Counting the stunted children in a population: a criticism of old and new approaches and a conciliatory proposal

C. A. Monteiro

Two methods for estimating the prevalence of growth retardation in a population are evaluated: the classical method, which is based on the proportion of children whose height is more than 2 standard deviations below the expected mean of a reference population; and a new method recently proposed by Mora, which is based on the whole height distribution of observed and reference populations. Application of the classical method to several simulated populations leads to the conclusion that in most situations in developing countries the prevalence of growth retardation is grossly underestimated, and reflects only the presence of severe growth deficits. A second constraint with this method is a marked reduction of the relative differentials between more and less exposed strata. Application of Mora’s method to the same simulated populations reduced but did not eliminate these constraints. A novel method for estimating the prevalence of growth retardation, which is based also on the whole height distribution of observed and reference populations, is also described and evaluated. This method produces better estimates of the true prevalence of growth retardation with no reduction in relative differentials.

Introduction

It is generally agreed that cross-sectional surveys of the growth of young children provide an excellent opportunity for determining the health and nutritional status that prevail in a community (1). It is also widely accepted that stunting—or the slowing in the skeletal growth of children—reflects poor overall economic conditions, particularly chronic or repeated infections and inadequate food intake (2). What is still a matter of dispute is how to estimate the prevalence of children whose attained height is lower than that expected for their age, sex, and genetic growth potential (3–5). The various alternative procedures used are critically reviewed in this article.

In practice it is impossible to determine the “expected” height of a single child. Usually, what is available is a distribution of heights of healthy children of the same age and sex from a reference population. If it is assumed that the group of children being examined is a random sample of the reference population, i.e., that the latter is an appropriate standard for the study population, it should be possible to make a probabilistic diagnosis of growth deficits.

Based on measurements of a specific child and of the corresponding standard distribution, we should therefore, with some chance of error, be able to assess his or her growth status.

Because the diagnosis is probabilistic, two kinds of errors have to be considered: assignment of “stunted” status to a normal child (type-1 error) and assignment of “normal” status to a stunted child (type-2 error). Since the height distribution of stunted children usually overlaps to some degree with that of healthy children, there is no cut-off point that separates stunted from nonstunted children. Therefore, lower (more restricted) cut-offs imply a reduction of type-1 error at the expense of an increase in type-2 error, and the converse for higher (more liberal) cut-offs.

When the purpose of the anthropometric evaluation is to estimate the prevalence of growth deficits, type-1 and type-2 errors should be balanced by taking into consideration the expected proportions of stunted and nonstunted children. For example, if the two proportions are expected to be similar, i.e., the true prevalence of growth deficits is around 50%, errors of the same magnitude will produce a similar number of false positives and false negatives, and the estimated prevalence will be equal to the true value.

---

1 Professor, Department of Nutrition, School of Public Health, University of São Paulo, Ave Dr Arnaldo 715—São Paulo 01255, Brazil. At present: Consultant, Nutrition Unit, World Health Organization, Geneva, Switzerland. Requests for reprints should be sent to São Paulo.


Reprint No. 5234

---

ª Type-1 and type-2 errors should be considered in a different perspective if the objective of the anthropometric evaluation is not the assessment of the prevalence of stunting but the clinical evaluation or screening of individuals for treatment.
If the proportion of normal children is expected to be greater than that of stunted children, an equal number of false positives and false negatives will be obtained with a cut-off point which implies that the magnitude of the type-I error is lower than that of type-2. Finally, if the proportion of stunted children is expected to be greater, the type-2 error should be lower. This can be expressed mathematically as follows:

\[ p = P(1 - E2) + (1 - P)E1 \]

where \( p \) is the estimated prevalence, \( P \) is the true prevalence, and \( E1 \) and \( E2 \) are the magnitude of type-1 and type-2 errors, respectively.

Since, by definition, the specificity (Sp) of any classification is \( 1 - E1 \) and the sensitivity (Se) is \( 1 - E2 \), this equation can be rewritten as

\[ p = P \cdot Se + (1 - P)(1 - Sp) \]

From the above discussion, the conclusions listed below can be drawn.

- There is no single exact cut-off that is appropriate for determining the prevalence of growth deficits in all populations (more restricted cut-offs are recommended for populations less affected by stunting, and more liberal cut-offs for those that are more affected).
- The choice of the exact cut-off for a specific population requires information about the distribution of type-1 and type-2 errors for various cut-offs.

