Does the use of mean or median Z-score of the thyroid volume indices provide a more precise description of the iodine deficiency disorder status of a population?

L C Foo and M Mafauzy

Institute for Medical Research, Kuala Lumpur, Malaysia and Department of Endocrinology, University of Science Malaysia Hospital, Kubang Kerian, Kelantan, Malaysia

(Correspondence should be addressed to L C Foo, Institute for Medical Research, Jalan Pahang, 50588 Kuala Lumpur, Malaysia; Email: lcfoo@imr.gov.my)

Abstract

Endemic iodine deficiency is largely an environmental problem affecting whole populations. Currently, thyroid volume data from a population are analyzed with the sole objective of obtaining an estimate of goitre prevalence using +97th percentile or +2 standard deviations of an appropriate reference as cut-off. This paper proposes an alternative approach to the analysis and presentation of thyroid volume data using Z-scores (standard deviation scores) of the thyroid volume indices such as thyroid volume-for-age or thyroid volume-for-body surface area. The calculation of the summary statistics of the Z-scores, such as mean or median, provides an alternative to the prevalence-based approach for expressing severity of iodine deficiency disorders (IDD). An advantage of the mean or median Z-score is that it describes the thyroid volume profile (and therefore the IDD status) of the entire population directly, unlike goitre prevalence which gives information only about the extremes of distribution. The frequency curve or histogram of the Z-scores provides a complete picture of the whole distribution. Although qualitatively similar conclusions on IDD severity can be drawn from both analytical approaches, only the Z-score system is able to capture adequately the trends or changes in thyroid size over time, and to establish whether a previously iodine-deficient community’s thyroid volume profile has returned to ‘normal’ (as indicated by a distribution that is not significantly different from that of the reference) following intervention. As a continuous variable, Z-scores are particularly useful for the analysis of data from populations where the sample size is relatively small, or where many individuals lie outside the extreme percentiles of the reference population. In view of its advantages in the context of activities based on single and multiple measurements, the Z-score system is to be preferred for the reporting and use of thyroid volume indices. A desirable consequence of this preference is that national goals will be oriented towards an improvement of the overall thyroid volume profile of the population, rather than just a reduction of the number of individuals at the extremes.

European Journal of Endocrinology 141 557-560

Introduction

Iodine deficiency endemia, which is largely an environmental problem, is presently defined by the goitre prevalence and the median urinary iodine concentration in a population. In the past, goitre prevalence was assessed by palpation using a five-grade classification system recommended by Perez et al. (1). In 1993, following a review of the system by a Joint WHO/UNICEF/ICCIDD consultation on indicators for assessing iodine deficiency disorders (IDD) and their control through salt iodization, a new three-grade classification system was introduced (2). At the same time, the previously recommended 10% cut-off level for indicating the presence of a public health problem was revised downwards to 5%. Recently, thyroid ultrasound has superseded palpation as the method of choice for the determination of thyroid size (3, 4). For consistency with clinical screening (by palpation), goitre prevalence is reported using a cut-off value, often +97th percentile (5) or +2 standard deviations (6) of an appropriate thyroid volume reference. This paper proposes an alternative to the prevalence-based approach for expressing severity of IDD using summary statistics, such as the mean or median, of the Z-scores (standard deviation scores) of the thyroid volume indices.
Application of the Z-score system to thyroid volume data

Transforming raw thyroid volume values into Z-scores of the thyroid volume indices

Z-scores are created for the purpose of combining scores or values coming from different scales or with different units of measurement. They provide information about the standing of an individual’s value relative to a reference population. The Z-score for an observed value is calculated as follows (9):

\[
Z\text{-score} = \frac{(\text{Observed value}) - (\text{Median reference value})}{\text{Standard deviation of reference population}}
\]

Thus, when the Z-score transformation is applied to thyroid volume data, the value of an individual is expressed as a number of standard deviations (or Z-scores) above or below the median value of the reference population. A fixed Z-score interval implies a fixed thyroid volume difference for children of a given age (or BSA, depending on whether the age/sex-specific or BSA/sex-specific thyroid volume reference is used) and sex. The advantage of this system is that, for population-based applications, a group of Z-scores can be subjected to summary statistics such as the mean and standard deviation. The mathematics of the Z-score transformation are such that if every value in a reference distribution is converted to its Z-score, the transformed scores will necessarily have a mean of 0 and a standard deviation of 1 (9).

Mean or median Z-score: description of the whole distribution

The mean or median Z-score provides information about a population’s standing relative to the reference population. It has the advantage of describing the thyroid volume status of the entire population directly. As an index of severity for IDD, it also addresses the conceptual shortcomings of the cut-off or prevalence-based approach by not assuming that only those individuals above the cut-off are affected. A mean or median Z-score significantly greater than 0 indicates that the entire distribution has shifted upwards compared with that of the reference, suggesting that most, if not all, individuals in the population have been affected (see Fig. 1). The larger the mean or median Z-score, the more severe the IDD. Because the Z-score can be calculated for any value of thyroid volume no matter how large, the mean or median Z-score is a more accurate estimate of poor IDD status than goitre prevalence in areas where many children lie outside the extreme percentiles of the reference population. A mean Z-score greater than 2 is indicative of the existence of severe IDD.

