, a high level of fieldworker reliability can be achieved within and across sites.

Acknowledgements

This paper was prepared by Trudy M.A. Wijnhoven, Mercedes de Onis, Reynaldo Martorell, Edward A. Frongillo and Gunn-Elin A. Bjoerneboe on behalf of the WHO Multicentre Growth Reference Study Group. The statistical analysis was conducted by Amani Siyam.

References

[1] de Onis M, Garza C, Victora CG, Onyango AW, Frongillo EA, Martines J, for the WHO Multicentre Growth Reference Study Group. The WHO Multicentre Growth Reference Study: Planning, study design and methodology. Food Nutr Bull 2004;25 Suppl 1:S15–26.

[2] Wijnhoven TM, de Onis M, Onyango AW, Wang T, Bjoerneboe GE, Bhandari N, et al., for the WHO Multicentre Growth Reference Study Group. Assessment of gross motor development in the WHO Multicentre Growth Reference Study. Food Nutr Bull 2004;25 Suppl 1:S37–45.

[3] Hindley CB, Filliozat AM, Klackenberg G, Nicolet-Meister D, Sand EA. Differences in age of walking in five European longitudinal samples. Hum Biol 1966;38:364–79.

[4] World Health Organization, Task Force for Epidemiological Research on Reproductive Health; Special Programme of Research, Development, and Research Training in Human Reproduction. Progestogen-only contraceptives during lactation: II. Infant development. Contraception 1994;50:55–68.

[5] Krebs DE. Measurement theory. Phys Ther 1987;67:1834–9.

[6] Russell DJ, Rosenbaum PL, Lane M, Gowland C, Goldsmith CH, Boyce WF, et al. Training users in the gross motor function measure: methodological and practical issues. Phys Ther 1994;74:630–6.

[7] Plewis A, Bax M. The uses and abuses of reliability measures in developmental medicine. Dev Med Child Neurol 1982;24: 388–90.

[8] Brazelton TB. Echelle d'évaluation du comportement néonatal. Neuropsychiatr Enfance Adolesc 1983;31:61–96.

[9] Mitchell SK. Interobserver agreement, reliability, and generalizability of data collected in observational studies. Psychol Bull 1979;86:376–90.

[10] Chmura Kraemer H, Periyakoil VS, Noda A. Kappa coefficients in medical research. Stat Med 2002;21:2109–29.

[11] Agresti A. An introduction to categorical data analysis. Wiley series in probability and statistics. New York: John Wiley & Sons, Inc.; 1996.

[12] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.

[13] Altman DG. Practical statistics for medical research. London: Chapman & Hall/CRC; 1991.

[14] Reed JF III. Homogeneity of Kappa statistics in multiple samples. Comput Methods Programs Biomed 2000;63:43–6.

[15] Stata/SE 8.0 for Windows. College Station, TX: Stata Corporation; 2003.

[16] Haley S, Harris SR, Tada WL, Swanson MW. Item reliability of the movement assessment of infants. Phys Occup Ther Pediatr 1986;61:21–39.

[17] Frankenburg WK, Dodds JB. The Denver Development Screening Test. J Pediatr 1967;71:181–91.

[18] Hammarlund K, Persson K, Sedin G, Stromberg B. A protocol for structured observation of motor performance in preterm and term infants. Interobserver agreement and intraobserver consistency. Ups J Med Sci 1993;98:77–82.

[19] Thomas SS, Buckon CE, Phillips DS, Aiona MD, Sussman MD. Interobserver reliability of the gross motor performance measure: preliminary results. Dev Med Child Neurol 2001;43: 97–102.

[20] Stuberg WA, White PJ, Miedaner JA, Dehne PR. Item reliability of the Milani-Comparetti Motor Development Screening Test. Phys Ther 1989;69:328–35.

[21] Gowland C, Boyce WF, Wright V, Russell DJ, Goldsmith CH, Rosenbaum PL. Reliability of the Gross Motor Performance Measure. Phys Ther 1995;75:597–602.

[22] Gross D, Conrad B. Issues related to reliability of videotaped observational data. West J Nurs Res 1991;13:798–803.

[23] Nordmark E, Hägglund G, Jarnlo GB. Reliability of the gross motor function measure in cerebral palsy. Scand J Rehab Med 1997;29:25–8.

[24] WHO Multicentre Growth Reference Study Group. Assessment of sex differences and heterogeneity in motor milestone attainment among populations in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450:66–75.

[25] WHO Multicentre Growth Reference Study Group. WHO Motor Development Study: Windows of achievement for six gross motor development milestones. Acta Paediatr Suppl 2006;450:86–95.

Acta Pædiatrica, 2006; Suppl 450: 56–65

Taylor & Francis
Taylor & Francis Group

Assessment of differences in linear growth among populations in the WHO Multicentre Growth Reference Study

WHO MULTICENTRE GROWTH REFERENCE STUDY GROUP[1,2]

[1]*Department of Nutrition, World Health Organization, Geneva, Switzerland, and* [2]*Members of the WHO Multicentre Growth Reference Study Group (listed at the end of the first paper in this supplement)*

Abstract

Aim: To assess differences in length/height among populations in the WHO Multicentre Growth Reference Study (MGRS) and to evaluate the appropriateness of pooling data for the purpose of constructing a single international growth standard. *Methods:* The MGRS collected growth data and related information from 8440 affluent children from widely differing ethnic backgrounds and cultural settings (Brazil, Ghana, India, Norway, Oman and the USA). Eligibility criteria included breastfeeding, no maternal smoking and environments supportive of unconstrained growth. The study combined longitudinal (birth to 24 mo) and cross-sectional (18–71 mo) components. For the longitudinal component, mother–infant pairs were enrolled at delivery and visited 21 times over the next 2 y. Rigorous methods of data collection and standardized procedures were applied across study sites. We evaluate the total variability of length attributable to sites and individuals, differences in length/height among sites, and the impact of excluding single sites on the percentiles of the remaining pooled sample. *Results:* Proportions of total variability attributable to sites and individuals within sites were 3% and 70%, respectively. Differences in length and height ranged from −0.33 to +0.49 and −0.41 to +0.46 standard deviation units (SDs), respectively, most values being below 0.2 SDs. Differences in length on exclusion of single sites ranged from −0.10 to +0.07, −0.07 to +0.13, and −0.25 to +0.09 SDs, for the 50th, 3rd and 97th percentiles, respectively. Corresponding values for height ranged from −0.09 to +0.08, −0.12 to +0.13, and −0.15 to +0.07 SDs.

Conclusion: The striking similarity in linear growth among children in the six sites justifies pooling the data and constructing a single international standard from birth to 5 y of age.

Key Words: *Childhood growth, growth curves, growth standards, height, length*

Introduction

Child growth charts are among the most commonly used tools for assessing the health and nutritional status of individual infants and children, and the general well-being of their communities [1]. They are useful in determining the degree to which physiological needs for growth and development are met during the fetal and childhood periods. Recognizing the shortcomings of the current National Center for Health Statistics/World Health Organization (NCHS/WHO) international growth reference [1,2], the WHO began planning in 1994 for new references that reflect how children *should* grow in all countries rather than merely describing how they grew at a particular time and place [3,4]. This prescriptive approach explicitly recognizes that growth references are often used as standards, that is, as tools that enable value judgments [5].

The WHO Multicentre Growth Reference Study (MGRS) collected primary growth data and related information from 8440 affluent children from widely differing ethnic backgrounds and cultural settings (Brazil, Ghana, India, Norway, Oman and the USA) [6]. An international sampling frame was selected on the basis of scientific and health advocacy considerations. Scientifically, it is well established that children from diverse ethnic groups grow very similarly during the first 5 y of life when their physiological needs are met and environments support healthy development [7–10]. Health advocacy considerations were also strong in the MGRS design. The development of a growth standard based on children from different world regions has the potential to yield an effective tool for child health advocacy by underscoring the fact that children in all countries *can* achieve their full growth potential when their nurturing follows health

Correspondence: Mercedes de Onis, Study Coordinator, Department of Nutrition, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland. Tel: +41 22 791 3320. Fax: +41 22 791 4156. E-mail: deonism@who.int

ISSN 0803-5326 print/ISSN 1651-2227 online
DOI: 10.1080/08035320500495514

recommendations and care practices associated with healthy outcomes [5].

This paper evaluates differences in length/height from birth to 5 y of age within and among the MGRS sites. It addresses two issues fundamental to the construction of the new standards: the potential for linear growth in diverse ethnic populations whose health and care needs are met, and the appropriateness of a single international standard for this age group. Length/height was selected as the most suitable measure to assess population differences of possible genetic or environmental origin among children of well-off families. Linear growth is normally distributed and resistant to skewing in response to excessive energy intakes, unlike weight which is more "plastic" in response to overnutrition. On the other hand, linear growth can be affected negatively and profoundly by environmental factors such as diet and infection, but it is unlikely that these would be relevant in the affluent populations selected for this study.

Methods

Design

The MGRS (July 1997–December 2003) was a population-based study covering the cities of Davis, California, USA; Muscat, Oman; Oslo, Norway; and Pelotas, Brazil; and selected affluent neighbourhoods of Accra, Ghana, and South Delhi, India. The MGRS protocol and its implementation in the six sites have been described in detail elsewhere [6,11–16]. Briefly, the MGRS combined a longitudinal study from birth to 24 mo with a cross-sectional study of children aged 18 to 71 mo. In the longitudinal study, mothers and newborns were screened and enrolled at birth and visited at home a total of 21 times on weeks 1, 2, 4 and 6; monthly from 2–12 mo; and bimonthly in the second year. Data were collected on anthropometry, motor development, feeding practices, child morbidity, perinatal factors, and socio-economic, demographic and environmental characteristics. The analyses in this paper focus on recumbent length measurements from the longitudinal sample and standing height measurements from the cross-sectional sample.

The study populations had socio-economic conditions favourable to growth and low mobility, with ≥20% of mothers following feeding recommendations and having access to breastfeeding support [6]. Individual inclusion criteria were: the absence of health or environmental constraints on growth, mothers willing to follow MGRS feeding recommendations (i.e. exclusive or predominant breastfeeding for at least 4 mo; introduction of complementary foods by the age of 6 mo; partial breastfeeding continued for at least 12 mo), no maternal smoking before and after delivery, single term birth, and absence of significant morbidity [6]. As part of the site-selection process in Ghana, India and Oman, surveys were conducted to identify socio-economic characteristics that could be used to select groups whose growth was not environmentally constrained [17–19]. Local criteria for screening newborns, based on parental education and/or income levels, were developed from those surveys [12,13,15]. Pre-existing survey data were available from Brazil, Norway and the United States for this purpose [11,14,16]. Term low-birthweight infants (2.3%) were not excluded since it is likely that, in well-off populations, such infants represent small but normal children and their exclusion would have artificially distorted the standards' lower percentiles. Eligibility criteria for the cross-sectional study were the same as those for the longitudinal study with the exception of infant feeding practices. A minimum of 3 mo of any breastfeeding was required for participants in the study's cross-sectional component.

The total sample size for the longitudinal and cross-sectional studies in all six sites was 8440 children. Length (longitudinal sample) and height (cross-sectional sample) were measured at all sites following standardized procedures using, respectively, a Harpenden Infantometer and Stadiometer. The detailed protocols followed to obtain anthropometric measurements and to ensure high-quality data are described elsewhere [6,20,21].

Analytical methods

The analyses of the MGRS longitudinal study are based on measurements taken at birth, and at 6, 12, 18 and 24 mo of all enrolled children. Analyses of the cross-sectional study were conducted at the following age intervals: 24–26 mo, 36–38 mo, 48–50 mo and 60–62 mo. Cross-sectional measurements obtained in the indicated age intervals were adjusted to the midpoint of each interval using linear regression and assumed equal growth rates for all children within each interval.

Heterogeneity in length among sites was assessed by comparing the percentages of the variance due to inter-individual and inter-site differences estimated by analysis-of-variance techniques that included adjustments for sex and age. For this analysis, the sample was restricted to those children followed for the entire period of 24 mo (88% of the enrolled sample, Table I) to permit measurement of variability within subjects using a balanced repeated-measures design. Variance components analyses [22] were based on a linear mixed-effect model. Analyses were done using SAS software, and restricted maximum likelihood was used for estimation. Age and sex were treated as fixed effects. Sites and individuals were treated as random effects. The repeated visits were also treated as random effects and represent the variability within subjects, estimated as the residual variance or random error.

The assessment of differences in length/height and the impact of individual sites on central values and selected percentiles was done by comparing each site's mean to the overall pooled mean and by comparing the effect of excluding single sites on the remaining pooled sample. Differences in length/height were expressed relative to the standard deviation (SD) of the all-site pooled sample, i.e. differences between individual site means and the pooled mean were divided by the pooled SD. These values are referred to as "standardized site effects". A similar approach was used when comparing the mean and selected percentiles calculated by excluding single sites with the corresponding pooled values. The magnitude and consistency of differences were used to assess the impact of site heterogeneity on the overall sample. According to Cohen [23], differences of 0.2 SD units are considered small, 0.5 SD medium and 0.8 SD large. In designing the MGRS, it had been decided that pooling would be appropriate if differences were less than medium in size.

Results

Table I presents the number of children in the longitudinal and cross-sectional samples and respective site-specific parental stature.

Results of variance components analyses for children in the longitudinal sample are summarized in Table II. After accounting for sex and age, variability among sites and among individuals within sites was, respectively, approximately 3% and 70% of the total variance. Thus, the percentage of the variation due to individuals was approximately 20 times greater than that due to sites.

Tables III and IV present mean lengths and heights, respectively, of the longitudinal and cross-sectional samples when all sites were pooled and for individual sites. They also present differences between individual site means and the overall pooled mean. These differences are expressed as standardized site effects, i.e. as fractions of the pooled standard deviation. Mean lengths and heights for the longitudinal and cross-sectional samples are presented graphically in Figures 1 and 2, respectively.

Differences in length (expressed as a fraction of the pooled sample SD) across sites at the indicated ages ranged from −0.33 to +0.49, most values being below 0.2 SD units (Table III). For height, values across sites at indicated ages ranged from −0.41 to +0.46 (Table IV). Although no site accounted for all the most positive or most negative differences, Oman accounted for the most negative values in seven of the nine ages and age intervals examined, and Norway and Brazil accounted most commonly for the most positive values.

Tables V and VI present, respectively, mean, 3rd percentile and 97th percentile values for length and height at the indicated ages when all sites were pooled and indicated sites excluded. Differences between values that resulted from the exclusion of single sites and the overall pooled value were also calculated. These, too, were expressed as standardized site effects, i.e. as fractions of the overall pooled standard deviation.

For length, differences between the 50th, 3rd and 97th percentiles calculated by excluding individual sites and the corresponding overall pooled values ranged from −0.10 to +0.07, −0.07 to +0.13, and −0.25 to +0.09, respectively.

For height, values ranged from −0.09 to +0.08, −0.12 to +0.13, and −0.15 to +0.07 for the 50th, 3rd and 97th percentiles, respectively. The wider ranges were observed for values at the 3rd and 97th

Table I. Sample size and parental stature in the longitudinal and cross-sectional samples.

	All sites	Brazil	Ghana	India	Norway	Oman	USA
Longitudinal sample:							
No. of enrolled children	1743	310	329	301	300	295	208
No. followed for 24 mo (% of total enrolled)	1542 (88)	287 (93)	292 (89)	269 (89)	262 (87)	260 (88)	172 (83)
Maternal stature (cm) (mean ± SD)	161.6 ± 7.2	161.1 ± 6.0	161.9 ± 5.2	157.6 ± 5.4	168.7 ± 6.6	156.6 ± 5.5	164.5 ± 6.9
Paternal stature (cm) (mean ± SD)	175.1 ± 7.9	173.6 ± 6.9	173.0 ± 6.6	172.7 ± 6.3	182.2 ± 6.7	170.4 ± 6.4	178.9 ± 7.4
Cross-sectional sample							
No. of enrolled children:	6697	487	1406	1490	1387	1447	480
Maternal stature (cm) (mean ± SD)	161.0 ± 7.2	160.0 ± 6.2	161.9 ± 5.7	157.6 ± 5.7	167.7 ± 6.5	156.6 ± 5.4	164.3 ± 6.7
Paternal stature (cm) (mean ± SD)	173.8 ± 7.9	173.2 ± 7.0	172.6 ± 6.6	172.1 ± 6.0	181.2 ± 7.2	169.2 ± 6.4	178.0 ± 7.4

Table II. Variance components analyses for length in the longitudinal sample [a].

Variance component	Estimate	Standard error (estimate)	Proportion (%)
Var(Site)	0.22	0.139	3.4
Var(Individual within site)	4.50	0.179	70.0
Var(Error)	1.71	0.032	26.6

[a] Age and sex as fixed effects.

percentiles. At the 3rd percentile, Oman's exclusion resulted in the most positive value in six of the nine ages and age intervals that were examined. Brazil's exclusion accounted for the most negative values in six of the nine ages and age intervals examined. The same pattern was observed at the 97th percentile.

Figures 3 and 4 illustrate the impact of excluding Brazil and Oman, respectively, on the 3rd, 25th, 50th, 75th and 97th length-for-age percentiles. Figures for Ghana, India, Norway and the USA are omitted because they had the least impact on the indicated percentiles when any of these sites was excluded.

Discussion

This study is the first to compare linear growth among affluent children aged 0–5 y using data collected in different countries according to a common protocol. Two lines of reasoning support the conclusion that all six MGRS sites can be used for the purpose of constructing a single international growth standard. The first relies on evidence provided by variance components analyses and, the second, on examining differences between individual site values and values derived from pooling all sites.

Variance components analyses demonstrated that variability in growth was due overwhelmingly to differences among individuals (70% of the total

Table III. Pooled and individual site sample sizes (n), means and standard deviations (SD) for length (cm).

Age	Sample	n	Mean (cm)	SD	Standardized site effects[a]
Birth	Pooled	1742	49.55	1.91	0.00
	Brazil	309	49.61	1.89	0.03
	Ghana	329	49.45	1.92	−0.05
	India	301	48.99	1.79	−0.29
	Norway	300	50.40	1.86	0.45
	Oman	295	49.18	1.72	−0.20
	USA	208	49.74	1.96	0.10
6 mo	Pooled	1648	66.72	2.35	0.00
	Brazil	296	66.75	2.35	0.01
	Ghana	306	66.57	2.29	−0.06
	India	287	66.60	2.28	−0.05
	Norway	286	67.88	2.37	0.49
	Oman	274	66.07	2.04	−0.27
	USA	199	66.30	2.39	−0.18
12 mo	Pooled	1594	75.02	2.62	0.00
	Brazil	290	75.39	2.69	0.14
	Ghana	301	75.16	2.69	0.05
	India	279	74.96	2.53	−0.02
	Norway	272	75.47	2.55	0.17
	Oman	265	74.43	2.41	−0.22
	USA	187	74.47	2.73	−0.21
18 mo	Pooled	1535	81.76	2.90	0.00
	Brazil	285	82.40	2.97	0.22
	Ghana	293	81.95	2.84	0.06
	India	268	81.50	2.86	−0.09
	Norway	255	82.06	2.77	0.10
	Oman	259	80.87	2.73	−0.31
	USA	175	81.70	3.01	−0.02
24 mo	Pooled	1524	87.40	3.18	0.00
	Brazil	280	88.35	3.17	0.30
	Ghana	289	87.48	3.04	0.03
	India	269	87.00	3.15	−0.13
	Norway	257	87.75	3.06	0.11
	Oman	260	86.36	3.08	−0.33
	USA	169	87.38	3.33	−0.01

[a] Standardized site effects are the differences between the indicated site means and the corresponding pooled (all sites) mean divided by the pooled standard deviation.

Table IV. Pooled and individual site sample sizes (*n*), means and standard deviations (SD) for height (cm).

Age	Sample	*n*	Mean (cm)	SD	Standardized site effects[a]
24–26 mo	Pooled	484	87.36	3.54	0.00
	Brazil	85	88.89	2.95	0.43
	Ghana	78	87.06	3.14	−0.08
	India	98	87.03	4.03	−0.09
	Norway	135	87.31	3.39	−0.01
	Oman	88	86.57	3.70	−0.22
	USA[b]	0			
36–38 mo	Pooled	502	96.26	4.04	0.00
	Brazil	91	97.91	4.04	0.41
	Ghana	85	96.34	3.95	0.02
	India	86	95.41	4.34	−0.21
	Norway	70	96.65	3.56	0.10
	Oman	83	95.26	3.84	−0.25
	USA	87	95.94	3.88	−0.08
48–50 mo	Pooled	478	103.52	4.23	0.00
	Brazil	71	104.87	4.84	0.32
	Ghana	94	104.29	4.56	0.18
	India	76	103.31	3.82	−0.05
	Norway	70	103.59	3.66	0.02
	Oman	80	101.78	4.31	−0.41
	USA	87	103.29	3.50	−0.05
60–62 mo	Pooled	465	110.32	4.86	0.00
	Brazil	91	111.15	4.98	0.17
	Ghana	76	112.55	6.00	0.46
	India	70	108.78	3.64	−0.32
	Norway	70	110.64	4.16	0.07
	Oman	73	109.00	4.07	−0.27
	USA	85	109.55	4.84	−0.16

[a] Standardized site effects are the differences between the indicated site means and the corresponding pooled (all sites) mean divided by the pooled standard deviation.
[b] The USA site did not enrol children in this age group for the cross-sectional study because the majority of that age cohort was participating in the longitudinal study.

variance) and only minimally to differences among sites (3% of the total variance). Thus, the percentage of the variability in length due to inter-individual differences was 20-fold greater than that due to differences among sites. Results from these analyses are consistent with genomic comparisons among

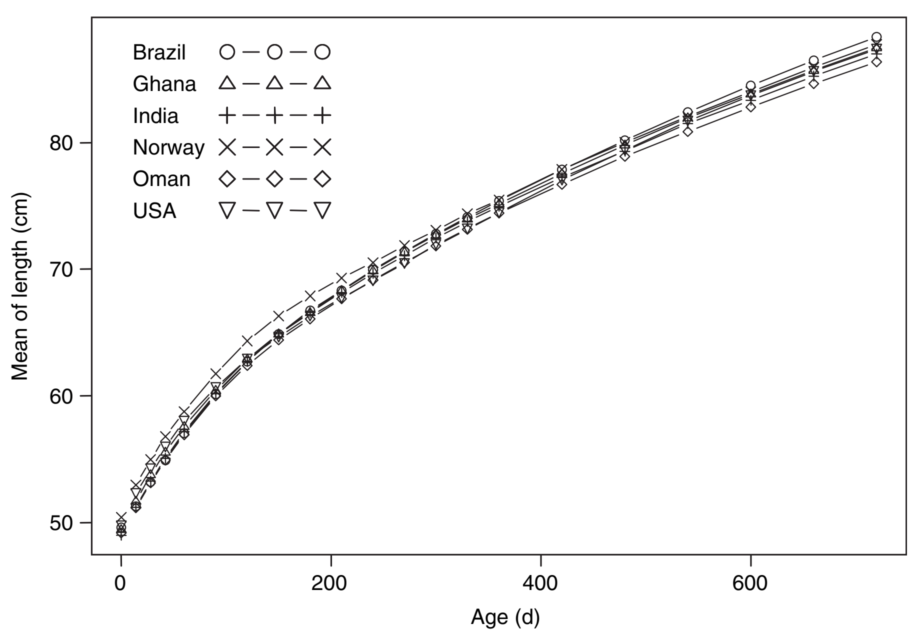


Figure 1. Mean length (cm) from birth through 2 y for each of the six sites.

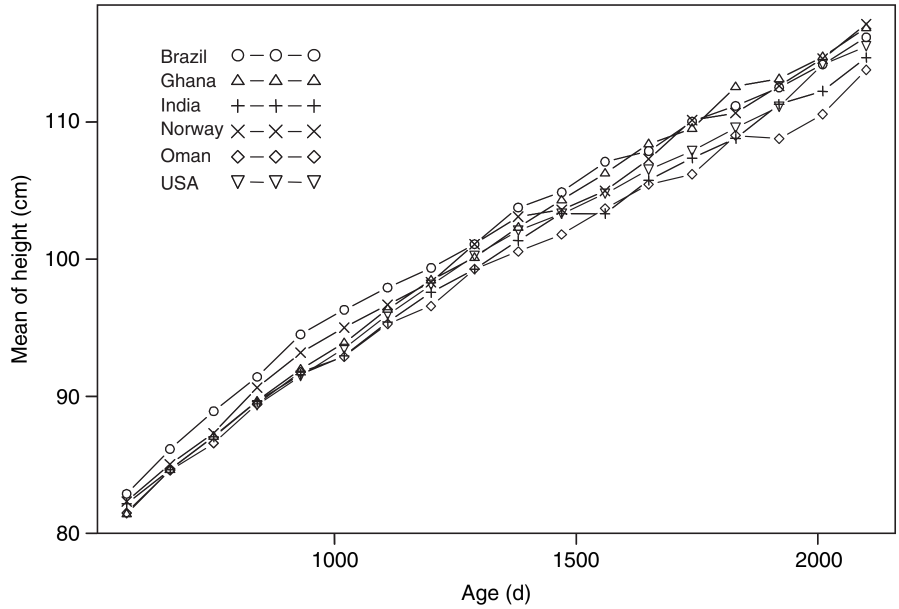


Figure 2. Mean height (cm) from 2 to 5 y of age for each of the six sites.

diverse continental groups reporting a high degree of inter-population homogeneity [24,25]. Current estimates suggest that 85 to 90% of total genetic variability resides within populations, whereas only 10% to 15% resides among populations [25]. Thus, it is unlikely that traits such as stature, which are continuous and multigenic, will differ significantly on the basis of genetics alone among large, non-isolated population groups [26]. The relatively small differences in child growth among sites, despite differences in parental stature, might decrease further in future studies. For example, the observed tendency towards smaller child size in Oman may be attributable to the shorter heights of mothers since maternal height influences birthweight and thus postnatal growth. Health conditions in Oman have improved in recent decades, and it is likely that the secular trend in adult stature will be sustained with continued economic development. Indeed, it took European populations several generations of prosperity to overcome the dire poverty and poor health that existed prior to the industrial revolution to reach their current stature [10,27].

