92.

In Human Growth + Devpt. Ed. Borms et al, 1982, Plenum, NY

CURRENT DEVELOPMENTS IN THE DESIGN AND ANALYSIS OF GROWTH STUDIES

H. Goldstein

Department of Mathematics, Statistics and Computing
Institute of Education, London, U.K.

INTRODUCTION

In a book published three years ago (Goldstein, 1979), an attempt was made to summarise the state of technical knowledge on the efficient design and analysis of longitudinal studies. In this paper we intend to review work which has been done since that book was written, indicating current areas of interest with an emphasis on those aspects of the subject which look most fruitful, and in particular where current practice appears to be weak.

SAMPLING

In a recent review, Mednick & Baert (1981) list nearly 70 European studies. While this is testimony to the popularity of such studies (despite the curious omission of most of the International Children's Centre Studies), a careful inspection of the list shows that not many studies have truly representative samples of well defined populations. In fact only 26 really qualify, and of these only 3, the 1946, 1958 and 1970 British Birth Cohorts are representative of a national child population. In North America the situation is perhaps somewhat better, but we have to go to Cuba to find the best current examples (Jordan et al., 1975).

Of course, it is usually more expensive to select and maintain a representative longitudinal sample than, for example, a local area or hospital based group. Nevertheless there is still a large apparent lack of interest, and even concern, over the use of non-representative samples. This partly may be due to a feeling that any resulting biases will not be serious when studying change over time. Such a view is erroneous however, and it is easy to demonstrate that serious biases can arise (Bailer, 1975). In fact there is now sufficient experience with representative longitudinal studies for us to expect their adoption as a matter of routine. Certainly in countries with near universal school enrolments, the selection and follow-up of representative samples does not present insurmountable problems. For pre-school children modern techniques for household surveys are quite feasible, and not necessarily very expensive. The Cuban Growth Study (Jordan et al., 1975) is a good example of the effective use of a multistage area household sample in order to obtain good population data. Rona and Altman (1977) describe the use of a complex sampling strategy to select schools within which children are further sampled. Likewise, births can be selected and followed, with sampling being based on maternity institutions and households. It is also worth pointing out, in view of the popularity of birth 'cohort' studies, that while a group born at the same time makes some of the administration easier, it will often be better to sample over a longer time period, say a year, so that seasonal effects can be studied.

Where a complex sampling scheme is used involving stratification and clustering, it is important to take the sample design into account when carrying out analyses. Thus if there is a lot of clustering the standard errors of average measurements will usually be larger than those from a simple random sample of equal size. These 'design effects' (Kish, 1965) will also affect model-based analyses such as multiple regression or log linear models for contingency tables and failure to take them into account can lead to misleading inferences. There has been recent theoretical work into this problem (Hold et al., 1980 a, b) and there exist computer programs to carry out correct analyses (see for example Hidiroglou, 1981).

NON RESPONSE

Biases from non response in surveys can be large and hence a serious threat to the validity of findings. The typical remedy is to make strenuos efforts to contact and obtain information from as many non-respondents as possible. By 'non-respondent' we exclude those who have refused to take part and include those who are still eligible for inclusion in the sample but have been missed, for example because they were away, ill, etc. Such individuals may have different characteristics from the remainder and it is important to ascertain how different they may be - for example they may be less healthy.

There is now a good deal of evidence that given sufficent resources and effort, non-response and failure to trace subjects in a longitudinal study can be kept to a minimum. It has been suggested for example, that money paid to subjects can be very effective and a study by Kerachsky and Mallor (1980) found that $ 5 per interview improved the response rate in a longitudinal study of disadvantaged youth and also helped in subsequent tracing. These authors even suggest that such expenditure at the outset may save money since it would lead to less expenditure on tracing, etc.

Where it is necessary to make further efforts to sample non-respondents, then in a longitudinal study those who are included in this way have to be followed-up in order to see whether estimates of change are the same for this group as for those who responded initially. The proportion of non-responders who should be chased-up will depend on the cost of the operation and Lessler et al. (1978) discuss various strategies. A good general introduction to the practicalities of surveys is the book by Hoinville and Jowell (1978).

PLANNING AND DATA PROCESSING

Not very much work of a systematic kind has been done on the problems of planning longitudinal studies, although a paper by Pelletier and Nolte (1978) describes a computerised data management system which keeps track of subjects, updates files, etc. Clearly, in a longitudinal study, such matters as good documentation, filing systems, etc., are rather more important than in a short term cross-sectional study and computerised systems have a lot to offer. Even with a good computer system, however, there are many detailed decisions which are made for which a practical account would be helpful. One interesting attempt to do this has been the creation of a video tape on which the three directors of the major British birth cohort studies, James Douglas, Mia Kellmer Pringle and Neville Butler, are interviewed by Jim Tanner and Michael Healy about the problems they faced and their general views on longitudinal studies (Three Generations of Children, 1982).

For the efficient and rapid processing of results good computing facilities are essential. There is now a choice of so called 'data base management systems' which will handle complex files of longitudinal data efficiently and easily and are capable of carrying out standard statistical analyses and interfacing with standard packages for further processing. S.I.R. (Robinson et al., 1977) is one of these. Some standard survey packages can also be used and Fendt et al. (1979) show how longitudinal data files can be manipulated with the S.A.S. package. Initial expenditure on a system with powerful facilities for manipulating longitudinal data is well worth while for a medium or large scale study.