Figure 1 Comparing distributions of thyroid volume-for-age Z-scores of two sample populations with that of the reference population. Solid line, reference; dashed line, sample population A; dotted line, sample population B.

Plotting the frequency curve or histogram of the Z-scores provides a complete picture of the whole distribution, which can be compared with that of the reference population (see Fig. 1). This method of analyzing and presenting thyroid volume data appears most appropriate for description of IDD, which is largely a generalized, geographical phenomenon affecting the whole population, for which population (rather than individual) interventions have long been the norm. In addition, a major objective of IDD surveillance and monitoring is to determine trends or changes in thyroid size and this is best achieved by examining the total distribution, rather than only the extremes. Strictly speaking, IDD elimination can be said to have been achieved only when the observed distribution is the same as the reference distribution (i.e. when the mean or median Z-score approaches 0), not just when the values of the individuals initially above the cut-off level have fallen below this level.
Other advantages of using Z-scores

Other advantages of the use of Z-scores are those associated with being a continuous variable, such as increased statistical power compared with the use of a cut-off, particularly where IDD is mild or when sample size is relatively small. Using the mean or median Z-score as an index of severity for IDD may also result in increased awareness among public health managers that, if a condition is severe, an intervention is required for the entire community, not just those who are classified as goitrous by the cut-off criteria (10). Plotting the entire distribution of the Z-scores against the reference distribution may also be helpful in distinguishing endemic IDD (as reflected by an upward shift in the entire distribution) from sporadic IDD (as reflected by a bimodal distribution with the major peak coinciding with that of the reference).

Z-scores are widely used in anthropometry for the assessment of children’s nutritional status (e.g. weight-for-age or weight-for-height Z-score) (11) and in bone densitometry (also T-score) for the diagnosis of osteopenia and osteoporosis (12).

Choice of appropriate thyroid volume index

In children, the two most commonly used indices for assessment of thyroid enlargement are thyroid volume-for-age (5–7), and thyroid volume-for-body surface area (BSA) (5, 7). When comparisons are made with a locally defined reference, the latter gives slightly higher goitre prevalence than the former in areas where protein-energy deficiency is prevalent but the difference is of little practical significance (7). Thus, the choice of which of these indices to use is dependent on practical considerations: whether weight and height (for calculation of BSA) or age are easier to obtain in the community to be investigated. In some developing areas, accurate ages of children are difficult to obtain so thyroid volume-for-BSA may have an advantage. A recent study has suggested that in adults thyroid volume is determined largely by lean body mass (8) but this has not yet been established in children. However, lean body mass is not easy to assess and practical constraints may limit its application in the field.

Major criticism of the use of cut-off point

A major criticism of the prevalence-based approach concerns the validity of using a cut-off value to divide a continuous distribution into goitrous and non-goitrous values. Ideally, cut-off points should be based on biological considerations, such as increased risk of functional impairment. The cut-off should distinguish an increase in thyroid size that matters from one that is of no real significance. However, as with most biological phenomena that are influenced by a multitude of factors, the practical problems of establishing a relation to risk are very great and difficult to surmount.

Use of cut-off points is easily justified in situations where the purpose is to identify individuals in a population as being in need of special consideration or treatment. However, this is generally not the practice in IDD where international agencies have recommended universal intervention if more than 5% of the population have values above the set cut-off.

Conclusion

Two approaches to the analysis and presentation of thyroid volume data are discussed. The first – the current approach – utilizes cut-off points to provide an estimate of the number or proportion outside the reference distribution and therefore gives information only about the extremes of the distribution; the second – proposed by us – uses Z-scores to describe the whole distribution. While prevalence estimate is helpful, and may even be pivotal, in securing priority of government intervention, it does not provide a complete picture of a population’s thyroid volume profile. A consequence of this narrow approach to data analysis is that national goals are being oriented towards the reduction of goitre prevalence based on an arbitrary cut-off, rather than towards an improvement of the overall thyroid volume profile. Although qualitatively similar conclusions on IDD severity can be drawn from the prevalence-based index and the mean or median Z-score, only the latter (and associated frequency curve or histogram of Z-scores) is able to capture adequately the trends or changes in thyroid size over time, and to establish whether the thyroid volume profile of a population is ‘normal’ (as indicated by a mean or median Z-score that is not significantly different from 0 and a standard deviation that is close to 1).

In conclusion, the mean or median Z-score of the thyroid volume indices is a more precise indicator of the IDD status of a population than goitre prevalence. In view of this and its other advantages in the context of activities based on single and multiple measurements, the Z-score system is to be preferred for the reporting and use of thyroid volume indices. However, for securing assignment of government priorities to IDD, the prevalence-based index (goitre prevalence) should continue to be reported.

References


Received 10 May 1999
Accepted 18 August 1999