STATISTICS IN MEDICINE, VOL. 11, 1289–1304 (1992)

THE AREA BETWEEN CURVES (ABC)—MEASURE IN NUTRITIONAL ANTHROPOMETRY

DANKMAR BÖHNING, AXEL HEMPFLING, FRANK-PETER SCHELP AND PETER SCHLATTMANN

Department of Epidemiology, Institute for Social Medicine, Free University Berlin, Augustastr. 37, 1000 Berlin 45, Germany

SUMMARY

This paper considers a statistic – recently suggested by Mora – for the deviation of a sample distribution from a reference distribution which typically arises in anthropometry when using the nutritional indicators height/age, weight/age or weight/height. The statistic measures the area between curves (ABC) and stands for the mass of the sample distribution which is not covered by the reference distribution. The paper provides a statistical framework for the ABC and includes some minor corrections of Mora's original paper. For the normal distribution situation with common or different variances, formulae are derived which include a partition of ABC into parts corresponding to malnourished and well-nourished groups. However, the *main* result is a non-parametric generalization of the ABC, motivated by the fact that the nutritional indicators often have skewed distributions with heavier left tails. Non-parametric statistical inference is provided by linking the ABC to the Kolmogorov–Smirnov statistic.

1. INTRODUCTION

In most countries in the tropics, although clinical cases of obvious forms of protein calorie malnutrition (PEM) in children are rare, subclinical forms of PEM are still widespread.[1]

> The most important nutritional problem in the world today is that of protein energy malnutrition (PEM): it is also the deficiency which, in general, supplementary feeding is intended to correct.
>
> For this reason it is suggested that the following three measurements are chosen to evaluate the presence or absence of nutritional impact in programmes aimed at feeding the vulnerable groups listed . . .: age, weight and height (length supine for all children < 2 years; height standing for children > 2 years).[2]

The nutritional status of populations, particularly those of infants and young children, can best be assessed through anthropometric measurements.[3] The recommended measurements are weight and height. Based on an international standard derived from the U.S. National Center for Health Statistics, Centers for Disease Control, the differences between the observed values in units of standard deviations (Z-scores) of the reference population might be calculated for weight-for-age (weight/age), weight-for-height (weight/height) and height-for-age (height/age).[4]

Nutritional anthropometry is a field of growing interest. Thus in the 47th session of the International Statistical Institute, two contributed papers sessions were devoted to the methodology for nutritional status surveys in developing countries, with at least four contributions directly connected to this topic.[5,16,17,19]

0277–6715/92/101289–16$13.00

Received May 1991
Revised February 1992

Recently Mora[6] drew attention to the problems arising from conflicting recommendations about the use of different cut-off points and classification systems. Z-scores are constructed to measure the nutritional status of a child against a reference group of children supposed to be healthy and well nourished. Then a certain value is chosen, also called *cut-off point*, and children having a Z-score below this cut-off point are considered to be malnourished or undernourished. The percentage of all children in the study population falling below the cut-off point defines the *prevalence rate* of malnourishment.

Different choices of cut-off values are commonly used ($-1{\cdot}5, -2, -3$), which makes the comparison of prevalence estimates difficult. This motivates the search for a measure which is independent of a cut-off value. However, if one is in a diagnostic situation to identify malnourished children in order to intervene, the selection of a cut-off value cannot be avoided.

In Mora's paper[6] a method is proposed to estimate the area between the study population density and the density of the reference population. He suggests the term *standardized prevalence* of malnutrition for this statistic and implies that his approach might be considered to replace the critical choice of cut-off points for prevalence estimates in cross-sectional population studies. The interpretation of the area between curves (ABC) as a measure of the deviation of the nutritional status from the reference population is suggested here and considered to be a helpful tool. A statistical framework for the ABC will be provided.

2. DEFINING NUTRITIONAL STATUS

2.1. Reference populations

The construction of the Z-scores assumes the availability of so-called reference or standard populations. A reference population is constructed for statistical comparisons, not as a norm for desired body growth. Some criteria for a reference population are that it is well organized, contains detailed information and is internationally available and used. Such a reference population is provided by the U.S. National Centre for Health Statistics and can be found, for example, in Reference 2. Table I and Table II are part of this reference population and are reproduced here for demonstration purposes. For example, Table I refers to the indicator height/age. Here, for each age group various statistical measures of the reference population are given. On the right, we find the median of each age group, as well as median $\pm$ 1, 2 or 3 SDs. To be more specific, for a boy of 7 months, we find a reference median of 69·5 cm, and median $-$ 2 SD is 64·1 cm. The left part of the table contains various percentiles. The computation of Z-scores involving height is somewhat confused by the fact that the height is measured differently as length (supine length) and stature (measured standing), and overlapping reference populations do exist. Obviously, both forms of computation lead to different height/age values. Which one to follow depends on the form of body height measurement.

Table II contains analogous statistical measures for the reference population referring to the indicator weight/height. Note that this reference population does *not* involve age. In this case, the grouping variable is body length having an increment of 0·5 cm.

2.2. Construction of Z-scores

Z-scores are constructed to measure the nutritional status of a child against a reference population of children. The measurements of the group of children under investigation are related to those of the reference with the same age and sex to yield a score independent of the child's age and sex. Usually, this goal is achieved for sex. With age, one often observes a drop-down effect for the