It is also important to have good data editing facilities which can detect 'suspect' values and present them graphically. Such facilities will be extremely useful in interactive systems, which are now becoming widespread, where editing is on-line. One example of a graphical procedure for use with growth data is given by Goldstein (1981).

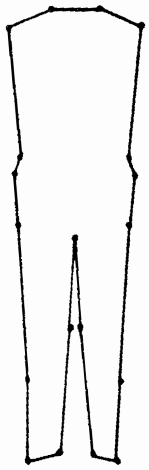

Fig. 1. Shape outline of a three year-old.

SHAPE GROWTH

For a long time auxologists have been concerned with the problem of measuring the shape as well as the size of individuals. For the most part however growth in size has been studied and there is extremely little work on growth in shape. This in some part seems to be due to the difficulty of actually defining shape and the complex statistical procedures needed. An outline of various approaches is given in Goldstein (1979). Since then one significant new development has occurred with the work of Bookstein (1978). Before describing this it will be useful to mention briefly the traditional methods. I shall deal with 2-dimensional shapes, obtained for example from photographs, and the move to three dimensions involves no new principles.

All methods begin with the identification of 'landmarks' on the shape image. Fig. 1. for example shows the shape outline of a 3 1/2 year old girl (Goldstein & Johnston, 1978) formed by joining landmarks such as the 'corner' of the shoulder.

In allometry, the ratios of distances between the points are used to define shape. The general procedure with more than two such distances is to deal with ratios of measurements by using additive functions of the logarithms of the measurements. Overall size is most conveniently defined as the average of the transformed measurements and the deviations from the average are used to study shape - typically by the methods of factor analysis or principal component analysis. A recent application of this method to adults is given by Healy and Tanner (1981).

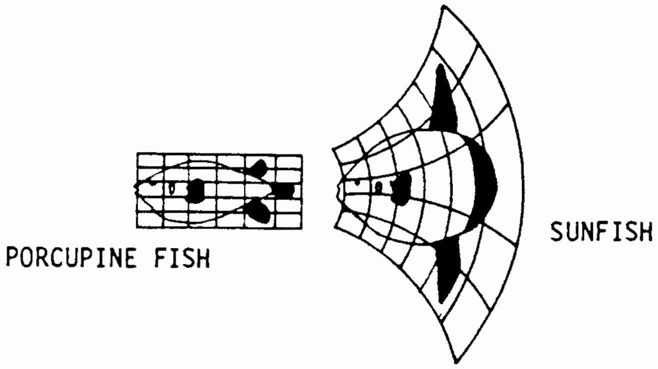

Fig. 2. Transformation grid.

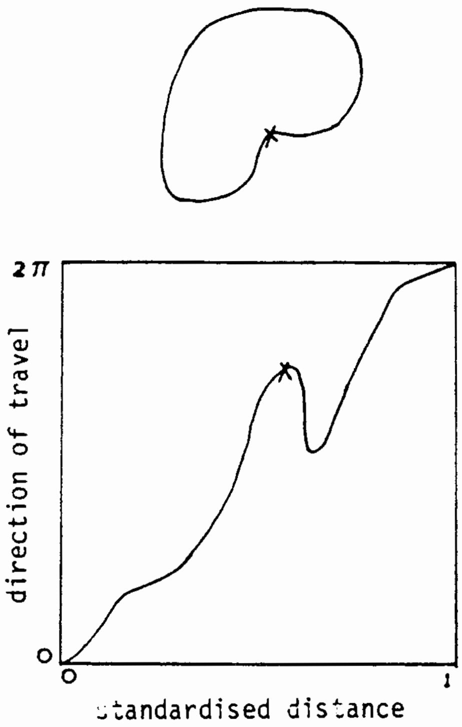

Fig. 3. Outline of human head and graphical representation.

The other main approach has been that indicated by D'Arcy Thompson with his method of transformation grids.

Fig. 2. adapted from Thompson (1961) shows a porcupine fish transformed into a sunfish by deforming the rectangular grid of the former so that landmarks are mapped onto corresponding landmarks. If such a transformation can be described by a relatively simple mathematical equation we clearly have a powerful method of describing shape change.

Bookstein (1978) develops his methods with particular reference to cephalometrics rather than body shape, so that some of his remarks have less force in the latter case. For example, he rightly points out that many so-called landmarks in cephalometrics, such as the 'Frankfurt Plane' are in fact defined in terms of relative lengths and hence dependent on shape itself and so not admissible in order to define it. When studying body shape however, this will generally be less of a problem.

Bookstein offers two major contributions. In the first he gives a method of describing the shape outline between landmarks, based on measuring the direction of travel and the distance travelled of a point as it moves along the outline curve. For closed curves he shows that this contains all the information necessary to describe shape. In particular it is a simple matter to adjust for size differences and the orientation of the figure is irrelevant.

Fig. 3. shows a side view of a human head together with a plot of the direction of travel against distance travelled (scaled to unity). Clearly, such plots can be averaged and compared in several different ways, for example by measuring discrepancies between individuals or groups. Bookstein shows how the smooth curves between landmarks can be drawn automatically using a digitizer to record co-ordinates and conic splines to join them up. This method seems to have a great deal of potential, but as yet little empirical exploitation.

Bookstein's other contribution is to develop Thompson's work on transformation grids. He suggests that for two shapes, all pairs of grid lines drawn through corresponding points should intersect at right angles, although curving in different ways and having different lengths between points.

