Epidemiologic Reviews
Vol. 16, No. 2

Printed in U.S.A.

Mathematical and Analytical Aspects of Tracking

J. W. R. Twisk,[1] H. C. G. Kemper,[1] and G. J. Mellenbergh[2]

INTRODUCTION

In the epidemiologic literature, tracking is used to describe the longitudinal development of a certain variable. There is available no single definition of tracking, but the following concepts are involved: 1) the relation/correlation between early measurements and measurements later in life or the maintenance of a relative position within a distribution of values in the observed population through time or, in other words, the longitudinal stability of a certain variable (1); and 2) the predictability of future values by early measurements (2).

To assess tracking, one needs a sample of N subjects from a population of interest, which is measured at different time points. So one measures $Y(i,j)$ for subject i, where $i = 1,2,\ldots,N$, at time points $t(j)$, where $j = 1,2,\ldots,T$.

First, a measurement ($Y(t)$) "tracks" if, for any two time points (t_1 and t_2) within that time period, there is a positive relation over subjects between $Y(t_1)$ and $Y(t_2)$. The degree of this tracking is operationalized in a certain tracking coefficient. Second, tracking deals with the predictability of an early measurement, $Y(t_1)$, for the value of a measurement later in life, $Y(t_2)$. The magnitude of this prediction is operationalized in a predictive value or risk measure, which is, of course, related to the tracking coefficient.

The measurement $Y(t)$ can be either ordinal or dichotomous and has to be measured with independent measurement errors, because otherwise the magnitude of the relation between $Y(t_1)$ and $Y(t_2)$ (i.e., the tracking between $Y(t_1)$ and $Y(t_2)$) can be influenced by bias of the observers. Furthermore, the measurement errors of $Y(t)$ have to be low, because unreliability of $Y(t)$ can conceal possible evidence of tracking.

Tracking is mostly used in relation to risk factors of chronic diseases (3–17). Early detection of these risk factors can lead to the possibility of early treatment. In this view, it is important to get an idea about the stability of a certain risk factor in time. What is the relation between measurements of risk factors early in life and values of the same risk factors later? In other words, how predictive are early measurements for values later in life?

A number of studies that involve tracking deal with the longitudinal development of such cardiovascular disease risk factors as hypertension (3–6), hypercholesteremia (7–9), or body fatness (10–13). However, tracking has been described not only in regard to chronic disease risk factors, but also in relation to growth parameters, e.g., body height and body weight (10, 13), and for such specific variables as dietary patterns (14, 15), physical activity (16, 17), and pulmonary function (18).

It is very difficult to compare the magnitudes of tracking indices or predictability

Received for publication December 23, 1993, and in final form August 18, 1994.

Abbreviations: GEE, generalized estimating equations; ICC, intraclass correlation coefficient; LPC, longitudinal principal component.

[1] Department of Health Science, Faculty of Human Movement Sciences, Vrije Universiteit, Amsterdam, The Netherlands.

[2] Department of Methodology, Faculty of Psychology, University of Amsterdam, Amsterdam, The Netherlands.

Reprint requests to Prof. Dr. H. C. G. Kemper, Dept. of Health Science, Faculty of Human Movement Sciences, Free University, vd Boechorststraat 9, 1081 BT Amsterdam, The Netherlands.

measures, because in most studies there are great differences in the length of the interperiod between the related measurements or in the ages of the subjects at which the initial measurement is carried out. Another, perhaps greater, problem is major differences in the methodology used to assess tracking and in the ways statistical tools are used to operationalize tracking.

In this review, a summary will be given of the basic methods used to assess tracking. Furthermore, a comparison will be made between the different methods by assessing these methods in terms of two different longitudinal data sets.

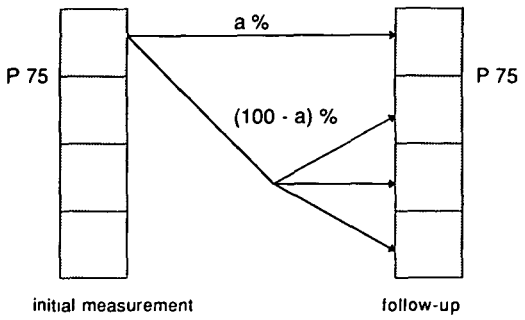

FIGURE 1. Graphical representation of the tracking coefficient calculated as the percent of subjects who stayed in the highest quartile of the distribution during follow-up.