Table I. Length (cm) by age of boys aged 0–36 months

Age (months)	Centiles													Standard deviations							Age (months)
	3rd	5th	10th	20th	30th	40th	50th	60th	70th	80th	90th	95th	97th	− 3SD	− 2SD	− 1SD	Median	+ 1SD	+ 2SD	+ 3SD	
0	46·2	46·7	47·6	48·6	49·3	49·9	50·5	51·1	51·7	52·4	53·4	54·2	54·8	43·6	45·9	48·2	50·5	52·8	55·1	57·4	0
1	49·9	50·5	51·4	52·5	53·3	53·9	54·6	55·2	55·9	56·6	57·7	58·6	59·2	47·2	49·7	52·1	54·6	57·0	59·5	61·9	1
2	53·2	53·9	54·8	55·9	56·7	57·4	58·1	58·7	59·4	60·2	61·4	62·3	62·9	50·4	52·9	55·5	58·1	60·7	63·2	65·8	2
3	56·1	56·8	57·7	58·9	59·7	60·4	61·1	61·8	62·5	63·3	64·5	65·5	66·1	53·2	55·8	58·5	61·1	63·7	66·4	69·0	3
4	58·6	59·3	60·3	61·4	62·3	63·0	63·7	64·4	65·1	66·0	67·1	68·1	68·7	55·6	58·3	61·0	63·7	66·4	69·1	71·7	4
5	60·8	61·5	62·5	63·6	64·5	65·2	65·9	66·6	67·3	68·2	69·4	70·3	71·0	57·8	60·5	63·2	65·9	68·6	71·3	74·0	5
6	62·8	63·4	64·4	65·6	66·4	67·1	67·8	68·5	69·2	70·1	71·3	72·2	72·9	59·8	62·4	65·1	67·8	70·5	73·2	75·9	6
7	64·5	65·1	66·1	67·2	68·1	68·8	69·5	70·2	70·9	71·7	72·9	73·9	74·5	61·5	64·1	66·8	69·5	72·2	74·8	77·5	7
8	66·0	66·6	67·6	68·7	69·6	70·3	71·0	71·6	72·4	73·2	74·4	75·3	76·0	63·0	65·7	68·3	71·0	73·6	76·3	78·9	8
9	67·4	68·0	68·9	70·1	70·9	71·7	72·3	73·0	73·7	74·6	75·7	76·7	77·3	64·4	67·0	69·7	72·3	75·0	77·6	80·3	9
10	68·7	69·3	70·2	71·4	72·2	73·0	73·6	74·3	75·0	75·9	77·0	78·0	78·6	65·7	68·3	71·0	73·6	76·3	78·9	81·6	10
11	69·9	70·5	71·5	72·6	73·5	74·2	74·9	75·6	76·3	77·1	78·3	79·3	79·9	66·9	69·6	72·2	74·9	77·5	80·2	82·9	11
12	71·0	71·6	72·6	73·8	74·7	75·4	76·1	76·8	77·5	78·4	79·5	80·5	81·2	68·0	70·7	73·4	76·1	78·8	81·5	84·2	12
13	72·1	72·7	73·7	74·9	75·8	76·5	77·2	77·9	78·7	79·5	80·7	81·7	82·4	69·0	71·8	74·5	77·2	80·0	82·7	85·5	13
14	73·1	73·8	74·8	76·0	76·9	77·6	78·3	79·1	79·8	80·7	81·9	82·9	83·6	70·0	72·8	75·6	78·3	81·1	83·9	86·7	14
15	74·1	74·7	75·8	77·0	77·9	78·7	79·4	80·1	80·9	81·8	83·1	84·1	84·8	70·9	73·7	76·6	79·4	82·3	85·1	88·0	15
16	75·0	75·7	76·7	78·0	78·9	79·7	80·4	81·2	82·0	82·9	84·2	85·2	85·9	71·7	74·6	77·5	80·4	83·4	86·3	89·2	16
17	75·9	76·6	77·6	78·9	79·9	80·7	81·4	82·2	83·0	83·9	85·3	86·3	87·0	72·5	75·5	78·5	81·4	84·4	87·4	90·4	17
18	76·7	77·4	78·5	79·8	80·8	81·6	82·4	83·2	84·0	85·0	86·3	87·4	88·1	73·3	76·3	79·4	82·4	85·4	88·5	91·5	18
19	77·5	78·2	79·4	80·7	81·7	82·6	83·3	84·1	85·0	86·0	87·3	88·4	89·2	74·0	77·1	80·2	83·3	86·4	89·5	92·7	19
20	78·3	79·0	80·2	81·6	82·6	83·4	84·2	85·0	85·9	86·9	88·3	89·5	90·2	74·7	77·9	81·1	84·2	87·4	90·6	93·8	20
21	79·1	79·8	81·0	82·4	83·4	84·3	85·1	85·9	86·8	87·8	89·3	90·4	91·2	75·4	78·7	81·9	85·1	88·4	91·6	94·8	21
22	79·8	80·6	81·8	83·2	84·3	85·2	86·0	86·8	87·7	88·7	90·2	91·4	92·2	76·1	79·4	82·7	86·0	89·3	92·5	95·8	22
23	80·6	81·3	82·6	84·0	85·1	86·0	86·8	87·7	88·6	89·6	91·1	92·3	93·1	76·8	80·2	83·5	86·8	90·2	93·5	96·8	23
24	81·3	82·1	83·3	84·8	85·9	86·8	87·6	88·5	89·4	90·5	92·0	93·2	94·0	77·5	80·9	84·3	87·6	91·0	94·4	97·7	24
25	82·1	82·9	84·1	85·6	86·7	87·6	88·5	89·3	90·2	91·3	92·8	94·0	94·8	78·3	81·7	85·1	88·5	91·8	95·2	98·6	25
26	82·8	83·6	84·9	86·4	87·5	88·4	89·2	90·1	91·0	92·1	93·6	94·9	95·7	79·0	82·4	85·8	89·2	92·7	96·1	99·5	26
27	83·6	84·4	85·6	87·1	88·2	89·2	90·0	90·9	91·8	92·9	94·4	95·7	96·5	79·8	83·2	86·6	90·0	93·4	96·9	100·3	27
28	84·4	85·2	86·4	87·9	89·0	89·9	90·8	91·7	92·6	93·7	95·2	96·4	97·2	80·5	83·9	87·4	90·8	94·2	97·6	101·1	28
29	85·1	85·9	87·2	88·7	89·8	90·7	91·6	92·4	93·3	94·4	95·9	97·2	98·0	81·3	84·7	88·1	91·6	95·0	98·4	101·8	29
30	85·8	86·7	87·9	89·4	90·5	91·4	92·3	93·2	94·1	95·2	96·7	97·9	98·7	82·0	85·4	88·9	92·3	95·7	99·2	102·6	30
31	86·6	87·4	88·6	90·1	91·2	92·2	93·0	93·9	94·8	95·9	97·4	98·7	99·5	82·7	86·2	89·6	93·0	96·5	99·9	103·3	31
32	87·3	88·1	89·3	90·9	91·9	92·9	93·7	94·6	95·5	96·6	98·2	99·4	100·2	83·4	86·9	90·3	93·7	97·2	100·6	104·1	32
33	88·0	88·8	90·0	91·6	92·6	93·6	94·5	95·3	96·3	97·4	98·9	100·1	100·9	84·1	87·6	91·0	94·5	97·9	101·4	104·8	33
34	88·6	89·4	90·7	92·2	93·3	94·3	95·2	96·0	97·0	98·1	99·6	100·9	101·7	84·7	88·2	91·7	95·2	98·6	102·1	105·6	34
35	89·3	90·1	91·4	92·9	94·0	95·0	95·8	96·7	97·7	98·8	100·3	101·6	102·4	85·4	88·8	92·3	95·8	99·3	102·8	106·3	35
36	89·9	90·7	92·0	93·5	94·7	95·6	96·5	97·4	98·4	99·5	101·0	102·3	103·2	85·9	89·4	93·0	96·5	100·1	103·6	107·1	36

Source: Reference 2, Table 18.