Fig. 4. illustrates for a simple case such a 'biorthogonal' pair of grids. The relative stretching of grid lines is given by the numbers. One line tends to be relatively squashed and the other extended at the ends, but meeting at 90° in both cases. Bookstein goes on to apply this procedure to some of Thomson's examples, typically obtaining different pictures of how the transformations work. He does not apply his method to developmental sequences in children, which would constitute a more stringent validation of the method than comparing evolutionary specimens.

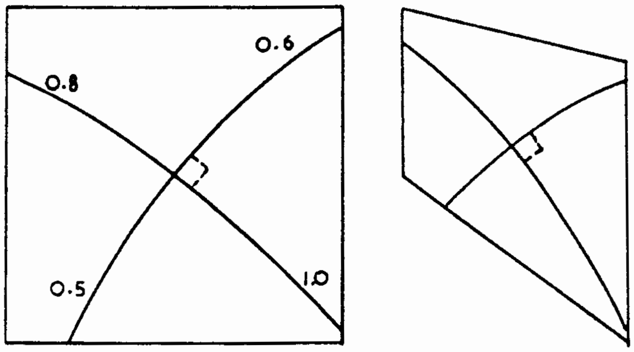

Fig. 4. Biorthogonal grid pair. Numbers show stretching from left to right figure.

Finally, it is worth noting that Bookstein's two methods can be put together to form a composite description of shape. Each basic feature, for example the eyes on a human body image, can be described as a shape outline. Each outline can then be used to define characteristic points and these then used as the basis for constructing grid lines. In this way a hierarchy of shapes can be produced. At each stage other than the first an outline would contain several smaller shapes with their own description and linked via a transformation grid.

MODELS FOR ANALYSIS

Categorical Data

Developments during the last 15 years have produced a number of models for handling categorical data in ways which are analgous to those for continuous data. These are described in Goldstein (1979), and briefly fall into two types. The one, path-analysis type of model, concerns relationships between occasions and the other resembles growth curve type models where e.g. a proportion is related to age. In the latter case, a methodology described by Koch et al. (1977) gives a particularly flexible model for analyzing all kinds of categorical data. In particular it can be applied to data derived from complex samples which the more common maximum likelihood log-linear models do not deal with properly. A computer program (Landis et al., 1976) is available for these models. Johnson et al. (1981) shows how this methodology can be applied to longitudinal data from finite populations. It is perhaps not readily appreciated that models where the dependent variable is categorical can involve continuous variables as predictors. This may be particularly relevant to growth data where categorical variables may be specific events, such as the occurrence of menarche which we may wish to relate to continuous variables such as height and weight. To illustrate, consider a model which could be used to study the controversial so called critical weight' hypothesis (Frish and Revelle, 1971; Billewicz et al., 1976) which suggested that menarche tended to be triggered when girls reached a certain weight. The following model will be kept relatively simple, although it could be elaborated by including further occasions or variables.

Suppose we have a sequence of yearly measurements at birthdays of weight on a sample of girls before and during asolescence and that we also record the presence or absence of menarche. The model relates the probability of menarche at any given measurement occasion, for those for whom menarche has not yet occurred at the previous occasion, to the weight at the previous occasion. Our basic hypothesis is that there is a strong positive relationship to this weight which is independent of the age at measurement.

The model is:

$$\log \frac{P_{it}}{1-P_{it}} = \alpha_t + \beta_t x_{t-1} \quad t = 2, \ldots . T \qquad (1)$$

where P_{it} is the probability of menarche having occurred by occasion t for subject i, x_{t-1} is the weight at occasion t-1 and α_t, β_t are constants. Our hypothesis is $\beta_t = \beta \neq 0$. The coefficients are estimated separately at each occasion and many standard program packages such as GLIM (Baker & Nelder, 1978) exist for this. The β_t estimates may then be tested for equality using their estimated standard errors, and their mean tested against zero. Alternatively the sum of the separate goodness of fit statistics for each equation (1) can be compared with that obtained from fitting a model for all occasions simultaneously with a constant value β. Model (1) can clearly be extended by including weight for more distant occasions, other variables such as height, bone age, etc. A detailed model of this type applied to a clinical trial is described by Wu and Ware (1979).

There are other topics in the analysis of categorical data for which there is insufficient space here. For example an important set of models is concerned with the study of sequences of categoric or state changes from one occasion to the next - for example in behaviour, attitudes or medical states. Plewis (1981) applies Markov models to the analysis of teacher ratings of children's behaviours, Korn and Whittemore (1979) give a medical example which incorporates covariates, and Altham (1981) discusses the analysis of long sequences of interactive behaviours between individuals.

Continuous Data: Individual Growth Curves

In fitting growth curves, usually of height, to sets of individual measurements there are two broad issues. The first is concerned with providing a useful summary of growth and the second related issue is in making comparisons between individuals and groups of individuals. I shall deal with the fitting of polynomial growth curves in the next section, here I will discuss the fitting of non-linear curves - a tradition which goes back to the beginning of this Century.

There appears still to be a misconception about the status of such curves as the logistic Jenns-Bayley, etc. These are sometimes referred to as having direct 'biological' interpretability as opposed to say polynomials (See e.g. Berkey, 1982). In fact, the biological justification is typically no more than a statement of how growth rate is related to growth achieved, with no more than a crude physical justification for these relationships, which are anyway chosen to represent the known shape of growth curves. Curves based on specific physiological models have been proposed (e.g. Weiss and Kavanu, 1957) but have had little success. Moreover we should expect useful 'biological' information only when we have other biological measurements available as well, e.g. hormone levels. Thus the usefulness of the curves ought to be judged on whether they provide adequate summaries of particular growth events, such as adult size, peak velocity, etc., as well as being good 'fits' in the statistical sense of minimising residual variation.