TRACKING COEFFICIENTS

Nonparametric approaches

The advantage of using a nonparametric approach to calculate a tracking coefficient is that there are no assumptions about the distribution of the measurement values.

When tracking is calculated between two measurements ($T = 2$), most studies have quantified the degree of tracking by calculating Spearman's rank correlation coefficient between $Y(t_1)$ and $Y(t_2)$ (7, 19–22). When the measurements $Y(t_1)$ and $Y(t_2)$ are bivariate normally distributed, the value of Spearman's rank correlation coefficient is similar to the value of Pearson's correlation coefficient.

Another way to describe tracking when $T = 2$ is the following: According to the initial measurement, the population is divided into percentile groups (quartiles, quintiles, or deciles). Then the percent of the subjects who stayed in the same upper or lower percentile group at one or more follow-up measurements is calculated. If that percent is more than the expected percent when the subjects are randomly divided into each percentile group (e.g., more than 25 percent for quartiles, more than 20 percent for quintiles, more than 10 percent for deciles), the population is said to track for that particular variable (1, 18, 23–25) (figure 1).

Sometimes the population is not divided into subgroups according to percentile ranking, but according to some predetermined cut point. The percent of subjects who stayed in the group above or under this specified level at a follow-up measurement is calculated (21).

The problem in calculating tracking coefficients based on the division into percentile groups is that the magnitude of the coefficients very much depends on the grouping of the data.

On the basis of the division of the population into percentile groups, two other nonparametric tracking coefficients have been used to describe tracking. Nishio et al. (26) calculated a tracking coefficient when $T = 2$. They tried to give different weights to movements to and from different percentile groups and calculated the following tracking index (TI):

$$TI = \frac{T(s)}{T(h)}, \qquad (1)$$

where: $T(h)$ = tracking of a hypothetical group whose values change randomly between the percentile groups during the measurement period; and $T(s)$ = tracking of the study group,

$$T(s) = \frac{(2x + y - z)}{(x + y + z)}, \qquad (2)$$

where: x = number of subjects who remained in the same percentile group; y = number of subjects who moved to a neigh-

boring percentile group; and z = number of subjects who moved to a remote percentile group.

If the population is divided into quintiles, then $T(h)$ takes a value of 0.24 and TI ranges between 1.0 and 8.3.

In this approach, not only the subjects who stayed in the same percentile group during follow-up but also the subjects who moved to a neighboring percentile group positively influence the tracking coefficient TI (figure 2).

Cohen's kappa (κ) (27) can be calculated in longitudinal studies where $T \geq 2$, and it is calculated as follows:

$$\kappa = \frac{\bar{p} - \hat{p}}{1 - \hat{p}}, \tag{3}$$

$$\bar{p} = \frac{1}{\mathrm{NT}(T-1)} \times \sum_{i=1}^{N} \sum_{g=1}^{G} n_{ig}(n_{ig} - 1), \tag{4}$$

$$\hat{p} = \frac{1}{\mathrm{NT}} \times \sum_{g=1}^{G} \sum_{i=1}^{N} n_{ig}^2, \tag{5}$$

where: G = number of groups (4 by using quartiles, 5 by quintiles, etc.); N = number of subjects; and n_{ig} = number of times individual i is in the gth percentile group.

$\bar{p}$, a measure of how well the group tracks as a whole, is compared with a value $\hat{p}$, which is expected if individuals are randomly assigned to the different groups at each time point. So, for each individual, the number of times that the individual is in each of the particular percentile groups is counted and compared with the value that is expected if the individuals are randomly assigned to the different percentile groups at each measurement.

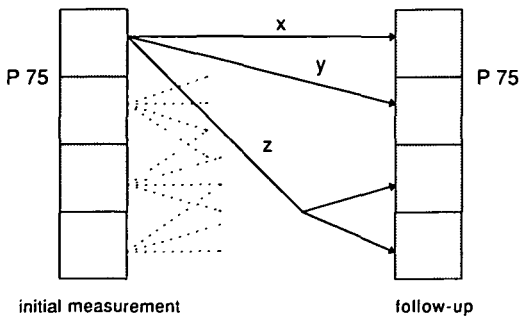

FIGURE 2. Graphical representation of the tracking index TI of Nishio et al. (26).

κ ranges from 0.0 to 1.0, and if $\kappa > 0.75$, then the variable tracks well. If $\kappa < 0.40$, then the variable tracks poorly, and if κ has a value in between these two, then there is moderate tracking for the variable of interest (28).