Table II. Weight (kg) by stature of boys 55–145 cm in height

Stature (cm)	Centiles													Standard deviations							Stature (cm)
	3rd	5th	10th	20th	30th	40th	50th	60th	70th	80th	90th	95th	97th	− 3SD	− 2SD	− 1SD	Median	+ 1SD	+ 2SD	+ 3SD	
55·0	2·9	3·1	3·3	3·7	3·9	4·1	4·3	4·6	4·9	5·3	5·9	6·3	6·6	2·0	2·8	3·6	4·3	5·5	6·7	7·9	55·0
55·5	3·0	3·2	3·5	3·8	4·1	4·3	4·5	4·8	5·1	5·5	6·0	6·5	6·8	2·2	2·9	3·7	4·5	5·7	6·9	8·1	55·5
56·0	3·2	3·4	3·7	4·0	4·3	4·5	4·7	5·0	5·3	5·7	6·2	6·7	6·9	2·3	3·1	3·9	4·7	5·9	7·1	8·3	56·0
56·5	3·3	3·5	3·8	4·2	4·4	4·7	4·9	5·2	5·5	5·9	6·4	6·8	7·1	2·4	3·2	4·1	4·9	6·1	7·3	8·4	56·5
57·0	3·5	3·7	4·0	4·3	4·6	4·8	5·0	5·3	5·7	6·0	6·6	7·0	7·3	2·6	3·4	4·2	5·0	6·2	7·4	8·6	57·0
57·5	3·6	3·8	4·1	4·5	4·8	5·0	5·2	5·5	5·8	6·2	6·7	7·2	7·4	2·7	3·5	4·4	5·2	6·4	7·6	8·8	57·5
58·0	3·8	4·0	4·3	4·7	4·9	5·2	5·4	5·7	6·0	6·4	6·9	7·3	7·6	2·8	3·7	4·5	5·4	6·6	7·8	9·0	58·0
58·5	3·9	4·1	4·4	4·8	5·1	5·3	5·5	5·8	6·2	6·5	7·1	7·5	7·8	3·0	3·8	4·7	5·5	6·7	7·9	9·1	58·5
59·0	4·1	4·3	4·6	5·0	5·2	5·5	5·7	6·0	6·3	6·7	7·2	7·7	7·9	3·1	4·0	4·8	5·7	6·9	8·1	9·3	59·0
59·5	4·2	4·4	4·7	5·1	5·4	5·6	5·9	6·2	6·5	6·9	7·4	7·8	8·1	3·2	4·1	5·0	5·9	7·1	8·2	9·4	59·5
60·0	4·4	4·6	4·9	5·3	5·6	5·8	6·0	6·3	6·6	7·0	7·6	8·0	8·3	3·4	4·3	5·1	6·0	7·2	8·4	9·6	60·0
60·5	4·5	4·7	5·0	5·4	5·7	6·0	6·2	6·5	6·8	7·2	7·7	8·1	8·4	3·5	4·4	5·3	6·2	7·4	8·6	9·8	60·5
61·0	4·6	4·8	5·2	5·6	5·9	6·1	6·3	6·6	7·0	7·3	7·9	8·3	8·6	3·6	4·5	5·4	6·3	7·5	8·7	9·9	61·0
61·5	4·8	5·0	5·3	5·7	6·0	6·3	6·5	6·8	7·1	7·5	8·0	8·5	8·7	3·8	4·7	5·6	6·5	7·7	8·9	10·1	61·5
62·0	4·9	5·1	5·5	5·9	6·2	6·4	6·6	6·9	7·3	7·6	8·2	8·6	8·9	3·9	4·8	5·7	6·6	7·8	9·0	10·2	62·0
62·5	5·0	5·3	5·6	6·0	6·3	6·5	6·8	7·1	7·4	7·8	8·3	8·8	9·1	4·0	4·9	5·9	6·8	8·0	9·2	10·4	62·5
63·0	5·2	5·4	5·7	6·1	6·4	6·7	6·9	7·2	7·6	7·9	8·5	8·9	9·2	4·1	5·1	6·0	6·9	8·1	9·3	10·6	63·0
63·5	5·3	5·5	5·9	6·3	6·6	6·8	7·1	7·4	7·7	8·1	8·6	9·1	9·4	4·3	5·2	6·1	7·1	8·3	9·5	10·7	63·5
64·0	5·4	5·7	6·0	6·4	6·7	7·0	7·2	7·5	7·8	8·2	8·8	9·2	9·5	4·4	5·3	6·3	7·2	8·4	9·6	10·9	64·0
64·5	5·6	5·8	6·1	6·5	6·8	7·1	7·3	7·7	8·0	8·4	8·9	9·4	9·7	4·5	5·5	6·4	7·3	8·6	9·8	11·0	64·5
65·0	5·7	5·9	6·3	6·7	7·0	7·2	7·5	7·8	8·1	8·5	9·1	9·5	9·8	4·6	5·6	6·5	7·5	8·7	9·9	11·2	65·0
65·5	5·8	6·0	6·4	6·8	7·1	7·4	7·6	7·9	8·3	8·7	9·2	9·7	9·9	4·7	5·7	6·7	7·6	8·9	10·1	11·3	65·5
66·0	5·9	6·2	6·5	6·9	7·2	7·5	7·7	8·1	8·4	8·8	9·3	9·8	10·1	4·9	5·8	6·8	7·7	9·0	10·2	11·5	66·0
66·5	6·1	6·3	6·6	7·1	7·4	7·6	7·9	8·2	8·5	8·9	9·5	9·9	10·2	5·0	6·0	6·9	7·9	9·1	10·4	11·6	66·5
67·0	6·2	6·4	6·8	7·2	7·5	7·8	8·0	8·3	8·7	9·1	9·6	10·1	10·4	5·1	6·1	7·0	8·0	9·3	10·5	11·8	67·0
67·5	6·3	6·5	6·9	7·3	7·6	7·9	8·1	8·5	8·8	9·2	9·8	10·2	10·5	5·2	6·2	7·2	8·1	9·4	10·7	11·9	67·5
68·0	6·4	6·7	7·0	7·4	7·8	8·0	8·3	8·6	8·9	9·3	9·9	10·4	10·7	5·3	6·3	7·3	8·3	9·5	10·8	12·1	68·0
68·5	6·6	6·8	7·1	7·6	7·9	8·1	8·4	8·7	9·1	9·5	10·0	10·5	10·8	5·5	6·4	7·4	8·4	9·7	10·9	12·2	68·5
69·0	6·7	6·9	7·3	7·7	8·0	8·3	8·5	8·8	9·2	9·6	10·2	10·6	10·9	5·6	6·6	7·5	8·5	9·8	11·1	12·4	69·0
69·5	6·8	7·0	7·4	7·8	8·1	8·4	8·6	9·0	9·3	9·7	10·3	10·8	11·1	5·7	6·7	7·7	8·6	9·9	11·2	12·5	69·5
70·0	6·9	7·1	7·5	7·9	8·2	8·5	8·8	9·1	9·4	9·8	10·4	10·9	11·2	5·8	6·8	7·8	8·8	10·1	11·4	12·7	70·0
70·5	7·0	7·3	7·6	8·0	8·4	8·6	8·9	9·2	9·6	10·0	10·5	11·0	11·3	5·9	6·9	7·9	8·9	10·2	11·5	12·8	70·5
71·0	7·1	7·4	7·7	8·2	8·5	8·7	9·0	9·3	9·7	10·1	10·7	11·2	11·5	6·0	7·0	8·0	9·0	10·3	11·6	12·9	71·0
71·5	7·2	7·5	7·8	8·3	8·6	8·9	9·1	9·4	9·8	10·2	10·8	11·3	11·6	6·1	7·1	8·1	9·1	10·4	11·8	13·1	71·5
72·0	7·4	7·6	8·0	8·4	8·7	9·0	9·2	9·6	9·9	10·3	10·9	11·4	11·7	6·3	7·2	8·2	9·2	10·6	11·9	13·2	72·0
72·5	7·5	7·7	8·1	8·5	8·8	9·1	9·3	9·7	10·0	10·5	11·1	11·5	11·9	6·4	7·4	8·3	9·3	10·7	12·0	13·4	72·5
73·0	7·6	7·8	8·2	8·6	8·9	9·2	9·5	9·8	10·2	10·6	11·2	11·7	12·0	6·5	7·5	8·5	9·5	10·8	12·1	13·5	73·0
73·5	7·7	7·9	8·3	8·7	9·0	9·3	9·6	9·9	10·3	10·7	11·3	11·8	12·1	6·6	7·6	8·6	9·6	10·9	12·3	13·6	73·5
74·0	7·8	8·0	8·4	8·8	9·2	9·4	9·7	10·0	10·4	10·8	11·4	11·9	12·2	6·7	7·7	8·7	9·7	11·0	12·4	13·8	74·0
74·5	7·9	8·2	8·5	9·0	9·3	9·5	9·8	10·1	10·5	10·9	11·5	12·0	12·4	6·8	7·8	8·8	9·8	11·2	12·5	13·9	74·5
75·0	8·0	8·3	8·6	9·1	9·4	9·7	9·9	10·3	10·6	11·1	11·7	12·2	12·5	6·9	7·9	8·9	9·9	11·3	12·7	14·0	75·0
75·5	8·1	8·4	8·7	9·2	9·5	9·8	10·0	10·4	10·7	11·2	11·8	12·3	12·6	7·0	8·0	9·0	10·0	11·4	12·8	14·2	75·5

Source: Reference 2, Table 27.

Z-scores height/age or weight/age during the first 6 months in data from developing countries.[7] Here, children are losing the nutritional status provided at birth within the first half-year of life.

Now the three Z-scores under consideration are defined more precisely. In general,

$$Z\text{-score} = \frac{\text{individual's value} - \text{median value of ref. pop.}}{\text{standard deviation value of ref. pop.}}.$$

Thus

$$Z_{HA} = \frac{H - m_{HA}}{SD_{HA}}, \quad Z_{WA} = \frac{W - m_{WA}}{SD_{WA}}, \quad Z_{WH} = \frac{W - m_{WH}}{SD_{WH}},$$

where H, W and A are the child's height, weight, and age, respectively; m_{HA} is the median height in the corresponding age and sex group of the reference population; m_{WA} is the median weight in the corresponding age and sex group of the reference population; m_{WH} is the median weight in the corresponding height and sex group of the reference population; and SD_{HA}, SD_{WA} and SD_{WH} are the corresponding standard deviations of the reference population. Different standard deviations are used in the reference population in connection with weight, since its distribution is skewed. To be more precise, there are two standard deviations, one for values above and one for values below the median. Suppose a boy is of 64·0 cm height. We find the lower standard deviation in Table II to be 0·9 kg, whereas the upper standard deviation is 1·2 kg. For the computation of Z_{WH} we have to take into consideration the actual weight of the boy. If it is above the median of the reference population, we use 1·2 kg SD; otherwise we use 0·9 kg SD. This is completely analogous for Z_{WA}. To demonstrate the computation with actual values, let us further assume the boy is 6 months old and weighs 5·8 kg. We find the three scores as $Z_{HA} = -1{\cdot}42$, $Z_{WA} = -1{\cdot}07$ and $Z_{WH} = -0{\cdot}44$.

