LINKED CROSS-SECTIONAL STUDY FOR DETERMINING NORMS AND GROWTH RATES—A PILOT SURVEY ON INDIAN SCHOOL-GOING BOYS

By M. N. RAO

All India Institute of Hygiene and Public Health, Calcutta

and

C. RADHAKRISHNA RAO

Indian Statistical Institute

SUMMARY. The paper deals with the statistical analysis of measurements obtained on school-going boys of various ages for estimating norms and growth rates, in what is termed as an LCS (Linked Cross-Sectional) study. This is a cross between a *longitudinal* study where some boys are chosen at a fixed young age and are measured periodically over a certain number of years covering the range of study, and a *cross-sectional* study where boys of different ages are chosen and each boy is measured only at one age and at one point of time. In an LCS study, boys of different ages are chosen and each boy is measured periodically over a comparatively smaller number of years. Thus, if one is interested in studying growth of boys between the ages of 5 and 16, a certain number of boys at each age (between 5 and 16) may be chosen and measured on three consecutive birth days. The study is thus restricted to a period of about 2 years, whereas in a longitudinal study the period will be about 11 years. The advantages of such a short range study compared to a purely longitudinal study extended over a larger number of years are explained.

1. Introduction

From one of these laboratories, two studies have been reported earlier on growth of children (M. N. Rao and Bhattacharjee, 1952) and (M. N. Rao and Bhattacharjee, 1953). The present study is made on a select sample of boys of the age group 5-16 with the following objectives :

(1) To study the pattern of physical growth of Indian children on a more satisfactory basis than attempted before.

(2) To study the dependence of blood pressure on age, weight and height among boys in the age group 5-16.

(3) To illustrate the use of appropriate statistical techniques in the estimation of norms and growth rates in what we call a linked cross-sectional study.

(4) To provide information for planning of future surveys for determining norms and growth rates.

Scientific studies on growth of children are of practical importance in school health programmes. In their monograph on *Health Observation of School Children*, Wheatley and Hallock (1951) have reviewed all the scientific data for the complete growth of a child—both physical and psychological. In some of the health clinics for children in the United States, comprehensive records are maintained to watch the growth and development of individual children. Wetzel-grid record is one such. Many variables are included to compute this grid, e.g., body build, development, basal metabolism, etc. Auxidromic progress ('age schedule of development') in the grid measured as the deviation of 'actual' from 'expected' gives a quantitative expression of the physical fitness of a child.

Public health and more so, school health programmes in India have not yet developed to such an extent as in some of the Western countries where such grids are used. But attempts have been made in the past to report on the health and nutrition status of school children.

The only comprehensive survey reported of anthropometric measurements in school-going children was by Someswar Rao (1961). From about 24 papers published between 1936 and 1954 he reviewed a sample of 41,587 boys and 6,355 girls, measured as a part of nutrition surveys. Someswar Rao says :

> 'The investigations were confined mainly to the lower economic groups... In most of the reports, no mention was made as to how the ages of the subjects were assessed. No effort was ever made to obtain the correct age. It was probably difficult to elicit any reliable information regarding age especially while dealing with the illiterate parents. No uniform method was adopted in expressing the age and in none of the above studies, however, was the sample suitable to provide the necessary basis for constructing normal standards.'
>
> 'In some of the surveys under review, even spring balances were used for obtaining weight and in the techniques of measurement and the presentation of results a lot was left to be desired.'

We have also examined a large number of isolated studies on growth of Indian children in recent journals, which suffer from the same defect as the earlier studies. There is a great need to make a systematic start and collect data according to well designed plans.

2. Sample and the collection of data in the present study

The sample of this report consisted of boys going to St. Xavier's School, Calcutta. This School was selected due to the excellent records maintained by them and the rigidity in admissions and birth certification. It is one of the oldest and well run schools of Calcutta. A large majority of the school children are from high socio-economic parentage able to afford a relatively costly education. It can be reasonably said that all the boys have comparatively healthy environments at home with adequate dietary and nutrition standards. The selection of such a sample would ensure the optimum expectations in health levels that can be reasonably reached through school health programmes in an Indian community within the genetic frame-work.

Moreover, this School is cosmopolitan and the non-Bengali students and their parents are migrants to Calcutta from different parts of India. Four zones have been arbitrarily categorized for studying regional differences in physical characteristics of boys :

(1) *Eastern Zone*, consisting of Assam, West Bengal, Bihar and Orissa.

(2) *Northern Zone*, consisting of the Punjab, Delhi and Northern Rajasthan,

(3) *Western Zone*, consisting of Maharashtra, Gujerat and Southern Rajasthan.

(4) *Southern Zone*, consisting of Andhra Pradesh, Madras, Mysore and Kerala.

All the children in this study are given a preliminary check-up clinically to eliminate obvious physical defects and deformities. The data were collected from each child during the month of its birth. Out of about 600 observations, more than 50 per cent were made within 11 days and more than 75 per cent, within 20 days of the date of birth.

The boys were selected at random from lists available in the School with the restriction that at best four boys for each zone in each year group are included in the sample.

TABLE 2.1. FREQUENCY DISTRIBUTION OF BOYS BY STATE AND MONTH OF BIRTH

month	code number of state						zone				total
	3	8	12	13	15	others	south	east	west	north	
January	0	3	3	3	3	7	3	4	6	0	10
February	0	0	0	2	0	10	2	8	7	1	18
March	0	0	0	2	8	2	0	0	3	0	12
April	1	3	3	2	9	9	7	10	6	4	27
May	1	3	0	1	5	2	3	5	3	1	12
June	1	1	3	2	5	1	2	5	3	3	13
July	2	0	6	4	22	4	0	23	7	8	38
August	2	3	9	8	8	13	9	0	14	11	43
September	4	4	7	2	7	8	5	8	10	9	32
October	3	4	2	3	5	9	7	6	8	5	26
November	4	1	7	2	3	8	3	3	9	10	25
December	4	8	2	1	5	8	10	6	9	4	28
total	22	30	42	32	86	81	51	95	85	62	293

Table 2.1 gives the frequency distributions of the boys by month of birth and State or Zone. The disproportionately high frequencies in the second half of the year (July–December) in general and for July in the State with the code number 15 could be explained as due to wrong reporting of month of birth by parents for some reason or other. However, we hope that the age as calculated from the reported month of birth is not in error by more than six months.