The major advantage of κ is that the index is very easy to compute, but one of the problems of κ is that all movements between two percentile groups are weighted equally, irrespective of the length of the movement. In order to overcome this drawback, Cohen (29) also developed a weighted κ, in which the lengths of movements are weighted unequally. However, to our knowledge, the weighted index has never been used in tracking analysis. Both κ and the weighted κ are interpretable as intraclass correlation coefficients (ICC) (30).

In situations where $T > 2$, Kendall's coefficient of concordance W is calculated to describe tracking (18). Kendall's W is not calculated on the basis of movements between different percentile groups, but is based on changes in individual rankings through time:

$$W = \frac{12}{T^2 N(N+1)(N-1)} \times \sum_{i=1}^{N} \left(R_i - \frac{T(N+1)}{2} \right)^2, \tag{6}$$

where: T = number of times a value is measured; N = number of subjects; and R_i = sum of all rankings at all measurements for individual i.

W is directly related to the average Spearman correlation coefficient:

$$\bar{\rho}_s = \frac{(WT - 1)}{(T - 1)}, \tag{7}$$

where: $\bar{\rho}_s$ = Spearman correlation coefficient averaged over T time points; W = Kendall's W; and T = number of time points.

W can take values between 0.0 and 1.0 and indicates the degree of association between the rankings at each of the repeated measurements. When W is calculated for random numbers, the coefficient is not equal to zero, but to a positive value that depends on the number of time points T. To compare W with other coefficients, W has to be rescaled so that W equals zero when applied to random numbers.

Lauer et al. (5) used a different definition of tracking. First, they expressed each value of a certain variable at a particular measurement as a percentile rank. For each individual, the average of the percentile ranks throughout time was calculated and was called level. For each individual, a regression line describing the change of the percentiles over time was calculated and the slope of this line was called trend. The goodness of fit of this line, measured as the residual standard deviation, was called variability (figure 3). Second, they defined not only subjects who maintain a high rank order with low variability to "track" toward future high values, but also subjects with lower levels, low variability but high trends. The tracking coefficient that Lauer et al. (5) calculated was the percent of subjects who belonged to these groups. Because this tracking coefficient is based on a different definition of tracking, it is not comparable to the other coefficients.

Table 1 summarizes the different nonparametric tracking coefficients used to describe tracking.

Parametric approaches

When T = 2, the most used parametric approach to calculate a tracking coefficient is the Pearson correlation coefficient (2, 31–34). The Pearson correlation coeffi-