The distribution of type-1 errors is independent of the population being examined and can readily be determined if the cut-offs are expressed as centiles (or Z-scores) of the reference population. For example, the cut-off corresponding to the third centile is associated with a type-1 error of 0.03, by definition, i.e., 3% of the nonstunted children will tend to be classified as stunted, while the tenth centile is associated with a type-1 error of 0.10.

In contrast, the distribution of type-2 errors is much more difficult to specify since it is dependent on the height distribution of the stunted children. Specification of this distribution requires application of an external criterion that should separate perfectly the stunted from nonstunted children. Since no such criterion is available, type-2 errors cannot be quantified and recourse has to be made to the use of more “restricted” cut-offs when lower prevalences are expected and more “liberal” cut-offs for higher prevalences.

The classical approach

In practice the prevalence of stunted children is determined by calculating the proportion of children whose heights are less than \(-2\) standard deviations (SD) of the mean value for the anthropometric standard, i.e., a fixed and fairly strict type-1 error of 0.023 is recommended and used. This has two main implications:

- The strict and specific cut-off suggests a strong imbalance between type-1 and type-2 errors, favouring the former, which is only appropriate when modest rates of growth deficits are expected. When this is not the case, and this probably applies to most populations submitted to anthropometric surveys, stunting may be significantly underestimated.
- The fixed cut-off implies that the degree of underestimation will not be the same for all populations, affecting the higher prevalences more and the lower ones less, and therefore producing an artificial contraction of real differences.

To examine these points, we have applied the cut-off of \(-2\) SD to hypothetical populations where the prevalence of stunting is 5%, 25% and 50%. The height distribution of the populations (Fig. 1 and 2) was established assuming the following:

- The height distribution of the nonstunted fraction of the population expressed in Z-scores reproduces the distribution in the reference population, i.e., the reference population is taken as a standard.
- The height distribution of the stunted children reproduces the standard distribution, but shifted to the left by 1 SD (Fig. 1) or 2 SD (Fig. 2). This is equivalent to assuming that all children belonging to the stunted fraction of the population deviate by 1 SD or 2 SD from their expected genetic height, i.e., that a uniform growth deficit of 1 SD or 2 SD is admitted for all stunted children (among under-5-year-olds this would mean a growth deficit of about 4 cm and 8 cm, respectively).

\[ \text{See footnote a, p. 761.} \]
Table 1 compares the true prevalences of growth deficits in the hypothetical populations with those resulting from application of a cut-off of −2 SD. The prevalences are considerably underestimated by this cut-off, with the magnitude of the underestimation being less pronounced for lower prevalences and more pronounced for higher.

The extent of the underestimation is always lower when a higher growth deficit is assumed. This is because the increase in the gap between genetic height and observed height displaces the stunted population to the left, reducing the overlapping with the normal fraction of the population and thus increasing the sensitivity, without changing the specificity of the classification. Fig. 3 illustrates this for the sensitivities obtained with a cut-off of −2 SD for growth deficits in the range 0–6 SD. Thus for growth deficits of 1–2 SD (probably not an uncommon range for stunting) the sensitivity attained by the cut-off of −2 SD is 15.1–50.0%. Such sensitivities are appropriate for prevalences of around 5% but extremely inadequate for higher prevalences (Fig. 4).

Table 1: Assessment of the use of the cut-off −2 SD for estimating the true prevalence of growth deficits in six hypothetical populations

<table>
<thead>
<tr>
<th>Population</th>
<th>True prevalence (%)</th>
<th>Estimated prevalence (%)</th>
<th>Degree of underestimation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>5.0</td>
<td>3.0</td>
<td>40.0</td>
</tr>
<tr>
<td>B1</td>
<td>25.0</td>
<td>5.7</td>
<td>77.2</td>
</tr>
<tr>
<td>C1</td>
<td>50.0</td>
<td>9.1</td>
<td>81.8</td>
</tr>
<tr>
<td>A2</td>
<td>5.0</td>
<td>4.7</td>
<td>6.0</td>
</tr>
<tr>
<td>B2</td>
<td>25.0</td>
<td>14.2</td>
<td>43.2</td>
</tr>
<tr>
<td>C2</td>
<td>50.0</td>
<td>26.1</td>
<td>47.8</td>
</tr>
</tbody>
</table>

*Expressed as ((True prevalence – estimated prevalence)/True prevalence) × 100. For A1, B1, and C1 a uniform growth deficit equivalent to −1 SD is assumed for all stunted children; for A2, B2, and C2 the uniform growth deficit is assumed to be equivalent to −2 SD (see text).

