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
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.

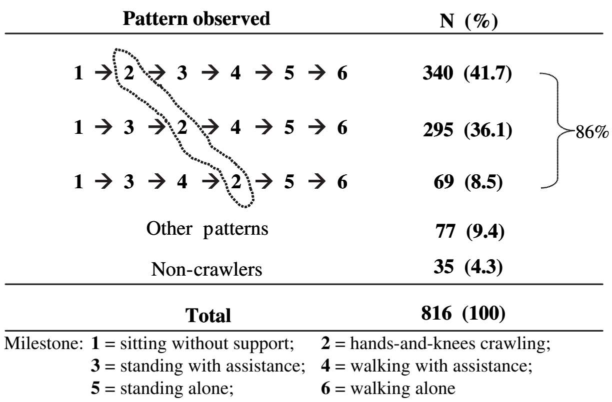

Figure 1. Observed sequences of attaining the six gross motor milestones.

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

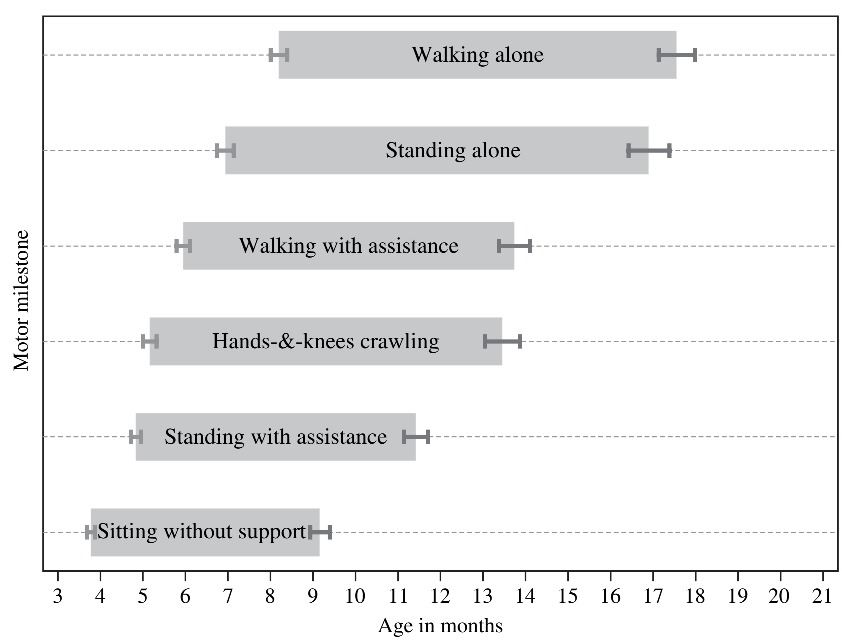

Figure 2. Windows of milestone achievement expressed in months.

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.