A particular problem of non-linear growth curves is that they tend to impose inflexible constraints on growth events. For example the logistic curve often fitted to adolescent growth in height constrains the ratio of height at peak height velocity (PHV) to be just half the adult height measurement from the height at the start of adolescent growth, whereas empirical results suggest a range of 0.35 - 0.45 (Goldstein, 1979). Thus Thissen et al. (1976) fit a double logistic model to height growth from 1 to 18 years. By careful comparison with the observed heights their results indicate that the estimates of adult height and final prepubertal height are good ones. This is to be expected and is one way in which these models are superior to polynomial ones. They do not however carry out such a comparison for their PHV estimates which one might expect to yield poorer results. These authors also compare different groups and in particular show how the above height estimates do provide useful summaries. More recently Bock and Thissen (1980) and in a paper to this conference Bock (1982) have elaborated this model by introducing a procedure for studying the correlations between the within-subject residual values predicted by the curves and by formally using information about the variation of the parameters between individuals to improve the efficiency of the individual estimates. Sandland and McGilchrist (1979) and Glaseby (1979) also propose models which incorporate a correlation structure for the residuals. While such procedures are statistically interesting, especially in that they can give more precise estimates for indivuduals with relatively few observations, they raise no new substantive issues.

The alternative approach of fitting general flexible curves which describe growth locally is exemplified by Stutzle et al. (1980). They assume no particular form of curve, but effectively allow the data to determine the local shape. There are two basic steps in this process. The first involves 'standardizing' each individual's growth curve and here since they assume two components of growth, each component curve is standardized. Thus for each component we work with the standardized ages.

$$t_{1i} = \frac{x-b_{1i}}{c_{1i}} \quad , \quad t_{2i} = \frac{x-b_{2i}}{c_{2i}}$$

where the b_{1i}, b_{2i}, c_{1i}, c_{2i} are to be estimated and the b's represent here the ages of maximum velocity and the c's its duration. This notion of an individualized developmental scale is not new, but seems not to have been incorporated previously in a formal model. Rather than work with growth achieved, observed velocities are modelled. The following 'switch off' model is found to perform well, for the jth velocity,

$$v_i(x_j) = a_{1i}S_1(t_{1i})\ T\ (t_{2i}) + a_{2i}S_2(t_{2i}) + e_{ij} \qquad (2)$$

where a_{1i}, a_{2i} are constants to be estimated, S_1, S_2 are the growth curves, T is a function which decreases to zero, thus inhibiting the expression of S_1 as S_2 increases. The residuals e_{ij} have a covariance structure which is known if certain simple assumptions about the independence and homoscedasticity of residuals for the corresponding distances are assumed.

The second step is to determine the functions S_1 and S_2 empirically by fitting to the pooled data spline curves which are local polynomials with smooth join points. The data are pooled using the standardized ages. Thus initial estimates of S_1 and S_2 are obtained (using logistic functions) from which each subjects parameters are estimated, and hence standardized ages. The pooled residuals from the curves, using these standardized ages then determine the spline adjustments to the initial curves to produce new estimates of S_1 and S_2 and so on until the process converges.

While there are still questions to be settled with the choice of model forms and particularly in the form of age standardization to be used, especially pre-puberty, this procedure, because of its ability smoothly to fit growth data locally using information from the entire sample would seem to have considerable potential for identifying subtle features of growth. Stutzle et al. (1980) for example seem to detect clearly a prepubertal midspurt.

Continuous Data: Polynomial Growth Curves

The other major approach to the analysis of growth curves has been to fit polynomial functions to a sample of individuals all measured at the same occasions. This can be viewed (Goldstein, 1979) as first fitting polynomials to each individual subject and then studying the resulting set of polynomial coefficients with a view to determining the order of polynomial necessary. We can compare groups, for example boys and girls, while the coefficients have straightforward interpretations in terms of growth rates, accelerations, etc. These curves do not tend to perform well when growth levels off, for example at adulthood. Nor do they incorporate an age standardization as described above, although polynomials could perhaps be used as the starting curves S_1, S_2. On the other hand, they do allow straigthforward application of well known multivariate techniques which can readily incorporate covariates and group differences and make no particular assumptions about the covariances of residuals.

The basic model is very general in the sense that it allows for any pattern of correlations between the measurements across occasions, or alternatively between polynomial coefficients. If some simplifying assumptions could be introduced to make these correlations dependent on a smaller number of parameters there would be a gain in precision of estimation and significance testing. This will be particularly useful in growth studies with small numbers of individuals. Rao (1965) suggests a reasonable model in which the constant, linear, etc. parameters of the polynomial growth curves are assumed to be random variables distributed independently of the residual error. This model is generalized by Reinsel (1982) to the case where there is more than one measurement and subjects are classified by independent variables such as sex, social class, etc. These models allow us to make sensible inferences about the behaviour of individuals' growth curves, recognizing for example that these may be of higher degree than the average curves. Unfortunately, large samples are required to see how far, in terms of low order polynomials, such a simplifying model is justified and there seem to be few substantial empirical investigations which attempt to do this (but see Ware & Wu, 1981). This seems to be another example where the theory, elegant as it may be, has somewhat outrun the practice.