In the setting of a developing country the age of the child is given by the mother and is very often imprecise, leading sometimes to a phenomenon called *age heaping*,[8] which means that the Z-score shows the behaviour of a wave if plotted against age. Therefore a Z-score, such as Z_{WH}, not involving age is often preferred. Also, the possible error for Z_{WA} when the age is not entirely correct is usually greater than for Z_{HA}. This is because weight increments are greater than height increments over time in a preschool child.

2.3. Computational aspects

It is clear that the calculation of Z-scores for an individual subject needs great care and, if done by hand, such computations can contain errors. It is therefore advisable whenever possible to use a computer program. Such software is available either in Epi-Info[9] in a submodule called Measure, or in a program specifically developed by Böhning and Schelp[10] to compute these scores and which is available on request.

2.4. The choice of cut-off values

The construction of Z-scores (as well as percentiles and median percentages) is done to achieve a measure for the nutritional status of a child. Typically, a low Z-score will indicate malnutrition. The question of a threshold or cut-off value arises and is still under discussion. However, quite frequently values of $Z < -2$ are judged to indicate a certain degree of malnutrition. Thus *stunted children* are defined as those with a $Z_{HA} < -2$, and *wasted children* with a $Z_{WH} < -2$. Presently the reasons for stunting are under active discussion, and the former implicit assumption that Z_{HA} represents the nutritional status in the past is no longer internationally accepted; however,

wasting represents acute malnutrition. Usually, Z_{HA} and Z_{WH} have a low correlation, whereas Z_{HA} and Z_{WA} as well as Z_{WH} and Z_{WA} are highly correlated.

A statistical way of understanding this choice is as follows. Suppose we want to decide between two alternative hypotheses: the first (H_0) that the child is not malnourished and effectively comes from the reference population, and the second (H_1) that the child is indeed malnourished and does not come from the reference population.

If H_0 is true, the Z-score follows a normal distribution with mean 0 and variance 1. Now, think of the cut-off value as a statistical test (decision for H_1 if Z-score is below -2, otherwise decision for H_0); then the choice of -2 corresponds to the usual *significance level* of 2·5 per cent. In other words, with this choice of cut-off value, we are willing to classify 2·5 per cent of our observations *falsely* to be malnourished. Of course, one could argue that it might be more appropriate to use a one-sided significance value of 5 per cent, which leads to a cut-off value of 1·645. In fact, the latter choice has been supported by Schelp *et al.*[11] indirectly by investigating the relationship of malnutrition to morbidity (cough, fever, running nose, etc.) in preschool children in a rural area in north-east Thailand. The optimal cut-off value was found using the maximally selected odds ratios technique.[12] Given a Z-score and a certain value for the cut-off point, the child can be above or below this value. Also, in terms of morbidity a child is either healthy or sick. This leads to 2×2 table with a certain odds ratio. Now the cut-off value is varied over the possible data values of Z-scores, and that value is chosen as cut-off where the odds ratio becomes maximal.

3. PARAMETRIC ESTIMATION OF ABC

We consider a nutritional indicator Z and assume that it is normally distributed with mean μ and SD σ in the study population and, by construction, with mean zero and SD unity in the reference population. Let φ_{obs} be the density of the study population and φ the standard normal density of the reference population. Typically, φ_{obs} will have its mass to the left of the mass of φ, so that there is some value z_0 such that $\varphi_{obs}(z) \geqslant \varphi(z)$ for z left of z_0 and $\varphi_{obs}(z) \leqslant \varphi(z)$ for z right of z_0. This implies (Figure 1) that there is a value at $z = z_0$ such that

$$\varphi_{obs}(z_0) - \varphi(z_0) = 0. \tag{1}$$

The *area between curves* (ABC) is thus defined as the area above φ and below φ_{obs} from $-\infty$ to z_0, that is $ABC = \Phi_{obs}(z_0) - \Phi(z_0)$, where Φ and Φ_{obs} are the cumulative distribution functions of the reference and study populations. It can be interpreted as the percentage of children or adolescents that do not fall under the curve of the reference population. Interpretations such as standardized prevalence as suggested by Mora[6] can be misleading, in particular, if a parametric normal assumption is used. In that case (parametric normal) standardized prevalence would be just a function of the mean, as we will see below, and thus would not provide new information. In many cases, however, the normal assumption for the study population is too strong, and the ABC statistic gives additional information about the deviation of the study population from the reference. Before turning to the more complex non-parametric case in Section 4, let us consider the simpler parametric normal situation.

3.1. Common variance σ^2

In this case, a unique value of z_0 satisfying (1) exists and is independent of the value of σ which we assume to be unity without loss of generality. Obviously $z_0 = \mu/2$ is just the arithmetic mean of the two populations under consideration. Consequently, $ABC = \Phi_{obs}(\mu/2) - \Phi(\mu/2) =$

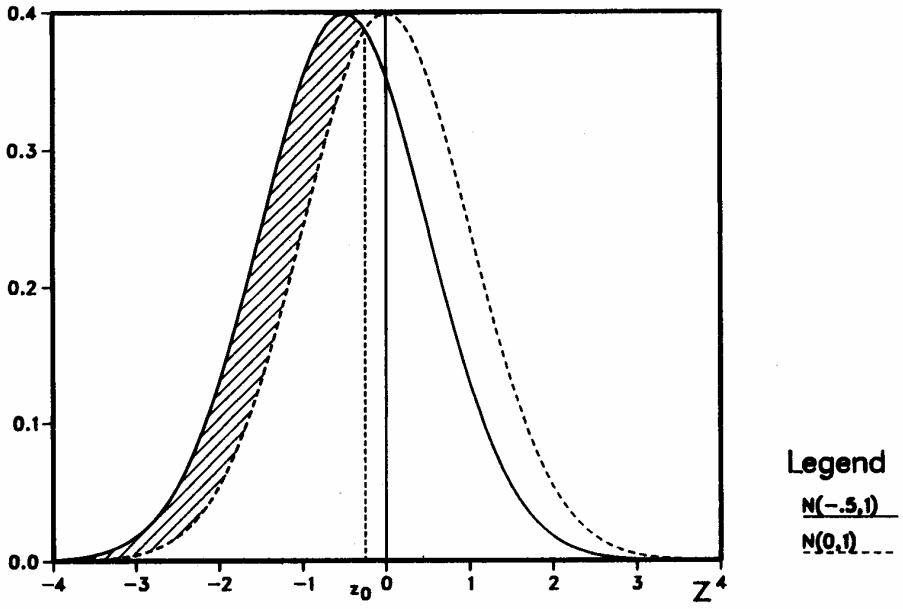

Figure 1. Area between curves (ABC)

$\Phi(\mu/2 - \mu) - \Phi(\mu/2) = \Phi(-\mu/2) - \Phi(\mu/2) = 1 - 2\Phi(\mu/2)$. This can be estimated by

$$\hat{ABC} = 1 - 2\Phi(\bar{z}/2), \tag{2}$$

where $\bar{z}$ is the mean of a sample from the study population. In Result 1 of the Appendix some facts are summarized which are useful for the statistical inference (P-value, confidence interval). For example, a large-sample 95 per cent confidence interval for ABC is given by

$$1 - 2\Phi((\bar{z} \pm 1{\cdot}96/\sqrt{n})/2),$$

that is replacing μ in $1 - \Phi(\mu/2)$ by its large-sample confidence interval.