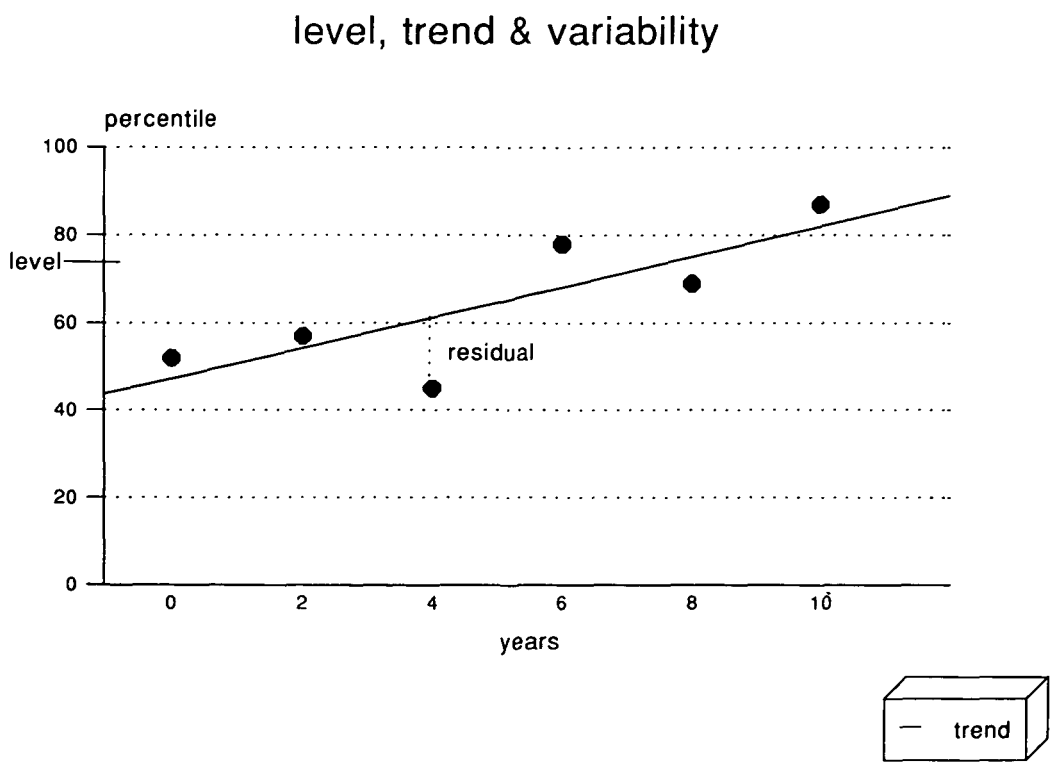

FIGURE 3. Graphical representation of level (average percentile), trend (slope of the least squares line), and variability (square root of the sum of the squared residuals divided by (N − 2)) according to Lauer et al. (5).

TABLE 1. Nonparametric tracking coefficients

Coefficient	Range*	Tracking criterion†	Data points‡
Spearman's ρ	−1. . . 1	Arbitrary	2
% at "high risk"	0% . . . 100%	>20% (quintiles) >25% (quartiles	2
Lauer's %	0% . . . 40%	Arbitrary	All
Cohen's κ	0 . . . 1	>0.75 good >0.40 moderate	All
Nishio's TI			
Quintiles	1 . . . 8.3	Arbitrary	2
Quartiles	1 . . . 4.0		
Kendall's W	0 . . . 1	Significance	All

* Range between the minimal and maximal value that the coefficient can take.
† *Criterion to decide whether or not a variable tracks.*
‡ Number of data points (measurements) used to calculate the coefficient.

cient, however, is only suitable when $Y(t_1)$, $Y(t_2)$ are bivariate normally distributed. When there are more than two longitudinal measurements, the use of correlation coefficients to describe tracking has the problem that it does not use all the available data. When $Y(t_1)$. . . .$Y(t_T)$ are T-variate normally distributed with equal variances and covariances, a possible solution to this problem is the calculation of the ICC, which is defined as:

$$\text{ICC} = \frac{(\sigma_B^2 - \sigma_W^2)}{(\sigma_B^2 + \sigma_W^2)}, \qquad (8)$$

where: σ_B^2 = between subjects variance; and σ_W^2 = within subjects variance.

When $T = 2$ and the values are bivariate normally distributed, the ICC is equivalent to the Pearson correlation coefficient. When $T > 2$, the ICC is in fact the parametric form of Kendall's W. The ICC, however, is never used in relation to tracking analysis.

Beckett et al. (35) and Guo et al. (36) have dealt with the problem of not using all the available data, and they have developed an alternative approach for studies where $T \geq 2$. Assuming that the longitudinal measurements are multivariate normally distributed, they used a general linear model, as specified by Jennrich and Schluchter (37) and implemented in BMDP5V (38), to describe the longitudinal pattern of change:

$$Y_i = X_i\beta + \epsilon_i, \qquad (9)$$

where: Y_i = vector of observed values of individual i; X_i = design matrix for individual i; β = vector of regression parameters; and ϵ_i = vector of measurement errors.

The estimation of the regression parameters of this longitudinal linear model includes an estimation of the correlation matrix between pairs of measurements. These estimated correlations are interpreted as tracking coefficients (Guo's ρ).

One of the problems with this approach is that the correlations between the repeated measurements were assumed to follow a certain structure. In the banded structure, the correlations between two measurements are assumed to depend only on the time interval (interperiod) between the two measurements. In the compound symmetry structure, all pairs of measurements, independent of the length of the time interval, are assumed to have the same correlation, and in the autoregressive structure the correlations between two measurements are assumed to follow an exponential structure. This means that the correlation between measurements one year apart is t^1, the correlation between measurements 2 years apart is t^2, and so on.

Foulkes and Davis (39) have developed a tracking coefficient (or index) γ for studies when $T \geq 2$. Their approach is based on a statistical model in which the longitudinal values of an individual change as a certain function of time. This function, which can be either polynomial, exponential, or anything else, has to be common for the population, whereby the certain parameters are assumed to vary from subject to subject. For each individual, the original values are replaced by the predicted values, based on this function of time.