Altogether, 293 boys have been measured; 101 of them only at one age, 66 at two consecutive ages, 113 at three consecutive ages and 13 at ages differing by two years. The actual age distributions of boys measured once, twice and thrice are given in Table 2.2. It is seen that the present study is not of a purely cross-sectional or purely longitudinal type. It is a mixture of the two spread over a period of maximum two years, which seems to be an ideal design to follow in projects for determining norms and growth rates. A full discussion of this problem, viz., the designing of surveys for studying growth of children is given in Section 5.

TABLE 2.2. NUMBERS OF INDIVIDUALS MEASURED ONCE, TWICE AND THRICE

age i	once at i	twice at $i, i+1$	thrice at $i, i+1, i+2$	twice at $i, i+2$	total number of measurments
5	5	1	1	1	8
6	3	4	15	1	25
7	7	5	9	—	42
8	3	7	16	2	58
9	2	7	12	1	54
10	8	3	13	2	62
11	1	5	14	2	50
12	8	5	11	1	58
13	4	8	11	—	55
14	6	8	6	1	52
15	17	6	4	1	53
16	17	4	2	1	43
17	10	3	—	—	30
18	4	—	—	—	10
total	101	66	113	13	600

On each visit observations were made on height, weight, systolic and diastolic blood pressures and the number of natural teeth. All the measurements were taken by the same experienced individual, a senior technician, and were regularly checked by one of the authors. The technique of measurements were as follows.

If the student visited had any acute illness or fever during the immediately past few days or if there was any evidence of convalescence, no readings were taken.

The records were taken usually at noon-time: about 3-4 hours after breakfast and just before lunch time.

Blood pressure is taken after the student is made to lie down. Measurement is repeated three times usually and very occasionally a few more times and the steady reading of the last two observations recorded ultimately.

The auscultatory method with a mercury baumanometer is used at the right forearm. It is ensured that the armband is smoothly wound round without folds and the shoulder and the elbow ensured to be resting at about the heart level.

The pressure is raised to nearly 200 mm to obliterate the brachial artery and to remove any possible spasm. The chest piece of the stethoscope is placed flat over the site of the brachial artery at the antecubital fossa, preliminarily located before the pressure is raised. The pressure in the arm bag is slowly released. The level at which a clear sound is suddenly heard is noted as the systolic pressure. In course of time, the crisp sound is muffled (2nd phase), but revives with a reverberation or a bang (3rd phase). Within another 15-20mm fall of mercury there is a sudden feeling of collapse of the sound which gets softish and muffled. This 4th phase is noted as the diastolic pressure. The 5th phase wherein the muffling fades away is ignored. The mercury is brought back to zero before the next reading is taken.

The blood pressure was always taken by the same person.

For other measurements the student is made to take off his shirt, pants, shoes and socks and examined with underwear only. Turbans were removed in the case of boys from the Punjab and the hair loosened over to the shoulders.

Height is taken as he stands at zero level on the measuring machine with heels touching the horizontal level and the vertical scale at right angles. The student stands erect with his sacral and thoracic spinal curves touching the vertical scale. The sliding head-rest is brought down to touch the skull through the crop of hair and the height read out from the side with the head-rest in position.

Weight is then taken by the student standing quiet on the machine kept in the centre of the room.

Teeth. The number of natural teeth, including the presence of stumps and freshly erupted teeth, have been recorded separately for the lower and upper jaws and on the left and right sides.

3. Preliminary reduction of data

We make a preliminary reduction of data to examine some salient features such as zonal differences, relationship between characters etc. In the next section we use some of the inferences drawn on the basis of a preliminary analysis of data and apply appropriate statistical techniques to obtain efficient estimates of norms and growth rates.

Mean values. Let us consider the observations on any character of the boys of a particular age and zone. By taking their simple average, we obtain an age-specific average for the particular character and zone. The age-specific averages so computed for height, weight, blood pressure (systolic and diastolic) are given in Tables 3.1, 3.2, 3.3 and 3.4 separately for the south, east, west and north zones and also for all the zones put together.

Differences between zones. For each age and character, differences between zones are tested by analysis of variance. The F ratios of 'between to within zones' are all small, indicating that age-specific difference, if any, between zones are likely to be small. The samples in our study are not numerous enough to establish significance for each age when the true differences are of a small order. We shall, therefore, try to compare the zones with respect to a suitable integrated measure computed for each zone as a whole for each character.

Standard population approach. In this method we consider a standard population of boys with specified frequencies for ages from 5-16. Applying the age specific averages of a character in a zone, the average for a standard population is calculated for each zone and character. In the present study, the frequencies in the standard population are taken to be proportional to the numbers of boys measured at different ages by pooling the observations over all the zones. The averages, so obtained, for height, weight and blood pressure with respect to the chosen standard population are given in Table 3.5 for the age groups 5-10, 11-16 and 5-16. It is seen that the boys of the North zone are, on the average, taller and heavier than those of the other zones, while the relative positions of the boys in the other zones are not clear-cut. There is, however, an indication that the boys of the South zone have the least average weight among the others. They are shorter than the boys of the East and nearly equal in height with those of the West. The boys of the West zone are shorter than those of the East although nearly equal in weight. The boys of the South zone seem to have higher systolic blood pressure than those of the other zones.

The average number of teeth in the four positions show no zonal differences. The average values for all zones taken together are given in Table 3.6.

The analysis of data, using a standard population, brings out clearly the zonal differences in norms and, perhaps, in some other aspects of the growth of boys. However, as observed earlier, larger samples than those available in the present study, will be needed to examine the age-specific differences in norms and growth-rates. Further analysis of data is carried out on the observations pooled over all the zones to obtain an insight into the average pattern of growth of boys and to illustrate the application of appropriate statistical methods. We believe that the broad conclusions drawn on the basis of pooled data are not very much vitiated, since the age-specific differences between the zones, if any, are likely to be small.