3.2. Different variances σ^2

In this case there are two solutions, as demonstrated in Figure 2, and equation (1) takes the form

$$\frac{1}{\sigma}\exp\left\{-\frac{1}{2}\left(\frac{z_0 - \mu}{\sigma}\right)^2\right\} - \exp(-\tfrac{1}{2}z_0^2) = 0 \tag{3}$$

or equivalently,

$$(1 - \sigma^2)z_0^2 - 2z_0\mu + \mu^2 - 2\sigma^2\ln(\sigma) = 0. \tag{4}$$

This is a quadratic equation in z_0 with two solutions,

$$z_0 = [\mu \pm \sigma\sqrt{\Re}]/(1 - \sigma^2), \quad \Re = \mu^2 + 2(\sigma^2 - 1)\ln(\sigma). \tag{5}$$

If $\sigma^2 > 1$, as is the case in Figure 2, then the left zero (the one of interest) is given by $[\mu + \sigma\sqrt{\Re}]/(1 - \sigma^2)$ and (5) is always a real root since $2(\sigma^2 - 1)\ln(\sigma)$ is always positive. It

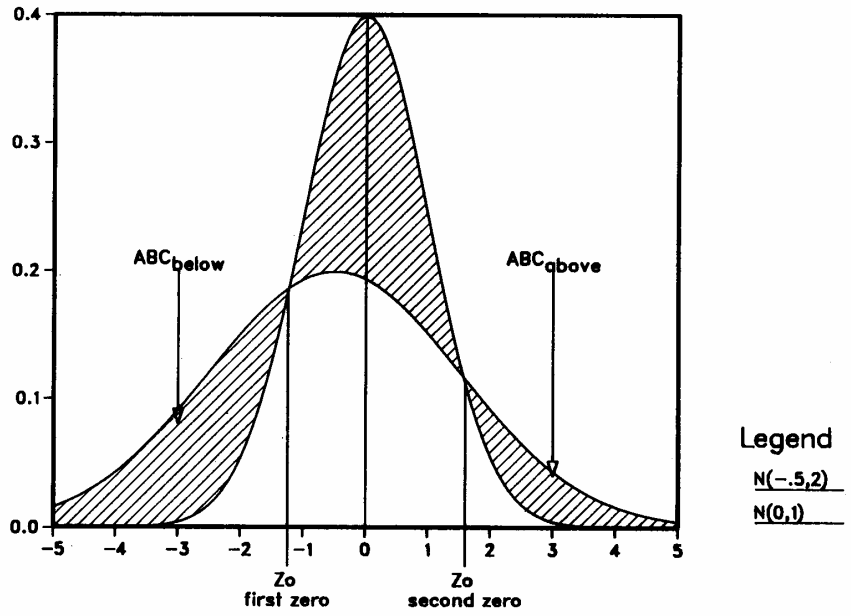

Figure 2. The role of different standard deviations: two points of intersection, two areas between the curves

should be noted that the left zero does not necessarily lie between the mean of the study population and the mean of the reference population.

Thus, the ABC is defined as the difference between the area below the density of the study population and the density of the reference population:

$$\begin{aligned} \text{ABC} &= \Phi_{\text{obs}}([\mu + \sigma\sqrt{\Re}]/(1-\sigma^2)) - \Phi([\mu + \sigma\sqrt{\Re}]/(1-\sigma^2)) \\ &= \Phi([\mu\sigma + \sqrt{\Re}]/(1-\sigma^2)) - \Phi([\mu + \sigma\sqrt{\Re}]/(1-\sigma^2)). \end{aligned} \tag{6}$$

This is the formula from which Table 2 in Mora[6] can be reproduced. The associated formula (Reference 6, p. 139) $\Phi([\mu\sigma + \sqrt{\Re}]/(1-\sigma^2)) + \Phi([\mu - \sigma\sqrt{\Re}]/(1-\sigma^2))$ is evidently incorrect. The problem remains to find a point estimate of ABC given in (6). This could be accomplished by simply replacing μ and σ in (6) by its sample estimates $\bar{z}$ and s, where s^2 is the sample variance. It should be noted, however, that replacing μ in (6) by its large-sample 95 per cent confidence interval, namely $\bar{z} \pm 1{\cdot}96s/\sqrt{n}$, will *not* provide a 95 per cent confidence interval for ABC, in full generality. It might be called a *bona fide* interval. Before this question is discussed any further, a different method of inference will be suggested.

3.3. Partitioning the nutritional status

Recall the interpretation of the ABC thus far: it is the percentage of those children in the sample that does not fall under the curve of the reference population. In contrast to the equal variance case, we have a second area, namely to the right of the second zero. This area could again be

interpreted as a percentage of children which does not fall under the reference density curve. However, these children would be described as *well nourished.*

In such circumstances it might be consistent to consider a partition of ABC on the following basis. Thus we calculate ABC for the malnourished and well nourished as

$$\mathrm{ABC}_{\mathrm{below}} = \Phi([\mu\sigma + \sqrt{\Re}]/(1-\sigma^2)) - \Phi([\mu + \sigma\sqrt{\Re}]/(1-\sigma^2))$$

$$\mathrm{ABC}_{\mathrm{above}} = \Phi([\mu - \sigma\sqrt{\Re}]/(1-\sigma^2)) - \Phi([\mu\sigma - \sqrt{\Re}]/(1-\sigma^2)),$$

and finally

$$\mathrm{ABC}_{\mathrm{overall}} = \mathrm{ABC}_{\mathrm{below}} - \mathrm{ABC}_{\mathrm{above}}. \tag{7}$$

The meanings of these measures are clear from Figure 2, in which $\mathrm{ABC}_{\mathrm{below}} = 24{\cdot}8$ per cent, $\mathrm{ABC}_{\mathrm{above}} = 9{\cdot}2$ per cent and $\mathrm{ABC}_{\mathrm{overall}} = 15{\cdot}6$ per cent. The first number is reported in Reference 6 (Table 2, p. 138) as the percentage of malnourished children. In our view, only $\mathrm{ABC}_{\mathrm{overall}} = 15{\cdot}6$ per cent can have this interpretation, leading to quite different numerical estimates. Again, we would stress that this partition will apply only rarely in practical situations and is therefore a theoretical rather than a practical difficulty.

Of more importance is that one might be tempted to ask: how realistic is this case that the standard deviation of the anthropometric indicator deviates drastically from 1? Our own experience supports the fact that the observational density deviates more in the direction of non-symmetry.

4. NON-PARAMETRIC ESTIMATION OF ABC

Typically, the empirical distributions of anthropometric indicators such as Z_{HA} or Z_{WH} have *heavier* left tails. In contrast, the parametric normal density estimator distributes the mass equally under each tail of the distribution, thus providing a biased estimate of ABC.

4.1. Non-parametric formulation

With this motivation in mind, we consider a non-parametric formulation of the problem. Suppose that the density intersecting point z_0, the one satisfying (1), is known. Then we could estimate ABC as $F(z_0) - \Phi(z_0)$, where F is the non-parametric estimator of Φ_{obs}, namely the empirical distribution function. The problem is the determination of z_0! To solve equation (1) we need to have a non-parametric estimator of φ_{obs}. However, it is not that easy to define the non-parametric density estimator. One way to approach the problem would be to consider non-parametric density estimation via kernels[13] or semi-parametric density estimation via mixtures of distributions.[14] Instead we will see that it is simpler and more promising to follow a different approach. We reformulate (1) as

$$\frac{\mathrm{d}}{\mathrm{d}z_0}[\Phi_{\mathrm{obs}}(z_0) - \Phi(z_0)] = 0, \tag{8}$$

or equivalently, find the maximum of

$$\Delta(z) = \Phi_{\mathrm{obs}}(z) - \Phi(z)$$

with respect to z. But the non-parametric estimator of $\Delta = \Phi_{\mathrm{obs}} - \Phi$ is $\hat{\Delta} = F - \Phi$, and ABC is estimated as the maximum of $F - \Phi$! In other words: given a sample $z_1, \ldots, z_n$ of Z-scores of the