The Foulkes and Davis tracking index γ, also known as the growth separation index (31), is used to determine the probability that two individuals selected at random will have curves that do not cross over the time

period under consideration. This probability is simply the number of curves in the population that do not cross divided by the number of ways in which two curves can be randomly selected from the population:

$$\gamma[T_1,T_2] = 1 - \sum_{i=1}^{N} \frac{m_i}{N(N-1)}, \quad (10)$$

where: m_i = number of times the growth curve of the ith individual crosses at least once with the growth curves of other individuals over the observed time period $[T_1, T_2]$; and N = number of subjects.

The index γ has always to be given with the observed time period $[T_1, T_2]$, because the value of γ highly depends on the length of the observed time period.

The coefficient γ can take values between 0.0 and 1.0. A value of 0.0 means that every individual curve crosses every other individual curve at least one time; a value of 1.0 indicates that none of the individual curves cross; and value > 0.5 indicates tracking because two individual curves chosen at random would be more likely to have curves that do not cross (13).

One of the advantages of the Foulkes and Davis tracking index γ is that there are no assumptions about the form of the curves. The assumptions made on the sample distribution, however, depend on the choice of the form of the curve. When the individual response patterns are simply drawn by connecting the successive time points without assuming any mathematical model, this procedure would be nonparametric.

Although there are no assumptions about the form of the curves, the simplicity of the curves used to describe the data is very important. In other words, the simpler the curve used to describe the data, the higher will be the value of the index (40). Another problem in interpreting γ is that individuals at the extremes of the distribution are less likely to cross the curves of other individuals than individuals who have curves near the mean curve (28).

If measurements are made only at two points in time, so that the curves for each individual are straight lines, then the Foulkes and Davis tracking index γ is an estimate of Kendall's rank correlation coefficient τ (39, 41).

McMahan (42) has developed a tracking coefficient (or index) τ based on all the available data. The coefficient is calculated under the assumption that $Y(t)$ are multivariate normally distributed with common covariance matrix. The general idea behind this index, which is also known as the growth constancy index (C) (40), is that a population tracks for a certain variable if, for each individual (growth) curve, the relative deviation from the population mean (growth) curve remains unchanged over time. For standardized variables, the coefficient is calculated as follows:

$$\tau = 1 - \frac{1}{(N-1)(T-1)} \sum_{i=1}^{N} S_i^2, \quad (11)$$

$$S_i^2 = \sum_{j=1}^{T} (y_{ij} - \bar{y}_i)^2, \quad (12)$$

$$\bar{y}_i = \frac{1}{T} \sum_{j=1}^{T} y_{ij}, \quad (13)$$

where: y_{ij} = observation for individual i at time point t; T = number of times a value is measured; and N = number of subjects.

In this way, τ is the average value of the $1/2T(T-1)$ Pearson correlation coefficients, where T is the number of times a value is measured. In this sense, the parametric tracking coefficient of McMahan is equivalent to the nonparametric tracking coefficient Kendall's W, which is the average value of the $1/2T(T-1)$ Spearman correlation coefficients.

When the measurements and consequently the tracking coefficient τ are biased by a considerable measurement error, the actual data can be replaced by predicted ones based on a function of time. If τ has the value 1.0, there is perfect tracking for that variable. If τ has the value 0.0, there is no tracking for that variable. McMahan's τ can take negative values, which indicates a reversal of the values between two observed time points.

In table 2, the different parametric tracking coefficients are summarized.

PREDICTION

Nonparametric approaches

For $T = 2$, Palti et al. (6) calculated a "predictive value," which is defined as the number of subjects who stayed in a certain high risk group during follow-up divided by the number of subjects in that high risk group at the initial measurement (figure 4).

Kemper et al. (25) also calculated for $T = 2$ a "relative probability." This "relative probability" is the percent of subjects who were in the high percentile group at the initial measurement as well as at the follow-up measurement divided by the percent of subjects who rose from one of the lower percentile groups at the initial measurement to the highest percentile group at the follow-up measurement (figure 5).

The name of this index is a bit misleading; the "relative probability" is not really a probability, because the value can be greater than one.

When the group division is based on dichotomization, both indices are related to kappa. The rescaled version of the "predictive value" (which is in fact the same as the percent of subjects who stayed in a certain high risk group during follow-up) is equivalent to kappa, while the relative probability divided by the number of subjects is equivalent to $\bar{p}$, which is used in the calculation of kappa (equation 4).