Continuous Data: Nonparametric Models

Where it seems unreasonable to assume normal distributions for growth measurements or their transforms, for example when dealing with a rating scale, non-parametric techniques can be used to compare the growth of groups of individuals. For example, polynomial curves can be calculated for each individual in the usual way and then each of the overall, linear, etc. coefficients can be ranked and mean rankings compared between groups for each coefficient separately, or perhaps jointly (Goldstein, 1979). In many cases it is not reasonable to fit polynomial curves - growth in skinfold is a case in point where there is generally no steady progression with age. If some other function can be found to represent such growth then a ranking analysis could be carried out on the corresponding individual coefficients. In general, however, we may simply wish to make an overall comparison and this may be done by ranking all the measurements separately at each occasion and calculating the mean ranks for each group. A joint test for equality of these means over all occasions can then be carried out (Koziol et al., 1981).

Finally, while on this point it seems worth remarking that nearly all the auxological applications of growth curve models have used height. We therefore know relatively little about how to deal with less regular measurements such as weight or skinfold, or even other linear measurements such as biocromial width. It would seem time we began developing models useful with these other measurements.

Continuous Data: Between-Occasion Models

As mentioned under categoric data, the other general class of longitudinal models is that for relating measurements between occasions. While not always to be recommended, these models do possess distinct advantages over time-related models in terms of scale invariance and an orientation towards causal interpretation (See Goldstein, 1979). In general, however, in growth studies there is relatively little work with these although they are rather popular in the social sciences. Two exceptions are models for predicting adult height from previous measurements (Tanner et al., 1975) and conditional growth standards (Cameron, 1980). I shall discuss growth standards below, and there is work in progress on improving height prediction methods (Tanner, personal communication). It is worth mentioning that the use of fitted growth curves to predict adult height is based on a time related model. It has been suggested by some authors (e.g. Ware and Wu, 1981) that growth curves fitted to a set of measurements can be used to predict later ones such as adult height, by utilising the summary of growth provided by these curves rather than the full set of original measurements. This is certainly an interesting suggestion, but there has been little attempt to compare such predictions with those from more straightforward multiple regression analysis.

Continuous Data: Growth Stability

It is often desired to estimate the regularity or stability of a growth curve for an individual or group of individuals. Goldstein (1981) applies two different measures to a sample of individuals. One method measures the relative constancy of growth by estimating an individual's variation about his or her average growth and the second measures growth separation by counting the number of times an individual's growth curve crosses other individual's growth curves. Provisional norms are provided for each index. These indices should prove useful in identifying individuals or groups with abnormal growth patterns.

POPULATION NORMS AND STANDARDS

The construction of growth norms or standards and their periodic updating is a common activity. For a sample of measurements over a wide age range there are problems of allocating measurements to age groups and producing smoothed percentile estimates. The former question can be answered for cross sectional norms by allocating the number of measurements proportional to the growth rate. In order to produce smoothed percentile lines the following procedure seems to work well (Goldstein, 1979). First, a smoothed estimate of the 50th percentile is obtained. This may be done in a number of ways. For example a regression line, possibly non linear is fitted within narrow age groups and the estimates at the centre of the group noted. The age groups are then all shifted one quarter of an age interval and the process repeated and so on. These overlapping estimates are finally joined together smoothly by eye or using splines. The next stage is to estimate the other percentiles. This can of course be carried out in similar fashion, using the residuals from the fitted 50th percentile curve within each age group to estimate the percentiles. This procedure should work well if the data are extensive, but it has a weakness; the more extreme percentiles, which are actually of more importance, are less precisely estimated. If we are prepared to assume a normal distribution of measurements then this can be used to improve precision but this will usually not be wise, especially for the extreme percentiles, so an alternative is needed. One possibility is to plot the residuals on probability paper and estimate percentiles from the plot, but this still effectively only uses measurements near to the estimated percentile. Instead we could set up a general relationship between the percentiles we want to estimate and the 50th.

Write, for every age group,

$$P_j(t) - P_{50}(t) = a_{j0} + a_{j1}t \qquad (3)$$

This states that the difference between the jth and 50th percentile is a simple linear function of age within the age group, and the constants a_{j0}, a_{j1} are to be estimated for every percentile. It is now possible to estimate the parameters a_{j0}, a_{j1} for all j so that the proportion of observations beyond each percentile correspond as closely as possible to the theoretical proportions. Note that the coefficients in equation (3) are the same for every age group, thus ensuring that since $P_{50}(t)$ is smooth over the whole age range the the other percentiles are also. Furthermore, all the percentiles are smoothly related to each other via (3) which makes efficient use of all the data. Furthermore, (3) can be extended by adding a quadratic term if required. It is straightforward to obtain estimates of the a_{ji} using weighted least quares (See Angers, 1979). It is also possible to carry out the estimation using weights for the residuals which reflect measurement accuracy, but for most growth data such a refinement should be unnecessary.

This method can also be adapted to the situation where old norms are being updated with new data. If we assume, for each percentile, that the relationships between the old values and the new ones are smooth then we can obtain the deviation of each new measurement from the old 50th percentile and use the residuals to estimate the new 50th percentile and other percentiles as before. Since the old 50th percentile is used as an 'anchor' we will also require fewer measurements to obtain the same accuracy as before. In addition we may be able to make further simplifying assumptions about the relationships between the old and new percentiles. For example, if we assume that the spacing between percentiles remains the same with only an overall change then in the estimation procedure the constants in (3) for all the percentiles except the 50th are set to known values. Alternatively, we may suppose that an overall increase in spread has occurred so that we have for the new values,

$$a^1_{ji} = ka_{ji}$$

and we will need to estimate, iteratively, the parameter k. The reasonableness of such assumptions would have to be tested on a real data set.