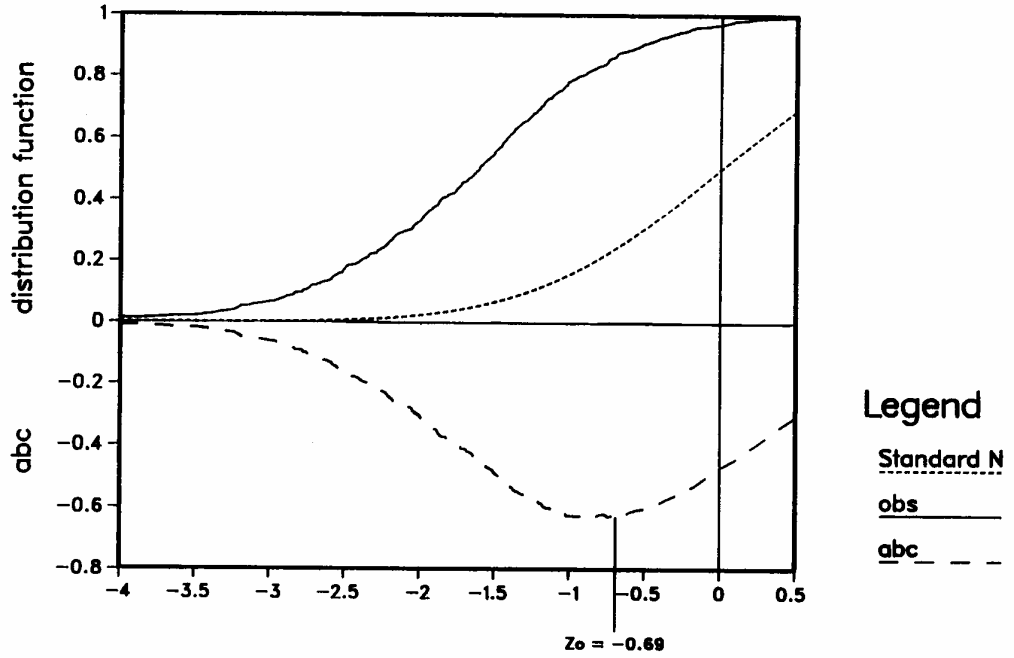

Figure 3. The ABC measure for height/age

study population, we could estimate ABC as

$$\hat{ABC}^{npar} = \max_{1 \leqslant i \leqslant n} \hat{\Delta}(z_i). \tag{9}$$

Figure 3 shows the empirical distribution function of height/age for 707 preschool children in north-east Thailand. As reference distribution we have plotted the standard normal. The third curve shows the negative difference (implying that the maximum we are searching for is the minimum in the figure) between these two cumulative distributions. The maximum is attained at $z_0 = -0{\cdot}69$, giving an estimate of $\hat{ABC}^{npar} = 0{\cdot}63$. The corresponding parametric estimate is 0·59 (assuming common variance) and 0·61 (assuming different variances). Evidently, the two parametric estimates differ from the non-parametric by 4 per cent and 2 per cent.

4.2. Non-parametric partition of the nutritional status

Clearly, the terms of ABC below, above and overall could be given analogously to Section 3.3. For example, the existence of a second zero could be detected as a minimum of $\hat{\Delta}$. Figure 4 shows the situation of Figure 2 in terms of F, Φ and $\hat{\Delta}$. Note again that $-\hat{\Delta}$ is plotted instead of $\hat{\Delta}$. Obviously,

$$\hat{ABC}^{npar}_{overall} = \hat{ABC}^{npar}_{below} - \hat{ABC}^{npar}_{above} = \max_{1 \leqslant i \leqslant n} \hat{\Delta}(z_i) - \left| \min_{1 \leqslant i \leqslant n} \hat{\Delta}(z_i) \right| = \max_{1 \leqslant i \leqslant n} \hat{\Delta}(z_i) + \min_{1 \leqslant i \leqslant n} \hat{\Delta}(z_i).$$

Again, from our experience in practical cases, such a partition will not be necessary.

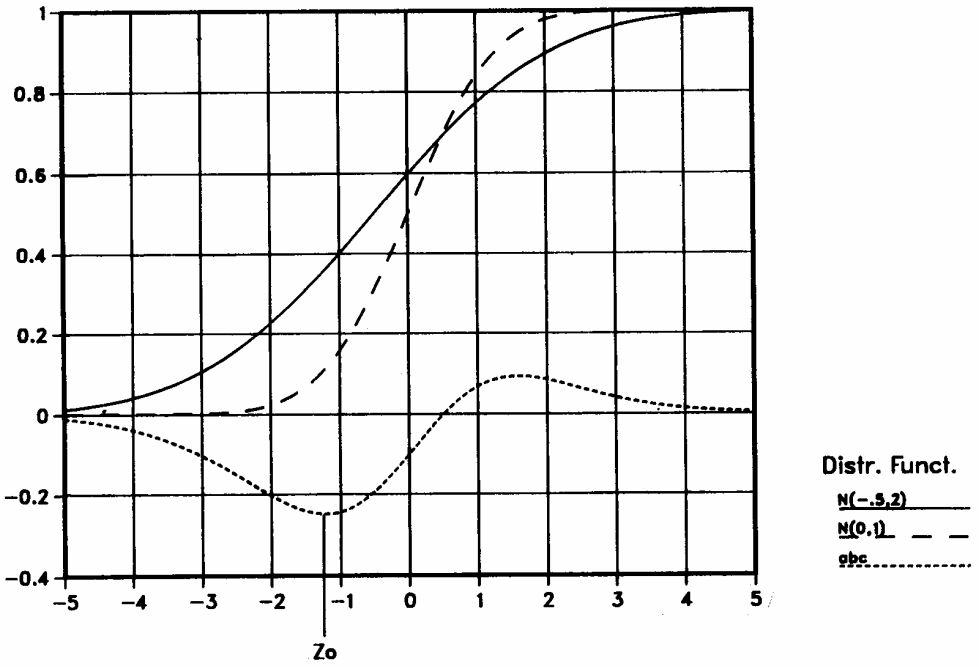

Figure 4. Area between curves: the role of different standard deviations

4.3. Statistical inference for the non-parametric estimator

It turns out to be surprisingly simple to give a confidence interval for the ABC parameter. Let Φ_{obs} denote the cumulative distribution function of the study population. Then we have

$$\max_z|D_n(z)| = \max_z|F(z) - \Phi_{obs}(z)|. \tag{10}$$

This is the usual Kolmogorov–Smirnov statistic.[15] In addition, (10) can be written as

$$\max_z|D_n(z)| = \max_z|F(z) - \Phi(z) - (\Phi_{obs}(z) - \Phi(z))| = \max_z|\hat{\Delta}(z) - \Delta(z)|. \tag{11}$$

Thus the distribution of (11) is known as the Kolmogorov–Smirnov distribution. This implies not only that $\hat{ABC}^{npar}$ is a consistent and unbiased estimator of ABC, but also that a $(1 - \alpha)$ confidence band for $\Delta(z)$ can be found as $\hat{\Delta}(z) \pm d_\alpha$, and in particular we find that the $(1 - \alpha)$ confidence interval for ABC^{npar} as

$$\hat{ABC}^{npar} \pm d_\alpha, \tag{12}$$

where d_α is the $(1 - \alpha)$ percentile of the Kolmogorov–Smirnov distribution. These results are derived in detail in Result 2 in the Appendix.

5. APPLICATION

We are interested in estimating the prevalence of stunting (Z_{HA}) and wasting (Z_{WH}) in preschool children in a region in north-east Thailand. The data stem from an intervention project aimed to improve the nutritional status of all preschool children up to 60 months of age out of six villages

in the north-east of Thailand. Anthropometric measurements were taken from 707 children every three months over a period of three years and their Z-scores computed. The following data set describes the situation at baseline. For Z_{HA} we find the following descriptive statistics and estimates for ABC (only one stationarity point of $\hat{\Delta}$ corresponding to a maximum):

$\hat{ABC}$(stunting) = $\max_i \hat{\Delta}(z_i) = \hat{\Delta}(z_0) = 0{\cdot}63$, $z_0 = -0{\cdot}69$

95% CI $\hat{ABC}$(stunting) $\pm d_\alpha$ = [0·58, 0·68]
parametric common variance 0·59[0·58, 0·62]
parametric different variance 0·61
$\hat{ABC}$(wasting) = $\max_i \hat{\Delta}(z_i) = \hat{\Delta}(z_0) = 0{\cdot}32$, $z_0 = -0{\cdot}32$
95% CI $\hat{ABC}$(wasting) $\pm d_\alpha$ = [0·27, 0·37]
parametric common variance 0·26[0·23, 0·29]
parametric different variance 0·30.