The second set of analyses evaluated inter-site differences in length/height and the impact on selected percentiles of omitting individual sites. Ghana and the USA tended to coincide most closely with the total pool's central tendencies and distribution. Omani and, to a lesser extent, Indian children were represented commonly at lower values, and Brazilian and Norwegian children were represented commonly at higher values. Inter-site differences, however, were relatively small. For the five ages examined in the longitudinal sample and the four age intervals examined in the cross-sectional sample, no site mean deviated by an absolute amount equal to or greater than 0.5 SD of the corresponding overall sample mean. Of 54 values examined, only 20 were above 0.2 SD units, a difference considered to be small by Cohen [23], and of these only 10 were above 0.3 SD units.

The impact of differences among sites on outer and intermediate percentiles was minimal. The percentile curves depicting length from birth to 2 y for the pooled sample are nearly indistinguishable from those that result when particular sites are excluded, as illustrated by Figures 3 and 4. These figures show the impact on various percentiles of excluding the two sites with the most divergent linear growth.

Among the most salient alternatives to using all sites for the purpose of developing a single international standard is to exclude a site or sites and/or adjust for other available measurements, e.g. maternal and/or paternal stature. The former would further reduce inter-site variability and regional representation and the latter inter-individual variability. Considering that the standard will be promoted for use worldwide, neither option is compelling technically or from a policy point of view.

Differences among sites were not consistent across the ages examined. This likely reflects relatively small age-specific sample sizes at each site, residual secular trends among sites, and possibly true inter-ethnic differences and inter-site differences in the implementation of the study protocol, despite the standardization efforts described elsewhere [20]. Most importantly, however, observed inconsistencies are relatively minor and are likely of little, if any, practical and/or clinical importance. Furthermore, the

Table V. Pooled and individual site exclusion sample sizes (*n*), means (P50), standard deviations (SD), 3rd percentiles (P3) and 97th percentiles (P97) for length (cm).

Age	Sample	*n*	Mean	SD	SSE P50 (SDs)[a]	P3	SSE P3 (SDs)[a]	P97	SSE P97 (SDs)[a]
Birth	Pooled	1742	49.55	1.91	0.00	46.10	0.00	53.14	0.00
	Excluding Brazil	1433	49.54	1.91	−0.01	46.10	0.00	53.15	0.01
	Excluding Ghana	1413	49.57	1.90	0.01	46.10	0.00	53.15	0.01
	Excluding India	1441	49.67	1.91	0.06	46.20	0.05	53.20	0.03
	Excluding Norway	1442	49.37	1.87	−0.09	46.01	−0.05	53.04	−0.05
	Excluding Oman	1447	49.63	1.93	0.04	46.10	0.00	53.15	0.01
	Excluding USA	1534	49.52	1.90	−0.01	46.15	0.03	53.05	−0.05
6 mo	Pooled	1648	66.72	2.35	0.00	62.32	0.00	71.25	0.00
	Excluding Brazil	1352	66.71	2.36	0.00	62.23	−0.04	71.20	−0.02
	Excluding Ghana	1342	66.75	2.37	0.01	62.36	0.02	71.25	0.00
	Excluding India	1361	66.75	2.37	0.01	62.25	−0.03	71.25	0.00
	Excluding Norway	1362	66.47	2.28	−0.10	62.19	−0.05	70.65	−0.25
	Excluding Oman	1374	66.85	2.39	0.05	62.45	0.05	71.47	0.09
	Excluding USA	1449	66.78	2.34	0.02	62.37	0.02	71.30	0.02
12 mo	Pooled	1594	75.02	2.62	0.00	70.24	0.00	79.92	0.00
	Excluding Brazil	1304	74.94	2.60	−0.03	70.05	−0.07	79.75	−0.07
	Excluding Ghana	1293	74.99	2.61	−0.01	70.25	0.00	80.05	0.05
	Excluding India	1315	75.03	2.64	0.00	70.07	−0.06	79.90	−0.01
	Excluding Norway	1322	74.93	2.63	−0.04	70.23	0.00	79.80	−0.05
	Excluding Oman	1329	75.14	2.65	0.04	70.25	0.00	80.16	0.09
	Excluding USA	1407	75.09	2.60	0.03	70.25	0.00	80.09	0.06
18 mo	Pooled	1535	81.76	2.90	0.00	76.30	0.00	87.25	0.00
	Excluding Brazil	1250	81.62	2.86	−0.05	76.12	−0.06	86.95	−0.10
	Excluding Ghana	1242	81.72	2.91	−0.01	76.30	0.00	87.25	0.00
	Excluding India	1267	81.82	2.90	0.02	76.45	0.05	87.25	0.00
	Excluding Norway	1280	81.70	2.92	−0.02	76.17	−0.05	87.25	0.00
	Excluding Oman	1276	81.94	2.90	0.06	76.55	0.09	87.39	0.05
	Excluding USA	1360	81.77	2.88	0.00	76.30	0.00	87.21	−0.01
24 mo	Pooled	1524	87.40	3.18	0.00	81.18	0.00	93.50	0.00
	Excluding Brazil	1244	87.19	3.15	−0.07	81.06	−0.04	93.25	−0.08
	Excluding Ghana	1235	87.38	3.22	−0.01	81.10	−0.03	93.50	0.00
	Excluding India	1255	87.48	3.19	0.03	81.20	0.00	93.52	0.01
	Excluding Norway	1267	87.33	3.20	−0.02	81.10	−0.03	93.50	0.00
	Excluding Oman	1264	87.61	3.16	0.07	81.60	0.13	93.60	0.03
	Excluding USA	1355	87.40	3.17	0.00	81.23	0.01	93.47	−0.01

[a] Standardized site effects (SSE) are the differences between the indicated site means and the corresponding pooled (all sites) mean divided by the pooled standard deviation.

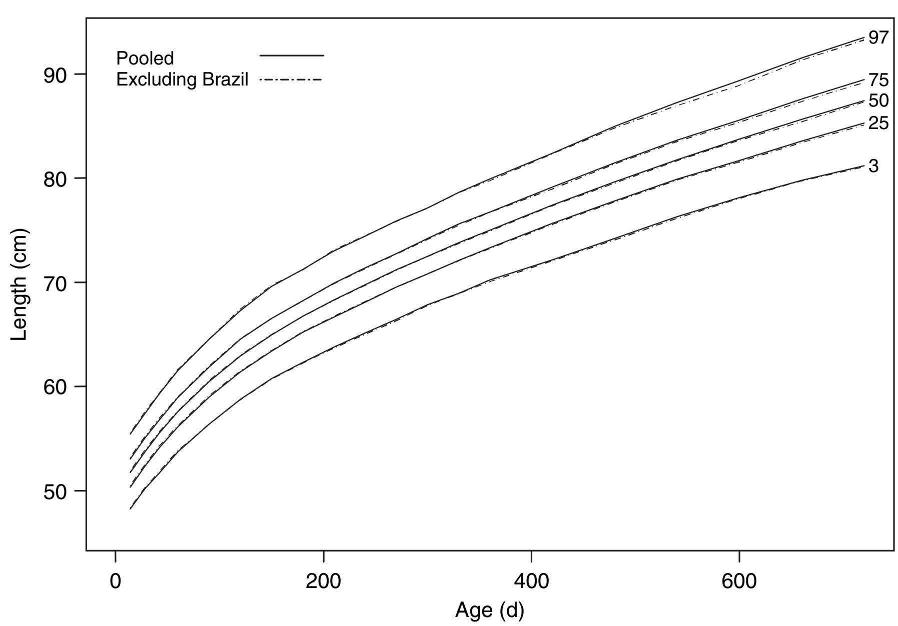


Figure 3. Length (cm) at selected percentiles for the pooled sample (solid line) and the sample following the exclusion of Brazil (dashed lines) from birth to 730 d.

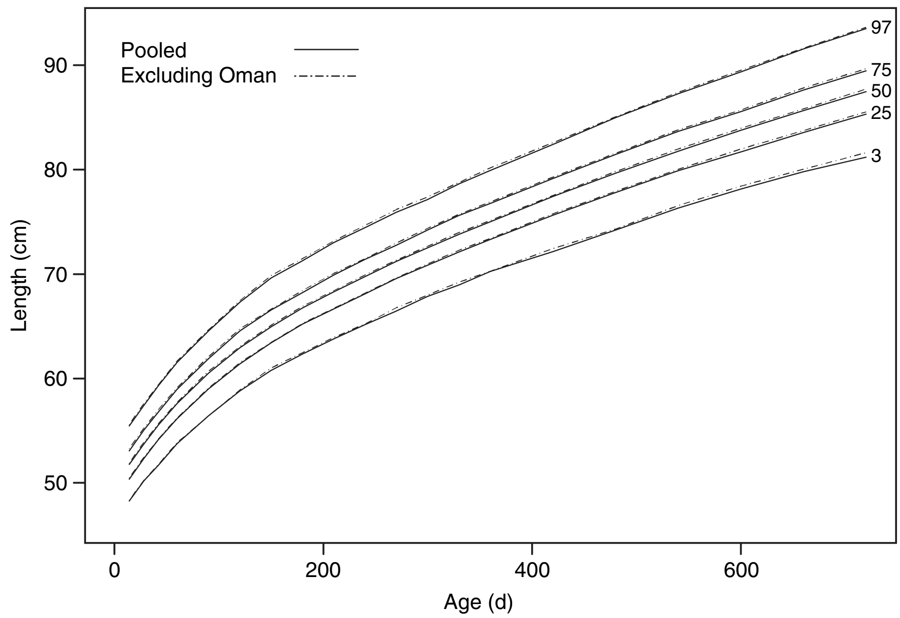


Figure 4. Length (cm) at selected percentiles for the pooled sample (solid line) and the sample following the exclusion of Oman (dashed lines) from birth to 730 d.

alternatives seem unworkable given existing ethnic diversity within countries and the evolution towards increasingly multiracial societies in the Americas and Europe as elsewhere in the world. Neither is it evident how one would adjust for children of mixed ethnicities.

Table VI. Pooled and individual site exclusion sample sizes (*n*), means (P50), standard deviations (SD), 3rd percentiles (P3) and 97th percentiles (P97) for height (cm).

Age	Sample	*n*	Mean	SD	SSE P50 (SDs) [a]	P3	SSE P3 (SDs) [a]	P97	SSE P97 (SDs) [a]
24–26 mo	Pooled	484	87.36	3.54	0.00	84.80	0.00	89.84	0.00
	Excluding Brazil	399	87.03	3.58	−0.09	84.38	−0.12	89.30	−0.15
	Excluding Ghana	406	87.41	3.62	0.02	84.60	−0.06	89.95	0.03
	Excluding India	386	87.44	3.41	0.02	84.98	0.05	90.07	0.06
	Excluding Norway	349	87.37	3.61	0.01	84.79	0.00	89.75	−0.02
	Excluding Oman	396	87.53	3.49	0.05	85.26	0.13	90.05	0.06
	Excluding USA	484	87.36	3.54	0.00	84.80	0.00	89.84	0.00
36–38 mo	Pooled	502	96.26	4.04	0.00	93.47	0.00	98.97	0.00
	Excluding Brazil	411	95.90	3.95	−0.09	93.06	−0.10	98.68	−0.07
	Excluding Ghana	417	96.25	4.06	0.00	93.45	0.00	99.06	0.02
	Excluding India	416	96.44	3.96	0.04	93.75	0.07	99.08	0.03
	Excluding Norway	432	96.20	4.11	−0.02	93.36	−0.03	98.93	−0.01
	Excluding Oman	419	96.46	4.05	0.05	93.87	0.10	99.08	0.03
	Excluding USA	415	96.33	4.07	0.02	93.56	0.02	99.05	0.02
48–50 mo	Pooled	478	103.52	4.23	0.00	100.68	0.00	106.26	0.00
	Excluding Brazil	407	103.28	4.08	−0.06	100.45	−0.05	106.18	−0.02
	Excluding Ghana	384	103.33	4.13	−0.04	100.56	−0.03	105.74	−0.12
	Excluding India	402	103.55	4.31	0.01	100.75	0.02	106.26	0.00
	Excluding Norway	408	103.50	4.33	0.00	100.56	−0.03	106.27	0.00
	Excluding Oman	398	103.87	4.14	0.08	101.18	0.12	106.50	0.06
	Excluding USA	391	103.57	4.38	0.01	100.53	−0.04	106.38	0.03
60–62 mo	Pooled	465	110.32	4.86	0.00	107.37	0.00	112.80	0.00
	Excluding Brazil	374	110.11	4.82	−0.04	107.04	−0.07	112.68	−0.02
	Excluding Ghana	389	109.88	4.49	−0.09	106.87	−0.10	112.32	−0.10
	Excluding India	395	110.59	5.01	0.06	107.49	0.03	113.18	0.08
	Excluding Norway	395	110.26	4.98	−0.01	107.03	−0.07	112.91	0.02
	Excluding Oman	392	110.56	4.97	0.05	107.49	0.03	113.14	0.07
	Excluding USA	380	110.49	4.86	0.04	107.51	0.03	113.06	0.05

[a] Standardized site effects (SSE) are the differences between the indicated site means and the corresponding pooled (all sites) mean divided by the pooled standard deviation.

In conclusion, these analyses document the strong similarity in linear growth from birth to 5 y in major ethnic groups living under relatively affluent conditions. They also support the inclusion of all six MGRS sites for the purpose of constructing a single international standard. The limitations of applying a prescriptive approach to free-living subjects and those imposed by a community-based sampling strategy likely preclude an error-free description of ideal growth patterns. Yet, despite those limitations and the marked differences among study sites in population and environmental characteristics, the similarity in linear growth among sites is striking. Most importantly, a single international standard for assessing the growth of all children embodies the very powerful message that when health and key environmental needs are met, the world's children grow very similarly.

The growth curves based on the pooled MGRS data for length/height-for-age, weight-for-age, weight-for-length/height and body mass index-for-age are presented in a companion paper in this supplement [28]. They represent the best description of physiological growth and should be applied to all children everywhere, regardless of ethnicity, socio-economic status and type of feeding.

Acknowledgements

This paper was prepared by Cutberto Garza, Mercedes de Onis, Reynaldo Martorell, Adelheid W. Onyango, Cesar G. Victora, Anna Lartey, Maharaj K. Bhan, Gunn-Elin A. Bjoerneboe, Deena Alasfoor, Kathryn G. Dewey, Edward A. Frongillo and Jose Martines on behalf of the WHO Multicentre Growth Reference Study Group. The statistical analysis was conducted by Elaine Borghi.

References

[1] WHO. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. Technical Report Series No. 854. Geneva: World Health Organization; 1995.

[2] de Onis M, Habicht JP. Anthropometric reference data for international use: recommendations from a World Health Organization Expert Committee. Am J Clin Nutr 1996;64: 650–8.

[3] de Onis M, Garza C, Habicht JP. Time for a new growth reference. Pediatrics 1997;100:E8.

[4] World Health Assembly. Resolution WHA47.5. Infant and young child nutrition. Geneva: World Health Organization; 1994.

[5] Garza C, de Onis M, for the WHO Multicentre Growth Reference Study Group. Rationale for developing a new international growth reference. Food Nutr Bull 2004;25 Suppl 1:S5–14.

[6] de Onis M, Garza C, Victora CG, Onyango AW, Frongillo EA, Martines J, for the WHO Multicentre Growth Reference Study Group. The WHO Multicentre Growth Reference Study: Planning, study design, and methodology. Food Nutr Bull 2004;25 Suppl 1:S15–26.

[7] 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.

[8] WHO Working Group on the Growth Reference Protocol and WHO Task Force on Methods for the Natural Regulation of Fertility. Growth patterns of breastfed infants in seven countries. Acta Paediatr 2000;89:215–22.

[9] Martorell R, Mendoza F, Castillo R. Poverty and stature in children. In: Waterlow JC, editor. Linear growth retardation in less developed countries. Nestlé Nutrition Workshop Series Vol. 14. New York: Raven Press; 1988. p. 57–73.

[10] Ulijaszek SJ. Ethnic differences in patterns of human growth in stature. In: Martorell R, Haschke F, editors. Nutrition and growth. Philadelphia: Lippincott-Williams and Wilkins; 2001. p. 1–20.

[11] Araujo CL, Albernaz E, Tomasi E, Victora CG, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in Brazil. Food Nutr Bull 2004;25 Suppl 1:S53–9.

[12] Lartey A, Owusu WB, Sagoe-Moses I, Gomez V, Sagoe-Moses C, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in Ghana. Food Nutr Bull 2004;25 Suppl 1:S60–5.

[13] Bhandari N, Taneja S, Rongsen T, Chetia J, Sharma P, Bahl R, et al., for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in India. Food Nutr Bull 2004;25 Suppl 1:S66–71.

[14] Baerug A, Bjoerneboe GE, Tufte E, Norum KR, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in Norway. Food Nutr Bull 2004;25 Suppl 1:S72–7.

[15] Prakash NS, Mabry RM, Mohamed AJ, Alasfoor D, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in Oman. Food Nutr Bull 2004;25 Suppl 1:S78–83.

[16] Dewey KG, Cohen RJ, Nommsen-Rivers LA, Heinig MJ, for the WHO Multicentre Growth Reference Study Group. Implementation of the WHO Multicentre Growth Reference Study in the United States. Food Nutr Bull 2004;25 Suppl 1:S84–9.

[17] Bhandari N, Bahl R, Taneja S, de Onis M, Bhan MK. Growth performance of affluent Indian children is similar to that in developed countries. Bull World Health Organ 2002;80:189–95.

[18] Owusu WB, Lartey A, de Onis M, Onyango AW, Frongillo EA. Factors associated with unconstrained growth among affluent Ghanaian children. Acta Paediatr 2004;93:1115–9.

[19] Mohamed AJ, Onyango AW, de Onis M, Prakash N, Mabry RM, Alasfoor DH. Socioeconomic predictors of unconstrained child growth in Muscat, Oman. East Mediterr Health J 2004;10:295–302.

[20] de Onis M, Onyango AW, Van den Broeck J, Chumlea WC, Martorell R, for the WHO Multicentre Growth Reference Study Group. Measurement and standardization protocols for anthropometry used in the construction of a new international growth reference. Food Nutr Bull 2004;25 Suppl 1:S27–36.

[21] Onyango AW, Pinol AJ, de Onis M, for the WHO Multicentre Growth Reference Study Group. Managing data for a multicountry longitudinal study: Experience from the WHO Multicentre Growth Reference Study. Food Nutr Bull 2004;25 Suppl 1:S46–52.

[22] Searle SR, Casella G, McCulloch CE. Variance components. New York: John Wiley and Sons; 1992.
[23] Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. New Jersey: Lawrence Erlbaum Associates; 1988. p. 24–7.
[24] King MC, Motulsky AG. Mapping human history. Science 2002;298:2342–3.
[25] Jorde LB, Wooding SP. Genetic variation, classification and'race'. Nat Genet 2004;36 Suppl 11:S28–33.
[26] Cooper RS, Kaufman JS, Ward R. Race and genomics. N Engl J Med 2003;348:1166–70.
[27] Tanner JM. A history of the study of human growth. Cambridge: Cambridge University Press; 1981.
[28] WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards based on length/height, weight and age. Acta Paediatr Suppl 2006;450:76–85.

Acta Pædiatrica, 2006; Suppl 450: 66–75

Taylor & Francis
Taylor & Francis Group

Assessment of sex differences and heterogeneity in motor milestone attainment among populations in the WHO Multicentre Growth Reference Study

WHO MULTICENTRE GROWTH REFERENCE STUDY GROUP[1,2]

[1]*Department of Nutrition, World Health Organization, Geneva, Switzerland, and* [2]*Members of the WHO Multicentre Growth Reference Study Group (listed at the end of the first paper in this supplement)*

Abstract
Aim: To assess the heterogeneity of gross motor milestone achievement ages between the sexes and among study sites participating in the WHO Multicentre Growth Reference Study (MGRS). *Methods:* Six gross motor milestones (sitting without support, hands-and-knees crawling, standing with assistance, walking with assistance, standing alone, and walking alone) were assessed longitudinally in five of the six MGRS sites, namely Ghana, India, Norway, Oman and the USA. Testing was started at 4 mo of age and performed monthly until 12 mo, and bimonthly thereafter until all milestones were achieved or the child reached 24 mo of age. Four approaches were used to assess heterogeneity of the ages of milestone achievement on the basis of sex or study site. *Results:* No significant, consistent differences in milestone achievement ages were detected between boys and girls, nor were any site–sex interactions noted. However, some differences among sites were observed. The contribution of inter-site heterogeneity to the total variance was <5% for those milestones with the least heterogeneous ages of achievement (hands-and-knees crawling, standing alone, and walking alone) and nearly 15% for those with the most heterogeneous ages of achievement (sitting without support, standing with assistance, and walking with assistance).

Conclusion: Inter-site differences, most likely due to culture-specific care behaviours, reflect normal development among healthy populations across the wide range of cultures and environments included in the MGRS. These analyses support the appropriateness of pooling data from all sites and for both sexes for the purpose of developing an international standard for gross motor development.

Key Words: *Gross motor milestones, longitudinal, motor skills, standards, young child development*

Introduction

The WHO Multicentre Growth Reference Study (MGRS) was designed to provide a description of the physical growth and gross motor development in healthy infants and children throughout the world. Previous efforts to develop growth references relied on data collected from infants and young children "free from disease" who were representative of defined geographical areas. When appropriately carried out, such studies provide accurate snapshots of how children grow and/or develop in a particular time and place. The MGRS, however, adopted a prescriptive approach designed to describe how children *should* grow independently of time and place. In so doing, it defined health not only as the absence of disease but also as the adoption of healthy practices known to promote health, e.g. breastfeeding. The rationale, design and protocol for the MGRS have been described in detail elsewhere [1,2].

The second unique feature of the MGRS is that it included children from many of the world's major regions: Brazil (South America), Ghana (Africa), India (Asia), Norway (Europe), Oman (the Middle East) and the USA (North America). This design feature tested the assertion that growth in infancy and early childhood is very similar among diverse ethnic groups when conditions that favour growth are met [1]. The MGRS also offered an opportunity to assess the heterogeneity/similarity in gross motor development across distinct cultures and environments.

Undoubtedly, MGRS participants from diverse sites differed genetically; however, it is unlikely that functions and traits such as motor development

Correspondence: Mercedes de Onis, Study Coordinator, Department of Nutrition, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland. Tel: +41 22 791 3320. Fax: +41 22 791 4156. E-mail: deonism@who.int

ISSN 0803-5326 print/ISSN 1651-2227 online
DOI: 10.1080/08035320500495530

and linear growth, which reflect the coordinated expression of multiple genes, differ substantially and systematically among large populations living in healthy environments. At the population level, it is likely that environmental disparities such as those seen in developing countries influence phenotypic expressions of multigenic functions and traits to a greater extent than genetic differences do [3].

The literature provides only a limited basis on which to directly evaluate how these views relate to motor development. A number of studies have considered relationships between general nutritional or specific nutrient status [4–9], feeding mode in early infancy [10,11], and specific disease states or conditions [12,13] and motor development. Some have examined differences in motor development among diverse cultural or ethnic groups in healthy and unhealthy states [14–18]. These interests are not new. For example, Garcia-Coll [19] reviewed early papers that evaluated potential aetiologies of the putative motoric precocity of African American infants and infants of African descent in developing and developed countries [19–23]. Clearly, there are significant difficulties associated with isolating biological from caretaker socio-economic and attitudinal/behavioural influences. Complexities such as these thus make it difficult to interpret results of evaluations of the role that ethnicity and culture play in motor development [19,24].

Although the literature includes a discussion of differences in motor development between boys and girls [16,18,25,26], findings are inconsistent in that either no differences are found between boys and girls, or boys are observed to be either more delayed or at risk of being delayed when faced with various forms of stress. Apparently, no study has evaluated potential interactions among sex, ethnicity and cultural background when assessing motor development in young children.

The aim of this paper is to assess the heterogeneity between the sexes and among MGRS study sites of gross motor milestone achievement ages. Analyses are carried out to evaluate the need for distinct standards for boys and girls and the appropriateness of pooling observations from all MGRS sites that performed motor development assessments.