TABLE 2. Parametric tracking coefficients

Coefficient	Range*	Tracking criterion†	Data points‡
Pearson's ρ	−1 . . . 1	Arbitrary	2
Guo's ρ	−1 . . . 1	Arbitrary	All
Foulkes and Davis's γ§	0 . . . 1	>0.5	All
McMahan's τ§	−1 . . . 1	>0.5	All

* Range between the minimal and maximal value that the coefficient can take.

† Criterion to decide whether or not a variable tracks.

‡ Number of data points (measurements) used to calculate the coefficient.

§ Foulkes and Davis and McMahan first replace the original values by predicted ones, based on a function in time, and they use the predicted curves to calculate their tracking coefficient.

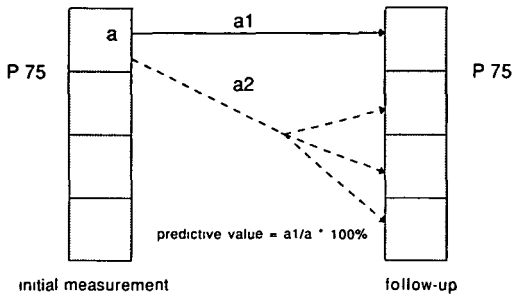

FIGURE 4. Graphical representation of the "predictive value" as calculated by Palti et al. (6).

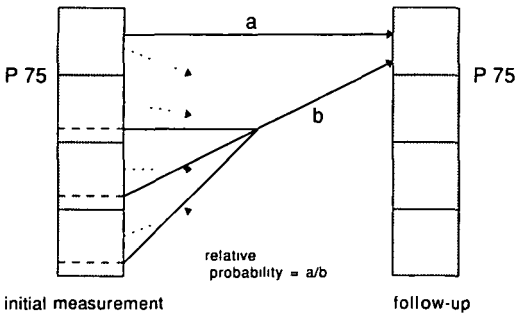

FIGURE 5. Graphical representation of the "relative probability" as calculated by Kemper et al. (25).

Parametric approaches

A stepwise linear regression analysis is mostly used to predict future values of certain variables from early measurements. In such an analysis, the value of the last measurement is the dependent variable and values of earlier measurements are the predictor variables (43). When $T = 2$, and there are no covariates in the statistical model involved, the estimated regression coefficient β is the same as the Pearson correlation coefficient between $Y(t_1)$ and $Y(t_2)$.

A different approach was used by Beckett et al. (35) and Guo et al. (36). As mentioned earlier, they used a general linear model to estimate the correlation coefficient between two repeated measurements. In addition, they calculated the probability of a follow-up measurement being above a certain predetermined cut point, given the initial value. The paired

measurements are assumed to be bivariate normally distributed:

$$\Pr(Y_2 > C|Y_1 = y) = \int_z^{\infty} (2\pi)^{1/2}\exp(-t^2/2)dt, \quad (14)$$

$$z = [C - (\mu_2 - \rho\sigma_2/\sigma_1(y_1 - \mu_1))] / \sigma_2\sqrt{1-\rho^2}, \quad (15)$$

where: Y_1, Y_2 = observed values; μ_1, μ_2 = mean values; σ_1, σ_2 = standard deviations; t = time between the related measurements; C = predetermined cut point; and ρ = correlation between the related measurements.

Lauer et al. (5) also estimated the risk (probability) of being above a certain predetermined cut point on the basis of the initial value. However, they used logistic regression analysis to calculate this probability. By means of this logistic regression analysis for each decile at the initial measurement, the probability of being above a predetermined cut point at follow-up can be calculated.

In many studies, the concept of tracking is related to modeling the longitudinal development of a certain variable $Y(t)$ (12, 44, 45). All of these studies have to do with growth in general and, although they are very interesting, they do not estimate some sort of predictive value. Therefore, they are not further discussed in this paper.

One of the latest innovations in predicting future values from early measurements is the use of longitudinal principal component (LPC) analysis (46). Assuming a linear relation between the longitudinal measurements, LPC analysis starts by finding the linear combination of the original variables, in this case the same variable measured on different occasions, which accounts for the maximum amount of variance. This linear combination is called the first principal component. The next step is to find a second linear combination, uncorrelated with the first principal component, which accounts for the next largest amount of variance, and so on. Although Berkey et al. (46) only used LPC analysis to model growth and not to construct a new tracking measure, the percent of variance (R^2) accounted for by the first principal component can be interpreted as one. When the measurements are multivariate normally distributed, the square root of this percent of variance (R) can be interpreted like an ICC.