TABLE 3.1. HEIGHTS OF BOYS : AGE SPECIFIC NORMS AND DIFFERENCES BETWEEN ZONES

age	south		east		west		north		all zones	
	N	mean	N	mean	N	mean	N	mean	N	mean ± s.e.
5	2	94.23	4	111.70	1	103.84	1	117.50	8	107.1238±3.8142
6	4	120.11	9	116.25	10	113.97	2	109.79	25	116.6208±1.0803
7	9	121.21	15	122.53	11	122.73	7	120.05	42	121.8874±1.0443
8	15	126.01	20	126.81	14	127.37	9	130.94	58	127.3809±0.9508
9	14	130.12	16	129.08	14	132.16	10	136.95	54	131.8750±0.8703
10	13	134.62	20	135.07	19	135.79	10	138.00	62	136.1048±0.9942
11	11	140.75	12	142.67	14	139.20	13	143.35	50	141.4750±1.0665
12	10	147.02	16	147.17	10	144.80	10	147.98	53	146.7147±0.9991
13	13	153.02	14	157.60	15	153.80	13	157.58	55	155.4491±1.0557
14	10	161.95	15	164.28	17	158.44	10	164.22	52	161.9135±1.0557
15	7	163.79	13	167.10	21	163.81	12	169.60	53	165.9243±1.0415
16	1	171.50	20	168.38	11	165.82	9	164.58	41	166.9354±1.1356

TABLE 3.2. WEIGHTS OF BOYS : AGE SPECIFIC NORMS AND DIFFERENCES BETWEEN ZONES

age	south		east		west		north		all zones	
	N	mean	N	mean	N	mean	N	mean	N	mean ± s.e.
5	2	15.00	4	18.22	1	16.40	1	17.60	8	17.3375±0.6364
6	4	22.12	9	21.59	10	19.15	2	20.80	25	20.6360±0.8027
7	9	22.50	16	23.00	11	24.43	7	21.26	42	23.3429±0.7040
8	15	23.52	20	24.28	14	25.01	9	27.23	58	24.8055±0.7020
9	14	24.54	16	25.14	14	28.61	10	29.53	54	26.7037±0.7380
10	13	26.64	20	28.51	19	29.26	10	31.29	62	28.8387±0.7340
11	11	31.88	12	33.42	14	31.64	13	35.71	50	33.1480±0.9245
12	10	35.94	16	36.03	16	33.18	16	37.39	58	35.8534±0.8784
13	13	40.32	14	44.77	15	40.56	13	45.63	55	42.7745±1.2123
14	10	46.85	15	49.16	17	46.80	10	50.87	52	48.2731±1.4753
15	7	47.03	13	48.65	21	52.01	12	56.94	53	51.6925±1.5656
16	1	49.40	20	49.21	11	56.11	9	58.40	41	53.0829±2.0592

TABLE 3.3. B.P.S. OF BOYS : AGE SPECIFIC NORMS AND DIFFERENCES BETWEEN ZONES

age	south		east		west		north		all zones	
	N	mean	N	mean	N	mean	N	mean	N	mean ± s.e.
5	2	75.00	4	77.50	1	65.00	1	75.00	8	75.0000±3.2588
6	4	86.75	9	85.56	10	81.50	2	87.50	25	84.6000±1.9747
7	9	86.67	14	86.14	11	89.45	7	91.43	41	88.0488±1.3411
8	15	86.00	19	89.47	14	89.64	9	90.56	57	88.7719±1.4577
9	14	93.21	16	89.00	14	93.93	10	90.50	54	91.6181±1.3227
10	13	94.62	20	89.75	19	96.21	10	97.00	62	93.9194±1.4475
11	11	101.55	12	93.67	14	96.57	13	93.15	50	96.0800±1.7207
12	10	105.30	16	96.62	16	95.75	16	100.94	58	99.0690±1.6927
13	13	104.62	14	106.57	15	102.60	13	105.38	55	104.8182±1.4985
14	10	113.50	15	110.60	17	105.12	10	108.50	53	108.0615±1.4829
15	7	108.57	13	110.85	21	111.19	12	113.58	53	111.3019±1.4178
16	1	130.00	18	108.00	11	115.91	9	112.11	30	111.7436±1.6285

TABLE 3.4. B.P.D. OF BOYS : AGE SPECIFIC NORMS AND DIFFERENCES BETWEEN ZONES

age	south		east		west		north		all zones	
	N	mean	N	mean	N	mean	N	mean	N	mean ± s.e.
5	2	32.50	4	36.25	1	35.00	1	35.00	8	35.0000±1.4286
6	4	47.50	9	48.33	10	41.00	2	42.50	25	44.8000±1.8745
7	9	46.07	14	50.21	11	53.45	7	50.71	41	50.3902±1.2247
8	15	47.00	19	51.58	14	50.36	9	52.22	57	50.1754±1.2175
9	14	47.86	16	45.44	14	52.50	10	50.00	54	48.7407±1.5807
10	13	53.08	20	50.50	19	54.26	10	54.50	62	52.8387±1.3375
11	11	56.00	12	53.33	14	56.14	13	58.31	50	56.0000±1.4379
12	10	56.00	16	57.19	16	54.69	16	56.25	58	56.0345±1.1217
13	13	60.85	14	60.43	15	58.47	13	60.31	55	59.9636±1.2183
14	10	64.10	15	61.67	17	61.29	10	58.50	52	61.4038±1.0011
15	7	67.14	13	63.00	21	60.95	12	65.50	53	63.3019±1.2357
16	1	65.00	18	63.50	11	65.91	9	65.11	39	64.5897±1.4941