Methods

General study design

The rationale, planning, design and methods of the MGRS, including its motor development component and site-specific protocol implementation, have been described in detail elsewhere [1,2,27].

Six distinct gross motor milestones were assessed: sitting without support, hands-and-knees crawling, standing with assistance, walking with assistance, standing alone, and walking alone. These were selected because they are considered universal, fundamental to the acquisition of self-sufficient locomotion, and simple to test and evaluate. These milestones were assessed longitudinally beginning at 4 mo of age on all children enrolled in the longitudinal sample in five of the six MGRS sites, namely Ghana, India, Norway, Oman and the USA. Motor development was not assessed in Brazil because most of that site's longitudinal sample was older than 4 mo when motor development was added to the MGRS protocol.

Using standardized testing procedures and criteria, study staff performed monthly assessments until 12 mo of age and bimonthly assessments thereafter until all milestones were achieved or the child reached 24 mo of age. No fixed milestone sequence was assumed and all milestones were assessed at each visit. Training and standardization procedures and data collection protocols, described in detail elsewhere [27,28], were similar among sites.

Sample used for analyses

Analyses of differences between the sexes or among sites in age of motor milestone achievement were based on the same sample of children included in assessments of inter-site heterogeneity in linear growth [29]. In the five study sites where motor development was assessed, 1433 children were enrolled in the MGRS longitudinal component. Because of missing data, 149 (10%) of these children were not included in the assessment of inter-site heterogeneity for linear growth. Of the children ($n = 1284$) included in the linear growth assessment, 75 (5%) did not participate in the MGRS motor development assessment component.

Variable numbers of motor milestone assessments by trained MGRS personnel were available for individual children in the remaining sample ($n = 1209$, 85%). This was mainly the result of late initiation of this MGRS component at the Norwegian and Ghanaian sites due to funding constraints, which meant that some children were too old to participate fully in motor assessments.

Statistical analyses

Estimation of ages of motor milestone achievement. The MGRS design [2] did not permit the determination of exact ages of milestone achievement because subjects were not supervised daily by trained staff. "True" ages of milestone achievement were linked to intervals between visits by staff documenting the first observed achievements of specific milestones and the most recent previous visit. Specific ages of achievement within those designated intervals were assigned randomly based on the assumption that achievement ages were distributed uniformly between scheduled visits. Detailed descriptions of the uses of fieldworker observations and caretaker reports of achievement ages are described in a companion paper in this supplement [30].

Evaluation of heterogeneity of milestone achievement ages between the sexes and among the MGRS sites. Two model-based approaches were used to characterize inter-site and inter-sex heterogeneity of the ages of milestone achievement.

A within-subject design ANOVA was used to assess proportional contributions of sex and site, both as main effects, to the total observed variation in ages of achievement of motor milestones and to evaluate site-sex interactions [31].

Another model-based approach applied a three-level variance components model (level 1: milestone indicator; level 2: individual child; and level 3: site). This model treated milestone achievement ages as successive occasions, assumed that achievement ages equalled fixed effects, and allowed for random perturbation on the normal scale [32]. To account for inter-level heterogeneity, a random effect was assigned to each clustering level. The percentages of the total variance attributable to each clustering level were calculated as fractions of the total variance [32,33]. Log-likelihood ratio was used to test the significance of sources of heterogeneity [32].

We also evaluated the magnitude of differences in ages of achievement of specific milestones between the sexes and among sites by calculating differences between the pooled mean age of achievement and the means for either sex or single sites as fractions of the pooled mean's standard deviation, i.e.

$$\frac{Y_A - Y}{SD} = Diff$$

where Y_A is the mean for *site A* or *sex A*, Y is the pooled mean, and SD is the standard deviation of the respective age of achievement corresponding to the pooled sample.

Site-specific and all-site average differences (in days) between boys' and girls' ages of achievement for each milestone were also calculated, and two-sample *t*-tests were performed to assess site- and milestone-specific differences in motor milestone achievement ages between boys and girls.

Lastly, the impact of inter-site heterogeneity was assessed further by evaluating the impact of excluding individual sites on percentile estimates. Differences were calculated between the 1st, 50th and 99th percentiles corresponding to "all-site" pooled values and the values calculated when single sites were individually omitted. Normalized differences were expressed as fractions of the standard deviations of the all-site pooled means.

Statistical significance was assigned to comparisons with p-values <0.05.

Results

Statistically significant differences in milestone achievement ages were not detected between boys and girls, nor were significant site-sex interactions noted (Table I) when a within-subject design ANOVA was applied. Figure 1 summarizes site-specific and overall differences in the ages of motor milestone achievement between boys and girls.

Two-sample *t*-tests assessing site- and motor milestone-specific differences between boys' and girls' ages of achievement detected statistically significant differences in five of 30 comparisons (Table II), namely sitting without support in India, walking with assistance in the USA, standing alone in Oman, and walking alone in Ghana and Oman. For all sites, statistically significant differences in the ages of achievement between boys and girls were detected for sitting without support (mean difference <5 d earlier for girls) and standing alone (mean difference of approximately 7 d earlier for girls).

Table I. Analysis of variance comparing the effect of sex, site and their interaction on milestone achievement ages.

Source of variation	Partial sum of squares	Degrees of freedom	p-value (prob >F)	Proportion of variance (%)
Among subjects:				
Site	1 119 723.3	4	0.0000	2.61
Sex	8626.0	1	0.2756	0.02
Interaction (site, sex)	50 649.9	4	0.1374	0.12
Residual (inter-subject)	8 686 429.2	1,198		20.22
Within subjects:				
Milestone	26 262 996.5	5	0.0000	61.12
Residual (intra-subject)	6 101 140.7	5,771		14.20
Total	42 970 018	6,983		100.00

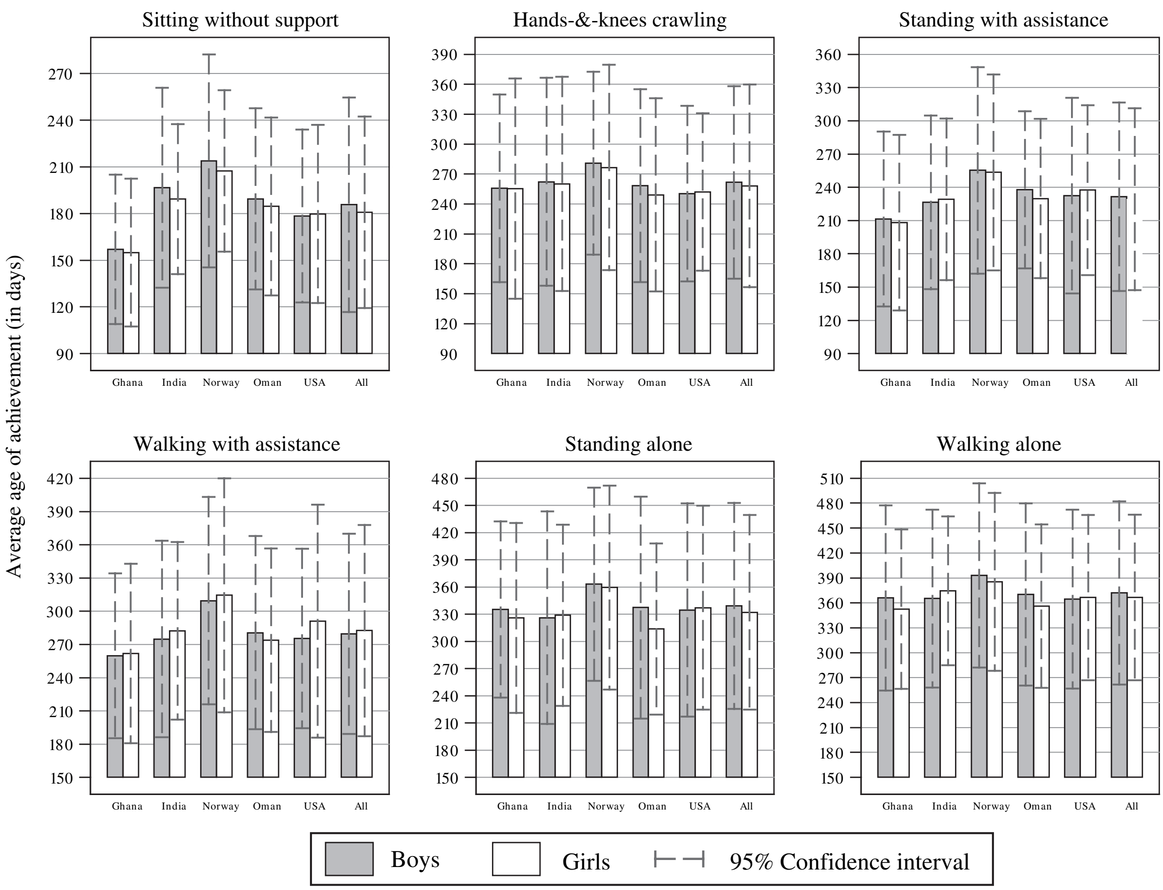


Figure 1. Average ages of gross motor milestone achievement in boys and girls.

Sitting without support exhibited the statistically most significant difference ($p = 0.0125$) in ages of achievement between boys and girls when all sites were pooled. Figure 2 illustrates the cumulative frequencies of the ages of achievement of sitting without support for boys and girls separately.

Small, though statistically significant, differences were observed among sites (sites accounted for 2.6% of the observed variance in Table I). Table III characterizes heterogeneity, by milestone, in the ages of milestone achievement. Ages of achievement for sitting without support demonstrated the greatest heterogeneity among sites. The least heterogeneity was observed for hands-and-knees crawling, standing alone and walking alone. *P*-values of log-likelihood ratio testing the significance of variance components due to site heterogeneity were <0.05. With the exception of standing alone ($p = 0.0298$), no evidence of heterogeneity due to sex, and no interaction of site and sex, were observed.

Estimates of the proportion of the total variance contributed by inter-site heterogeneity and inter-individual differences are summarized in Table IV. Inter-site heterogeneity contributed the least to the total variance (8.3%). Table IV also summarizes the contributions of inter-site heterogeneity to total variance when milestones with the greatest and

Table II. *P*-values of the two-sample *t*-tests on the equality of means between boys and girls.

Site	Sitting without support	Hands-and-knees crawling	Standing with assistance	Walking with assistance	Standing alone	Walking alone
Ghana	0.4665	0.9614	0.4885	0.6831	0.1377	0.0376*
India	0.0423*	0.7579	0.5608	0.1582	0.6988	0.1304
Norway	0.1730	0.5437	0.7861	0.4570	0.6073	0.2865
Oman	0.1781	0.1303	0.0798	0.2089	0.0008*	0.0371*
USA	0.7591	0.7860	0.4326	0.0348*	0.7718	0.8135
Total	0.0125*	0.2254	0.3900	0.3184	0.0297*	0.0654

*Statistically significant ($p < 0.05$).

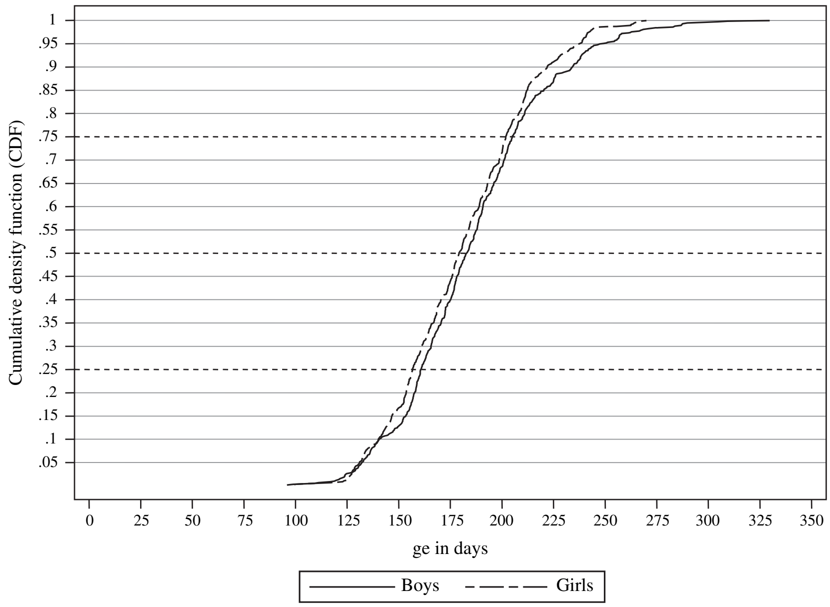


Figure 2. Cumulative frequency of motor achievement of sitting without support for boys and girls.

least heterogeneity were grouped. The contribution of inter-site heterogeneity to the total variance was <5% for those milestones with the least heterogeneous ages of achievement (hands-and-knees crawling, standing alone, and walking alone) and nearly 15% for those with the most heterogeneous ages of achievement (sitting without support, standing with assistance, and walking with assistance). P-values of log-likelihood ratios testing the significance of variance components due to site heterogeneity were <0.05. No evidence of heterogeneity due to sex or significant interaction of site and sex was observed.

Site-specific mean achievement ages and pooled means are presented in Table V. Normalized differences (expressed as fractions of the standard deviation of the pooled means) between site-specific means and the pooled mean varied by milestone. The Ghanaian sample exhibited the earliest mean ages of achievement for sitting without support (−0.82), standing with assistance (−0.49), walking with assistance (−0.43) and walking alone (−0.19). Normalized differences for all other sites with mean ages of achievement below the all-site pooled mean ranged from −0.17 to −0.05.

The Norwegian sample exhibited the latest mean ages of achievement for all six milestones (Table V). Normalized differences for all other sites with mean ages of achievement greater than the all-site pooled mean varied from 0.01 to 0.29.

Table VI summarizes the impact of eliminating single sites on the mean, 1st, 50th and 99th age of achievement percentiles. The impact of site elimination was assessed by comparing the "single-site elimination" values with "all-site" pooled values. Excluding the Ghanaian site increased the remaining site pooled mean (and corresponding percentiles) for sitting without support, standing with assistance,

Table III. Variance components two-level model comparing site heterogeneity by milestone.

Milestone	Variance component[a]	Estimate	Standard error (estimate)	p-value	Proportion of variance (%)
Sitting without support	Var(Site)	438.4	279.5	<0.000	34.8
	Var(Error)	823.1	33.6		65.2
Hands-and-knees crawling	Var(Site)	87.1	61.7	<0.000	3.5
	Var(Error)	2382.1	99.1		96.5
Standing with assistance	Var(Site)	255.8	166.1	<0.000	13.9
	Var(Error)	1584.8	64.6		86.1
Walking with assistance	Var(Site)	289.5	188.5	<0.000	12.8
	Var(Error)	1976.8	80.8		87.2
Standing alone	Var(Site)	177.2	120.4	<0.000	5.5
	Var(Error)	3042.3	125.1		94.5
Walking alone	Var(Site)	123.1	85.5	<0.000	4.3
	Var(Error)	2776.4	114.4		95.7

[a]"Site" as a random effect.

walking with assistance and walking alone by 9, 7, 6 and 3 d, respectively. Excluding Norway decreased the remaining site pooled mean for all six milestones by 5, 4, 5, 7, 6 and 5 d, respectively. As absolute values, these differences represent less than 0.3 of the pooled mean's SD for all estimated differences. Of the 30 "single-site exclusion" means calculated for all milestones, 23 differed from the pooled mean by ≤ 0.1 of the all-site pooled mean's SD; six were between 0.1 and 0.2, and one was between 0.2 and 0.3.

Discussion

These findings support the conclusion that MGRS gross motor development data from female and male infants and toddlers should be pooled for the purpose of constructing standards. The statistical insignificance of sex as a source of variability in the ages of milestone achievement that is documented in Tables I, III and IV is underscored by Figure 2.

This view is justified despite sporadic statistically significant differences in the ages of motor milestone achievement between boys and girls when two-sample *t*-tests were applied (Table II). These differences were small, i.e. 7 d or less, and inconsistent. Also, they should be interpreted cautiously given that the study's large sample size and the large number of two-sample *t*-tests performed increase the possibility of alpha errors. As reported in other studies [25,26], girls in the MGRS tended to achieve milestones at earlier ages than did boys. The tendency of girls to achieve motor milestones earlier than boys observed in Figure 1 is of interest from a developmental perspective; however, the magnitude of observed differences is too small to justify sex-specific norms.

The absence of any site–sex interaction is also reassuring. Its absence discounts the possibility that boys and girls were treated differentially in diverse sites in a manner that operated across sites to obscure sex-based differences. The paucity of other information evaluating differences in gross motor development between male and female infants and toddlers raised in diverse cultural settings and environments makes this finding particularly valuable to the construction of an international standard. These findings also support the view that any disparities between boys and girls in gross motor development likely reflect dissimilarities in care practices and/or other factors, which is to say that it is unlikely they are due to physiological sex-based differences.

These analyses found statistically significant inter-site differences in the ages of motor milestone achievement. This finding is generally consistent with another WHO collaborative study designed to develop and standardize culturally appropriate scales of psychosocial development [18]. That study included a wide array of developmental assessments. Although specific tests of inter-site differences were not included in the cited reference, tabulated information documents homogeneity in ages of achievement among some milestones but not among others. These findings suggest that environmental diversity may have accounted for the lack of homogeneity across all measures, which is consistent with observations made by others. For example, Lima et al. [25] reported that environments influence mental and motor development to a much greater degree than do biological factors (e.g. birthweight).

Analyses summarized in Table I indicate that sites contributed $<3\%$ of the variability observed in the MGRS. This estimate merits close examination. The variability and error introduced by the random point determination of ages of milestone achievement and the likelihood of uneven susceptibility of different milestones to caretaker influences (discussed further below) may have decreased the proportional contribution of inter-site differences. The most important challenge presented by statistically significant inter-site differences and considerations of the determinants of variability is assessing their implications for the purpose of constructing an international standard. Three aspects of the analyses addressed this point. The first assessed the magnitude of differences among

Table IV. Variance components three-level model comparing site heterogeneity by milestones combined.

Milestones grouped	Variance component [a]	Estimate	Standard error (estimate)	p-value	Proportion of variance (%)
All six milestones	Var(Site)	192.4	125.0	<0.000	8.3
	Var(Child)	1067.3	50.9	<0.000	46.1
	Var(Error)	1057.6	19.4		45.6
Sitting without support, standing with assistance, walking with assistance	Var(Site)	248.7	159.9	<0.000	14.5
	Var(Child)	690.7	39.5	<0.000	40.1
	Var(Error)	781.6	22.5		45.4
Hands-and-knees crawling, standing alone, walking alone	Var(Site)	129.6	87.5	<0.000	4.5
	Var(Child)	1701.9	85.1	<0.000	59.1
	Var(Error)	1046.6	31.0		36.4

[a]"Site" as a random effect.

Table V. Site-specific and "all-site" achievement ages (in days) by milestone.

	n	Mean	SD	Diff. in SD		*n*	Mean	SD	Diff. in SD
Sitting without support					Hands-and-knees crawling				
Pooled estimate	1139	183.3	33.4	0.00	Pooled estimate	1128	260.0	50.4	0.00
Estimate for Ghana	280	156.0	24.1	−0.82	Estimate for Ghana	261	255.7	51.7	−0.09
Estimate for India	262	193.1	29.0	0.29	Estimate for India	244	261.1	53.3	0.02
Estimate for Norway	173	210.8	30.8	0.82	Estimate for Norway	203	278.8	48.8	0.37
Estimate for Oman	258	187.1	29.2	0.12	Estimate for Oman	255	253.9	49.1	−0.12
Estimate for USA	166	179.1	28.3	−0.12	Estimate for USA	165	251.3	41.9	−0.17
Standing with assistance					Walking with assistance				
Pooled estimate	1169	230.5	42.6	0.00	Pooled estimate	1185	281.1	47.3	0.00
Estimate for Ghana	280	209.8	39.9	−0.49	Estimate for Ghana	278	260.9	39.3	−0.43
Estimate for India	262	227.8	38.3	−0.06	Estimate for India	262	278.6	42.9	−0.05
Estimate for Norway	203	254.5	45.7	0.56	Estimate for Norway	224	311.9	50.1	0.65
Estimate for Oman	258	234.0	36.2	0.08	Estimate for Oman	255	277.4	43.1	−0.08
Estimate for USA	166	235.0	41.4	0.11	Estimate for USA	166	283.3	47.8	0.05
Standing alone					Walking alone				
Pooled estimate	1182	335.6	56.4	0.00	Pooled estimate	1182	369.3	53.6	0.00
Estimate for Ghana	268	330.5	51.2	−0.09	Estimate for Ghana	266	359.2	52.8	−0.19
Estimate for India	262	327.4	55.2	−0.14	Estimate for India	261	369.7	50.1	0.01
Estimate for Norway	231	361.3	55.1	0.46	Estimate for Norway	236	389.3	55.1	0.37
Estimate for Oman	255	325.8	56.6	−0.17	Estimate for Oman	255	363.3	53.1	−0.11
Estimate for USA	166	335.9	57.7	0.01	Estimate for USA	164	365.4	52.0	−0.07

sites. As noted in the results section, the largest deviations from all-site pooled values were observed for Ghana and Norway. Those deviations were large in several instances, but neither Ghana nor Norway consistently accounted for the largest deviations (Table V).

Other analyses examined the consequences of specific single-site elimination on the resulting pooled means and selected percentiles. The greatest impact was observed when either Ghana or Norway was excluded from the sample. However, the exclusion of either country did not result consistently in the largest deviations from all-site pooled values. Also, as summarized in Table VI, the exclusion of any single site seldom resulted in normalized differences greater than 0.2 SD between corresponding means and the 1st, 50th and 99th centile values. Normalized differences most often were below 0.1 SD.

The contributions of inter-site differences to the total variability of specific milestones were also examined. Among the statistically significant sources of variation, sites contributed least to the variability in ages of achievement for hands-and-knees crawling (3.5%), standing alone (5.5%) and walking alone (4.3%). The most marked contribution to total variability by inter-site differences was observed for sitting without support (35%). Inter-site contributions to the total variability were intermediate in magnitude for the milestones standing with assistance (13.9%) and walking with assistance (12.8%).

Among the inferences that may be drawn from these differences is that developmental domains governing milestone achievement are influenced significantly by environmental and/or genetic factors specific to individual sites. Theories of motor development and skill acquisition and of genetic controls of development [34–37] make it unlikely that genetic factors linked to ethnicity determine the ability to sit without support to a greater extent than they do hands-and-knees crawling. The involvement of multiple gene networks seems unavoidable in the orchestration of anatomical, cognitive and other changes linked to development [38]. Thus, environmental influences appear to provide the more parsimonious explanation for observed differences. The two most relevant potential environmental influences relate to distinct gestational and/or perinatal conditions among participants and/or childcare practices in the various sites. It seems unlikely that unspecified gestational and/or perinatal site-specific conditions carry over only to the "earliest" motor milestone that was examined, but such possibilities cannot be discounted based on data collected by this study.

Although neither genetic nor environmental influences can be discounted completely as explanations for observed inter-site differences, inconsistencies within and among sites (e.g. children in Ghana did not always demonstrate the earliest ages of achievement for all milestones) and field observations suggest that childcare practices likely explain observed inter-site differences. As indicated earlier, inter-site differences were greatest between Ghana and Norway. Field reports indicate that Ghanaian caretakers commonly engaged in practices consistent with the training of infants so as to accelerate their achievement of motor milestones.