It can be seen that there is a drastic difference between the parametric ABC estimate of wasting when a common variance is assumed. The difference from the non-parametric ABC estimate is less strong if a different variance is allowed. To judge the possible deviations it might be valuable to mention the corresponding points in Mora.[6] He was obviously aware of the issue, since he writes: 'The frequent skewness in the distributions observed in developing countries may introduce some underestimation in the calculations; however, exact estimations applying our method to actual data from nutrition surveys in developing countries showed that the magnitude of the error is negligible (under 10 per cent of the total prevalence)' (p. 139). What is remarkable is not the fact that there is a bias up to 10 per cent if a parametric normal is assumed (which is also supported by the study above) but rather the opinion that a bias up to 10 per cent is negligible. Mora continues: 'Although adjusting for skewness in prevalence estimates is theoretically feasible, for practical purposes this would be an unnecessary sophistication.' It should be pointed out that our non-parametric approach is a simple alternative as well as having the statistical advantages of being always unbiased and providing a valid confidence interval.

6. DISCUSSION

This paper has provided a statistical concept for the area between the density curve of a sample from the study population and the reference population. Statistical inference is provided for both the parametric normal and the non-parametric situation. The question arises as to how much the parametric and non-parametric approaches differ. To answer this, let us assume that 95 per cent of our study population follows a normal distribution with mean $-1{\cdot}5$ and variance 1, and 5 per cent of the population is 'contaminated' by a normal distribution with mean $-4{\cdot}5$ and variance 1. This subpopulation could be thought of as a group of children suffering severe malnourishment. In other words we are assuming that the study population follows a mixture of two normals:

$$0{\cdot}95\varphi(x + 1{\cdot}5) + 0{\cdot}05\varphi(x + 4{\cdot}5). \tag{13}$$

If the reference population is again the normal with mean 0 and variance 1 there is a population ABC of 0·558 at $z_0 = -0{\cdot}78$. In the parametric normal case with common variance it is falsely assumed that the data are coming from a normal with variance 1 and mean $0{\cdot}95 \times (-1{\cdot}5) + 0{\cdot}05 \times (-4{\cdot}5) = -1{\cdot}65$. From Figure 5 it is clear that the corresponding ABC is larger than the true ABC, leading to an overestimation. In the parametric normal case with different variances it is falsely assumed that the data are coming from a normal, again with

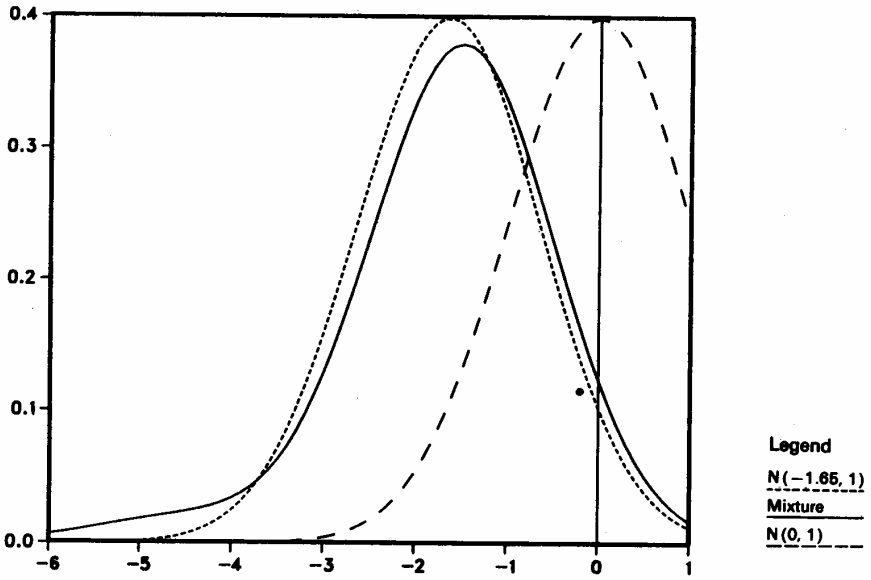

Figure 5. Overestimation of ABC by parametric estimation (common variance) for contaminated normal study population

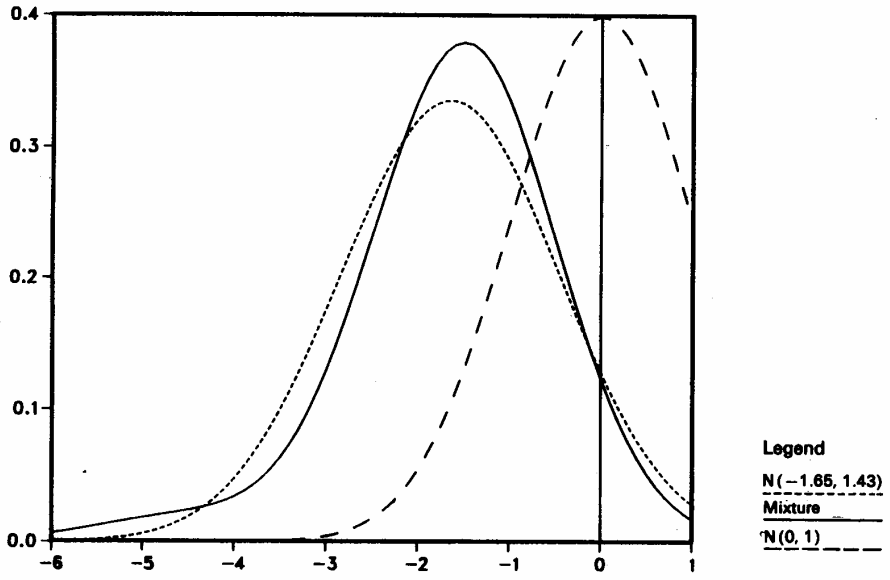

Figure 6. Underestimation of ABC by parametric estimation (different variances) for contaminated normal study population

mean $-1{\cdot}65$ but with overdispersed variance $1 + 0{\cdot}95(1{\cdot}65 - 1{\cdot}5)^2 + 0{\cdot}05(1{\cdot}65 - 4{\cdot}5)^2 = 1{\cdot}43$. From Figure 6 it becomes clear that this time there will be an underestimation bias. A simulation demonstrates this point. If a sample of size 10,000 is taken from the mixture (13) there is a non-parametric estimate $\hat{ABC} = 0{\cdot}556$ at $z_0 = -0{\cdot}766$ with confidence interval $(0{\cdot}542, 0{\cdot}569)$ which is already quite close to the true ABC. For the parametric with common variance unity an overestimating $\hat{ABC} = 0{\cdot}587$ with confidence interval $(0{\cdot}582, 0{\cdot}593)$ can be observed, and for the different variance situation there is an underestimating $\hat{ABC} = 0{\cdot}548$ with *bona fide* confidence interval $(0{\cdot}541, 0{\cdot}554)$. It should be noted that in both parametric cases the confidence intervals do not cover the true ABC of 0·558.

Sometimes samples from different populations have to be compared in their deviation from the same population. Then the two-sample Kolmogorov–Smirnov test could be used to test whether the two samples have the same ABC parameter.

Also, there is the possibility of applying this approach to other problems. It is a widely accepted technique in medicine and epidemiology to use a cut-off value to detect individuals at risk. Typical examples would be blood pressure, cholesterine level, intraocular pressure and blood glucosis. In many situations one is interested in giving a prevalence estimate of the people suffering from high blood pressure, having a high cholesterine level etc. If a reference population is available, the above technique can be applied similarly.

APPENDIX

Result 1

Here we summarize a helpful distributional property of (2). Under the assumptions of Section 3.1 the following statements hold:

(a) The cumulative distribution function of $\hat{ABC}$ is

$$\Phi_{\hat{ABC}}(x) = 1 - \Phi(\sqrt{(n)}\{2\Phi^{-1}[(1-x)/2] - \mu\}).$$

(b) The P-value of $\hat{ABC}$, namely $\Pr_{H_0}\{\hat{ABC} \geqslant \hat{abc}\}$, is given as $\Phi(\sqrt{(n)}\{2\Phi^{-1}[(1-\hat{abc})/2]\})$, where $\hat{abc}$ is the observed value of $\hat{ABC}$, and we are testing H_0: ABC = 0 against H_1: ABC > 0.