TABLE 3.5. COMPARISON OF ZONES WITH RESPECT TO STANDARD POPULATION AVERAGES

age group	character	south	east	west	north
5—10	Height	126.62±0.93	127.82±0.70	127.62±0.89	129.83±1.22
	Weight	24.02±0.68	25.01±0.50	26.03±0.62	26.78±0.82
	B.P.S.	89.77±1.36	88.10±1.13	90.67±1.23	91.49±1.67
	B.P.D.	48.24±1.30	48.91±1.09	50.87±1.18	50.52±1.58
11—16	Height	155.71±1.35	157.39±0.82	153.85±0.70	157.59±0.80
	Weight	41.69±2.07	43.49±1.01	42.79±0.99	40.99±1.16
	B.P.S.	109.71±1.85	104.19±1.23	103.92±1.17	105.35±1.30
	B.P.D.	61.31±1.04	59.70±0.98	59.23±0.94	60.41±1.06
5—16	Height	142.73±0.87	144.10±0.57	142.15±0.59	145.20±0.73
	Weight	33.75±1.10	35.24±0.61	35.31±0.61	37.07±0.74
	B.P.S.	100.82±1.19	97.02±0.84	97.97±0.85	99.17±1.04
	B.P.D.	55.48±1.08	54.89±0.73	55.50±0.74	56.00±0.93

TABLE 3.6. MEAN VALUES OF THE NUMBER OF TEETH IN FOUR POSITIONS

age	n	positions			
		1	2	3	4
5	8	5.00	5.00	5.13	5.25
6	25	5.40	5.44	5.40	5.48
7	41	5.83	5.85	5.80	5.90
8	54	5.98	5.94	5.98	5.94
9	60	5.92	5.92	5.98	5.94
10	57	6.02	6.02	5.91	5.96
11	45	6.11	6.07	6.18	6.15
12	56	6.52	6.48	6.45	6.45
13	50	6.72	6.70	6.78	6.78
14	51	6.92	6.88	6.88	6.88
15	52	6.98	6.94	6.92	6.96
16	41	7.02	7.02	7.02	6.97

Standard deviations and correlation. Age-specific standard deviations are calculated for height, weight, $\log_e$ weight*, B.P.S. and B.P.D using all the available measurements at each age. They are given in Table 3.7. It is seen that while the age specific standard deviations for height are nearly the same for all ages, those for weight tend to increase with age. On the other hand the age-specific standard deviations of $\log_e$ weight are nearly the same for all ages. For this reason we choose $\log_e$ weight instead of weight for the regression analysis of blood pressure on the physical characteristics and also for the efficient estimation of norms.

Table 3.7 also contains the correlation coefficients of heights at ages i and $i+1$ for $i = 5, \ldots, 15$ and at ages i and $i+2$ for $i = 5, \ldots, 14$. Similar correlation coefficients are computed also for $\log_e$ weight. Each coefficient is computed from pairs of measurements made on boys at two different ages. The number of observations on which each correlation coefficient is based is also given in Table 3.7. It may be noted that the correlation coefficient of measurements at ages i and $i+1$ is of the same order of magnitude for all i and so also the correlation coefficients of measurements at ages

*$\log_e$ weight represents the logarithm of weight to the Naperian base e. For purposes of computations on an electronic machine it is convenient to use the base e. Many logarithmic tables provide logarithms of numbers to the base 10. Given $\log_{10} w$ we can find $\log_e w$ by multiplying $\log_{10} w$ by 2.3026.

i and $i+2$. The pooled values of these coefficients for height and $\log_e$ weight are also given in Table 3.7.

TABLE 3.7. AGE-SPECIFIC STANDARD DEVIATIONS AND CORRELATIONS BETWEEN AGES ONE YEAR APART (r_1) AND TWO YEARS APART (r_2)

		standard deviation					height				$\log_e$ weight**	
age	n	height	weight	$\log_e$ weight	B.P.D.*	B.P.S.*	n	r_1	n	r_2	r_1	r_2
5	8	10.786	1.800	.008	5.076	3.780	2	1.0000	2	1.0000	1.0000	1.0000
6	25	8.402	4.014	.181	9.674	9.183	20	0.8985	16	0.0204	0.0525	0.0340
7	41	6.708	4.568	.184	8.482	7.740	20	0.0794	0	0.0910	0.0790	0.0837
8	58	7.241	5.346	.103	10.000	9.111	32	0.9625	18	0.9773	0.9699	0.9515
9	54	6.395	5.423	.185	9.629	11.507	35	0.9795	13	0.9908	0.9729	0.9797
10	62	7.829	5.770	.194	11.305	10.446	27	0.9705	14	0.9614	0.9883	0.9737
11	60	7.541	6.537	.200	12.108	10.065	31	0.9527	16	0.9102	0.9665	0.7164
12	58	7.000	6.089	.193	12.780	8.409	30	0.9101	12	0.9460	0.9390	0.9046
13	55	7.829	8.990	.213	11.012	8.053	30	0.9260	11	0.8734	0.9662	0.9630
14	52	7.612	10.636	.220	10.590	7.792	25	0.9410	7	0.9638	0.9719	0.9747
15	53	7.531	11.398	.211	10.224	8.911	16	0.9694		—	0.9787	—
16	41	7.271	13.185	.217	10.039	9.210	—	—		—	—	—
pooled		7.503	8.004	.199	—	—		0.9440		0.9185	0.9650	0.9320

*The values of n for the ages 8 and 16 are 57 and 30 respectively.

**The values of n on which r_1 and r_2 are based are same as those for height.

Blood pressure. The age-specific mean values of B.P.S. and B.P.D. are already reported in Tables 3.3 and 3.4 respectively. It is seen that both B.P.S. and B.P.D. increase with age as in the case of height and weight. Since weight increases with age, we also expect weight and blood pressure to be highly associated. It is, therefore, pertinent to raise the following questions.

Is the relationship between blood pressure and weight solely due to their individual relationships (i.e., both increasing) with age? Or is there an intrinsic relationship between the two independently of age ?

Is blood pressure related to height in addition to weight ?

In order to answer the first question we obtain the regression of B.P.S and B.P.D on height and weight *within age groups* and test their significance. In the following analysis we have made use of all the observations on blood pressure, height and weight available on boys of a given age. Since some boys are common for measurement at different ages the observations at different ages are not independent and on this account the statistical regression analysis used by us may be in error to a small extent. But we hope the broad conclusions are correct.

The regression analysis within age groups is given in Table 3.8. The variance ratio for regression on weight is high and that for the partial regression on height eliminating weight is low. This suggests that for *any given age*, blood pressure depends on weight of the boy and there is no further dependence on height. Consequently norms of blood pressure by ages without reference to weight are not meaningful. For any given age the regression of B.P.S on $\log_e$ weight is 24.8655 and that of B.P.D on $\log_e$ weight is 13.7224.