Table VI. Comparisons of achievement ages (days) by milestones when all sites are pooled and when single sites are excluded.

	n	Mean	SD	Diff. in SD	P1	Diff. in SD	P50	Diff. in SD	P99	Diff. in SD
Sitting without support										
Pooled estimate	1139	183.3	33.4	0.00	121.2	0.00	181.0	0.00	270.0	0.00
Excluding Ghana	859	192.2	31.1	0.27	127.3	0.18	190.2	0.28	282.9	0.39
Excluding India	877	180.3	34.1	−0.09	117.1	−0.12	177.4	−0.11	270.9	0.03
Excluding Norway	966	178.3	31.4	−0.15	118.9	−0.07	176.8	−0.12	265.4	−0.14
Excluding Oman	881	182.2	34.5	−0.03	117.1	−0.12	179.6	−0.04	268.5	−0.04
Excluding USA	973	184.0	34.2	0.02	122.8	0.05	181.6	0.02	274.6	0.14
Hands-and-knees crawling										
Pooled estimate	1128	260.0	50.4	0.00	169.4	0.00	254.2	0.00	410.4	0.00
Excluding Ghana	867	261.3	50.0	0.03	170.0	0.01	255.6	0.03	409.9	−0.01
Excluding India	884	259.7	49.6	−0.01	167.7	−0.03	254.4	0.00	415.2	0.10
Excluding Norway	925	255.8	49.9	−0.08	165.4	−0.08	251.1	−0.06	405.1	−0.11
Excluding Oman	873	261.7	50.7	0.04	167.7	−0.03	255.5	0.03	415.2	0.10
Excluding USA	963	261.5	51.6	0.03	170.0	0.01	255.3	0.02	417.1	0.13
Standing with assistance										
Pooled estimate	1169	230.5	42.6	0.00	153.1	0.00	227.0	0.00	351.5	0.00
Excluding Ghana	889	237.0	41.3	0.15	156.0	0.07	233.8	0.16	357.2	0.13
Excluding India	907	231.3	43.7	0.02	153.1	0.00	228.6	0.04	353.6	0.05
Excluding Norway	966	225.5	40.1	−0.12	150.2	−0.07	224.0	−0.07	340.9	−0.25
Excluding Oman	911	229.5	44.2	−0.02	150.2	−0.07	225.5	−0.04	353.6	0.05
Excluding USA	1003	229.8	42.7	−0.02	153.8	0.02	226.5	−0.01	351.5	0.00
Walking with assistance										
Pooled estimate	1185	281.1	47.3	0.00	190.6	0.00	275.4	0.00	423.7	0.00
Excluding Ghana	907	287.3	47.8	0.13	195.0	0.09	281.8	0.14	426.0	0.05
Excluding India	923	281.8	48.5	0.02	190.6	0.00	276.1	0.01	424.6	0.02
Excluding Norway	961	273.9	43.6	−0.15	189.8	−0.02	269.6	−0.12	406.1	−0.37
Excluding Oman	930	282.1	48.4	0.02	190.7	0.00	275.5	0.00	424.6	0.02
Excluding USA	1019	280.8	47.2	−0.01	190.6	0.00	275.1	−0.01	420.6	−0.06
Standing alone										
Pooled estimate	1182	335.6	56.4	0.00	230.7	0.00	329.9	0.00	491.0	0.00
Excluding Ghana	914	337.1	57.8	0.03	230.7	0.00	331.2	0.02	491.0	0.00
Excluding India	920	337.9	56.6	0.04	233.9	0.06	333.2	0.06	491.0	0.00
Excluding Norway	951	329.3	54.9	−0.11	221.6	−0.16	323.7	−0.11	486.0	−0.09
Excluding Oman	927	338.3	56.1	0.05	230.7	0.00	333.2	0.06	487.7	−0.06
Excluding USA	1016	335.5	56.2	0.00	232.3	0.03	329.7	0.00	491.0	0.00
Walking alone										
Pooled estimate	1182	369.3	53.6	0.00	256.8	0.00	361.2	0.00	517.0	0.00
Excluding Ghana	916	372.2	53.5	0.05	264.7	0.15	363.5	0.04	515.0	−0.04
Excluding India	921	369.2	54.6	0.00	256.7	0.00	360.1	−0.02	521.0	0.07
Excluding Norway	946	364.3	52.1	−0.09	255.8	−0.02	357.5	−0.07	513.6	−0.06
Excluding Oman	927	370.9	53.7	0.03	256.8	0.00	363.8	0.05	515.0	−0.04
Excluding USA	1018	369.9	53.9	0.01	257.1	0.01	361.9	0.01	517.0	0.00

For example, Ghanaian mothers often propped infants in a variety of ways to assist the infant's assumption of an upright sitting position. Norwegians, on the other hand, were encouraged by paediatric care norms not to push children to perform but to rely on a child's spontaneous interest and development, e.g. allowing infants to achieve an upright sitting position without assistance or prompting. The greater homogeneity in ages of achievement for milestones that require the most coordinated movements and control, namely hands-and-knees crawling and standing and walking alone, thus may be the least amenable to trainer "interference". However, this explanation merits further investigation.

Although the origins of inter-site heterogeneity in the ages of milestone achievement and differences in the degree of heterogeneity in the ages of achievement among the six milestones remain unclear, the implications of these analyses for the purposes of the MGRS appear straightforward. The ranges of observed ages of achievement amply document the variability of normal development in diverse cultural and environmental settings. Thus, given the health and environmental advantages inherent in the MGRS sample, pooling observations from all five sites appears to be the most appropriate manner to reflect the range of normal development. This and other considerations led to the formulation of "windows of achievement" for specific milestones [30] that reflect the range of ages of achievement of motor milestones observed in the MGRS population. For reasons described in a companion paper in this supplement [30], these "windows" were estimated conservatively (as bounded by the 1st to 99th percentile age interval for individual milestone achievement).

Lastly, consideration is given to the relative contributions of inter-site and inter-individual differences to the total variability in ages of milestone achievement. The detailed evaluations of the roles of inter-site and inter-individual differences summarized in Tables III and IV are particularly informative. Clearly, the heterogeneity in ages of milestone achievement differs markedly among milestones (Table III). We suggest that milestones with the most homogeneous ages of achievement are likely to provide the most robust assessments of inherent inter-site differences, i.e. those that are least influenced by caretaker behaviours. Partitioning of variability for milestones with the most homogenous ages of achievement (hands-and-knees crawling, standing alone, and walking alone) attributes approximately 4% of the total variability to site differences and approximately 60% to inter-individual differences. The remaining 36% is ascribed to other sources of variation and random error, a proportion likely to be inflated by the random point method of determining ages of achievement and the inability to partition out a reasonable estimate of intra-individual variability. The 15-fold difference in the proportional contributions of inter-site and inter-individual differences are consistent with estimates of human genetic variability across and within populations. Population genomic analyses suggest that 85 to 90% of genetic variation resides within populations, whereas approximately 10 to 15% resides among populations [38]. The likely multigenic control of motor development suggests that variability between and within populations should be distributed similarly.

In summary, since these analyses found only small and sporadic differences in ages of achievement of gross motor milestones due to sex, we conclude they are of no practical relevance to the construction of gross motor development standards. Similarly, no significant site–sex interactions were observed. Significant differences among sites, however, were observed. Inter-site differences most likely reflect factors related to culture-specific care behaviours, but the aetiology of those differences cannot be discerned adequately from these analyses. Most importantly, however, these differences reflect the range of normal development among healthy populations across the relatively wide range of cultures and environments included in the MGRS, and they provide a useful basis for assessing motor development in populations. Lastly, the relative contributions of between- and within-site variability to the total variability across all six milestones are consistent with the relative contributions of those sources of variability to the total variability in child length discussed in a companion paper in this supplement [29]. These analyses support the appropriateness of pooling data from all sites for the purposes of developing an international standard for the six motor development milestones assessed by the MGRS.

Acknowledgements

This paper was prepared by Cutberto Garza, Mercedes de Onis, Reynaldo Martorell, Kathryn G. Dewey and Maureen Black on behalf of the WHO Multicentre Growth Reference Study Group. The statistical analysis was conducted by Amani Siyam.

References

[1] Garza C, de Onis M, for the WHO Multicentre Growth Reference Study Group. Rationale for developing a new international growth reference. Food Nutr Bull 2004;25 Suppl 1:S5–14.

[2] de Onis M, Garza C, Victora CG, Onyango AW, Frongillo EA, Martines J, for the WHO Multicentre Growth Reference Study Group. The WHO Multicentre Growth Reference Study: Planning, study design and methodology. Food Nutr Bull 2004;25 Suppl 1:S15–26.

[3] Cooper RS, Kaufman JS, Ward R. Race and genomics. N Engl J Med 2003;348:1166–70.

[4] Kariger PK, Stoltzfus RJ, Olney D, Sazawal S, Black R, Tielsch JM, et al. Iron deficiency and physical growth predict attainment of walking but not crawling in poorly nourished Zanzibari infants. J Nutr 2005;135:814–9.

[5] Black MM, Baqui AH, Zaman K, Ake Persson L, El Arifeen S, Le K, et al. Iron and zinc supplementation promote motor development and exploratory behavior among Bangladeshi infants. Am J Clin Nutr 2004;80:903–10.

[6] Kuklina EV, Ramakrishnan U, Stein AD, Barnhart HH, Martorell R. Growth and diet quality are associated with the attainment of walking in rural Guatemalan infants. J Nutr 2004;134:3296–300.

[7] Lozoff B, De Andraca I, Castillo M, Smith JB, Walter T, Pino P. Behavioral and developmental effects of preventing iron-deficiency anemia in healthy full-term infants. Pediatrics 2003; 112:978.

[8] Black MM. The evidence linking zinc deficiency with children's cognitive and motor functioning. J Nutr 2003;133: 1473S–6S.

[9] Jahari AB, Saco-Pollitt C, Husaini MA, Pollitt E. Effects of an energy and micronutrient supplement on motor development and motor activity in undernourished children in Indonesia. Eur J Clin Nutr 2000;54 Suppl 2:S60–8.

[10] Dewey KG, Cohen RJ, Brown KH, Rivera LL. Effects of exclusive breastfeeding for four versus six months on maternal nutritional status and infant motor development: results of two randomized trials in Honduras. J Nutr 2001;131:262–7.

[11] Lucas A, Fewtrell MS, Morley R, Singhal A, Abbott RA, Isaacs E, et al. Randomized trial of nutrient-enriched formula versus standard formula for postdischarge preterm infants. Pediatrics 2001;108:703–11.

[12] Carrel AL, Moerchen V, Myers SE, Bekx MT, Whitman BY, Allen DB. Growth hormone improves mobility and body composition in infants and toddlers with Prader-Willi syndrome. J Pediatr 2004;145:744–9.
[13] Chase C, Ware J, Hittelman J, Blasini I, Smith R, Llorente A, et al. Early cognitive and motor development among infants born to women infected with human immunodeficiency virus. Pediatrics 2000;106:E25.
[14] Campbell SK, Hedeker D. Validity of the Test of Infant Motor Performance for discriminating among infants with varying risk for poor motor outcome. J Pediatr 2001;139:546–51.
[15] Treuth MS, Sherwood NE, Butte NF, McClanahan B, Obarzanek E, Zhou A, et al. Validity and reliability of activity measures in African-American girls for GEMS. Med Sci Sports Exerc 2003;35:532–9.
[16] Allen MC, Alexander GR. Gross motor milestones in preterm infants: correction for degree of prematurity. J Pediatr 1990; 116:955–9.
[17] Garn SM, Petzold AS, Ridella SA, Johnston M. Effect of smoking during pregnancy on Apgar and Bayley scores. Lancet 1980;2:912–3.
[18] Lansdown RG, Goldstein H, Shah PM, Orley JH, Di G, Kaul KK, et al. Culturally appropriate measures for monitoring child development at family and community level: a WHO collaborative study. Bull World Health Organ 1996;74:283–90.
[19] Garcia Coll CT. Developmental outcome of minority infants: a process-oriented look into our beginnings. Child Dev 1990; 61:270–89.
[20] Pasamanick B. A comparative study of the behavioral development of Negro infants. J Genet Psychol 1946;69:3–44.
[21] Williams JR, Scott RB. Growth and development of Negro infants: IV. Motor development and its relationship to child rearing practices in two groups of Negro infants. Child Dev 1953;24:103–21.
[22] Geber M, Dean RF. The state of development of newborn African children. Lancet 1957;272:1216–9.
[23] Super CM. Cross-cultural research on infancy. In: Triandis HC, Heros A, editors. Handbook of cross-cultural psychology: Developmental psychology, vol. 4. Boston: Allyn & Bacon; 1981. p. 17–53.
[24] Pachter LM, Dworkin PH. Maternal expectations about normal child development in 4 cultural groups. Arch Pediatr Adolesc Med 1997;151:1144–50.
[25] Lima MC, Eickmann SH, Lima AC, Guerra MQ, Lira PI, Huttly SR, et al. Determinants of mental and motor development at 12 months in a low income population: a cohort study in northeast Brazil. Acta Paediatr 2004;93:969–75.
[26] To T, Guttmann A, Dick PT, Rosenfield JD, Parkin PC, Cao H, et al. What factors are associated with poor developmental attainment in young Canadian children. Can J Public Health 2004;95:258–63.
[27] Wijnhoven TM, de Onis M, Onyango AW, Wang T, Bjoerneboe GE, Bhandari N, et al. for the WHO Multicentre Growth Reference Study Group. Assessment of gross motor development in the WHO Multicentre Growth Reference Study. Food Nutr Bull 2004;25 Suppl 1:S37–45.
[28] WHO Multicentre Growth Reference Study Group. Reliability of motor development data in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450:47–55.
[29] WHO Multicentre Growth Reference Study Group. Assessment of differences in linear growth among populations in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450:56–65.
[30] WHO Multicentre Growth Reference Study Group. WHO Motor Development Study: Windows of achievement for six gross motor development milestones. Acta Paediatr Suppl 2006;450:86–95.
[31] Kirk RE. Experimental design. 2nd ed. Belmont, CA: Brooks/Cole; 1992.
[32] Goldstein H. Multilevel statistical models. 2nd ed. Kendall's Library of Statistics 3. London: Arnold Publications; 1995.
[33] Johnson NL, Kotz S. Continuous univariate distributions. In: The Houghton Mifflin Series in Statistics, vol. 1. New York: John Wiley and Sons; 1970.
[34] Lockman JJ, Thelen E. Developmental biodynamics: brain, body, behavior connections. Child Dev 1993;64:953–9.
[35] Thelen E. Motor development. A new synthesis. Am Psychol 1995;50:79–95.
[36] Gibson G, Wagner G. Canalization in evolutionary genetics: a stabilizing theory? Bioessays 2000;22:372–80.
[37] Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh CH, et al. A genomic regulatory network for development. Science 2002;295:1669–78.
[38] Jorde LB, Wooding SP. Genetic variation, classification and 'race'. Nat Genet 2004;36 Suppl 11:S28–33.

Acta Pædiatrica, 2006; Suppl 450: 76–85

Taylor & Francis
Taylor & Francis Group

WHO Child Growth Standards based on length/height, weight and age

WHO MULTICENTRE GROWTH REFERENCE STUDY GROUP[1,2]

[1]*Department of Nutrition, World Health Organization, Geneva, Switzerland, and* [2]*Members of the WHO Multicentre Growth Reference Study Group (listed at the end of the first paper in this supplement)*

Abstract

Aim: To describe the methods used to construct the WHO Child Growth Standards based on length/height, weight and age, and to present resulting growth charts. *Methods:* The WHO Child Growth Standards were derived from an international sample of healthy breastfed infants and young children raised in environments that do not constrain growth. Rigorous methods of data collection and standardized procedures across study sites yielded very high-quality data. The generation of the standards followed methodical, state-of-the-art statistical methodologies. The Box-Cox power exponential (BCPE) method, with curve smoothing by cubic splines, was used to construct the curves. The BCPE accommodates various kinds of distributions, from normal to skewed or kurtotic, as necessary. A set of diagnostic tools was used to detect possible biases in estimated percentiles or z-score curves. *Results:* There was wide variability in the degrees of freedom required for the cubic splines to achieve the best model. Except for length/height-for-age, which followed a normal distribution, all other standards needed to model skewness but not kurtosis. Length-for-age and height-for-age standards were constructed by fitting a unique model that reflected the 0.7-cm average difference between these two measurements. The concordance between smoothed percentile curves and empirical percentiles was excellent and free of bias. Percentiles and z-score curves for boys and girls aged 0–60 mo were generated for weight-for-age, length/height-for-age, weight-for-length/height (45 to 110 cm and 65 to 120 cm, respectively) and body mass index-for-age.

Conclusion: The WHO Child Growth Standards depict normal growth under optimal environmental conditions and can be used to assess children everywhere, regardless of ethnicity, socio-economic status and type of feeding.

Key Words: *Body mass index, growth standards, height, length, weight*

Introduction

Nearly three decades ago, an expert group convened by the World Health Organization (WHO) recommended that the National Center for Health Statistics (NCHS) reference data for height and weight be used to assess the nutritional status of children around the world [1]. This recommendation was made recognizing that not all of the criteria the group used to select the best available reference data had been met. The reference became known as the NCHS/WHO international growth reference and was quickly adopted for a variety of applications regarding both individuals and populations.

The limitations of the NCHS/WHO reference are well known [2–5]. The data used to construct the reference covering birth to 3 y of age came from a longitudinal study of children of European ancestry from a single community in the United States. These children were measured every 3 mo, which is inadequate to describe the rapid and changing rate of growth in early infancy. Also, shortcomings inherent to the statistical methods available at the time for generating the growth curves led to inappropriate modelling of the pattern and variability of growth, particularly in early infancy. For these likely reasons, the NCHS/WHO curves do not adequately represent early childhood growth.

The origin of the WHO Multicentre Growth Reference Study (MGRS) [6] dates back to the early 1990s when the WHO initiated a comprehensive review of the uses and interpretation of anthropometric references and conducted an in-depth analysis of growth data from breastfed infants [2,7]. This analysis showed that breastfed infants from well-off households in northern Europe and North America (i.e. the WHO pooled breastfed data set) deviated negatively and significantly from the NCHS/WHO reference [2,7]. Moreover, healthy breastfed infants from Chile, Egypt, Hungary, Kenya and Thailand

Correspondence: Mercedes de Onis, Study Coordinator, Department of Nutrition, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland. Tel: +41 22 791 3320. Fax: +41 22 791 4156. E-mail: deonism@who.int

ISSN 0803-5326 print/ISSN 1651-2227 online
DOI: 10.1080/08035320500495548

showed similar deviations when compared to the NCHS/WHO reference but not when compared to the WHO pooled breastfed group [2]. Finally, the variability of growth in the pooled breastfed data set was significantly lower than that of the NCHS/WHO reference [2]. It was unclear whether the reduced variability reflected homogeneity in the WHO pooled breastfed group—perhaps because of uniformity in infant feeding patterns—or unphysiological variability in the NCHS/WHO reference. The data for infants used in the NCHS/WHO reference were collected between 1929 and 1975. The majority of these infants were fed artificial milks which, with increasing knowledge about the nutritional needs of infants, changed in formulation over time. It is thus possible that the greater variability in the current international reference reflects responses to formulas of varying nutritional quality over four decades.

The review group concluded from these and related findings that new references were necessary because the current international reference did not adequately describe the growth of children. Under these circumstances, its uses to monitor the health and nutrition of individual children or to derive population-based estimates of child malnutrition are flawed. The review group recommended a novel approach: that a standard rather than a reference be constructed. Strictly speaking, a reference simply serves as an anchor for comparison, whereas a standard allows both comparisons and permits value judgments about the adequacy of growth. The MGRS breaks new ground by describing how children *should* grow when not only free of disease but also when reared following healthy practices such as breastfeeding and a non-smoking environment.

The MGRS is also unique because it includes children from around the world: Brazil, Ghana, India, Norway, Oman and the USA. In a companion paper in this volume [8], the length of children is shown to be strikingly similar among the six sites, with only about 3% of variability in length being due to inter-site differences compared to 70% for individuals within sites. Thus, excluding any site has little effect on the 3rd, 50th and 97th percentile values, and pooling data from all sites is entirely justified. The striking similarity in growth during early childhood across human populations means either a recent common origin as some suggest [9] or a strong selective advantage across human environments associated with the current pattern of growth and development.

The key objectives of this article are 1) to provide an overview of the methods used to construct the standards for length/height-for-age, weight-for-age, weight-for-length/height and BMI-for-age, and 2) to present some of the resulting curves. Complete details and a full presentation of charts and tables pertaining to the standards are available in a technical report [10] and on the Web: www.who.int/childgrowth/en

Methods

Description and design of the MGRS

The MGRS (July 1997–December 2003) was a population-based study taking place in the cities of Davis, California, USA; Muscat, Oman; Oslo, Norway; and Pelotas, Brazil; and in selected affluent neighbourhoods of Accra, Ghana, and South Delhi, India. The MGRS protocol and its implementation in the six sites are described in detail elsewhere [6]. Briefly, the MGRS combined a longitudinal component from birth to 24 mo with a cross-sectional component of children aged 18–71 mo. In the longitudinal component, mothers and newborns were screened and enrolled at birth and visited at home a total of 21 times on weeks 1, 2, 4 and 6; monthly from 2–12 mo; and bimonthly in the second year. In the cross-sectional component, children aged 18–71 mo were measured once, except in the two sites (Brazil and the USA) that used a mixed-longitudinal design in which some children were measured two or three times at 3-mo intervals. Both recumbent length and standing height were measured for all children aged 18–30 mo. Data were collected on anthropometry, motor development, feeding practices, child morbidity, perinatal factors, and socio-economic, demographic and environmental characteristics [11].

The study populations lived in socio-economic conditions favourable to growth, where mobility was low, $\geq$20% of mothers followed WHO feeding recommendations and breastfeeding support was available [11]. Individual inclusion criteria were: no known health or environmental constraints to growth, mothers willing to follow MGRS feeding recommendations (i.e. exclusive or predominant breastfeeding for at least 4 mo, introduction of complementary foods by 6 mo of age, and continued partial breastfeeding to at least 12 mo of age), no maternal smoking before and after delivery, single term birth, and absence of significant morbidity [11].

As part of the site-selection process in Ghana, India and Oman, surveys were conducted to identify socio-economic characteristics that could be used to select groups whose growth was not environmentally constrained [12–14]. Local criteria for screening newborns, based on parental education and/or income levels, were developed from those surveys. Pre-existing survey data for this purpose were available from Brazil, Norway and the USA. Of the 13741 mother–infant pairs screened for the longitudinal component, about 83% were ineligible [15]. A family's low socio-economic status was the most common reason for ineligibility in Brazil, Ghana, India and Oman, whereas parental refusal was the main reason for non-participation in Norway and the USA [15]. For the cross-sectional component, 69% of the 21510 subjects screened were excluded for reasons similar to those observed in the longitudinal component.

Term low-birthweight (<2500 g) infants (2.3%) were *not* excluded. Since it is likely that, in well-off populations, such infants represent small but normal children, and their exclusion would have artificially distorted the standards' lower percentiles. Eligibility criteria for the cross-sectional component were the same as those for the longitudinal component with the exception of infant feeding practices. A minimum of 3 mo of any breastfeeding was required for participants in the study's cross-sectional component.

Anthropometric methods

Data collection teams were trained at each site during the study's preparatory phase, at which time measurement techniques were standardized against one of two MGRS anthropometry experts. During the study, bimonthly standardization sessions were conducted at each site. Once a year, the anthropometry expert visited each site to participate in these sessions [16]. Results from the anthropometry standardization sessions are reported in a companion paper in this volume [17]. For the longitudinal component of the study, screening teams measured newborns within 24 h of delivery, and follow-up teams conducted home visits until 24 mo of age. The follow-up teams were also responsible for taking measurements in the cross-sectional component involving children aged 18–71 mo [11].

The MGRS data included weight and head circumference at all ages, recumbent length (longitudinal component), height (cross-sectional component), and arm circumference, triceps and subscapular skinfolds (all children aged $\geq$3 mo). However, here we report on only the standards based on length or height and weight. Observers working in pairs collected anthropometric data. Each observer independently measured and recorded a complete set of measurements, after which the two compared their readings. If any pair of readings exceeded the maximum allowable difference for a given variable (weight 100 g; length/height 7 mm), both observers once again independently measured and recorded a second and, if necessary, a third set of readings for the variable(s) in question [16].

All study sites used identical measuring equipment. Instruments needed to be highly accurate and precise, yet sturdy and portable to enable them to be carried back and forth on home visits. Length was measured with the Harpenden Infantometer (range 30–110 cm for portable use, with digit counter readings precise to 1 mm). The Harpenden Portable Stadiometer (range 65–206 cm, digit counter reading) was used for measuring both adult and child heights. Portable electronic scales with a taring capability and calibrated to 0.1 kg (i.e. UNICEF Electronic Scale 890 or UNISCALE) were used to measure weight. Length and height were recorded to the last completed unit rather than to the nearest unit. To correct for the systematic negative bias introduced by this practice, 0.05 cm (i.e. half of the smallest measurement unit) was added to each measurement before analysis. This correction did not apply to weight, which was rounded off to the nearest 100 g. Full details of the instruments used and how measurements were taken are provided elsewhere [16].

Criteria for including children in the sample used to generate the standards

The total sample size for the longitudinal and cross-sectional studies from all six sites was 8440 children. A total of 1743 children were enrolled in the longitudinal sample, six of whom were excluded for morbidities affecting growth (four cases of repeated episodes of diarrhoea, one case of repeated episodes of malaria and one case of protein-energy malnutrition), leaving a final sample of 1737 children (894 boys and 843 girls). Of these, the mothers of 882 children (428 boys and 454 girls) complied fully with the MGRS infant-feeding and no-smoking criteria and completed the follow-up period of 24 mo. The other 855 children contributed only their birth records, as they either failed to comply with the study's criteria or dropped out before 24 mo. The total number of records for the longitudinal component was 19 900. The cross-sectional sample comprised 6697 children. Of these, 28 were excluded for medical conditions affecting growth (20 cases of protein-energy malnutrition, five cases of haemolytic anaemia G6PD deficiency, two cases of renal tubulo-interstitial disease and one case of Crohn disease), leaving a final sample of 6669 children (3450 boys and 3219 girls) with a total of 8306 records.

Data cleaning procedures and exclusions applied to the data

The MGRS data management protocol [18] was designed to create and manage a large databank of information collected from multiple sites over a period of several years. Data collection and processing instruments were prepared centrally and used in a standardized fashion across sites. The data management system contained internal validation features for timely detection of data errors, and its standard operating procedures stipulated a method of master file updating and correction that maintained a clear trail for data-auditing purposes. Each site was responsible for collecting, entering, verifying and validating data, and for creating site-level master files. Data from the sites were sent to the WHO every month for master file consolidation and more extensive quality-control checking. All errors identified were communicated to the site for correction at source.