We often require separate norms for population subgroups, for example boys and girls. Normally, these are estimated separately but we can adapt the above method to give efficient estimates in this case. If we assume that the percentiles differ smoothly, then we first estimate the joint smooth 50th percentile and then for each group we can estimate the 50th and other percentiles as above. In this case, simplifying assumptions will involve relationships between the a_{ji} for different groups. For example if we have two sex groups with coefficients a_{1ji}, a_{2ji} then we might assume, as before, having estimated the 50th percentile, that

$$a_{1ji} = ka_{2ji}$$

Finally, it is worth pointing out that these procedurs can be automated, thus saving much of the tedium of hand plotting, etc. as well as providing more efficient estimates.

LONGITUDINAL AND BIVARIATE NORMS

Cameron's (1980) conditional standards of height seem to be the first development of longitudinal standards since the introduction of velocity standards. He calculates the percentile distribution of height at one age for each possible value of height at the age one year previously, this being done using simple regression techniques. For a variable like height where the year to year correlations are very high, this allows very precise judgement of 'typicality'. In fact, with such high correlations, there is little practical advantage over using simple velocities, but this will not be the case with other variables like weight where the correlations are lower. Unlike simple velocity standards the method can be extended to standards conditional on two or more preceeding occasions.

Conditional standards, of course, are not new. Standards of birthweight for gestation length or weight for height are ubiquitous. While all these are indeed useful they also have to be interpreted with care. Thus a child may not have an extreme weight for his height, yet may be below the first percentile for height on its own and so require attention. Likewise in industrialized countries, a 2000g baby at 36 weeks would be above the 5th percentile of birthweight for gestation length, yet his risk of dying in the perinatal period would be more than five times the average (Hellier & Goldstein, 1979). This example illustrates neatly the need to take account of both variables in such situations, or in other words to consider a bivariate standard rather than a single univariate standard whether a simple or a conditional one. In the birthweight-gestation case Hellier and Goldstein (1979) show how these two factors act jointly to determine perinatal mortality and are able to construct 'contours' of equal mortality which can be used as a screening device as in Fig. 5.

In cases such as this, where an appropriate 'outcome' variable is available, there is no difficulty in principle in providing useful bivariate standards. More generally, however, we have no ready outcome variable and, as with univariate standards, we have to choose an 'atypical' region on other grounds. Of course with univariate standards it is fairly clear that the atypical region should certain those measurements farthest away from the average. With a pair of bivariate measurements the usual procedure for normally distributed variables has been to draw 'equiprobability' ellipses, so that for example the 97th percentile ellipse is such that 97% of measurements lie inside it and the probability of a randomly chosen individual having a measurement anywhere inside the ellipse is greater than the probability of having a measurement anywhere outside the ellipse.

Fig. 6. shows the 95% ellipse based on sitting height (x) and lower limb (y) measurements from data of Harrison and Marshall (1970)

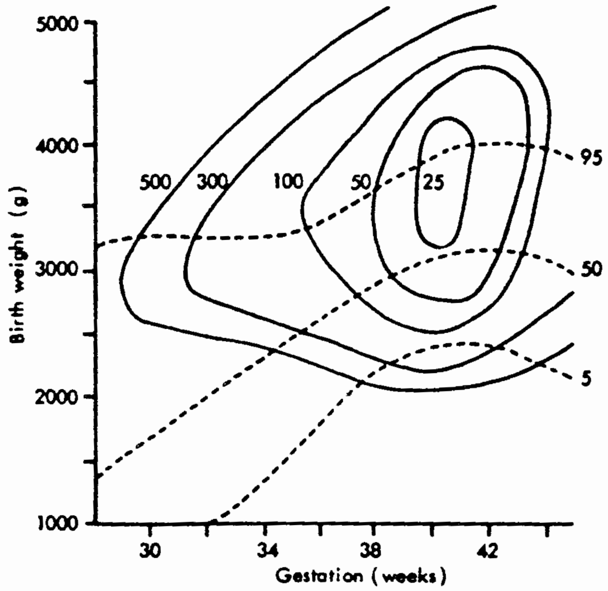

Fig. 5. Constant perinatal mortality contours (100 = average) and three selected birth-weight-for-gestation percentiles. (Based on 1973 data from Cuba, New Zealand and Sweden: see Hellier & Goldstein, 1979).