Only when the growth deficits are equivalent to 4 SD (or about 16 cm for under-5-year-olds) does the sensitivity associated with the cut-off −2 SD reach 97.7%, i.e., the value required to estimate properly prevalences of around 50%.

Therefore in most situations the proportion of children with heights (Z-scores) less than −2 SD covers only part of the problem, those children with severe growth retardation.

Mora's approach

Mora has recently proposed a new method of estimating a standardized prevalence of growth deficits, i.e., a standardized prevalence of child malnutrition determined from anthropometric indicators. This approach is based on the following equation:

\[ SP = MP - FP + FN \]

where \( SP \) is the standardized prevalence, \( MP \) is the measured prevalence calculated from the observed population (the proportion of children under a given cut-off point of the reference population) and \( FP \) and \( FN \), respectively, are the false positives and false negatives.

The false positives are taken to be the proportion of cases present in the reference population.
below the cut-off employed, an approach already proposed by WHO that poses no major problems if the reference population is taken as a standard. What is new and polemic in Mora's approach is the method of estimating the false negatives, which he defines as the excess proportion of cases in the observed population above the cut-off point compared with that in the reference population (see Fig. 5). Unfortunately, however, he presents no real argument to justify this. With this definition the standardized prevalence of growth deficits (or the true prevalence) is independent of the cut-off employed and can be easily calculated from the proportion of individuals in the observed population who are outside the normal distribution of the reference population, i.e., the area of no overlapping between the two populations.

Until now, no real criticism has been raised against Mora's approach. Bohning et al. have criticized the method used by Mora to calculate the area of no overlapping, and have suggested a nonparametric estimation procedure that avoids Mora's assumption of a Gaussian distribution for the anthropometric indicator.°

The difficulty with Mora's approach is his definition of false negatives, since the area of no overlapping considered represents only a fraction of the false negatives. This can readily be demonstrated by examining the height distribution (Z-scores) that corresponds to the normal individuals according Mora, or the area of overlapping between the reference and the observed populations ($TN + FP$ in Fig. 5). Compared with the reference population, the height distribution of these normal individuals is still shifted to the left. Thus, by definition, this fraction of the observed population still contains a proportion of stunted individuals. Such individuals are precisely those whose growth deficits are not sufficient enough to displace them from the curve of the reference population: by being shifted to the left they simply replace other stunted individuals who were displaced to the area of no overlapping. These individuals, although affected by growth deficits, are incorrectly taken to be normal (or true negatives) in Mora's method. This confirms that without a "golden standard" we cannot obtain the height distribution of the stunted, from which it follows that we are not able to determine type-2 errors, which in turn implies that no estimation of false negatives is really possible.

Table 2 shows the performance of Mora's approach in estimating the prevalence of growth deficits for the same hypothetical populations as in Table 1.

Although attenuated, the same trends that were detected using a cut-off of $-2$ SD result also with Mora's approach: gross underestimation of the real prevalences (mainly for higher values) and an artificial reduction in the contrasts between populations.

### Estimating the minimum prevalence of growth deficits

Since no assumptions can be made about the stunted fraction of the population (other than that, compared to the reference population, it should have a height distribution shifted in some degree to the left), the most logical starting point is the identification of the nonstunted or the true normal individuals. By taking the reference population as a standard, we can assume that in any observed population the height distribution of the subgroup of nonstunted individuals should be similar to that of the reference population, or in terms of Z-score a Gaussian distribution with a mean of zero and $SD = 1$. By using the same area approach proposed by Mora (Fig. 6), we see that several areas could represent that fraction of the population: the minimum area, corre-

---

° See footnote a, p. 761.

responding to a proportion close to zero, and the maximum, corresponding to approximately twice the proportion of Z-scores \( \geq 0 \). Therefore, the true prevalence of growth deficits can vary from a maximum of 100% to a minimum of approximately

\[ 100\% - 2(\% \text{ Z-scores} > 0) \]

This last formula, which is easy to evaluate, represents an estimate of what could be referred to as the minimum prevalence of growth deficits in any population.