After data collection was completed at a given site, a period of about 6 mo was dedicated to in-depth data quality checking and master file cleaning. The WHO produced detailed validation reports, descriptive statistics and plots from the site's master files. For the longitudinal component, each anthropometric measurement was plotted for every child from birth to the end of his/her participation. These plots were examined individually for any questionable patterns. Query lists from these analyses were sent to the site for investigation and correction, or confirmation, as required. As with the data collection process, the site data manager prepared correction batches to update the master files. The updated master files were then sent to the WHO, and this iterative quality assurance process continued until both the site and WHO were satisfied that all identifiable problems had been detected and corrected. The rigorous implementation of what was a highly demanding protocol yielded very high-quality data.

To avoid the influence of unhealthy weights for length/height, prior to constructing the standards, observations falling above +3 SD and below −3 SD of the sample median were excluded. For the cross-sectional sample, the +2 SD cut-off (i.e. 97.7 percentile) was applied instead of +3 SD as the sample was exceedingly skewed to the right, indicating the need to identify and exclude high weights for height. This cut-off was considered to be conservative given that various definitions of overweight all apply lower cut-offs than the one we used [19,20]. The procedure by which this was done is described in the technical report outlining the construction of the standards [10]. The number of observations excluded for unhealthy weight-for-length/height was 185 (1.4%) for boys and 155 (1.1%) for girls, most of which were in the upper end of the cross-sectional sample distribution. In addition, a few influential observations for indicators other than weight-for-height were excluded when constructing the individual standards: for boys, four (0.03%) observations for weight-for-age and three (0.02%) observations for length/height-for-age; and for girls, one (0.01%) and two (0.01%) observations for the same indicators, respectively.

Statistical methods for constructing the WHO child growth curves

The construction of the child growth curves followed a careful, methodical process. This involved a) detailed examination of existing methods, including types of distributions and smoothing techniques, in order to identify the best possible approach; b) selection of a software package flexible enough to allow comparative testing of alternative methods and the actual generation of the curves; and c) systematic application of the selected approach to the data to generate the models that best fit the data.

A group of statisticians and growth experts met at the WHO to review possible choices of methods and to define a strategy and criteria for selecting the most appropriate model for the MGRS data [21]. As many as 30 methods for attained growth curves were examined. The group recommended that methods based on selected distributions be compared and combined with two smoothing techniques for fitting its parameter curves to further test and provide the best possible method for constructing the WHO child growth standards.

Choice of distribution. Five distributions were identified for detailed testing: the Box-Cox power exponential [22], the Box-Cox t [23], the Box-Cox normal [24], the Johnson's SU [25] and the modulus-exponential-normal [26]. The first four distributions were fitted using the GAMLSS (Generalized Additive Models for Location, Scale and Shape) software [27] and the last using the "xriml" module in the STATA software [28]. The Box-Cox power exponential (BCPE) with four parameters—μ (for the median), σ (coefficient of variation), ν (Box-Cox transformation power) and τ (parameter related to kurtosis)—was selected as the most appropriate distribution for constructing the curves. The BCPE is a flexible distribution that simplifies to the normal distribution when $\nu = 1$ and $\tau = 2$. Also, when $\nu \neq 1$ and $\tau = 2$, the distribution is the same as the Box-Cox normal (LMS method distribution). The BCPE is defined by a power transformation (or Box-Cox transformation) having a shifted and scaled (truncated) power exponential (or Box-Tiao) distribution with parameter τ [22]. Apart from other theoretical advantages, the BCPE presents as good as or better goodness of fit than the modulus-exponential-normal or the SU distribution.

Choice of smoothing technique. Two smoothing techniques were recommended for comparison by the expert group: cubic splines and fractional polynomials [21]. Using GAMLSS, comparisons were carried out for length/height-for-age, weight-for-age and weight-for-length/height. The cubic spline smoothing technique offered more flexibility than fractional polynomials in all cases. For the length-for-age and weight-for-age standards, a power transformation applied to age prior to fitting was necessary to enhance the goodness of fit by the cubic splines technique.

Choice of method for constructing the curves. In summary, the BCPE method, with curve smoothing by cubic splines, was selected as the approach for constructing the growth curves. This method is included in a broader methodology, the GAMLSS [29], which offers a general framework that includes a wide range of known methods for constructing growth curves. The GAMLSS allows for modelling the mean (or location) of the growth variable under consideration as well as other parameters of its distribution that determine scale and shape. Various kinds of distributions can be assumed for each growth variable of interest, from normal to highly skewed and/or kurtotic distributions. Several smoothing terms can be used in generating the curves, including cubic splines, lowess (locally weighted least squares regression), polynomials, power polynomials and fractional polynomials.

Process and diagnostic criteria for selecting the best model to construct the curves. The process for selecting the best model to construct the curves for each growth variable involved selecting first the best model *within* a class of models and, second, the best model *across* different classes of models. The Akaike Information Criteria [30] and the generalized version of it [22] were used to select the best model *within* a considered class of models. In addition, worm plots [31] and Q-tests [32] were used to determine the adequate numbers of degrees of freedom for the cubic splines fitted to the parameter curves. In most cases, it was necessary to transform age before fitting the cubic splines to "stretch" the age scale during the neonatal period when growth is rapid and the rise in percentile curves is steep. Thus, selecting the best model within the same class of models involved finding the best choice for degrees of freedom for the parameter curves, determining whether age needed to be transformed and finding the best power (λ). In selecting the best model *across* different classes of models, we started from the simplest class of models (i.e. the normal distribution) and proceeded to more complex models when necessary. The goal was to test the impact of increasing the model's complexity on its goodness of fit. The same set of diagnostic tools/tests was used at this stage.

Two diagnostic tools were used to detect possible biases in estimated percentile or z-score curves. First, we examined the pattern of differences between empirical and fitted percentiles; second, we compared observed and expected proportions of children with measurements below selected percentiles or z-score curves.

A more detailed description of the statistical methods and procedures that were followed to construct the WHO Child Growth Standards is provided elsewhere [10].

Types of curves generated

Percentile and z-score curves were generated ranging from the 99th to the 1st percentile and from +3 to −3 standard deviations, respectively. Due to space constraints, we present in this article only the z-score curves for the following lines: 3, 2, 1, 0, −1, −2 and −3 standard deviations. An extensive display of the standards' charts and tables containing such information as means and standard deviations by age and sex, percentile values and related measures is provided in the technical report [10] and on the Web: www.who.int/childgrowth/en

Results

The specifications of the BCPE models that provided the best fit to generate specific standards are summarized in Table I. These are specific values for the age power transformation and the degrees of freedom for the cubic spline functions fitting the four parameters that define the BCPE distribution selected for each standard. Age needed to be transformed for boys and girls except for weight-for-length/height and BMI curves from 24 to 60 mo. There was wide variability in the degrees of freedom that were necessary for the cubic splines to achieve the best fit for modelling the median (μ) and its coefficient of variation (σ). In the case of length/height-for-age for boys and girls, the normal distribution (i.e. when ν takes the value of 1 and τ is 2) proved to be the parsimonious option. In all other cases, it was necessary to model skewness (ν) but not kurtosis (i.e. τ was 2 for all standards), which simplified the model considerably. One to three degrees of freedom for the ν parameter were sufficient in all cases where the distribution was skewed (Table I). The degrees of freedom chosen for boys and girls were often the same or similar.

It was possible to construct both length-for-age (0 to 2 y) and height-for-age (2 to 5 y) standards fitting a unique model, yet still reflecting the difference between recumbent length and standing height. The cross-sectional component included the measurement of both length and height in children 18 to 30 mo old ($n = 1625$ children), and from these data it was estimated that length was the larger measure by 0.7 cm [10]. To fit a single model for the whole age range, 0.7 cm was therefore added to the cross-sectional height values. After the model was fitted, the final curves were shifted downwards by 0.7 cm for ages 2 y and above to create the height-for-age standards. Coefficient of variance values were adjusted to reflect this back transformation using the shifted medians and standard deviations. The length-for-age (0 to 24

Table I. Degrees of freedom for fitting the parameters of the Box-Cox power exponential (BCPE) distribution for the models with the best fit to generate standards based on age, length and weight in children 0–60 mo of age.

Standards	Sex	λ[a]	df(μ)[b]	df(σ)[c]	df(ν)[d]	τ[e]
Length/height, 0–60 mo	Boys	0.35	12	6	0[f]	2
Length/height, 0–60 mo	Girls	0.35	10	5	0[f]	2
Weight, 0–60 mo	Boys	0.35	11	7	2	2
Weight, 0–60 mo	Girls	0.35	11	7	3	2
Weight-for-length/height, 0–60 mo	Boys	None	13	6	1	2
Weight-for-length/height, 0–60 mo	Girls	None	12	4	1	2
BMI, 0–24 mo	Boys	0.05	10	4	3	2
BMI, 0–24 mo	Girls	0.05	10	3	3	2
BMI, 24–60 mo	Boys	None	4	3	3	2
BMI, 24–60 mo	Girls	None	4	4	1	2

[a] Age transformation power.
[b] Degrees of freedom for the cubic splines fitting the median (μ).
[c] Degrees of freedom for the cubic splines fitting the coefficient of variation (σ).
[d] Degrees of freedom for the cubic splines fitting the Box-Cox transformation power (ν).
[e] Parameter related to the kurtosis fixed ($\tau = 2$).
[f] $\nu = 1$: normal distribution.

mo) standard was derived directly from the fitted model. A similar approach was followed in generating the weight-for-length (45 to 110 cm) and weight-for-height (65 to 120 cm) standards. In the generation of the length/height-for-age standards, data up to 71 mo of age were used and the fitted model truncated at 60 mo in order to control for edge effects. For the weight-for-length/height standards, data up to 120 cm height were used to fit the model to prevent the fitting from being influenced by the portion of the data presenting instability [10].

In addressing the differences between length and height, a different approach for the BMI-for-age standards was followed because BMI is a ratio with length or height squared in the denominator. After adding 0.7 cm to the height values, it was not possible, after fitting, to back-transform lengths to heights. The solution adopted was to construct the standards for younger and older children separately based on two sets of data with an overlapping range of ages below and above 24 mo. To construct the BMI-for-age standard using length (0–2 y), the longitudinal sample and the cross-sectional height data up to 30 mo were used after adding 0.7 cm to the height values. Analogously, to construct the standard from 2 to 5 y, the cross-sectional sample plus the longitudinal length from 18–24 mo were used after subtracting 0.7 cm from the length values. Thus, a common set of data from 18 to 30 mo was used to generate the BMI standards for younger and older children.

The concordance between smoothed percentile curves and observed or empirical percentiles was remarkably good. As examples, we show comparisons for the 3rd, 10th, 50th, 90th and 97th percentiles for length-for-age for boys (Figure 1) and for weight-for-height for girls (Figure 2). Overall, the fit was best for length and height-for-age standards, but it was almost as good for the standards based on combinations of weight and length [10]. The average absolute difference between smoothed and empirical percentiles was small: 0.13 cm for length-for-age in boys 0 to 24 mo (Figure 1) and 0.16 kg for weight-for-height for girls 65 to 120 cm (Figure 2). Taking the sign into account, the average differences are close to zero: -0.03 cm and -0.02 kg in Figures 1 and 2, respectively, which indicates lack of bias in the fit between smoothed and empirical percentiles.

Z-score curves are given for length/height-for-age for boys and girls from birth to 60 mo of age (Figures 3 and 4), weight-for-age for boys and girls from birth to 60 mo (Figures 5 and 6), weight-for-length for boys and girls 45 to 110 cm (Figures 7 and 8), weight-for-height for boys and girls 65 to 120 cm (Figures 9 and 10) and BMI-for-age for boys and girls from birth to 60 mo (Figures 11 and 12). The last are in addition to the previously available set of indicators in the NCHS/WHO reference.

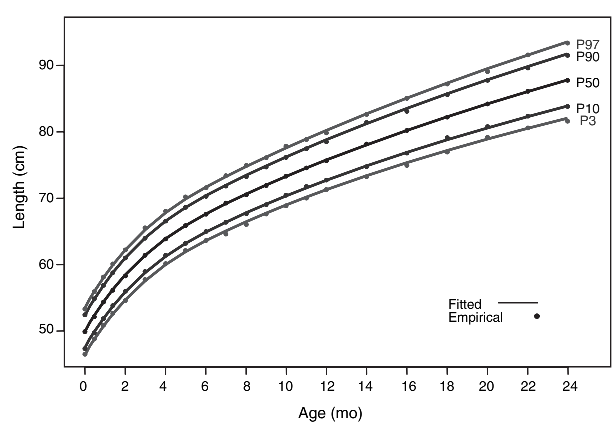


Figure 1. Comparisons between 3rd, 10th, 50th, 90th and 97th smoothed percentile curves and empirical values for length-for-age for boys.

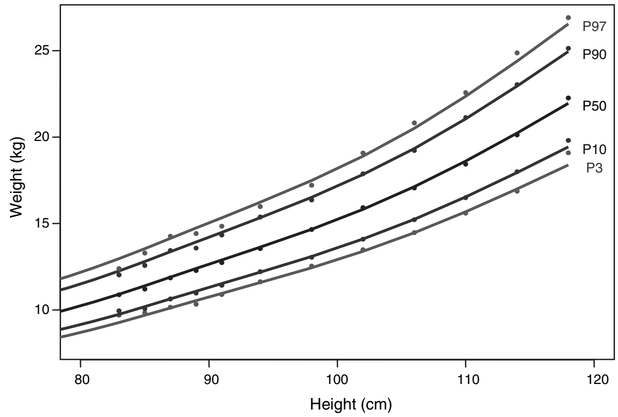


Figure 2. Comparisons between 3rd, 10th, 50th, 90th and 97th smoothed percentile curves and empirical values for weight-for-height for girls.

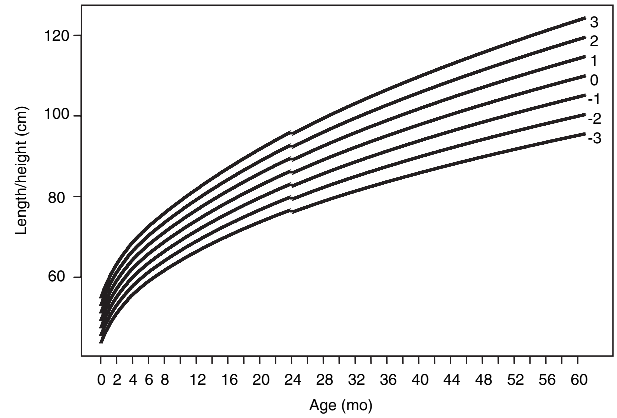


Figure 4. Z-score curves for length/height-for-age for girls from birth to 60 mo. Length from birth to 23 completed months; height from 24 to 60 completed months.

Discussion

The goal of the MGRS was to describe the growth of healthy children. Criteria were applied in the study design to achieve this aim. Screening at enrolment using site-specific socio-economic criteria and maternal non-smoking status excluded children likely to experience constrained growth. Morbidities that affect growth (e.g. repeated bouts of infectious diarrhoea and Crohn disease) were identified, and affected children were excluded from the sample. Application of these criteria resulted in no evidence of under-nutrition in either the longitudinal or cross-sectional samples.

In the longitudinal sample, the behavioural criteria of breastfeeding through 12 mo and its close monitoring throughout data collection yielded a sample of children with no evidence of over-nutrition (i.e. no excessive right skewness). In the cross-sectional sample, however, despite the criterion of at least 3 mo of any breastfeeding, the sample was exceedingly skewed to the right, indicating the need to identify and exclude excessively high weights for heights if the goal of constructing a standard was to be satisfied. A similar prescriptive approach was taken by the developers of the 2000 CDC growth charts for the USA when excluding data from the last national survey (i.e. NHANES III) for children aged $\geq$6 y from the revised weight and BMI growth charts [33]. Without this exclusion, the 95th and 85th percentile curves of the CDC charts would have been higher, and fewer children would have been classified as overweight or at risk of overweight.

Rigorous methods of data collection, standardized across sites, were followed during the entire study. Sound procedures for data management and cleaning were applied. As a result, the anthropometric data available for analysis were of the highest possible quality. A process of consultation with experts in statistical methods and growth was followed, and methodical, state-of-the-art statistical methodologies were employed to generate the standards [21]. The fit between the smoothed curves and empirical or observed percentiles was excellent and free of bias at

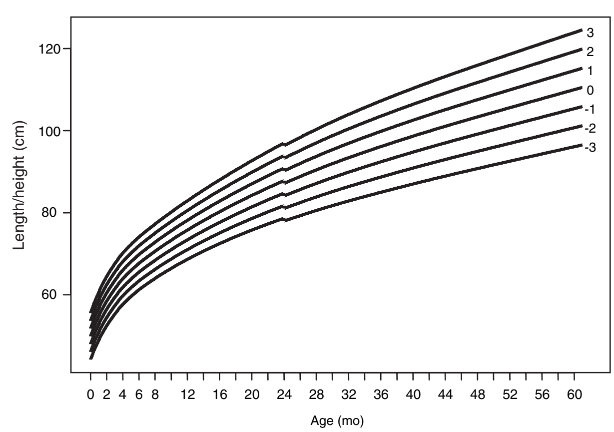


Figure 3. Z-score curves for length/height-for-age for boys from birth to 60 mo. Length from birth to 23 completed months; height from 24 to 60 completed months.

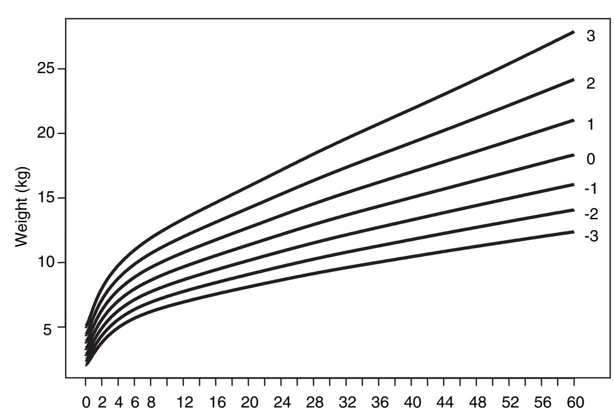


Figure 5. Z-score curves for weight-for-age for boys from birth to 60 mo.

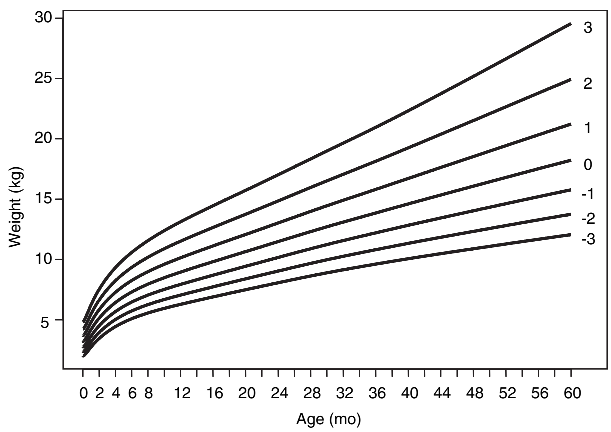


Figure 6. Z-score curves for weight-for-age for girls from birth to 60 mo.

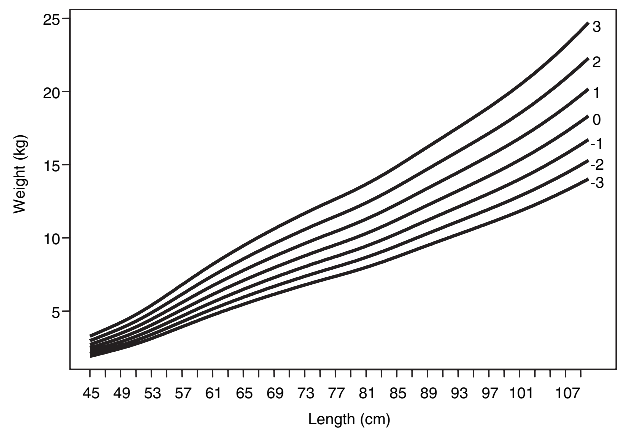


Figure 8. Z-score curves for weight-for-length for girls from 45 to 110 cm.

both the median and the edges, indicating that the resulting curves are a fair description of the true growth of healthy children. Thus, the MGRS can serve as a model of how studies of this type should be carried out and analysed.

The technical report, of which this article is a summary, includes a comparison of the new WHO standards to the previously recommended NCHS/WHO international reference [10]. As expected, there are important differences. However, these vary—by anthropometric measure, sex, specific percentile or z-score curve, and age—in ways that are not easily summarized. Differences are particularly important in infancy. Impact on population estimates of child malnutrition will depend on age, sex, anthropometric indicator considered and population-specific anthropometric characteristics. Thus, it will not be possible to provide an algorithm that will convert new prevalence values from old ones. A notable effect is that stunting will be greater throughout childhood when assessed using the new WHO standards compared to the previous international reference. The growth pattern of breastfed infants compared to the NCHS/WHO reference will result in a substantial increase in underweight rates during the first half of infancy (i.e. 0–6 mo) and a decrease thereafter. For wasting, the main difference between the new standards and the old reference is during infancy (i.e. up to about 70 cm length) when wasting rates will be substantially higher using the new WHO standards. With respect to overweight, use of the new WHO standards will result in a greater prevalence that will vary by age, sex and nutritional status of the index population.

The WHO Child Growth Standards were derived from children who were raised in environments that minimized constraints to growth such as poor diets and infection. In addition, their mothers followed healthy practices such as breastfeeding their children and not smoking during and after pregnancy. The standards depict normal human growth under optimal environmental conditions and can be used to assess children everywhere, regardless of ethnicity, socio-economic status and type of feeding. It would be as inappropriate to call for separate standards to be developed for children whose mothers smoked during pregnancy as it would be for children who are fed a

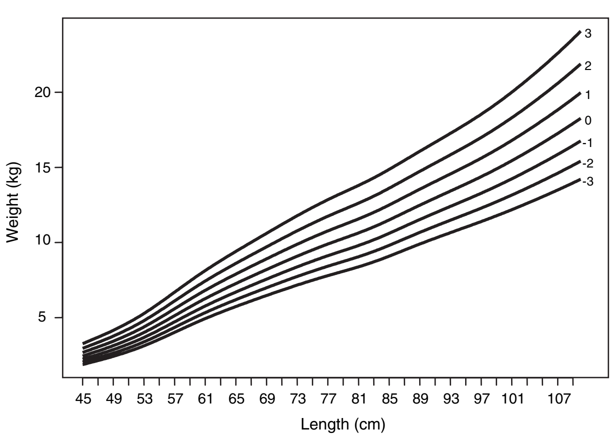


Figure 7. Z-score curves for weight-for-length for boys from 45 to 110 cm.

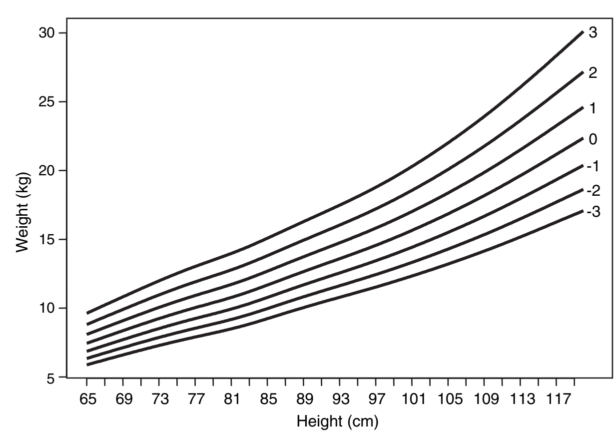


Figure 9. Z-score curves for weight-for-height for boys from 65 to 120 cm.

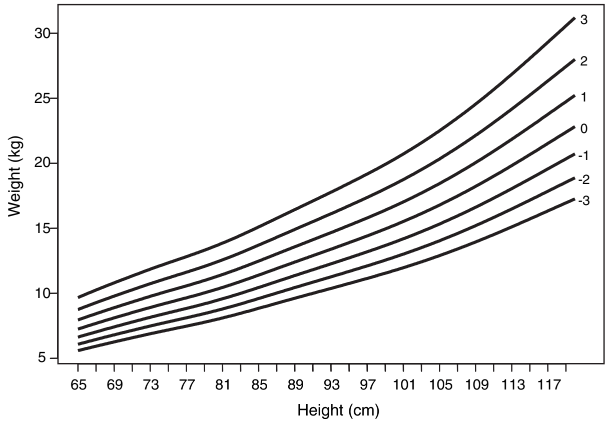


Figure 10. Z-score curves for weight-for-height for girls from 65 to 120 cm.

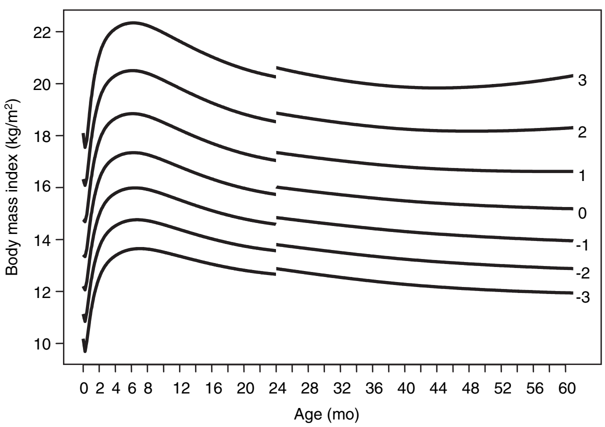


Figure 11. Z-score curves for BMI-for-age for boys from birth to 60 mo. BMI based on length from birth to 23 completed months; BMI based on height from 24 to 60 completed months.