TABLE 3.8. REGRESSION ANALYSIS WITHIN AGE GROUPS

due to	d.f.	B.P.S.		B.P.D.	
		s.s.	m.s.	s.s.	m.s.
regression on weight	1	13138	13138	4001	4001
partial regression on height	1	125	125	30	30
total regression on height and weight	2	13263	6631.5	4031	2015.5
residual	539	49466	91.77	40224	74.63
total (within age groups)	541	62729		44255	

Having established the dependence of blood pressure on weight at any given age, we raise the following question. Is the observed increase in blood pressure with age solely due to increase in weight with age or are other physiological factors which independently influence blood pressure with increasing age? For this purpose we carry out a regression analysis covering all ages. The results are reported in Table 3.9. Although weight is the most predominant factor influencing blood pressure, we observe that the variance ratio for partial regression on age is high indicating the effects of other physiological factors. Thus age in addition to weight would be useful in predicting blood pressure.

The regression equations for B.P.S. and B.P.D. on $\log_e$ weight ($\log_e w$) and age (a) are as follows :

$$\text{B.P.S.} = 4.0126 + 25.1788 \log_e w + 0.4875a \qquad \dots \quad (3.1)$$

$$\text{B.P.D.} = 0.1925 + 13.9744 \log_e w + 0.5664a \qquad \dots \quad (3.2)$$

Using these formulae the norms for B.P.S. and B.P.D. for a given combination of weight and age can be found. Thus for age 6 and weight of 20.64 (with $\log_e 20.64 = 2.3026 \times \log_{10} 20.64 = 2.3026 \times 1.3147 = 3.0272$) the norm for B.P.S. is

$$4.0126 + 25.1788 \times 3.0272 + 0.4875 \times 6 = 83.7589.$$

Similarly the norm for B.P.D at age 6 and weight 20.64 is

$$0.1925 + 13.9744 \times 3.0272 + 0.5664 \times 6 = 45.8942.$$

At the same age of 6, for 1 kilogram more in weight the B.P.S. would be more by 1.1960 and for 1 kilogram less, the B.P.S. would be less by 1.2514; and so on.

TABLE 3.0. REGRESSION ANALYSIS OVER ALL AGE GROUPS

due to	d.f.	B.P.S.		B.P.D.	
		s.s.	m.s.	s.s.	m.s.
partial regression on age	1	346	346	103	103
linearity of partial regression on age	10	467	46.70	996	99.6
effect of age	11	813	73.87	1099	100.0
residual	540	49592	91.84	40254	74.54
deviation from regression	551	50405	91.48	41353	75.05
regression on weight	1	58823	58823	23060	23060
total	552	109228		64413	

4. Efficient estimates of norms and growth rates

In Section 3, age-specific norms were obtained by averaging all the available observations at each age. These are only crude estimates obtained in the preliminary reduction of data to draw inferences on zonal differences and second order moments (standard deviations and correlations), which would be useful in the efficient estimation of norms. It is known that when each boy is measured at only one age, the best method of estimating the norm for any age is to take the average of all the observations at that age. Since some boys are measured at two or three different ages, the observations at different ages are not independent. It is therefore possible that the observations at any age provide information on the norms at other ages, through their correlations with the observations at the other ages. We, therefore, explore the possibility of jointly estimating all the norms for ages 5-16 by utilising all the available data. This is done by the method of least squares as explained below.

Let us observe that on each individual either one, two or three measurements are taken. If there is only one measurement x_i say at age i, then we consider the squared deviation

$$(x_i-\mu_i)^2 \qquad \dots \quad (4.1)$$

where μ_i is the norm to be estimated for age i.

If there are two measurements x_i and x_{i+1} at ages i and $i+1$, on the same boy then we consider the quadratic form

$$\lambda^{11}(x_i-\mu_i)^2+2\lambda^{12}(x_i-\mu_i)(x_{i+1}-\mu_{i+1})+\lambda^{22}(x_{i+1}-\mu_{i+1})^2 \qquad \dots \quad (4.2)$$

where the matrix (λ^{ij}) is the reciprocal of the matrix

$$\begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix} \qquad \dots \quad (4.3)$$

where ρ_1 is the correlation between measurements at ages i and $i+1$. It is assumed that ρ_1 is independent of i, which appears to be valid on the basis of the estimates reported in Table 3.7.

If there are two measurements at ages i and $i+2$ on the same boy, we consider the quadratic form

$$\xi^{11}(x_i-\mu_i)^2+2\xi^{12}(x_i-\mu_i)(x_{i+2}-\mu_{i+2})+\xi^{22}(x_{i+2}-\mu_{i+2})^2 \qquad \dots \quad (4.4)$$

where the matrix (ξ^{ij}) is the reciprocal of the matrix

$$\begin{pmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{pmatrix} \qquad \dots \quad (4.5)$$

where ρ_2 is the correlation between the measurements at ages i and $i+2$. Again ρ_2 is assumed to be independent of i.

If there are three measurements at ages i, $i+1$ and $i+2$ on the same boy, we consider the quadratic form

$$\eta^{11}(x_i-\mu_i)^2+2\eta^{12}(x_i-\mu_i)\,(x_{i+1}-\mu_{i+1})+2\eta^{13}(x_i-\mu_i)(x_{i+2}-\mu_{i+2})+\eta^{22}(x_{i+1}-\mu_{i+1})^2$$
$$+2\eta^{23}(x_{i+1}-\mu_{i+1})(x_{i+2}-\mu_{i+2})+\eta^{33}(x_{i+2}-\mu_{i+2})^2 \qquad \dots \quad (4.6)$$

where the matrix (η^{ij}) is the reciprocal of the matrix

$$\begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{pmatrix} \qquad \dots \quad (4.7)$$

where ρ_1, ρ_2 are as defined earlier.