EXAMPLE

To illustrate the diversity of the tracking indices, we calculated the different coefficients for two separate data sets. The data used came from the Amsterdam Growth and Health Study (47). This longitudinal study started in 1977 with four annual measurements on boys and girls with a mean age of 13 years. The study continued with a 5th measurement at age 21 years (1985) and the subjects returned for a 6th measurement at age 27 years (1991). At each year of measurement, the investigators assessed anthropometric parameters (body height, body weight, and body composition), biologic parameters (lipoprotein levels, blood pressure, and physical fitness), psychological parameters (personality and achievement motivation), and lifestyle parameters (nutritional habits, smoking behavior, and daily physical activity).

For the purpose of this example, the longitudinal measurements of the anthropometric variable body height and the biologic variable total serum cholesterol will be used. The reason for choosing these two variables is the accuracy with which both variables are measured, so measurement errors can not influence the calculation of tracking coefficients. To avoid potential confounding effects of gender, only the female data were used. The total number of females with a complete longitudinal data set is 98.

Most tracking indices are not influenced by the shape of the longitudinal curves. For other parameters, however, the first step in calculating the different indices of tracking is to assess the best way to describe

the population longitudinal development. Therefore, for both body height and serum cholesterol, different linear models with time or a higher order of time as the independent variable are compared with each other. According to this comparison, the longitudinal development of serum cholesterol was best described by a straight line model, while the development of body height was best described by a third-degree linear model. In calculating Foulkes and Davis's tracking index γ and McMahan's tracking index τ, the observed values for each individual are replaced by the predicted ones according to the above models. In figure 6, for one subject, the observed values of body height are plotted against the individual predicted values, while, in figure 7, the observed values of serum cholesterol are plotted against the individual predicted values.

TRACKING COEFFICIENTS

Some of the tracking coefficients can only be calculated for $T = 2$. With the data set of the Amsterdam Growth and Health Study (where $T = 6$), these coefficients are calculated over a 15-year period between the first and last measurement. In table 3, both parametric and nonparametric tracking coefficients are shown for body height and serum cholesterol. Besides the absolute values, the rescaled values of the different coefficients are also presented. The coefficients are rescaled so that when they are calculated for random numbers the value of the coefficient equals zero. For body height, Spearman's ρ is almost equivalent to Pearson's ρ, while for serum cholesterol there is some disagreement between the two coefficients. A 6×6 matrix of paired time point correlation coefficients with Spearman's ρ above the diagonal and Pearson's ρ below the diagonal shows the same result. For body height (table 4), the coefficients are almost the same, while for serum cholesterol (table 5), there is some minor disagreement.

It is difficult to compare both correlation coefficients with the coefficients based on the division into subgroups (in the example calculated for quartiles), because the magnitude of the coefficients is highly influenced by the choice of the division.

Table 6 shows both the absolute and rescaled tracking coefficients calculated with all the available longitudinal data. For body height, all coefficients show good to

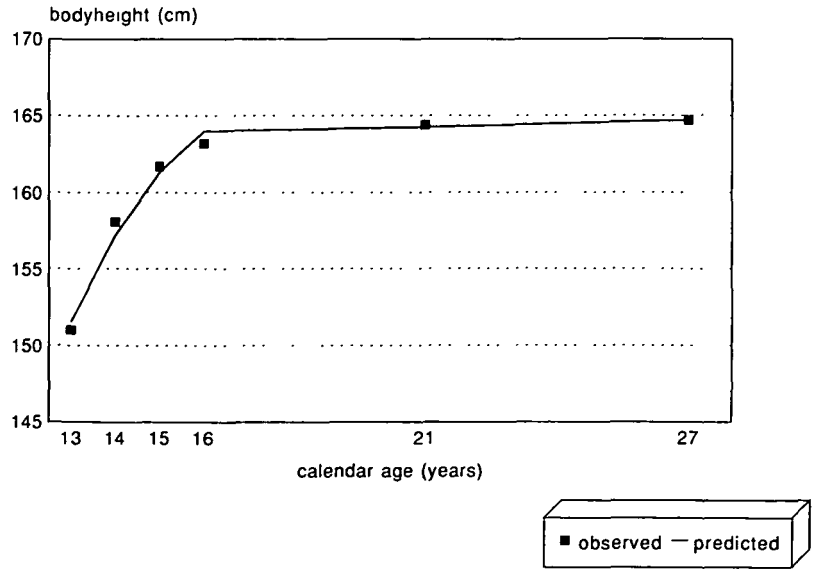

FIGURE 6. The observed values for body height of one subject and the predicted linear model with time to the third degree for that subject.