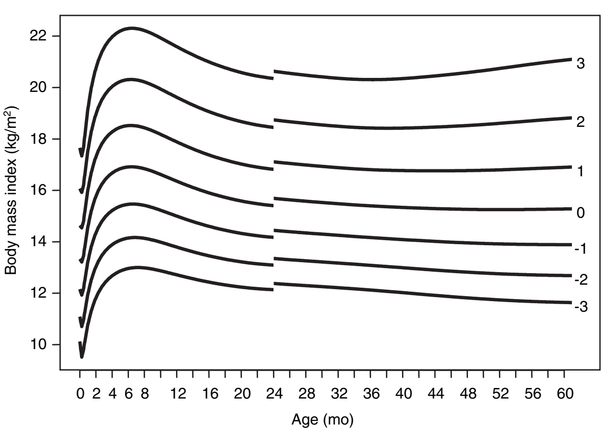


Figure 12. Z-score curves for BMI-for-age for girls from birth to 60 mo. BMI based on length from birth to 23 completed months; BMI based on height from 24 to 60 completed months.

breast-milk substitute. Rather, deviations from any area in the world in the patterns described by the standards, such as a high proportion of children with short heights or high weight-for-heights, when properly assessed and interpreted, should be seen as representing abnormal growth and taken as evidence of stunting and obesity, respectively, in these examples.

Acknowledgements

This paper was prepared by Mercedes de Onis, Reynaldo Martorell, Cutberto Garza and Anna Lartey on behalf of the WHO Multicentre Growth Reference Study Group. The statistical analysis was conducted by Elaine Borghi.

References

[1] Waterlow JC, Buzina R, Keller W, Lane JM, Nichaman MZ, Tanner JM. The presentation and use of height and weight data for comparing the nutritional status of groups of children under the age of 10 years. Bull World Health Organ 1977;55:489–98.

[2] WHO Working Group on Infant Growth. An evaluation of infant growth. Geneva: World Health Organization; 1994.

[3] de Onis M, Yip R. The WHO growth chart: historical considerations and current scientific issues. Bibl Nutr Dieta 1996;53:74–89.

[4] de Onis M, Habicht JP. Anthropometric reference data for international use: Recommendations from a World Health Organization Expert Committee. Am J Clin Nutr 1996;64:650–8.

[5] Garza C, de Onis M, for the WHO Multicentre Growth Reference Study Group. Rationale for developing a new international growth reference. Food Nutr Bull 2004;25 Suppl 1:S5–14.

[6] de Onis M, Garza C, Victora CG, Bhan MK, Norum KR, editors. WHO Multicentre Growth Reference Study (MGRS): Rationale, planning and implementation. Food Nutr Bull 2004;25 Suppl 1:S1–89.

[7] WHO Working Group on Infant Growth. An evaluation of infant growth: The use and interpretation of anthropometry in infants. Bull World Health Organ 1995;73:165–74.

[8] WHO Multicentre Growth Reference Study Group. Assessment of differences in linear growth among populations in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450:56–65.

[9] Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. Genetic structure of human populations. Science 2002;298:2381–5.

[10] WHO Multicentre Growth Reference Study Group. 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: World Health Organization; 2006.

[11] de Onis M, Garza C, Victora CG, Onyango AW, Frongillo EA, Martines J, for the WHO Multicentre Growth Reference Study Group. The WHO Multicentre Growth Reference Study: Planning, study design and methodology. Food Nutr Bull 2004;25 Suppl 1:S15–26.

[12] Owusu WB, Lartey A, de Onis M, Onyango AW, Frongillo EA. Factors associated with unconstrained growth among affluent Ghanaian children. Acta Paediatr 2004;93:1115–9.

[13] Bhandari N, Bahl R, Taneja S, de Onis M, Bhan MK. Growth performance of affluent Indian children is similar to that in developed countries. Bull World Health Organ 2002;80:189–95.

[14] Mohamed AJ, Onyango AW, de Onis M, Prakash N, Mabry RM, Alasfoor DH. Socioeconomic predictors of unconstrained child growth in Muscat, Oman. East Mediterr Health J 2004;10:295–302.
[15] WHO Multicentre Growth Reference Study Group. Enrolment and baseline characteristics in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450:7–15.
[16] de Onis M, Onyango AW, Van den Broeck J, Chumlea WC, Martorell R, for the WHO Multcentre Growth Reference Study Group. Measurement and standardization protocols for anthropometry used in the construction of a new international growth reference. Food Nutr Bull 2004;25 Suppl 1:S27–36
[17] WHO Multicentre Growth Reference Study Group. Reliability of anthropometric measurements in the WHO Multicentre Growth Reference Study. Acta Paediatr 2006;Suppl 450:38–46.
[18] Onyango AW, Pinol AJ, de Onis M, for the WHO Multicentre Growth Reference Study Group. Managing data for a multicountry longitudinal study: Experience from the WHO Multicentre Growth Reference Study. Food Nutr Bull 2004;25 Suppl 1:S46–52.
[19] Daniels SR, Arnett DK, Eckel RH, Gidding SS, Hayman LL, Kumanyika S, et al. AHA scientific statement. Overweight in children and adolescents: Pathophysiology, consequences, prevention and treatment. Circulation 2005;111:1999–2012.
[20] Koplan JP, Liverman CT, Kraak VI, editors. Preventing childhood obesity: Health in the balance. Washington, DC: National Academies Press; 2005.
[21] Borghi E, de Onis M, Garza C, Van den Broeck J, Frongillo EA, Grummer-Strawn L, et al., for the WHO Multicentre Growth Reference Study Group. Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. Stat Med 2006;25:247–65.
[22] Rigby RA, Stasinopoulos DM. Smooth centile curves for skew and kurtotic data modelled using the Box-Cox power exponential distribution. Stat Med 2004;23:3053–76.
[23] Rigby RA, Stasinopoulos DM. Box-Cox t distribution for modelling skew and leptokurtotic data. Technical report 01/04. London: STORM Research Centre, London Metropolitan University; 2004.
[24] Cole TJ, Green PJ. Smoothing reference centile curves: The LMS method and penalized likelihood. Stat Med 1992;11:1305–19.
[25] Johnson NL. Systems of frequency curves generated by methods of translation. Biometrika 1949;36:149–76.
[26] Royston P, Wright EM. A method for estimating age-specific reference intervals ('normal ranges') based on fractional polynomials and exponential transformation. J R Stat Soc Ser A 1998;161:79–101.
[27] Stasinopoulos DM, Rigby RA, Akantziliotou C. Instructions on how to use the GAMLSS package in R. Technical report 02/04. London: STORM Research Centre, London Metropolitan University; 2004.
[28] Wright E, Royston P. Age-specific reference intervals ("normal ranges"). Stata Technical Bulletin 1996;34:24–34.
[29] Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. J R Stat Soc Ser C Appl Stat 2005;54:507–44.
[30] Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr 1974;19:716–23.
[31] van Buuren S, Fredriks M. Worm plot. A simple diagnostic device for modelling growth reference curves. Stat Med 2001;20:1259–77.
[32] Royston P, Wright EM. Goodness-of-fit statistics for age-specific reference intervals. Stat Med 2000;19:2943–62.
[33] Kuczmarski RJ, Ogden CL, Guo SS, Grummer-Strawn LM, Flegal KM, Mei Z, et al. 2000 CDC growth charts for the United States: Methods and development. National Center for Health Statistics. Vital Health Stat Series 11 No. 246, 2002:1–190.

Acta Pædiatrica, 2006; Suppl 450: 86–95

Taylor & Francis
Taylor & Francis Group

WHO Motor Development Study: Windows of achievement for six gross motor development milestones

WHO MULTICENTRE GROWTH REFERENCE STUDY GROUP[1,2]

[1]*Department of Nutrition, World Health Organization, Geneva, Switzerland, and* [2]*Members of the WHO Multicentre Growth Reference Study Group* (*listed at the end of the first paper in this supplement*)

Abstract

Aim: To review the methods for generating windows of achievement for six gross motor developmental milestones and to compare the actual windows with commonly used motor development scales. *Methods:* As part of the WHO Multicentre Growth Reference Study, longitudinal data were collected to describe the attainment of six gross motor milestones by children aged 4 to 24 mo in Ghana, India, Norway, Oman and the USA. Trained fieldworkers assessed 816 children at scheduled visits (monthly in year 1, bimonthly in year 2). Caretakers also recorded ages of achievement independently. Failure time models were used to construct windows of achievement for each milestone, bound by the 1st and 99th percentiles, without internal demarcations. *Results:* About 90% of children achieved five of the milestones following a common sequence, and 4.3% did not exhibit hands-and-knees crawling. The six windows have age overlaps but vary in width; the narrowest is sitting without support (5.4 mo), and the widest are walking alone (9.4 mo) and standing alone (10.0 mo). The estimated 1st and 99th percentiles in months are: 3.8, 9.2 (sitting without support), 4.8, 11.4 (standing with assistance), 5.2, 13.5 (hands-and-knees crawling), 5.9, 13.7 (walking with assistance), 6.9, 16.9 (standing alone) and 8.2, 17.6 (walking alone). The 95% confidence interval widths varied among milestones between 0.2 and 0.4 mo for the 1st percentile, and 0.5 and 1.0 mo for the 99th.

Conclusion: The windows represent normal variation in ages of milestone achievement among healthy children. They are recommended for descriptive comparisons among populations, to signal the need for appropriate screening when individual children appear to be late in achieving the milestones, and to raise awareness about the importance of overall development in child health.

Key Words: *Gross motor milestones, longitudinal, motor skills, standards, young child development*

Introduction

The WHO Multicentre Growth Reference Study (MGRS) had as its primary objective the construction of curves and related tools to assess growth and development in children from birth to 5 y of age [1]. The MGRS is unique in that it was designed to produce a standard rather than a reference. Standards and references both serve as bases for comparison, but differences with respect to their curves result in different interpretations. A standard defines how children should grow, and thus deviations from the pattern it sets should be taken as evidence of abnormal growth. A reference, on the other hand, is not a sound basis for such judgements, although in practice references are often misused as standards.

The MGRS data provide a solid basis for developing a standard because they concern healthy children living under conditions that are highly unlikely to constrain growth. Moreover, the mothers of the children selected for the construction of the standards followed certain healthy practices, namely breastfeeding their children and not smoking [2]. A second feature of the MGRS that makes it attractive as a standard for children everywhere is that it included healthy children from six geographically diverse countries: Brazil, Ghana, India, Norway, Oman and the USA. Thus, the study design has considerable built-in ethnic or genetic variability but reduces some aspects of environmental variation by including only privileged, healthy populations [2]. On the other hand, along with ethnic variation comes cultural variation, including the way children are nurtured.

Another distinguishing feature of the MGRS is that it included the collection of ages of achievement of

Correspondence: Mercedes de Onis, Study Coordinator, Department of Nutrition, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland. Tel: +41 22 791 3320. Fax: +41 22 791 4156. E-mail: deonism@who.int

ISSN 0803-5326 print/ISSN 1651-2227 online © 2006 Taylor & Francis
DOI: 10.1080/08035320500495563

motor milestones in five of the six study sites. The WHO has in the past issued recommendations concerning reference curves for assessing attained growth [3], but it has not made any with respect to motor development. The MGRS curves were designed to replace the previously recommended reference curves for child growth (i.e. the NCHS/WHO growth reference), which are now known to suffer from a number of deficiencies. A companion paper in this volume [4] shows that differences among MGRS sites in linear growth are minor compared to inter-individual variation and residual error, and concludes that pooling data across sites is justified. The physical growth standards are presented in a second paper in this volume [5], and this is done separately for boys and girls because patterns of growth differ importantly by sex. A third paper [6] considers variability in ages of achievement of motor milestones and concludes that, in contrast to physical growth, the differences between the sexes in motor development are trivial and do not justify separate standards for boys and girls. Furthermore, the paper calls for pooling of the information across sites in generating the standards for motor development and does so despite some evidence of modest heterogeneity across sites in ages of achievement for some of the milestones [6]. Since the children were healthy and showed similar growth in length, the variation observed across sites in ages of achievement of motor milestones is best viewed as normal variation. The differences possibly reflect cultural variations in childrearing, but ethnic or genetic causes cannot be ruled out. An additional article in this volume [7] shows that there is little or no relationship between physical growth and motor development in the population studied. The literature indicates that growth retardation is related to delayed motor development, perhaps because of common causes such as nutritional deficiencies and infections, but in healthy children, as we have found, size and motor development are not linked.

The above considerations led to different approaches in the construction of standards for physical growth compared to motor development. In the case of physical growth, curves were generated that depict gradations of the distribution surrounding the median, such as percentile or z-score lines [5], and software was developed to estimate z scores for individual children. An expert group convened to review the potential uses of the motor development data and methods for generating a standard on their basis recommended that "windows of achievement" be used rather than percentile curves [8]. These windows, the experts recommended, should be bounded by the 1st and 99th percentiles of the pooled distribution of all sites and should be interpreted as normal variation in ages of achievement among healthy children. The concept of a "window" offers a simple tool that can be easily used to assess children since it requires no calculations, an aspect to which we will return later.

The objectives of this paper are to review methods for generating the windows of achievement and to present the actual windows for all six milestones considered. We also compare the MGRS windows of achievement to commonly used scales of motor development.

Methods

Description of data collection for achievement of motor milestones

The design and general methods of the MGRS, and the training and standardization of fieldworkers and data collection procedures in the area of motor development, are described in detail elsewhere [2,9]. The recruitment criteria, sample characteristics and reliability of the motor development assessments are presented in companion papers in this supplement [5,10,11]. Motor development data were collected in five sites: Ghana, India, Norway, Oman and the USA. The study was already well under way in Brazil when the decision to add this component was made.

Data were collected monthly from 4 to 12 mo of age and bimonthly thereafter until all milestones were achieved or the child reached 24 mo of age. Trained fieldworkers assessed children directly at the scheduled home visits, and mothers also independently recorded ages of achievement (see below). Six milestones were selected for study: sitting without support, hands-and-knees crawling, standing with assistance, walking with assistance, standing alone and walking alone. These milestones were considered to be universal, fundamental to the acquisition of self-sufficient erect locomotion, and simple to test and evaluate. The description, criteria and testing procedures used to judge whether a child demonstrated achievement of a milestone are given elsewhere [9]. The child's performance was recorded as follows: a) tried but failed to perform the milestone, b) refused to perform despite being alert and calm, c) was able to perform the milestone, and d) could not be tested because of irritability, drowsiness or sickness. In practice, it proved difficult to distinguish between this last category and refusals. On average, it took about 10 min to assess motor development in a child [9].

An important feature of data collection is that there was no progression or hierarchy assumed among the milestones. Performance was assessed on each examination date for all six milestones. Each examination was carried out independently of all previous assessments, although it is likely that fieldworkers, who knew the families and children intimately, remembered some or all previous results. Whenever possible, the number of people present was limited to the caretaker, child and fieldworker. Efforts were made to keep the floor clean and free of clutter, and mothers were asked to select no more than three of the child's toys to use in the testing. Since it was important that the child remained calm and cheerful during the assessment, the motor assessments were made at the most opportune moments, often after completing the anthropometric assessment. After each examination, the fieldworker rated the child's state of wakefulness as either awake and alert or drowsy, and of irritability as calm, fussy or upset (crying) [9].

Caretakers were also instructed on the criteria for each milestone's achievement and the correct procedures for testing them, and they were encouraged to observe and assess the child's performance. Caretakers were provided a record form with drawings of each milestone and boxes for recording the first date the child achieved the milestone. In the second year, when home visits occurred every 2 mo, caretakers of children who had not yet achieved certain milestones were telephoned during the unvisited months and reminded to assess their children.

The fieldworker noted any date written by the caretaker. If, upon examination by the fieldworker, the performance of a milestone was confirmed, the fieldworker recorded the date of achievement observed by the caretaker. Every time a date of achievement was recorded, caretakers were also asked whether the date was obtained by actual testing and recording or simply by recall, and this information was recorded as well. If, on the other hand, the child was not able to perform the milestone during the examination by the fieldworker, a discussion took place with the caretaker during which the criteria for that milestone's achievement were again reviewed. If the caretaker insisted that the child had indeed met the criteria, the fieldworker accepted and recorded the date reported by the caretaker. If the caretaker acknowledged that the criteria were not met, a new line was added to the form, and the caretaker was encouraged to monitor the child's progress, repeat the assessment and note the actual date of milestone achievement. The fieldworker took the form from the caretaker when all six milestones had been achieved. The data recording form and other details of data collection are provided elsewhere [9].

Selecting the method of estimation for generating the windows of achievement

Estimating the windows of achievement requires estimates of the lower and upper margins of the window, specifically the 1st and 99th percentiles of ages of achievement. There are two basic approaches to estimating percentiles from data such as the motor development data of the MGRS: logistic marginal models and failure time models [12]. A disadvantage of logistic marginal models is that they do not account adequately for age-related changes in the likelihood of achieving targeted milestones. The expert group [8] recommended failure time models for the analysis because these models allow probabilities (or hazards) of achieving milestones to vary with age. The application of failure time models requires that a date of achieving the milestone be provided or otherwise interval censoring methods of estimation be used. We describe below the methodical process followed to estimate the lower and upper bounds of the interval based on the fieldworkers' and caretakers' reports. Once the bounds were defined, a single date within the interval was selected at random.

Combining fieldworker and caretaker information to define the most probable intervals for the first occurrence of milestones

There are two independent sources of information about the achievement of motor milestones in the MGRS. The first, by the caretaker, provides the actual date when the milestone was first observed and/or tested. The second, by the fieldworker, provides a date when the performance was first demonstrated on a scheduled visit.

The fieldworkers were trained carefully, and standardization exercises were held frequently. Assessments made by the fieldworkers were highly concordant with those of the MDS coordinator and were consistently concordant across observers, milestones and sites [11]. Although fieldworkers instructed caretakers in the correct assessment of motor milestones, the caretakers' reports are likely to be biased toward earlier dates. Thus, the estimation of the dates of achievement relied primarily on the information generated by the fieldworkers.

In most cases, the fieldworkers' reports provided a definitive window during which the milestone must have been performed for the first time. For example, if the child could not walk alone at 11 mo but did so at 12 mo, then it is likely that the child first walked alone between 11 and 12 mo. However, the child might have been uncooperative or sick, and thus relying only on the fieldworkers' reports may have resulted in too-broad intervals. In the foregoing example, had the child been uncooperative at the 11-mo assessment, we would have been forced to accept the 10-mo examination as the lower bound of the interval or, if this was also unavailable, a still earlier one, thus diminishing precision in measurement. While biased towards earlier dates, we reasoned that the caretakers' reports could nevertheless be used in selecting the most probable lower bound. In the above example, if the child was uncooperative at the

11-mo examination, we could examine when the caretaker reported that the child walked alone in deciding the most likely lower bound. If, for example, the caretaker gave a date between 11 and 12 mo, then we could, with confidence, accept 11 mo as the most probable lower bound. On the other hand, if the caretaker gave a date between 10 and 11 mo and the fieldworker had not observed that the child walked at 10 mo, then 10 mo was accepted as the lower bound. Thus, in these and other types of cases, the information from the caretaker was very helpful in selecting the most probable lower bound of the age interval during which the milestone was achieved. However, we used only those records based on testing by the caretaker, i.e. we disregarded reports that were based on recall.

The sample from the five sites that collected motor development data used to generate the windows of achievement consisted of 816 children whose mothers complied with the MGRS feeding and no-smoking criteria and were followed until 24 mo of age. These, together with similarly compliant children from Brazil, were included in the sample for generating the physical growth standards [5].

In 69.5% of cases for sitting without support, and 78 to 90% of cases for the other milestones, available data indicated that the milestone observed in visit X (index visit) had been absent in visit X-1 (immediate prior visit). This established with a high degree of certainty that the milestone was achieved sometime between these two visits, an interval of approximately 1 mo prior to age 12 mo and 2 mo thereafter, reflecting the data collection schedule. In these cases, there was no need to consider the caretakers' reported dates to define the interval. Conversely, all other types of cases described below required the use of the caretakers' reports.

In a few instances, the assessment at visit X-1 was coded as "refusal" (1–12% of cases) or "unable to test" (1–7% of cases). In these instances, if the caretaker's date was after the X-1 examination, then the date of the X-1 examination was accepted as the lower bound of the interval, or if the caretaker's report preceded the X-1 examination, the date of the X-2 examination was taken as the lower bound.

In 2 to 3% of cases, the immediate prior assessment, X-1, was missing but X-2 was available. In these instances, the caretaker's report was used to determine whether the examination date for X-1 or X-2 should be used as the lower bound, depending upon whether the caretaker's reported date was after or before the date of the X-1 visit, respectively. In less than 1% of cases, the earliest available examination was X-3 or even earlier; the same procedure was followed as in the case where X-2 was the earliest examination available for selecting the lower bound.

The last type of situation is where the milestone was observed on the very first examination made of the child. This occurred in 26.5% of cases for sitting without support and in 0.1 to 5% of children for the other milestones. Many children demonstrated the ability to sit without support by 5 mo, the age at which the motor assessments by the fieldworkers began. For the other milestones, the cases in this category include a few instances of precocious performances, but mostly they were situations where the first assessment occurred between 6 and 14 mo of age because, due to funding constraints, the motor development assessments began later than other components of the MGRS in some sites (Ghana and Norway). At the 4-mo visit, the caretakers were informed about the motor development study, instructed on the criteria for assessing the milestones and given the form for recording the dates of achievement [9]. Only four caretakers claimed at the 4-mo visit that their children could already sit without support, which was verified and recorded by the fieldworkers. We used 3 mo as the lower bound in these cases since, based on the literature [13–17], it is highly unlikely that the child would have sat without support earlier than 3 mo of age. In cases where the motor milestone was demonstrated in the first visit at 5 mo, we accepted 4 mo as the lower bound because 99% of caretakers reported a date of achievement after 4 mo. In instances where the milestone was exhibited at the first testing occurring at 6 mo of age or later, we used the caretaker's report of a tested performance to select the lower bound in the manner described previously.

Some 35 children (4.3%) were never observed to crawl on hands and knees, and thus were not included in the analysis of this milestone. Other studies also report that this milestone is sometimes not performed and that instead some other type of locomotion is used, such as bottom shuffle or crawling on the stomach, as was observed in the MGRS [18–20]. There were also a few children who had still not met the criteria for certain milestones at 24 mo; in other words, who were right censored when the motor milestone assessment ended. This occurred in five children (0.6%) for walking with assistance, 17 (2.1%) for standing alone and 22 (2.7%) for walking alone. An age of achievement could not be calculated for these children because they are right censored; however, they were coded as such and included in the analysis to generate the windows of achievement.

The results of the above procedures are summarized in Table I. It was possible to define an interval for 97 to 100% of cases depending on the milestone. Also shown are the cases that were right censored.

Selecting failure time models with the best fit for the estimation of percentiles

Failure time models were applied to estimate percentiles using the cases shown in Table I. The hazard function in failure time models specifies instantaneous expected rates of achievement for children with an unachieved targeted milestone at age t. The hazard function fully specifies the distribution of t and simultaneously determines both density and survivor functions. There are five possible specifications of the distribution that are commonly evaluated. The simplest approach is to assume that the "hazard" is constant over time, and thus that failure times have an exponential distribution. Other approaches are the Weibull and the generalized gamma distributions, which are generalizations of the exponential distribution, and the log-normal and log-logistic distributions that use the log transformation of the failure (achievement) time. This set of five distributions is commonly referred to as a family of parametric failure time models [12]. They allow closed-form expressions of tail probabilities, provide simple formulae for survivor and hazard functions (e.g. exponential and Weibull), and can adapt to a diverse range of distributional shapes (e.g. generalized gamma). Also, these parametric models can estimate survival (achievement) times and residuals, i.e. differences between observed and predicted values [12].

The LIFEREG procedure in SAS was used to fit all the models. When using the interval-censoring estimation, an iterative algorithm developed by Turnbull [21] was used to compute a non-parametric maximum likelihood estimate of the cumulative distribution function.

Goodness-of-fit criteria were used in selecting the best models (i.e. the best distribution) for each milestone. One approach applied the Akaike-information (AIC) [22] and Bayesian-information criteria (BIC) [23] to assess goodness of fit, and the other applied Cox and Snell model diagnostics, which are the most widely used diagnostic residuals in the analysis of survival data [24,25]. In the case of the AIC and BIC criteria, the model providing the smallest values of these criteria is considered to have the best fit. If an appropriate model is selected, the Cox-Snell residuals should have a standard exponential distribution, i.e. with hazard function (λ) equal to one for all ages, and their cumulative hazard should be described by a straight 45° line [24]. For each milestone, the closer the residuals' fit to the straight line, the better the fit of the survival distribution to the empirical data [24,25].