(c) Let $z \pm t_{\alpha/2}/\sqrt{n}$ denote the usual $(1-\alpha)$ confidence interval for μ. Then $1 - 2\Phi[(\bar{z} \mp t_{\alpha/2}/\sqrt{n})/2]$ is a $(1-\alpha)$ confidence interval for ABC. Note that the signs are reversed in the ABC confidence interval estimator.

Proof

To prove (a) and (b):

$$\Pr\{\hat{ABC} \leqslant x\} = \Pr\{1 - 2\Phi(\bar{z}/2) \leqslant x\} = \Pr\left\{\frac{1-x}{2} \leqslant \Phi(\bar{z}/2)\right\} = \Pr\left\{\Phi^{-1}\left(\frac{1-x}{2}\right) \leqslant \bar{z}/2\right\}$$

$$= \Pr\left\{2\Phi^{-1}\left(\frac{1-x}{2}\right) \leqslant \bar{z}\right\} = 1 - \Phi\left(\sqrt{n}\left\{2\Phi^{-1}\left(\frac{1-x}{2}\right) - \mu\right\}\right).$$

To prove (c) we note the equivalence of the following statements:

$$\bar{z} - t_{\alpha/2}/\sqrt{n} \leqslant \mu \leqslant \bar{z} + t_{\alpha/2}/\sqrt{n}$$

$$\Phi((\bar{z} - t_{\alpha/2}/\sqrt{n})/2) \leqslant \Phi(\mu/2) \leqslant \Phi((\bar{z} + t_{\alpha/2}/\sqrt{n})/2)$$

$$1 - 2\Phi((\bar{z} - t_{\alpha/2}/\sqrt{n})/2) \geqslant \mathrm{ABC} \geqslant 1 - 2\Phi((\bar{z} + t_{\alpha/2}/\sqrt{n})/2)$$

This completes the proof. □

Result 2

Let $\Delta(z) = \Phi_{obs}(z) - \Phi(z)$ and let $\hat{\Delta}(z) = F(z) - \Phi(z)$ be its pointwise estimate. Also, let d_α denote the $1 - \alpha$ fractile of the Kolmogorov–Smirnov distribution, for example $d_\alpha = \Phi_{KS}^{-1}(1 - \alpha)$, where Φ_{KS} is the cumulative distribution function of the Kolmogorov–Smirnov statistic. Then

(a) $\hat{\Delta} = F - \Phi$ estimates Δ unbiasedly and consistently.
(b) $\hat{\mathrm{ABC}}^{npar} = \max_{z_i} \hat{\Delta}(z_i)$ is a consistent estimate of ABC.
(c) $F(z) - \Phi(z) \pm d_\alpha$ is a $(1 - \alpha)$ confidence band for $\Phi_{obs}(z) - \Phi(z)$.
(d) $\hat{\mathrm{ABC}}^{npar} \pm d_\alpha = \max_{z_i} \hat{\Delta}(z) \pm d_\alpha$ is a $(1 - \alpha)$ confidence interval for $\mathrm{ABC} = \max_z \Phi_{obs}(z) - \Phi(z)$.

Proof

The following three statements are equivalent and hold with probability $1 - \alpha$:

$$\max_z |F(z) - \Phi(z) - (\Phi_{obs}(z) - \Phi(z))| \leqslant d_\alpha$$

$$\Leftrightarrow -d_\alpha \leqslant F(z) - \Phi(z) - (\Phi_{obs}(z) - \Phi(z)) \leqslant d_\alpha \quad \text{for all } z$$

$$\Leftrightarrow F(z) - \Phi(z) - d_\alpha \leqslant \Phi_{obs}(z) - \Phi(z) \leqslant F(z) - \Phi(z) + d_\alpha \quad \text{for all } z.$$

The last statement holds for all z, and thus it is true also for the maximum:

$$\max_z [F(z) - \Phi(z) - d_\alpha] \leqslant \max_z [\Phi_{obs}(z) - \Phi(z)] \leqslant \max_z [F(z) - \Phi(z) + d_\alpha]. \qquad \square$$

ACKNOWLEDGEMENTS

This work is supported by the German Research Foundation under various grants. Part of this work was completed while the first author was visiting the Faculty of Tropical Medicine and the Department of Biostatistics, School of Public Health, Mahidol University, Bangkok with support from the German Academic Exchange Service (DAAD). He would like to express his thanks to the faculty for their support during his stay.

REFERENCES

1. United Nations Administrative Committee on Coordination, Subcommittee on Nutrition (UN ACC/SCN). *Update on the Nutrition Situation. Recent Trends in Nutrition in 33 Countries*, ACC/SCN Secretariat, Geneva, 1989.
2. World Health Organization. *Measuring Change in Nutritional Status*, WHO, Geneva, 1983.
3. Keller, W., Donoso, G. and De Maeyer, E. M. 'Anthropometry in nutritional surveillance: a review based on nutritional anthropometry', *Nutrition Abstracts and Reviews*, **46**, 591–609 (1976).
4. Waterlow, J. C., Buzina, R., Keller, W., Lane, J. M., Nichaman, M. Z. and Tanner, J. M. 'The presentation and use of height and weight data for comparing the nutritional status of groups of children under the age of 10 years', *Bulletin of the World Health Organization*, **55**, 489–498 (1977).

5. Carlson, B. A. 'Assessing the nutritional status of young children: how to do it with national household sample surveys: a methodological overview', *Bulletin of the International Statistical Institute, Contributed Papers of the 47th Session*, **I**, 183–184 (1989).
6. Mora, J. O. 'A new method for estimating a standardized prevalence of child malnutrition from anthropometric indicators', *Bulletin of the World Health Organization*, **67**, 133–142 (1989).
7. Wolff, M. C., Perez, L., Gibson, J. G., Lopez, L. S., Peniston, B. and Wolff, M. M. 'Nutritional status of children in the health district of Cusco, Peru', *The American Journal of Clinical Nutrition*, **42**, 531–541 (1985).
8. Heitjan, D. F. and Rubin, D. B. 'Inference from coarse data via multiple imputation with application to age heaping', *Journal of the American Statistical Association*, **85**, 304–314 (1990).
9. Dean, A. D., Dean, J. A., Burton, A. H. and Dicker, R. C. *Epi Info, Version 5: a Word Processing, Database, and Statistics Program for Epidemiology on microcomputers*, USD Incorporated, Stone Mountain, Georgia, 1990.
10. Böhning, D. and Schelp, F.-P. 'A FORTRAN subroutine for computing indicators of the nutritional status of children and adolescents', *Statistical Papers*, **27**, 141–150 (1985).
11. Schelp, F.-P., Sormani, S., Pongpaew, P., Vudhivai, N., Egormaiphol, S. and Böhning, D. 'Seasonal variation of wasting and stunting in preschool children during a three-year community-based nutritional intervention study in northeast Thailand', *Tropical Medicine and Parasitology*, **41**, 279–285 (1990).
12. Miller, R. and Siegmund, D. 'Maximally selected χ^2-statistics', *Biometrics*, **38**, 1011–1016 (1982).
13. Thisted, R. A. *Elements of Statistical Computing*, Chapman and Hall, London, 1988.
14. Titterington, D. M., Smith, A. F. M. and Makov, U. E. *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York, 1985.
15. Mood, A. M., Graybill, F. A. and Boes, D. C. *Introduction to the Theory of Statistics*, 3rd edn, McGraw-Hill, New York, 1974.
16. Baker, S. K., Duran-Bordier, A. and Mahamadou, O. 'Rapid nutritional assessment in Niger: a practical approach', *Bulletin of the International Statistical Institute, Contributed Papers of the 47th Session* **I**, 91–92 (1989).
17. Gorstein, J. and Pradilla, A. 'Interpreting anthropometric measures—issues of standardization in analysis and presentation', *Bulletin of the International Statistical Institute, Contributed Papers of the 47th Session* **I**, 389–390 (1989).
18. Lindgren, B. W. *Statistical Theory*, 3rd Edition, Macmillan Publishing Co., New York, 1976.
19. Rutstein, S. and Sommerfelt, E. 'The nutritional status of small children: methods and results from the demographic and health surveys', *Bulletin of the International Statistical Institute, Contributed Papers of the 47th Session* **II**, 278 (1989).