If we substitute for ρ_1 and ρ_2 the pooled estimates as shown in Table 3.7 then corresponding to each boy in the sample we have a quadratic form in the unknowns $\mu_5, \dots, \mu_{16}$ of the type (4.1) or (4.2) or (4.4) or (4.6). We take the sum of quadratic forms over all the boys, which is a quadratic function in $\mu_5, \dots, \mu_{16}$. Minimising this function with respect to $\mu_5, \dots, \mu_{16}$ we obtain least square estimates. The normal equations leading to a minimum may be written

$$A\mu = d \qquad \dots \quad (4.8)$$

where μ is the vector of parameters $\mu_5, \dots, \mu_{16}$ and A is the matrix of normal equations. If $(a^{ij}) = A^{-1}$, then the least square estimates are

$$\mu = A^{-1}\, d. \qquad \dots \quad (4.9)$$

The variance-covariance matrix of the estimates is

$$\sigma^2 A^{-1} \qquad \dots \quad (4.10)$$

assuming that the standard deviation of measurements at any age i is σ, independent of i and ignoring errors in the estimation of ρ_1 and ρ_2. The pooled values of ρ_1, ρ_2 and σ^2 given in Table 3.7 have been used in the estimation of parameters and their standard errors.

The normal equations, the inverse matrix A^{-1} and the solutions were computed on the IBM computer 1401. The estimates so obtained for the norms of height and log weight and their standard errors are given in Table 4.1. The same table contains estimates of growth rate (increase in one year) and also the differential rate of growth (differential increase). The norms for weight are derived from those of log weight by taking antilogs.

A comparison of the standard errors of the norms given in Tables 3.1, 3.2 and 4.1 shows the advantages of Linked Cross-Sectional Study (LCS). The standard errors in Table 4.1 are nearly half of those in Tables 3.1 and 3.2, which shows that LCS study with only 100 observations at each age is equivalent to a cross-sectional or longitudinal study with 350 observations at each age. Thus, the LCS study results in considerable saving of observations, besides providing information on other aspects of growth, such as rate and differential rate of growth.

The growth curves for height and weight show a similar pattern. There is a constant increase per year (growth rate) in the first few years followed by sudden acceleration in growth rate. This acceleration is maintained for a few years and then a gradual decline in growth rate takes place. This sudden acceleration in growth rate of height occurs in the 11th year and that of weight follows a year later. There is also a slight acceleration in the growth rate of weight one year preceding that in height.

TABLE 4.1. IMPROVED ESTIMATES OF NORMS, GROWTH RATES AND DIFFERENTIAL GROWTH RATE

age	height			$\log_e$ weight			weight			age
	norm ± s.e.	increase in one year	differential increase	norm ± s.e.	increase in one year	differential increase	norm	increase in one year	differential increase	
5	108.25±1.475			2.840±0.033			17.11			5
		5.83			0.107			1.94		
6	114.08±0.806		0.20	2.947±0.020		0.031	19.05		0.88	6
		6.03			0.138			2.82		
7	120.11±0.706		0.01	3.085±0.018		−0.039	21.87		0.52	7
		6.04			0.100			2.30		
8	126.15±0.638		−0.38	3.185±0.016		−0.007	24.17		−0.05	8
		5.66			0.093			2.35		
9	131.80±0.613		−0.39	3.278±0.015		0.001	26.52		0.28	9
		5.26			0.094			2.63		
10	137.07±0.591		−0.66	3.372±0.015		0.017	29.14		0.78	10
		4.61			0.111			3.41		
11	141.68±0.592		1.72	3.483±0.015		−0.012	32.55		−0.04	11
		6.33			0.098			3.37		
12	148.01±0.587		0.19	3.581±0.016		0.025	35.93		1.37	12
		6.53			0.124			4.74		
13	154.53±0.595		−0.34	3.705±0.015		−0.013	40.66		0.06	13
		6.18			0.111			4.79		
14	160.72±0.615		−1.30	3.817±0.013		−0.008	45.46		0.18	14
		4.88			0.104			4.97		
15	165.60±0.654		−1.56	3.921±0.016		−0.045	50.43		−1.94	15
		3.32			0.050			3.04		
16	168.91±0.700			3.970±0.019			53.47			16

The standard errors of increase in height in one year are generally of the order 0.4—0.5 and those in $\log_e$ weight, of the order 0.04—0.05.

Charts. Charts 1 and 2 give a graphical representation of the records of height and weight on boys measured on three consecutive birth days. These charts provide a visual picture of how different sections of individual growth curves behave and also the variation between individuals.

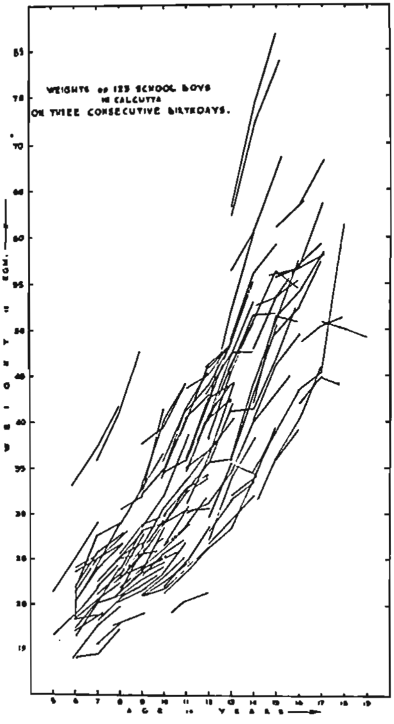

Chart 1.

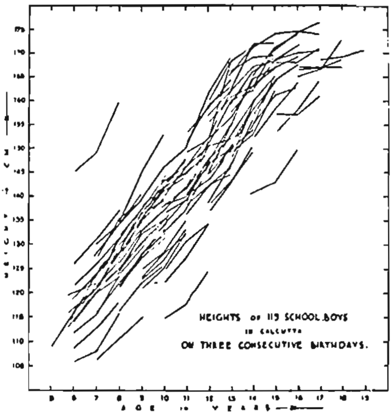

Chart 2.

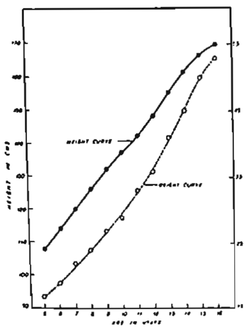

Chart 3.