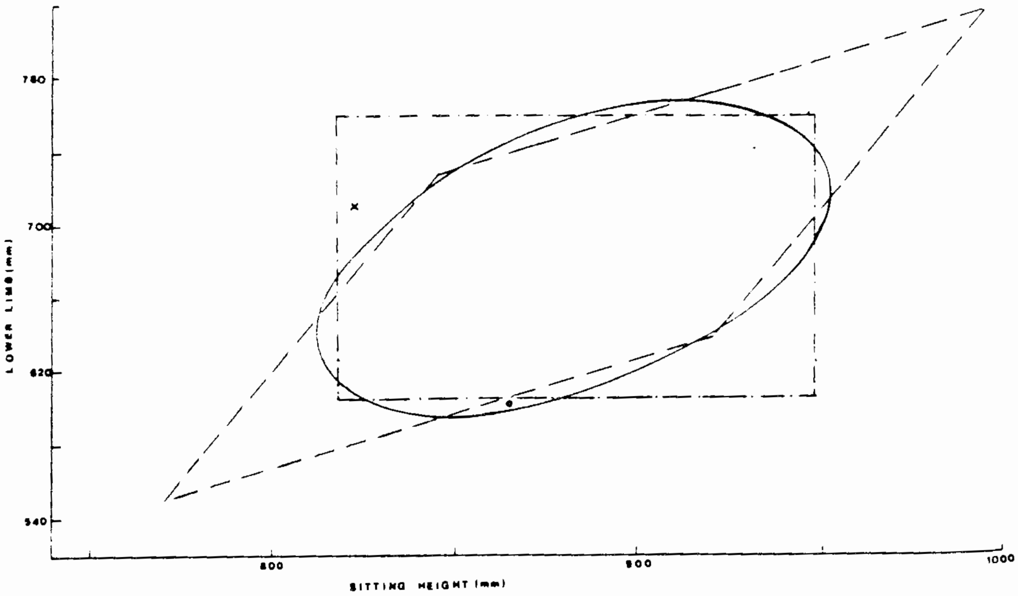

Fig. 6. Bivariate Norms 5% regions.

on adult women. In addition the rectangle gives the separate symmetrical univariate percentiles such that it encloses 95% of the measurements and the parallelogram gives the symmetrical y on x and x on y conditional standards such that it also encloses 95% of measurements.* Indeed, this figure could be regarded as a 'shape' norm for these two measurements.

The point is that in each case a different 5% of measurements is classified as atypical. The cross for example is a measurement which would be judged atypical by the ellipse and parallelogram but not by the rectangle, while the small circle would be judged atypical by the parallelgoram and rectangle but not by the ellipse. There are of course an infinity of shapes which enclose 95% of the measurements and the problem is which one to choose. My own view is that both the rectangle and the parallelogram are reasonable choices since they are direct extensions of univariate standards and atypical measurements can readily be interpreted. The circle for example is an individual with an extremely short lower limb. The ellipse on the other hand seems to have no particular justification, other than a certain mathematical elegance perhaps, and seems to allow no sensible interpretation to be placed on atypical measurements, since it pays no attention to the direction of atypicality.

In Fig. 6. the rectangle and parallelogram have been drawn so that the same percentage lie beyond the boundaries for each measurement. In general however we can choose to give more weight to one measurement rather than another. Thus in height and weight standards we might decide to make the boundary lines for height exclude more individuals than those for weight, consistent with a given overall percentage excluded, so that a measurement would need to be relatively more extreme for weight to be beyond that boundary line. In the case of weight and height we might also choose to have simple (non-conditional) boundaries for height together with conditional weight for height ones. This would then be exactly equivalent to first judging whether an individual was atypical for height in which case he is followed up. If he is typical then we judge the typicality of his wieght for height. Such a procedure seems a reasonable one to adopt in practice and the above discussion provides a formal framework for doing so. It can also, of course, readily be extended to the multivariate case.

CONCLUSIONS

In this necessarily brief summary of recent developments we have not attempted to cover every published contribution. Rather we have concentrated on certain contributions which we believe address themselves to the existing practicalities of growth studies, and which if followed should improve their effectiveness.

*The calculations have been carried out assuming a bivariate normal distribution.

REFERENCES

Altham, P.M.E., 1979, Detecting relationships between categorical variables observed over time: a problem of deflating a chi-squared statistic. Applied Statistics, 28: 115-125.

Angers, D., 1979, Simultaneous estimation of percentile curves with application to salary data. J. of American Statistical Association, 74: 621-625.

Bailar, B.A., 1975, The effects of rotation group bias in estimates from panel surveys. J. of American Statistical Association, 70: 23-30.

Baker, R.J. and Nelder, J.A., 1978, Generalized linear interactive modelling (GLIM). Rothamsted Experimental Station, Harpenden, Herts.

Berkey, C.S., 1982, Comparison of two longitudinal growth models for pre-school children. Biometrics, 38: 221-234.

Billewicz, W.Z., Fellowes, H.M. and Hytten, C.A., 1976, Comments on the critical metabolic mass and age of menarche. Ann. Human Biology, 3: 51-60.

Bock, R.D. and Thissen, D., 1980, Statistical problems of fitting individual growth curves, in: Human Growth and Motivation: Methodologies and Factors, by Johnston, F.E., Roche, A.F. and Susanne, C. (Eds.), Plenum, New-York.

Bock, R.D., 1982, Prediction of mature stature. Golden Jubilee Celebrations of the Indian Statistical Institute.

Bookstein, F.C., 1978, The measurement of biological shape and shape change. Springer-Verlag, New York.

Cameron, N., 1980, Conditional standards for growth in height of British cildren from 5.0 to 15.99 years of age. Ann. Human Biology, 7: 331-338.

Fendt, K., Helms, R.W. and Morrisey, L.J., 1979, Some examples of implementing longitudinal data management in SAS. Proceedings of Statistical Computing Section of American Statistical Association, 1979, 170-173.

Frisch, R.C. and Revelle, R., 1971, Height and weight at menarche and a hypothesis of menarche. Arch. Dis. Child., 46: 695-701.

Glazeby, C.A., 1979, Correlated residuals in non-linear regression applied to growth data. Applied statistics, 28: 251-259.

Goldstein, H., 1979, The design and analysis of longitudinal studies. Academic Press, London.

Goldstein, H., 1981, Measuring the stability of individual growth patterns. Ann. Human Biology, 8: 549-557.

Goldstein, H., 1981, Some graphical procedures for the preliminary processing of longitudinal data, in: Interpreting Multivariate data. Ed. Barnett, V. Wiley, New York.