The "best-fit" regression models were then used to estimate the cumulative distribution of ages of milestone achievement (measured in days) and their corresponding standard deviations using the single-draw random method to generate an age of achievement for each case where the interval was known, or to code the case as censored where an interval was not known. Achievement values for the 1st, 3rd, 5th, 10th, 25th, 50th, 75th, 90th, 95th, 97th and 99th percentiles and their corresponding 95% confidence intervals were estimated. Values corresponding to the 1st and 99th percentiles were used to construct the windows of achievement.

Results

Figure 1 presents the observed sequences of attaining the six motor milestones. In about 90% of the cases, the pattern observed followed a fixed sequence for five of the milestones (namely, sitting without support, standing with assistance, walking with assistance, standing alone and walking alone) with only hands-and-knees crawling shifting in between the earlier milestones. Of the total sample, 35 children (4.3%) did not exhibit hands-and-knees crawling.

Using the criteria of the smallest AIC and BIC values, the log-normal distribution provided the best fit for sitting without support and standing with assistance, and the log-logistic distribution provided the best fit for hands-and-knees crawling. The generalized gamma distribution fitted best for walking with assistance, standing alone and walking alone. However, use of the generalized gamma distribution led to wide 95% confidence intervals for the highest percentiles because of the distribution's high degree of sensitivity to right-censored values. This led us to turn to the second-best model, the log-logistic distribution, which had only slightly greater AIC and BIC values, and which did not result in wide confidence intervals; the log-logistic distribution also produced Cox-Snell residual plots that were nearly identical to those of the

Table I. Children for whom it was possible to define an interval, or who were right censored.

Number of children	Sitting without support	Hands-and-knees crawling	Standing with assistance	Walking with assistance	Standing alone	Walking alone
Interval defined	816	781	816	811	799	794
Right-censored interval	0	0	0	5	17	22
Total number of children	816	781 [a]	816	816	816	816

[a] Including the number of "non-crawlers" (35), the total is 816.

Pattern observed	N (%)	
1 → 2 → 3 → 4 → 5 → 6	340 (41.7)	86%
1 → 3 → 2 → 4 → 5 → 6	295 (36.1)	
1 → 3 → 4 → 2 → 5 → 6	69 (8.5)	
Other patterns	77 (9.4)	
Non-crawlers	35 (4.3)	
Total	**816 (100)**	

Milestone: **1** = sitting without support; **2** = hands-and-knees crawling; **3** = standing with assistance; **4** = walking with assistance; **5** = standing alone; **6** = walking alone

Figure 1. Observed sequences of attaining the six gross motor milestones.

generalized gamma distribution (data not shown). In summary, based on these considerations, the log-normal distribution was selected for the models for sitting without support and standing with assistance, and the log-logistic distribution was selected for the models for all other milestones.

The percentile values along with the 95% confidence intervals are given in Table II, and the windows of achievement bounded by the 1st and 99th percentiles are displayed in Figure 2. The windows of achievement overlap across the six milestones but vary in width. They are narrowest for sitting without support (5.4 mo) and standing with assistance (6.6 mo), intermediate for walking with assistance (7.8 mo) and hands-and-knees crawling (8.3 mo), and widest for walking alone (9.4 mo) and standing alone (10.0 mo). The widths of the 95% confidence intervals varied between 0.2 and 0.4 mo for the estimates of the 1st percentile and between 0.5 and 1.0 mo for the 99th percentile.

Discussion

The motor milestone study was a belated but very useful addition to the MGRS. The collection of motor development data was added to a predefined data collection scheme, specifically to the home visits programmed to collect anthropometric and related data. The periodicity of the home visits was meant to capture the faster growth in length and weight during infancy and the slower growth in the second year. It would have been more consistent also to have monthly assessments of motor development in the second year, but this would have significantly increased the data collection workload. Monthly data collection after 12 mo would have been particularly relevant for standing alone and walking alone, which were achieved later than the other milestones. While sitting without support, a milestone which all study children achieved by 9 mo and was therefore entirely monitored at monthly intervals, had the smallest 95% confidence

Table II. Estimated percentiles and mean (SD) in days and months for the windows of milestone achievement.

	Sitting without support	
Percentile	Days (95% CI)	Months [a] (95% CI)
1st	115 (112, 118)	3.8 (3.7, 3.9)
3rd	125 (123, 128)	4.1 (4.0, 4.2)
5th	131 (128, 134)	4.3 (4.2, 4.4)
10th	140 (138, 143)	4.6 (4.5, 4.7)
25th	158 (155, 160)	5.2 (5.1, 5.3)
50th	179 (177, 181)	5.9 (5.8, 6.0)
75th	204 (201, 207)	6.7 (6.6, 6.8)
90th	229 (225, 233)	7.5 (7.4, 7.6)
95th	245 (240, 250)	8.0 (7.9, 8.2)
97th	256 (251, 262)	8.4 (8.2, 8.6)
99th	279 (272, 286)	9.2 (8.9, 9.4)
Mean (SD)	182 (35)	6.0 (1.1)
	Standing with assistance	
Percentile	Days (95% CI)	Months [a] (95% CI)
1st	147 (144, 151)	4.8 (4.7, 5.0)
3rd	160 (156, 163)	5.2 (5.1, 5.4)
5th	167 (164, 170)	5.5 (5.4, 5.6)
10th	178 (175, 182)	5.9 (5.8, 6.0)
25th	200 (197, 203)	6.6 (6.5, 6.7)
50th	226 (223, 229)	7.4 (7.3, 7.5)
75th	256 (253, 260)	8.4 (8.3, 8.5)
90th	287 (282, 292)	9.4 (9.3, 9.6)
95th	307 (301, 313)	10.1 (9.9, 10.3)
97th	320 (314, 327)	10.5 (10.3, 10.7)
99th	348 (339, 356)	11.4 (11.1, 11.7)
Mean (SD)	230 (43)	7.6 (1.4)
	Hands-and-knees crawling	
Percentile	Days (95% CI)	Months [a] (95% CI)
1st	157 (152, 162)	5.2 (5.0, 5.3)
3rd	177 (172, 181)	5.8 (5.7, 5.9)
5th	187 (183, 191)	6.1 (6.0, 6.3)
10th	202 (198, 206)	6.6 (6.5, 6.8)
25th	226 (223, 229)	7.4 (7.3, 7.5)
50th	254 (250, 257)	8.3 (8.2, 8.4)
75th	284 (280, 289)	9.3 (9.2, 9.5)
90th	319 (313, 325)	10.5 (10.3, 10.7)
95th	345 (337, 352)	11.3 (11.1, 11.6)
97th	364 (355, 373)	12.0 (11.7, 12.3)
99th	409 (397, 422)	13.5 (13.0, 13.9)
Mean (SD)	259 (51)	8.5 (1.7)
	Walking with assistance	
Percentile	Days (95% CI)	Months [a] (95% CI)
1st	181 (176, 186)	5.9 (5.8, 6.1)
3rd	200 (196, 205)	6.6 (6.4, 6.7)
5th	210 (206, 214)	6.9 (6.8, 7.0)
10th	225 (222, 229)	7.4 (7.3, 7.5)
25th	249 (246, 252)	8.2 (8.1, 8.3)
50th	275 (272, 278)	9.0 (8.9, 9.1)
75th	304 (300, 308)	10.0 (9.9, 10.1)
90th	336 (331, 341)	11.0 (10.9, 11.2)
95th	360 (353, 367)	11.8 (11.6, 12.0)
97th	378 (370, 386)	12.4 (12.1, 12.7)
99th	418 (407, 429)	13.7 (13.4, 14.1)
Mean (SD)	279 (45)	9.2 (1.5)

Table II (*Continued*)

Percentile	Days (95% CI)	Months [a] (95% CI)
	Standing alone	
1st	211 (205, 217)	6.9 (6.7, 7.1)
3rd	235 (230, 241)	7.7 (7.6, 7.9)
5th	248 (243, 253)	8.1 (8.0, 8.3)
10th	266 (262, 271)	8.8 (8.6, 8.9)
25th	296 (292, 300)	9.7 (9.6, 9.9)
50th	330 (326, 333)	10.8 (10.7, 11.0)
75th	367 (362, 371)	12.0 (11.9, 12.2)
90th	408 (401, 415)	13.4 (13.2, 13.6)
95th	438 (429, 447)	14.4 (14.1, 14.7)
97th	461 (451, 472)	15.2 (14.8, 15.5)
99th	514 (500, 529)	16.9 (16.4, 17.4)
Mean (SD)	334 (57)	11.0 (1.9)
	Walking alone	
Percentile	Days (95% CI)	Months [a] (95% CI)
1st	250 (244, 256)	8.2 (8.0, 8.4)
3rd	274 (269, 279)	9.0 (8.8, 9.2)
5th	286 (281, 291)	9.4 (9.2, 9.6)
10th	304 (300, 309)	10.0 (9.9, 10.1)
25th	333 (330, 337)	11.0 (10.8, 11.1)
50th	365 (362, 369)	12.0 (11.9, 12.1)
75th	400 (395, 404)	13.1 (13.0, 13.3)
90th	438 (432, 444)	14.4 (14.2, 14.6)
95th	466 (458, 474)	15.3 (15.0, 15.6)
97th	487 (478, 497)	16.0 (15.7, 16.3)
99th	534 (521, 547)	17.6 (17.1, 18.0)
Mean (SD)	368 (54)	12.1 (1.8)

[a] The calculation in months involves the division of the estimate in days by 30.4375.

intervals around percentile estimates, the confidence intervals for all other milestones were similar, suggesting that a 2-mo interval did not introduce much error variance relative to monthly assessments.

The data generated by our design were analysed using appropriate statistical methods and employed failure time models that fitted the data appropriately. To prepare the data for analysis, an approach was followed that took into account the strengths and weaknesses of the two available sources of information: the fieldworkers' assessments and the caretakers' reports. The fieldworkers' examinations only established whether or not the children met the performance criteria for a milestone on given days. However, the fieldworkers were very well trained and standardized, and their assessments were consequently very reliable [11]. The caretakers reported an "exact" date when they observed a child perform a milestone. The level of error was reduced by accepting only those reports that were backed by a direct assessment by caretakers. Despite efforts to standardize the study's hundreds of caretakers involved in the assessment of motor milestones, their reports were likely biased towards earlier dates of achievement. This is understandable because caretakers take great pleasure in and are reassured by their children's development. Hence, it would have been inappropriate to accept the caretakers' dates as true dates. Instead, we used the caretakers' reports in selecting the probable lower bound of the interval during which the milestone must have occurred in cases where either we lacked a lower bound (left censored) or an examination was not available at the home visit immediately preceding the assessment by the fieldworker. To have ignored the caretakers' reports would have led to wider intervals than were used in the analyses and to less precise estimates of percentiles. The approach followed effectively

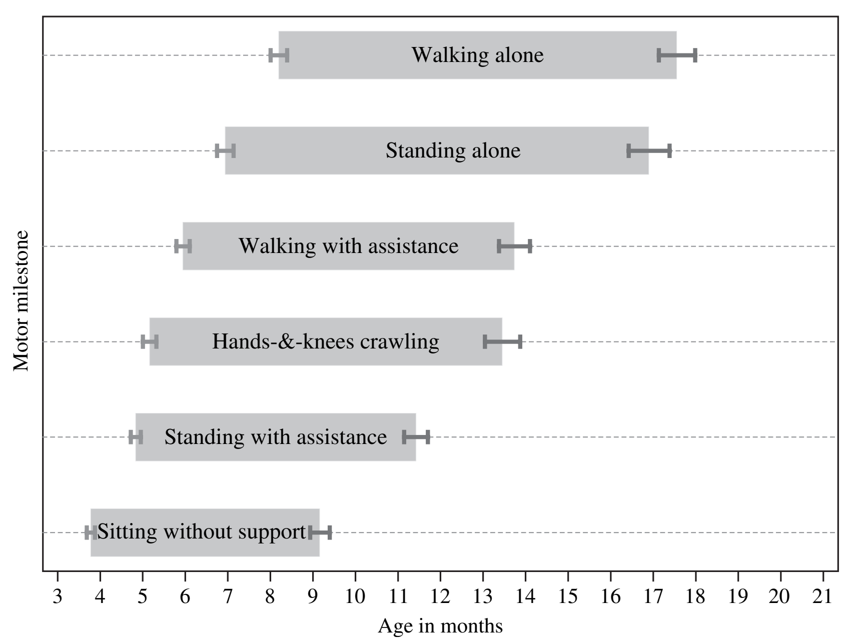


Figure 2. Windows of milestone achievement expressed in months.

integrated the two sources of information, such that the resulting pooled information is superior to what would have been obtained had we relied on either one alone.

Having identified the most likely interval within which a milestone was first exhibited, we were confronted with several data specification alternatives. One approach was to pick the mid-point of the interval as an estimate of the date of achievement. We explored this option, but it concentrated achievement ages at mid-month dates in the first year and at the odd months in the second year; as a result, the cumulative distribution functions had a stairway shape, which is an unnatural distribution. This led us to select the random draw method but using a single draw because averages of many draws will centre on the mid-point and also lead to stairway distributions. We also explored the use of interval censoring techniques, which require that only the lower and upper bounds of the interval be specified, in addition to left- and right-censored cases; we found that the model parameters generated were similar to those obtained using the random draw methods. An advantage of the single draw method is that it provides dates of achievement for each child, except for those who had not reached certain milestones by 24 mo, when the motor development study ended. These dates are convenient for many kinds of analyses.

The main products of the MDS are the windows of achievement, bounded by the 1st and 99th percentiles only and without any internal demarcations. This is to emphasize that variations within these windows, 5 to 10 mo wide, are to be taken as normal variation. All normal children will eventually reach these milestones within these windows (except for those few that will not crawl on hands and knees). We also provide estimates for other percentiles as these may be useful to researchers. We report median ages of achievement and corresponding standard deviations, which will allow the calculation of population z scores (i.e. (median age of achievement in index population - MGRS median age of achievement) / standard deviation of the MGRS). These z scores will describe differences in median ages of achievement with respect to the WHO standard and facilitate comparisons across study populations.

The foregoing reference to the MGRS windows as a standard is twofold. First, the windows have been constructed using a healthy sample, selected according to the same criteria that would ensure overall health and well-being, optimal growth and, presumably, development. Second, it avoids confusion in the use of terminology that would likely result from positioning them as a reference within the WHO Child Growth Standards. However, as explained later in this discussion, their proposed application is more restrictive than that of the physical growth standards. The windows are recommended for descriptive comparisons among populations, to signal the need for appropriate screening when individual children appear to be late in achieving the milestones, and to call attention to the importance of overall development in child health.

A number of motor development screening scales are available in the literature [13–17,26–28]. Comparing those with the MGRS windows of achievement proved to be a difficult task as the screening scales varied considerably in study design (most being based on cross-sectional studies), method of data collection, periodicity of assessments, measurement of the milestones (e.g. pass/fail versus a grading scale of achievement), criteria for defining milestone achievement, origin of study population, sample size and statistical procedures for estimating percentiles. For example, Griffith's developmental scale for the first 2 y of life was based on a small cross-sectional observational study conducted in the early 1950s [13]. The DENVER II [16] study used quota sampling to select 2096 healthy full-term children, sampled in 12 age groups between 1 wk through to age 6 y, recruited from well-child clinics, paediatricians, family physicians, hospital birth records, childcare centres and private sources. Very few studies assessed children's achievement longitudinally; the most relevant is a 3-y follow-up from birth conducted in 1960–1962 by Neligan and Prudham [15] that included two of the MGRS motor milestones: sitting without support ($n = 3831$) and walking alone ($n = 3554$). The average frequency of contact by health visitors was about six times during the first year and twice during each of the next two years. The percentile values were calculated on the assumption that the recorded age was the mid-point of the actual age interval during which the child performed the milestone.

Differences in the methods applied to report milestone achievement are also important. Some studies report cumulative frequencies (i.e. percentage of infants who pass an item at a given age) as empirical estimates [13], while others derive model-based estimates [16] with corresponding 95% confidence limits [17].

More recent scales [26,27] have been designed to provide a combined evaluation of a child's status of mental and psychomotor development. Similarly, AIMS [28] has four separate sets of items corresponding to four positions in which infants are assessed (i.e. prone, supine, sitting and standing). Such scales assess items based on *a priori* criteria, added up to provide a quantitative summary score that is compared against "cut scores" or boundaries to determine the child's level of risk. Sometimes scores are also converted to percentile ranks, indicating the infant's position relative to the normative sample; the lower the percentile, the less mature the infant's motor development. Although these scales are based on items used extensively in longitudinal research studies, they require careful observation of the child's behaviour by examiners who must be thoroughly trained to use the materials and procedures of the scale tests. Moreover, interpretation of the scores is often not straightforward.

Despite these methodological differences across studies, there are noteworthy commonalities between existing scales and the MGRS windows of achievement. All of them could not identify appreciable differences between boys and girls, and consequently pooled the sexes in reporting results. Similarly, where available, the average ages of milestone achievement are comparable to those of the MGRS, except for Griffith [13] which has later median ages of achievement than all other scales. For example, median ages in months for sitting without support are 8.0 [13], 6.6 [14], 6.4 [15], 5.9 [16] and 6.5 [17] compared to 5.9 mo in the MGRS. For walking alone, median ages in months are 14 [13], 11.7 [14], 12.8 [15], 12.3 [16] and 12.4 [17] compared to 12.0 mo in the MGRS. Despite different percentile ranges available from published sources, it would appear that the MGRS windows (1st to 99th percentile) are the widest, except again for Griffith's [13]. A noteworthy feature of some of the distributions is the marked skewness of the upper tail for some of the available scales. For example, for walking alone in the Bayley-I [14] and the Neligan and Prudham [15] scales, the difference between the 50th and the 95th percentiles is about double the difference between the 5th and 50th percentiles.

The MGRS windows of achievement have been constructed to depict the range in ages of achievement of key motor milestones in healthy children from around the world. Surveys of child health rarely collect data on motor milestones, and this information is not routinely assessed in child growth clinics. We hope that interest in the motor development of children will increase now that the MGRS windows of achievement are available for surveillance and monitoring of individuals and populations. At the individual level, the windows can be used to detect, on a single visit or on repeated assessments, whether substantial developmental delays occur, as indicated by ages of achievement outside the windows. At any age after 9 mo, one can easily compare a child's actual performance to what should have been demonstrated at that age using the windows of achievement. The reason why such comparisons cannot be carried out earlier than 9 mo is that the earliest closure of an achievement window, specifically for sitting without support, occurs at 9.4 mo. From a population point of view, the analyses are more complex and will depend on whether the data are longitudinal or, more commonly, cross-sectional. Cross-sectional surveys of young children, preferably from 3 to 24 mo of age, or even later if growth retardation is significant in the population under study, can collect data on which milestones are demonstrated by each child in the survey; statistical methods can then be applied to these cross-sectional data to generate windows of achievement for the population under study and compare them to those of the MGRS. The greater the displacement to the right relative to the MGRS windows, the greater will be the degree of motor development retardation in the population under study. For research purposes, population z scores—estimated as the difference in the 50th percentile of the population under study with respect to the MGRS median, relative to the MGRS SD—would be a useful metric for analysing population surveys that collect motor development data. On the other hand, for reasons discussed below, we do not recommend calculating z scores for individuals.

More simply, the percentage of children failing to achieve one or more milestones expected for their age can be reported. This last analysis will be very sensitive to the ages of the children included in the survey. By definition, children younger than 9 mo will never be found to fail; at the other extreme, inclusion of many older children, for example 24- to 36-mo-olds, will also lower the percentage of children with delays, as even children who are significantly retarded in motor development eventually perform them. A reasonable age range for this type of simpler analysis and reporting is 9 to 24 mo of age. For obvious reasons, comparisons of populations, such as those representing different regions of a country, will be valid only if the same age range of children is used in sampling all populations under consideration.

While it is simple to calculate a percentile or z-score value for the physical growth indicators, percentile values or z scores of motor development for an individual child would be extremely difficult, if not impossible, to generate. This is because the MGRS standards depict the variation in first age of achievement, something one cannot measure in a survey. If a child has not reached a milestone on the date of a survey, we do not know when he or she will, and thus we have only limited information about the child's motor development. If we assess two children today who have not reached a particular milestone, one might reach the milestone tomorrow and the other in 3 mo, but they would appear identical to us today with respect to the milestone of interest. Also, for any two children who exhibit the milestone on the day of the survey, we would be unable to differentiate between them with regard to development because we would not know when they performed the milestone for the first time. In contrast, z scores are easy to estimate for physical growth for any individual child and can be assessed at any age. This is because measures such as length or weight are measures of achieved status on any particular day. The use of windows of achievement, therefore, leads us simply to compare the child's performance today to the windows of achievement and to ask the most meaningful question possible: What milestones should a child of this age have reached by now? Concern would be expressed only if the child has not performed one or more milestones that he or she should have and, ideally, the assessment should be based on repeated evaluations over time.

Acknowledgements

This paper was prepared by Reynaldo Martorell, Mercedes de Onis, Jose Martines, Maureen Black, Adelheid Onyango and Kathryn G. Dewey on behalf of the WHO Multicentre Growth Reference Study Group. The statistical analysis was conducted by Amani Siyam.

References

[1] de Onis M, Garza C, Victora CG, Bhan MK, Norum KR, editors. WHO Multicentre Growth Reference Study (MGRS): Rationale, planning and implementation. Food Nutr Bull 2004;25 Suppl 1 :S1–89.
[2] de Onis M, Garza C, Victora CG, Onyango AW, Frongillo EA, Martines J, for the WHO Multicentre Growth Reference Study Group. The WHO Multicentre Growth Reference Study: Planning, study design, and methodology. Food Nutr Bull 2004;25 Suppl 1:S15–26.
[3] WHO. Measuring change in nutritional status. Geneva: World Health Organization; 1983.
[4] WHO Multicentre Growth Reference Study Group. Assessment of differences in linear growth among populations in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450:56–65.
[5] WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards based on length/height, weight and age. Acta Paediatr Suppl 2006;450:76–85.
[6] WHO Multicentre Growth Reference Study Group. Assessment of sex differences and heterogeneity in motor milestone attainment among populations in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450:66–75.
[7] WHO Multicentre Growth Reference Study Group. Relationship between physical growth and motor development in the WHO Child Growth Standards. Acta Paediatr Suppl 2006; 450:96–101.
[8] WHO. Motor Development Study. Report of an expert meeting. Geneva: World Health Organization; 2004.
[9] Wijnhoven TM, de Onis M, Onyango AW, Wang T, Bjoerneboe GE, Bhandari N, et al., for the WHO Multicentre Growth Reference Study Group. Assessment of gross motor development in the WHO Multicentre Growth Reference Study. Food Nutr Bull 2004;25 Suppl 1:S37–45.
[10] WHO Multicentre Growth Reference Study Group. Enrolment and baseline characteristics in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450: 7–15.
[11] WHO Multicentre Growth Reference Study Group. Reliability of motor development data in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450:47–55.
[12] Kalbfleisch JD, Prentice R L. The statistical analysis of failure time data. 2nd ed. Wiley series in probability and theory. New Jersey: John Wiley & Sons, Inc.; 2002.
[13] Griffiths R. The abilities of babies. New York: McGraw-Hill Book Co. Inc; 1954.
[14] Bayley N. Manual of the Bayley Scales of Infant Development. San Antonio: Psychological Corporation; 1969.
[15] Neligan G, Prudham D. Norms for four standard developmental milestones by sex, social class and place in family. Dev Med Child Neurol 1969;11:413–22.
[16] Frankenburg WK, Dodds J, Archer P, et al. The DENVER II training manual. Denver: Denver Developmental Materials, Inc.; 1992.
[17] Lejarraga H, Krupitzky S, Kelmansky D, Martínez E, Bianco A, Pascucci MC, et al. Edad de cumplimiento de pautas de desarrollo en niños argentinos sanos menores de seis años. J Pediatr (Rio J) 1997;73 Suppl 1:521–32.
[18] Baker D, Vanace P W. Motor and intellectual development. In: Kaye R, Oski FA, Barness LA, editors. Core textbook of pediatrics. 3rd ed. Philadelphia: J. B. Lippincott Company; 1988. p. 23–41.
[19] Hellbrugge T, Pohl P. Munich functional diagnostic tests and early behavioural diagnosis. In: Udani PM, editor. Textbook of pediatrics. With special reference to problems of child health in developing countries. New Delhi: Jaypee Brothers; 1991. p. 136–43.
[20] Illingworth R S. The normal child: Some problems of the early years and their treatment. 10th ed. Edinburgh: Churchill Livingston; 1991. p. 127–65.
[21] Turnbull B W. The empirical distribution function with arbitrarily grouped, censored and truncated data. J R Stat Soc Ser B 1976;38:290–5.
[22] Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr 1974;19:716–23.
[23] Schwarz G. Estimating the dimension of a model. Ann Stat 1978;6:461–4.
[24] Collett D. Modelling survival data in medical research. Texts in statistical science. London: Chapman & Hall; 1994.
[25] Cox DR, Snell E J. A general definition of residuals. J R Stat Soc Ser B 1968;30:248–75.
[26] Bayley N. Bayley Scales of Infant Development. 2nd ed. manual. San Antonio: The Psychological Corporation; 1993.
[27] Aylward G P. The Bayley Infant Neurodevelopmental Screener. San Antonio: The Psychological Corporation; 1995.
[28] Piper MC, Darrah J. Motor assessment of the developing infant. Philadelphia: WB Saunders Co; 1994.