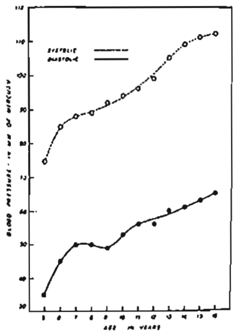

Chart 4.

Chart 3 gives the graphs of norms for height and weight and Chart 5 gives a comparison of growth rates for height and weight. Slight increase in the growth rate of weight at the age of 10 prior to a sudden high increase in the growth rate of height at the age of 11, which is further followed by a high increase in the growth rate of weight at the age of 12 are noticeable from Chart 5.

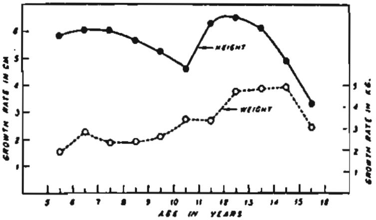

Chart 5.

Chart 4 provides the graph of the observed averages of systolic and diastolic blood pressures at different ages. As observed earlier, the norms depend on both weight and age and they are calculable from the formulae (3.1) and (3.2) given in Section 3.

5. Design of surveys for estimating norms and growth rates

In growth studies, two types of approaches are generally followed. One is a cross-sectional study, involving the measurements of individuals of different ages at one point (or period) of time. Another is a longitudinal study, where a certain number of individuals of a fixed age are chosen and they are measured periodically over a given number of years. It is well known that the longitudinal study, which is spread over a number of years, involves more expense and careful planning of collection of observations than the cross-sectional study. Yet, it is recommended on the grounds that it provides more information on different aspects of growth than what could be obtained from a cross-sectional study. In the present survey, we tried a mixed approach, which may be called a linked cross-sectional (LCS) study. Some individuals have been measured only at one age, some others at two consecutive ages differing by one year, and a third group at three consecutive ages. In a few cases of the third group, a second measurement at the end of one year could not be obtained, while the third measurement at the end of two years was available. The numbers of individuals and the ages at which they were measured have already been given in Table 2.2. It is seen that the period of survey is extended only over two years and the survey actually consists of short periods of longitudinal studies, overlapping with each other. We shall examine the theoretical consequences of such a study.

Norms. Suppose that we are interested in estimating the norms of some character at two different ages say of 6 and 7 years by obtaining a total number n of measurements at each age. Let us suppose that a total of πn individuals are measured only at age 6, πn are measured only at age 7 and $(1-\pi)n$ are measured at both the ages. In such a scheme the numbers of individuals measured at each age is n and $(1-\pi)$ represents the proportion of individuals common for both the measurements (at ages 6 and 7). The value of π is 1 in a cross-sectional study, and the value is zero for a longitudinal study. Thus the two different approaches refer to two extreme values of π. Is there an optimum value of π which gives the maximum precision to the estimates of norms ?

To investigate this problem, we compute the variance of the estimate of the norm at any age for a chosen value of π, which is found to be

$$\frac{\sigma^2}{n}\,\frac{1-\pi\rho^2}{1-\pi^2\rho^2} \qquad \dots \quad (5.1)$$

where ρ is the correlation between the measurements at the two ages and σ^2 is the variance of the measurements at any age. We then find the best value of π for which the estimated norms have the least variance or maximum precision. By minimising the expression (5.1) for variance, we find the optimum choice of π as

$$\pi = \frac{1}{1+\sqrt{1-\rho^2}} \qquad \dots \quad (5.2)$$

indicating that the best plan is to have a certain proportion of individuals common for measurement at both the ages. The optimum value of the proportion of individuals, on whom only one measurement is to be taken, is given in Table 5.1 for different values of ρ. We find that for the values of ρ upto 0.60, the number of individuals to be measured twice should be about 50 per cent of the individuals supplying measurements at any age; for higher values of ρ the optimum proportion of individuals supplying measurements at both ages becomes less, being of the order 40 per cent for ρ between 0.60 and 0.84, 30 per cent for ρ between 0.84 and 0.95 and 20 per cent for ρ between 0.95 and 0.98.

In our present study, the computed correlations between successive ages are greater than 0.95 (see Table 3.7) suggesting that the otpimum proportion of individuals to be measured twice is about 25 per cent for determining physical norms.

Growth rates. Let us suppose that our primary interest is not in determining norms but growth rates at different periods. The growth rate depends on the difference between the norms of two successive years, so that we need to collect the

TABLE 5.1. OPTIMUM VALUE OF π FOR GIVEN ρ FOR ESTIMATING NORMS, AND THE ASSOCIATED PRECISIONS OF NORM AND GROWTH RATE ESTIMATES

	optimum	precision of		maximum precision for growth rate
		norm	growth rate	
0	0.500	1.000	2.000	2.00
.1	0.501	0.997	1.895	1.80
.2	0.505	0.990	1.780	1.60
.3	0.512	0.977	1.654	1.40
.4	0.522	0.958	1.517	1.20
.5	0.536	0.933	1.366	1.00
.6	0.556	0.900	1.200	0.80
.7	0.583	0.857	1.014	0.60
.75	0.602	0.831	0.911	0.50
.80	0.625	0.800	0.800	0.40
.82	0.636	0.786	0.752	0.36
.84	0.648	0.771	0.702	0.32
.86	0.662	0.755	0.650	0.28
.88	0.678	0.737	0.595	0.24
.90	0.696	0.718	0.535	0.20
.91	0.707	0.707	0.536	0.18
.92	0.718	0.696	0.505	0.16
.93	0.731	0.684	0.472	0.14
.94	0.746	0.671	0.438	0.12
.95	0.762	0.656	0.401	0.10
.96	0.781	0.640	0.362	0.08
.97	0.804	0.621	0.320	0.06
.98	0.834	0.599	0.273	0.04
.99	0.875	0.571	0.210	0.02

observations in such a way that the difference between the norms of two successive years is estimable with maximum precision. For given π and ρ, the variance of the estimated difference is

$$\frac{2\sigma^2}{n}\,\frac{1-\rho}{1-\pi\rho} \qquad \ldots \quad (5.3)$$

The expression (5.3) attains the minimum value of $2\sigma^2(1-\rho)/n$ when $\pi = 0$, i.e., when all the individuals are measured at two successive ages.