Harrison, J.M. and Marshall, W.A.M., 1970, Normal standards for the relationships between the lengths of limbs and limb segments in young British women. Human Biology, 42: 90-104.

Healy, M.J.R. and Tanner, J.M., 1981, Size and shape in relation to growth and form. Symposium of Zoological Society of London, 46: 19-35.

Hellier, J. and Goldstein, H., 1979, Use of Birthweight and Gestation to assess perinatal mortality risk. J. of Epidemiology & Community Health, 33: 183-185.
Hidiroglou, M.A., 1981, Computerization of complex survey estimates. Proc. Statistical Computing Section of American Statistical Association, 1981, 1-7.
Hoinville, G. and Jowell, R., 1978, Survey research practice, Heinemann, London.
Holt, D., Scott, A.J. and Ewings, P.D., 1980a, Chi-squared tests with survey data. J. Royal Statist. Soc. A, 143: 303-320.
Holt, D., Smith, T.M.F. and Winter, P.D., 1980b, Regression analysis of data from complex surveys. J. Royal Statist. Soc. A, 143: 474-487.
Johnson, C. Amara, I., Edwards, S. and Koch, G.G., 1981, Some computational strategies for the analysis of longitudinal categorical data. Proceedings of Statistical Computing Section of American Statistical Association, 1981, 293-298.
Jordan, J., Ruben, M., Hernandez, J., Bebelagua, A., Tanner, J.M. and Goldstein, H., 1975, The 1972 Cuban national child growth study as an example of population health monitoring: Design and Methods. Ann. of Human Biology, 2: 153-171.
Kerachisky, S.M. and Mallor, C.D., 1981, The effect of monitoring payments on survey responses: Experimental evidence from a longitudinal study of economically disadvantaged youths. Proceedings of Section on Survey Research Methods of American Statistical Association, 1981, 258-263.
Kish, L., 1965, Survey Sampling. Wiley, New York.
Koch, G.G., Landis, J.R., Freeman, J.L., Freeman, D.H. and Lehnen, R.G., 1977, A general methodology for the analysis of experiments with repeated measurement of categorical data. Biometrics, 33: 133-158.
Korn, E.L. and Whittemore, A.S., 1979, Methods for analysing Panel studies of acute health effects of air polution. Biometrics, 35: 795-802.
Koziol, J.A., Maxwell, D.A., Fukushima, M., Colmerauer, M.A.M. and Pilch, Y.H., 1981, A distribution-free test for tumer-growth curve analyses with application to an animal immunotherapy experiment. Biometrics, 37: 383-390.
Landis, J.R., Standish, N.M. and Koch, G.G., 1976, A Computer Program for the Generalized Chi-Square Analysis of Categorical Data Using Weighted Least Squares to Compute Wald Statistics (GENCAT). Department of Biostatistics, University of Michigan.
Lessler, J.T., Maron, R.E., Rosthal, R.A., Pugh, G.E. and Caldwell, J.G., 1978, Non-respondent subsampling schemes for longitudinal surveys. Proceedings of Section on Survey Research Methods of American Statistical Association, 1978, 617-620.
Mednick, S.A. and Baert, A.E., 1981, Prospective longitudinal Research. Oxford University Press, Oxford.
Pelletier, P.A. and Nolte, M.A., 1978, Data Processing and Data-Base Management - The Panel Study of Income Dynamics. Proceedings of Section on Survey Research Methods of American Statis-

tical Association, 1978, 660-666.

Plewis, I., 1981, Using longitudinal data to model teachers' ratings of classroom behaviour as a dynamic process. J. of Educational Statistics, 6: 237-255.

Rao, C.T., 1965, The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. Biometrika, 52: 447-458.

Reinsel, G., 1982, Multivariate repeated-measurement or growth curve models with multivariate random-effects covariance structure. J. of American Statistical Association, 77, 190-195.

Robinson, B.N., Anderson, G.D., Cohen, E. and Gazdzik, W.F., 1977, Scientific information retrieval; Users Manual. S.I.R. Inc., Evanston, Illinois, U.S.A.

Rona, R.J. and Altman, D.G., 1977, National study of health and growth: standards of attained height, weight and triceps skinfold in English children 5 to 11 years old. Ann. of Human Biology, 4: 501-523.

Sandland, R.L. and McGilchrist, C.A., 1979, Stochastic growth curve analysis. Biometrics, 35: 255-271.

Stutzle, W., Gasser, T., Molinari, I., Largo, R.M. Prader, A. and Huber, P.J., 1980, Shape invariant modelling of human growth. Ann. of Human Biology, 7: 507-528.

Tanner, J.M., Whitehouse, R.H., Marshall, W.A., Healy, M.J.R. and Goldstein, H., 1975, Assessment of skeletal maturity and prediction of adult height. Academic Press, London.

Three Generations of children, 1982, Colour Video Film. 55 Minutes. Audio Visual Department, University of Bristol, Bristol, England.

Thompson, D.A.W., 1961, On growth and form. Ed. J.T. Bonner, Cambridge University Press.

Ware, J.M. and Wu, M.C., 1981, Tracking: Prediction of future values from serial measurements. Biometrics, 37: 427-437.

Weiss, P and Kavanu, J.L., 1957, A model for growth and growth control in mathematical terms. J. of General Physiology, 41: 1-47.

Wu, M. and Ware, J.M., 1979, On the use of repeated measurements in regression analysis with dichotomous responses. Biometrics, 35: 513-521.