Acta Pædiatrica, 2006; Suppl 450: 96–101

Relationship between physical growth and motor development in the WHO Child Growth Standards

WHO MULTICENTRE GROWTH REFERENCE STUDY GROUP[1,2]

[1]*Department of Nutrition, World Health Organization, Geneva, Switzerland, and* [2]*Members of the WHO Multicentre Growth Reference Study Group (listed at the end of the first paper in this supplement)*

Abstract

Aim: To examine relationships among physical growth indicators and ages of achievement of six gross motor milestones in the WHO Child Growth Standards population. *Methods:* Gross motor development assessments were performed longitudinally on the 816 children included in the WHO Child Growth Standards. Six milestones (sitting without support, hands-and-knees crawling, standing with assistance, walking with assistance, standing alone, walking alone) were assessed monthly from 4 until 12 mo of age and bimonthly thereafter until children could walk alone or reached 24 mo. Failure time models were used 1) to examine associations between specified ages of motor milestone achievement and attained growth z scores and 2) to quantify these relationships as delays or accelerations in ages of milestone achievement. *Results:* Statistically significant associations were noted between ages of achievement of sitting without support and attained weight-for-age, weight-for-length and BMI-for-age z scores. An increase of one unit z score in these indicators was associated with 3 to 6 d acceleration in the respective achievement age. Statistically significant associations also were noted between various milestone achievement ages and growth when 3- or 6-mo and birth length-for-age z scores were entered jointly in the failure time models. In these analyses, one unit z-score increase in length-for-age was associated with 1 to 3 d delay in the respective achievement age.

Conclusion: Sporadic, significant associations were observed between gross motor development and some physical growth indicators, but these were quantitatively of limited practical significance. These results suggest that, in healthy populations, the attainment of these six gross motor milestones is largely independent of variations in physical growth.

Key Words: *Childhood growth, gross motor milestones, growth standards, young child development*

Introduction

The WHO Child Growth Standards include descriptions of the physical growth and ages of achievement of universally recognized gross motor milestones in healthy infants and children throughout the world. The sample used to construct the growth standards consists of a sub-sample of children included in the WHO Multicentre Growth Reference Study (MGRS). The MGRS adopted a "prescriptive" approach designed to describe how children *should* grow rather than *how* children grew at a particular time and place. In so doing, it broadened the definition of health to include the adoption of several practices associated with healthy outcomes, e.g. breastfeeding and non-smoking. The rationale for the MGRS and its design and protocol are described in detail elsewhere [1,2].

The second unique feature of the MGRS was its inclusion of children from six of the world's major regions, i.e. Brazil (South America), Ghana (Africa), India (Asia), Norway (Europe), Oman (the Middle East) and the USA (North America). This design feature tested the assumption that growth in infancy and early childhood is very similar among diverse ethnic groups when conditions that do not constrain growth are met [3,4]. The MGRS also offered the possibility to assess the heterogeneity/similarity in gross motor development across distinct cultural groups and environments. It demonstrated that, although some differences were observed in the ages of gross motor milestone achievement among study sites, they were not consistent and likely reflected diverse culture-specific care practices rather than inherent biological differences [5].

Correspondence: Mercedes de Onis, Study Coordinator, Department of Nutrition, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland. Tel: +41 22 791 3320. Fax: +41 22 791 4156. E-mail: deonism@who.int

ISSN 0803-5326 print/ISSN 1651-2227 online
DOI: 10.1080/08035320500495589

The longitudinal and simultaneous assessments of physical growth and gross motor development also provided an opportunity to examine associations between physical growth and gross motor development in healthy children. Studies are published demonstrating the effects of diverse diseases and conditions on motor development [6–9], links between motor delays and various forms of general and specific under-nutrition [10–16], positive links between motor development and exclusive breastfeeding [17], and improved linear growth in undernourished children who undergo nutritional rehabilitation coupled with a physical activity regimen rather than only nutritional rehabilitation [18,19]. Published assessments of associations between physical growth and motor milestone achievement among well-nourished, healthy children are fewer (e.g. [17]), and to our knowledge none has a sample as large or as diverse as that of the WHO Child Growth Standards.

The objective of this paper is to examine relationships among attained weight-for-age, length-for-age, body mass index (BMI)-for-age, and weight-for-length z scores and ages of achievement of specified gross motor milestones in the growth standards' sample of healthy breastfed infants and young children who enjoyed living conditions that did not constrain linear growth.

Methods

Study design

The rationale, planning, design and methods of the MGRS, including its motor development component and site-specific protocol implementation, are described in detail elsewhere [1,2,20]. Briefly, in five of the six MGRS sites, i.e. Ghana, India, Norway, Oman and the USA, gross motor development assessments were performed longitudinally beginning at 4 mo of age on subjects enrolled in the MGRS longitudinal component. Motor development assessments were not performed in Brazil because most of this site's longitudinal sample was older than 4 mo when motor development was added to the MGRS protocol. Six distinct gross motor milestones were assessed: sitting without support, hands-and-knees crawling, standing with assistance, walking with assistance, standing alone and walking alone. These were selected because they are considered universal, fundamental to the acquisition of self-sufficient locomotion, and simple to test and evaluate. All milestones were assessed using standardized testing procedures and criteria [20], and were performed by study staff monthly from 4 until 12 mo of age and bimonthly thereafter until children could walk alone or reached 24 mo of age. Training and standardization procedures were similar among sites. The criteria used to document the attainment of the six milestones were applied with equally high levels of reliability among observers within a site, among milestones within a site, and among sites across milestones [21]. No fixed milestone sequence was assumed, and all milestones were assessed at each visit.

Study sample

The sample used for these analyses consisted of 816 children included in the generation of the physical growth standards [22]. By study site, the sample included 227 children from Ghana, 173 from India, 148 from Norway, 149 from Oman and 119 from the USA.

Statistical analyses

Ages of gross motor milestone achievement. The MGRS design [2] did not permit the determination of exact ages of milestone achievement because subjects were not supervised daily by trained staff. "True" ages of milestone achievement were linked to intervals between the visit documenting the achievement of specific milestones and the most recent previous visit. Specific ages of achievement within designated intervals were assigned randomly based on the assumption that achievement ages were distributed uniformly within the interval [23].

Ages of milestone achievement were modelled using failure time analysis and either log-normal or log-logistic distributions, as appropriate. For the small percentages of children who were not observed by field staff to have achieved specific milestones (walking with assistance 0.6%; standing alone 2.1%; and walking alone 2.7%) by 24 mo, i.e. the visit date that terminated longitudinal follow-up, the ages of milestone achievement were right censored. For these same children, primary caretakers had reported that, by 24 mo, 80% walked with assistance, 94% stood alone and 55% walked alone. However, only the information reported by trained staff was used in estimating ages of achievement [23].

Estimation of attained weight-for-age, length-for-age, weight-for-length and BMI-for-age at milestone achievement. Z scores for attained weight-for-age, length-for-age, BMI-for-age and weight-for-length were based on the WHO Child Growth Standards [22]. Z scores corresponding to specific anthropometric measurements at ages of milestone achievement were estimated by linear interpolation of weight or length. Interpolations were bounded by the intervals used to assign ages of milestone achievement as described above. Z scores were calculated for interpolated weight and length values.

Analyses of links among gross motor milestones and growth. Failure time models were used to examine associations between assigned ages of achievement of gross motor milestones and attained growth z scores. Z scores for weight, length, BMI and weight-for-length at birth, 3 mo, 6 mo and at the ages of achievement of the gross motor milestones were added individually or jointly to "best-fitting" failure time models [23].

Associations were evaluated between z scores of attained anthropometric measurements at birth and ages of gross motor milestone achievement, and between z scores at birth and 3 mo (for the milestones sitting without support, hands-and-knees crawling and standing with assistance) or 6 mo (for the milestones walking with assistance, standing alone and walking alone) and ages of milestone achievement. These ages were selected arbitrarily to assess relationships among ages of milestone achievement and growth attained both at the age of achievement and at ages proximal to the attainment of the respective milestones.

Achievement ages were considered as failure times. The log-normal distribution provided the best fit for sitting without support and standing with assistance, and the log-logistic distribution the best fit for hands-and-knees crawling, walking with assistance, standing alone and walking alone. Failure time models were also used to quantify the relationships as delays or accelerations in ages (in days) of gross motor milestone achievement. Statistical significance was set at $\alpha = 0.05$.

Results

Table I summarizes the statistical significance of associations between ages of achievement of the six gross motor milestones and weight-for-age, length-for-age, weight-for-length and BMI-for-age z scores at birth and/or at the ages of milestone achievement. Significant associations were observed only for sitting without support and limited to anthropometric indicators that included weight. Thus, associations were noted between ages of achievement of sitting without support and attained weight-for-age, weight-for-length and BMI-for-age z scores. The table also includes estimates of the increments (+) or decrements (−) in average ages of achievement (in days) per one unit z-score increase in the respective anthropometric indicator for which statistically significant associations were detected (see also Figure 1).

Table II summarizes associations between ages of achievement of the six gross motor milestones and weight-for-age, length-for-age, weight-for-length and BMI-for-age z scores at birth and/or at 3 mo for the milestones sitting without support, hands-and-knees crawling and standing with assistance, or birth and/or 6 mo for the milestones walking with assistance, standing alone and walking alone. Statistically significant associations were noted most often for sitting without support; however, unlike associations summarized in Table I, when z scores at birth and 3 or 6 mo were added jointly in the analytical model, statistically significant associations with length-for-age z scores were also noted for all milestones but walking

Table I. Associations between attained growth and ages of motor milestone achievement at birth and ages of milestone achievement.

Z scores based on the WHO Child Growth Standards	Sitting without support	Hands-and-knees crawling	Standing with assistance	Walking with assistance	Standing alone	Walking alone
Weight-for-age						
At birth (a)	×	×	×	×	×	×
At achievement (b)	√	×	×	×	×	×
(a) + (b)	×, √ [a]	×, ×	×, ×	×, ×	×, ×	×, ×
Length-for-age						
At birth (a)	×	×	×	×	×	×
At achievement (b)	×	×	×	×	×	×
(a) + (b)	×, ×	×, ×	×, ×	×, ×	×, ×	×, ×
Weight-for-length						
At birth (a)	×	×	×	×	×	×
At achievement (b)	√	×	×	×	×	×
(a) + (b)	×, √ [b]	×, ×	×, ×	×, ×	×, ×	×, ×
BMI-for-age						
At birth (a)	×	×	×	×	×	×
At achievement (b)	√	×	×	×	×	×
(a) + (b)	×, √ [c]	×, ×	×, ×	×, ×	×, ×	×, ×

√: $p < 0.05$; ×: $p > 0.05$.
[a] One z-score increase in weight-for-age *reduces* the expected achievement age of sitting without support by approximately 3 d (2.9 d).
[b] One z-score increase in weight-for-length *reduces* the expected achievement age of sitting without support by approximately 5 d (5.1 d).
[c] One z-score increase in BMI-for-age *reduces* the expected achievement age of sitting without support by approximately 6 d (6.2 d).

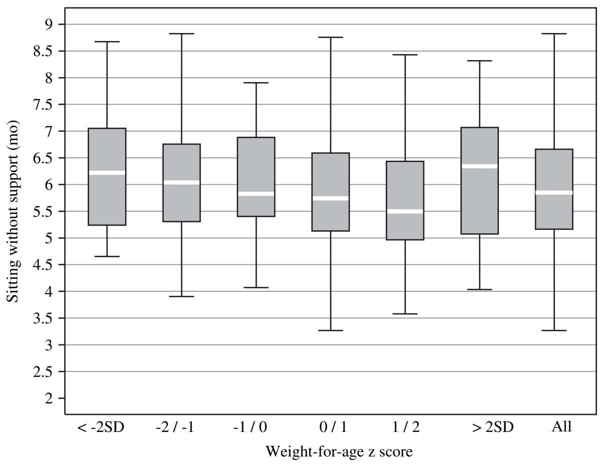


Figure 1. Ages of achievement of sitting without support for children grouped by weight-for-age z scores at achievement.[a]

[a]Horizontal bars within the respective boxes represent median ages of achievement, and the upper and lower boundaries for each box represent the 75th (P75) and 25th (P25) percentiles, respectively. The upper whisker is set at the sum of P75 and 1.5 times the difference between P75 and P25. The lower whisker is set at the difference between P25 and 1.5 times the difference between P75 and P25.

alone. The table also includes estimates of the increments (+) or decrements (−) in the average ages of achievement (in days) per one unit z-score increase in the respective anthropometric indicator for statistically significant associations, e.g. one unit z-score increase in length-for-age was associated with 1 to 3 d delay in the respective achievement age.

Discussion

These results indicate that associations between ages of gross motor milestone achievement and attained growth in healthy infants and toddlers are limited primarily to the milestone sitting without support. The exceptions to this generalization are statistically significant associations among length-for-age z scores at birth and at 3 mo of age and ages of achievement of sitting without support, hands-and-knees crawling and standing with assistance; and associations between length-for-age z scores at birth and at 6 mo and ages of achievement of walking with assistance and standing alone when these were entered jointly in failure time models. In each of those cases, however, significant associations were of limited practical significance (e.g. approximately 1 to 3 d delay in achievement ages for those milestones for which length-for-age was found to be related to ages of achievement). The increments/decrements in ages of milestone achievement associated with increments in z scores were small in both absolute terms and relative to the wide variability in the ages of milestone

Table II. Associations between attained growth and ages of motor milestone achievement at birth and 3 mo or 6 mo.

Z scores based on the WHO Child Growth Standards	Sitting without support	Hands-and-knees crawling	Standing with assistance	Walking with assistance	Standing alone	Walking alone
Weight-for-age						
At birth (a)	×	×	×	×	×	×
At age X mo (b) [a]	√	×	×	×	×	×
(a) + (b)	×,√ [b]	×, ×	×, ×	×, ×	×, ×	×, ×
Length-for-age						
At birth (a)	×	×	×	×	×	×
At age X mo (b)	×	×	×	×	×	×
(a) + (b) [c]	√, ×	√, ×	√, √	√,√	×,√	×, ×
Weight-for-length						
At birth (a)	×	×	×	×	×	×
At age X mo (b)	√	×	×	×	×	×
(a) + (b)	×,√ [d]	×, ×	×, ×	×, ×	×, ×	×, ×
BMI-for-age						
At birth (a)	×	×	×	×	×	×
At age X mo (b)	√	×	×	×	×	×
(a) + (b)	×,√ [e]	×, ×	×, ×	×, ×	×, ×	×, ×

√: $p < 0.05$; ×: $p > 0.05$

[a] Three months for milestones sitting without support, hands-and-knees crawling and standing with assistance; 6 mo for milestones walking with assistance, standing alone and walking alone.

[b] One z-score increase in weight-for-age at age 3 mo *reduces* the expected achievement age of sitting without support by approximately 4 d (3.5 d).

[c] One z-score increase in length-for-age (at birth and/or at age 3 mo) *extends* the expected achievement age of sitting without support, hands-and-knees crawling and standing with assistance by 1.4, 0.9 and 2.6 d, respectively. One z-score increase in length-for-age (at birth and/or at age 6 mo) *extends* the expected age of achievement for walking with assistance and standing alone by 2.1 and 1.5 d, respectively.

[d] One z-score increase in weight-for-length at age 3 mo *reduces* the expected achievement age of sitting without support by approximately 6 d (6.1 d).

[e] One z-score increase in BMI-for-age at age 3 mo *reduces* the expected achievement age of sitting without support by approximately 6 d (6.3 d).

achievement observed in the WHO Child Growth Standards population [23].

Relationships among anthropometric indicators and accelerations in ages of milestone achievement (related to weight-based indicators) or delays (related to length-for-age), even if small, appear to vary qualitatively in healthy populations with respect to specific motor milestones. This may reflect greater weight/length helping to sustain the balance and control necessary for sitting without support, whereas greater stature may not be advantageous with respect to mobility at later ages. Although these relationships are of inherent biological interest, their quantitative impact is likely to be of minimal practical significance in non-research settings.

These findings, coupled with published associations between motor development and states of under-nutrition [10–16] or the presence of specific diseases or conditions [6–9], suggest that observations of links between growth performance and motor development often signal past or ongoing stresses that should be evaluated and addressed. They also indicate that population-level motor development can be a robust functional indicator of various forms of stress during vulnerable developmental periods. Such population delays, however, must be assessed with care to determine possible influences of locally recommended care practices (see below).

The consistent achievement of gross motor milestones at later ages within normal "windows of achievement" likely has limited predictive value of good or bad outcomes in motor and other developmental domains for individuals within healthy populations [24,25]. The exceptions to this are infants in populations with severe deficits [26–28] such as those in special categories, e.g. extremely low-birthweight infants [29].

Equally importantly, there is no conclusive evidence in the literature that significant population-level motor delays are independently predictive of future functional delays or of other adverse outcomes (e.g. poorer cognitive performance or motor agility). For example, motor delays associated with under-nutrition may not be any more or any less predictive of other problems in subsequent development than direct measures of the severity of the co-existing under-nutrition. Motor delays thus may signal only the active impairment of normal development and not necessarily future impaired functional capacities [30]. There is ample evidence that regenerative, redundant and/or degenerative pathways often correct functional delays or may positively influence future attainment of motor capabilities [26,31,32]. Enabling regenerative, redundant and/or degenerative pathways, however, may require actively addressing under-nutrition or other aetiologies responsible for developmental delays.

It is also important to point out that consistent "delays" or "accelerations" in milestone achievement can occur among milestones that are especially susceptible to caretaker training [5]. There is no direct evidence that apparent milestone achievement delays or accelerations enabled by training have any functional significance beyond the specific milestone's achievement. Nonetheless, the acceleration of motor skill acquisition may hasten the development of other functional domains through a child's enhanced abilities to interact with the immediate environment [33,34]. Also, the accelerated attainment of certain skills may be of cultural value, e.g. field reports from Ghana in this study suggesting that mothers used several strategies to accelerate the ability of infants to sit without support so as to increase their time to attend to other tasks without having to carry the baby around [5].

In summary, growth and motor development are largely independent in healthy populations. Associations between motor development and attained growth parameters were restricted principally to sitting without support and were quantitatively of limited practical significance. Nonetheless, the universality of gross motor development and its reliable attainment within predictable age ranges among healthy populations have positive implications for using motor development standards to assess gross motor development in children at the population level and perhaps as an educational tool to reinforce the importance of development dimensions other than physical growth.

Acknowledgements

This paper was prepared by Cutberto Garza, Mercedes de Onis, Reynaldo Martorell, Anna Lartey and Kathryn G. Dewey on behalf of the WHO Multicentre Growth Reference Study Group. The statistical analysis was conducted by Amani Siyam.

References

[1] Garza C, de Onis M, for the WHO Multicentre Growth Reference Study Group. Rationale for developing a new international growth reference. Food Nutr Bull 2004;25 Suppl 1:S5–14.

[2] de Onis M, Garza C, Victora CG, Onyango AW, Frongillo EA, Martines J, for the WHO Multicentre Growth Reference Study Group. The WHO Multicentre Growth Reference Study: Planning, study design and methodology. Food Nutr Bull 2004;25 Suppl 1:S15–26.

[3] 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.

[4] Martorell R, Habicht JP. Growth in early childhood in developing countries. In: Falkner F, Tanner JM, editors.

Human growth: A comprehensive treatise. 2nd ed. New York: Plenum Press; 1986. p. 241–62.

[5] WHO Multicentre Growth Reference Study Group. Assessment of sex differences and heterogeneity in motor milestone attainment among populations in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450:66–75.

[6] Carrel AL, Moerchen V, Myers SE, Bekx MT, Whitman BY, Allen DB. Growth hormone improves mobility and body composition in infants and toddlers with Prader-Willi syndrome. J Pediatr 2004;145:744–9.

[7] Cooke RW, Foulder-Hughes L. Growth impairment in the very preterm and cognitive and motor performance at 7 years. Arch Dis Child 2003;88:482–7.

[8] Hediger ML, Overpeck MD, Ruan WJ, Troendle JF. Birthweight and gestational age effects on motor and social development. Paediatr Perinat Epidemiol 2002;16:33–46.

[9] Chase C, Ware J, Hittelman J, Blasini I, Smith R, Llorente A, et al. Early cognitive and motor development among infants born to women infected with human immunodeficiency virus. Pediatrics 2000;106:E25.

[10] Kariger PK, Stoltzfus RJ, Olney D, Sazawal S, Black R, Tielsch JM, et al. Iron deficiency and physical growth predict attainment of walking but not crawling in poorly nourished Zanzibari infants. J Nutr 2005;135:814–9.

[11] Kuklina EV, Ramakrishnan U, Stein AD, Barnhart HH, Martorell R. Growth and diet quality are associated with the attainment of walking in rural Guatemalan infants. J Nutr 2004;134:3296–300.

[12] Lozoff B, De Andraca I, Castillo M, Smith JB, Walter T, Pino P. Behavioral and developmental effects of preventing iron-deficiency anemia in healthy full-term infants. Pediatrics 2003; 112:846–54.

[13] Black MM. The evidence linking zinc deficiency with children's cognitive and motor functioning. J Nutr 2003;133: 1473S–6S.

[14] Cheung YB, Yip PS, Karlberg JP. Fetal growth, early postnatal growth, and motor development in Pakistani infants. Int J Epidemiol 2001;30:66–72.

[15] Grantham-McGregor S. A review of studies of the effect of severe malnutrition on mental development. J Nutr 1995;125: 2233S–8S.

[16] Jahari AB, Saco-Pollitt C, Husaini MA, Pollitt E. Effects of an energy and micronutrient supplement on motor development and motor activity in undernourished children in Indonesia. Eur J Clin Nutr 2000;54 Suppl 2:S60–8.

[17] Dewey KG, Cohen RJ, Brown KH, Rivera LL. Effects of exclusive breastfeeding for four versus six months on maternal nutritional status and infant motor development: results of two randomized trials in Honduras. J Nutr 2001;131:262–7.

[18] Torun B, Viteri FE. Influence of exercise on linear growth. Eur J Clin Nutr 1994;48 Suppl 1:S186–9.

[19] Viteri FE, Torun B. Nutrition, physical activity, and growth. In: Ritzen M, Aperia A, Hall K, Larson A, Zetterberg A, Zetterström R, editors. The biology of normal human growth. New York: Raven Press; 1981. p. 265–73.

[20] Wijnhoven TM, de Onis M, Onyango AW, Wang T, Bjoerneboe GE, Bhandari N, et al. for the WHO Multicentre Growth Reference Study Group. Assessment of gross motor development in the WHO Multicentre Growth Reference Study. Food Nutr Bull 2004;25 Suppl 1:S37–45.

[21] WHO Multicentre Growth Reference Study Group. Reliability of motor development data in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450: 47–55.

[22] WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards based on length/height, weight and age. Acta Paediatr Suppl 2006;450:76–85.

[23] WHO Multicentre Growth Reference Study Group. WHO Motor Development Study: Windows of achievement for six gross motor development milestones. Acta Paediatr Suppl 2006;450:86–95.

[24] Darrah J, Hodge M, Magill-Evans J, Kembhavi G. Stability of serial assessments of motor and communication abilities in typically developing infants – implications for screening. Early Hum Dev 2003;72:97–110.

[25] Bartlett DJ, Okun NB, Byrne PJ, Watt JM, Piper MC. Early motor development of breech- and cephalic-presenting infants. Obstet Gynecol 2000;95:425–32.

[26] Blasco PA. Pitfalls in developmental diagnosis. Pediatr Clin North Am 1991;38:1425–38.

[27] Erikson C, Allert C, Carlberg EB, Katz-Salamon M. Stability of longitudinal motor development in very low birthweight infants from 5 months to 5.5 years. Acta Paediatr 2003;92: 197–203.

[28] Campbell SK, Hedeker D. Validity of the Test of Infant Motor Performance for discriminating among infants with varying risk for poor motor outcome. J Pediatr 2001;139:546–51.

[29] Burns Y, O'Callaghan M, McDonell B, Rogers Y. Movement and motor development in ELBW infants at 1 year is related to cognitive and motor abilities at 4 years. Early Hum Dev 2004; 80:19–29.

[30] Majnemer A, Snider L. A comparison of development assessments of the newborn and young infant. Ment Retard Dev Disabil Res Rev 2005;11:68–73.

[31] Black MM, Dubowitz H, Hutcheson J, Berenson-Howard J, Starr RH Jr. A randomized clinical trial of home intervention for children with failure to thrive. Pediatrics 1995;95:807–14.

[32] Hadders-Algra M, Brogren E, Forssberg H. Nature and nurture in the development of postural control in human infants. Acta Paediatr Suppl 1997;422:48–53.

[33] Butler C. Effects of powered mobility on self-initiated behaviors of very young children with locomotor disability. Dev Med Child Neurol 1986;28:325–32.

[34] Butler C, Darrah J. Effects of neurodevelopmental treatment (NDT) for cerebral palsy: an AACPDM evidence report. Dev Med Child Neurol 2001;43:778–90.