The optimum values of π for the estimation of norms and growth rates are somewhat conflicting and the choice of π, therefore, depends on the purpose of the study, i.e., on the desired relative precisions for the estimates of norms and growth rates.

Differential growth rate. Suppose that we are interested in estimating the differential rate of growth, i.e. the difference in growth during successive years. Consider the two alternative schemes.

(a) Each of n individuals is measured at three successive ages (say at 6, 7 and 8).

(b) A batch of n individuals is measured at the two successive ages of 6 and 7 and another batch of n individuals is measured at the two successive ages of 7 and 8.

Although the actual number of measurements in the second scheme is somewhat large, the period of survey is restricted to only one year, instead of two years as in the first scheme.

Let ρ_1, ρ_2 be the correlations between measurements with one year and two-year intervals respectively. The precision of the estimate of difference in growth rates between '6 and 7' and '7 and 8', for the first scheme is

$$\frac{4\sigma^2}{n}(1-\rho_1)(2-2\rho_1+\rho_2) \quad \ldots \quad (5.4)$$

and that for the second scheme is

$$\frac{4\sigma^2}{n}(1-\rho_1). \quad \ldots \quad (5.5)$$

The first scheme is better than the second iff

$$2-2\rho_1+\rho_2 < 1 \quad \text{or} \quad 2\rho_1 > 1+\rho_2 \quad \ldots \quad (5.6)$$

which is not true for the values of ρ_1, ρ_2 observed in our study (see Table 3.7). Thus, between the two alternative schemes the second is better than the first when the correlations are as high as those observed in the present study.

So it appears that the optimum design for studying the norms, growth rates and differential growth should be of an LCS (linked cross-sectional) type. The optimum proportions of individuals to be measured once, twice and thrice, will depend on the relative precisions desired for the estimation of norms, growth rates and differential growth. In the present survey, we have not used optimum proportions in determining the numbers of individuals measured twice and thrice as the basic parameters such as the correlations ρ_1 and ρ_2 were unknown. The project was initiated to throw information on future surveys and the investigator was asked to contact as many individuals as he could for measuring twice and thrice. While our estimates of norms and growth rates would be more efficient than those resulting from a purely cross-sectional study involving the same number of measurements at each age, better results could be achieved by suitably determining the proportions of individuals to be measured twice, thrice, and so on.

Since the estimation of norms is of main interest and some data are needed to determine the correlations for planning of efficient surveys we make the following recommendations :

(i) Attempts should be made to measure some boys at three consecutive ages, i.e., over a period of two years. The number of boys who are repeatedly measured should be about 50 per cent of those measured only at one age.

(ii) It is preferable to have nearly equal numbers of measurements at each age for given amount of expenditure on field work.

Phenomenon of catching up. The fact noted above that the second scheme is better than the first for the estimation of differential growth leads us to an interesting phenomenon which may be called 'catching up'. Let x_i, x_{i+1} and x_{i+2} represent the measurements of an individual at ages i, $i+1$ and $i+2$ respectively. The covariance between the growths $x_{i+1}-x_i$ and $x_{i+2}-x_{i+1}$ during successive years is σ^2 times

$$(2\rho_1-1-\rho_2) \qquad \text{... (5.7)}$$

where σ^2 is the variance of measurements at any age, ρ_1 is the correlation between the measurements at ages differing by one year and ρ_2 is the correlation between the measurements at ages differing by two years.

The second scheme is better when the expression (5.7) is negative, i.e., when the covariance between growths in successive years is negative. Using the values of ρ_1, ρ_2 given in Table 3.7 it is seen that the value of (5.7) for height is -0.0293 and that for weight is -0.1014. Both the covariances are negative indicating a tendency for poor growth in one year being compensated by better growth next year, which may be described as the phenomenon of catching up.

6. Conclusion

The following may be stated in support of an LCS study in preference to a longitudinal study :

1. A longitudinal study being spread over several years poses a number of organisational problems and involves more expense. On the other hand, an LCS study spread over a maximum period of 2 years is a more feasible project, less expensive and permits a more carefully-planned collection of data.

2. An LCS study provides quick estimates of age-specific growth-rates which are nearly as efficient as those provided by a longitudinal study. Other aspects of growth such as differential growth-rate are estimable with greater efficiency by LCS study when the growth-process exhibits what we call the phenomenon of catching up (as found in the present data) that is, poor growth in one year being compensated by better growth in the next year.

3. It should be mentioned that the age-specific growth rates provided by LCS and longitudinal studies may be qualitatively different. The former reflects the effect of factors affecting growth at different ages of individuals of a population at a specified point (or short period) of time, while the latter provides information on factors which have affected growth of individuals at different ages but at different periods of time. Both these studies yield the same results if external factors affecting growth are relatively stable over a long period of time. But in a rapidly developing country, estimates provided by LCS study are more relevant in judging improvement in standards of living over short periods of time.

Acknowledgements

We wish to thank Dr. A. B. Sur who took the blood pressure measurements and Shri S. Samanta who took the observations on height, weight, teeth etc.

The programmes on the 1401 IBM computer for the various complex computations involved in our statistical analysis were supplied by Dr. J. Roy and Shri R. G. Naha, for which we are thankful to them.

We are also grateful to the Director, All India Institute of Hygiene and Public Health for permission to use the data collected under the auspices of the Institute and to the authorities of the St. Xavier's College for their kind cooperation and scientific interest in our project.

References

Rao, M. N. and Bhattacharjee, B. (1952): Growth curve of children in first month. *Ind. J. Pediatrics*, 19, 1-11.

——— (1953): Growth curve of children in first year. *Ind. J. Pediatrics*, 29, 249-260.

Rao, Someswar (1961): *Special Report*, Series No. 36, Indian Council of Medical Research.

Wheatley, G. M. and Hallock, G. T. (1951): *Health Observations of School Children*, McGraw Hill and Company.

Paper received : March, 1966.