Table 3 shows to what extent the minimum prevalence deviates from the true prevalence in each of the six hypothetical populations. As expected, the estimates obtained using the technique are lower than the true prevalences. Nevertheless, the degree of underestimation is smaller than that obtained either with a cut-off of \(-2 \text{ SD}\) or from Mora’s technique, particularly for the higher prevalences. Another important advantage of the approach is that the degree of underestimation is not dependent on the level of prevalence, which therefore maintains the true prevalence gradients between populations. In this case the degree of underestimation is a function of the assumed intensity of growth deficits (Fig. 7); for growth deficits of intensity equivalent to \( >2 \text{ SD}\) the degree of underestimation is insignificant. This indicates that the minimum prevalence of growth deficits is a reasonable estimate of the true prevalence if the majority of stunted individuals have deficits equivalent to \( \geq 1.5 \text{ SD}\), i.e., about 6 cm for children aged up to 5 years. Considering this, and taking into account the usual scenarios where anthropometric surveys are performed, the figures provided by the minimum prevalence technique are in most situations very close to the real values.

**Recommendations**

Based on the above analysis and also on practical considerations, the recommendations outlined below can be made.

- All growth deficit prevalence studies should determine at the outset the complete Z-score distribution of the indicator used in the assessment of growth status. The plot of this distribution against that of the reference distribution (as in Fig. 1 and 2), although it does not provide a direct measure of the prevalence of growth deficits, \(^7\) is the most appropriate way of gauging the extent of the problem in the study population, taking into consideration both the frequency and intensity of growth retardation.
- Since use of the cut-off \(-2 \text{ SD}\) is so widespread and the term “stunted” is so closely associated with this cut-off, and since it is important to have an indicator of severe growth retardation, it seems advisable to retain both the term and the cut-off. Nevertheless it should be borne in mind that in most situations use of the \(-2 \text{ SD}\) cut-off will result in a measure of the prevalence of severe growth retardation and not the total prevalence of growth retardation.

\(^7\) In extreme cases, where the area of overlapping with the reference population corresponds to 100% or 0% of the observed population, the prevalences of growth retardation are 0% and 100%, respectively.
Résumé

Dénombrement des enfants présentant un retard de croissance dans une population: critique de l'ancienne et de la nouvelle approche et nouvelle méthode proposée

Deux méthodes d'estimation de la prévalence du retard de croissance dans une population sont ici évaluées: la méthode classique, basée sur la proportion d'enfants dont la taille est inférieure à deux écarts types (ET) au-dessous de la moyenne attendue d'une population de référence, et une nouvelle méthode récemment proposée par Mora, basée sur la distribution totale de la taille de la population observée et de la population de référence.

L'application de la méthode classique à plusieurs populations théoriques conduit à conclure que dans la plupart des situations rencontrées dans les pays en développement, la prévalence du retard de croissance est grossièrement sous-estimée et ne reflète que la présence de déficits sévères. Cette méthode présente le deuxième inconvénient de réduire fortement la différence relative entre les strates les plus et les moins exposées. L'application de la méthode de Mora à ces mêmes populations théoriques réduit, mais n'élimine pas ces problèmes. Une nouvelle méthode d'estimation de la prévalence du retard de croissance, également basée sur la distribution totale de la taille dans la population observée et dans la population de référence, est décrite et évaluée. Cette méthode permet d'obtenir une meilleure estimation de la prévalence vraie du retard de croissance, sans réduction des différences relatives.

Il est recommandé que toutes les études transversales portant sur la croissance fournissent dans la mesure du possible les informations suivantes:

- la distribution complète du Z-score de l'indicador utilisé pour l'évaluation de la croissance (si possible comparée graphiquement à celle de la population de référence);
- le pourcentage d'individus situés au-dessous de la valeur limite -2 ET, défini comme présentant un retard de croissance, interprété comme indication de la fréquence du retard de croissance sévère dans la population;
- le pourcentage donné par la formule 100 - 2 (% des scores > 0), concernant les individus classés comme ayant une insuffisance de croissance, interprété comme indication de la fréquence totale des déficits de croissance dans la population.

References