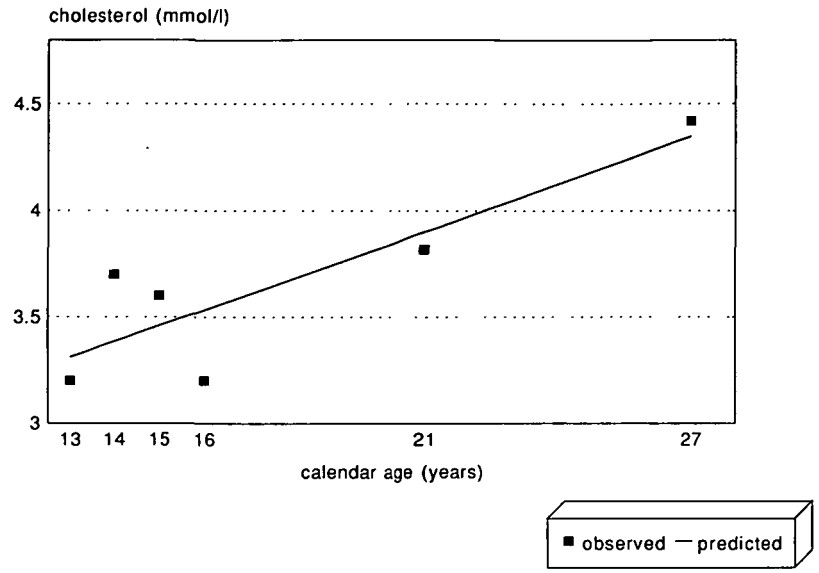

FIGURE 7. The observed values for serum cholesterol of one subject and the predicted linear regression line for that subject.

TABLE 3. Absolute values and rescaled* values of tracking coefficients calculated over a 15-year period between two longitudinal measurements for body height and serum cholesterol

Coefficient	Body height		Cholesterol	
	Absolute	Rescaled	Absolute	Rescaled
Spearman's ρ	0.66	0.66	0.60	0.60
Pearson's ρ	0.66	0.66	0.54	0.54
% at "high risk"†	50	0.33	39	0.19
Nishio's TI†	2.49	2.49	2.28	2.28

* Each coefficient is rescaled so that calculated for random numbers the coefficient equals zero.
† Values of the coefficients are based on the division of the population into quartiles.

TABLE 4. Correlation matrix* of all paired measurements of body height

Points of measurement	Spearman					
	1	2	3	4	5	6
Pearson						
1		0.94	0.85	0.77	0.67	0.66
2	0.95		0.97	0.92	0.85	0.84
3	0.84	0.96		0.98	0.94	0.94
4	0.76	0.91	0.99		0.98	0.97
5	0.67	0.85	0.95	0.99		0.99
6	0.66	0.84	0.94	0.98	0.99	

* Above the diagonal are shown the Spearman correlation coefficients, while below the diagonal are shown the Pearson correlation coefficients.

moderate tracking, while for the serum cholesterol measurements there is some disagreement. Kendall's W is highly significant ($p \leq 0.01$) and Foulkes and Davis's index γ is 0.67 (>0.5), which means that for both indices the serum cholesterol values seem to track. Cohen's κ for quartiles, however, is low (<0.4) and applying the decision criterion leads to the conclusion that the serum cholesterol values do not track. McMahan's τ indicates good tracking, while the value for Guo's ρ is a bit low. According to this coefficient, one may conclude that serum cholesterol tracks moderately. One of the problems with the method used by Beckett et al. (35) and Guo et al. (36) is that this method is not really designed to estimate a correlation coefficient, e.g., the method is not robust against the choice for a different correlation structure. To illustrate this, the correlation coefficients assuming different correlation structures are calculated for both serum cholesterol and body height (table 7). The estimated correlation coefficients are highly influenced by the choice of a particular structure. This indicates that the value of the tracking coefficient depends on the choice of a certain correlation structure and is therefore not useful as a tracking coefficient.

TABLE 5. Correlation matrix* of all paired measurements of serum cholesterol

Points of measurement	Spearman					
	1	2	3	4	5	6
Pearson						
1		0.83	0.78	0.74	0.65	0.62
2	0.84		0.82	0.83	0.65	0.58
3	0.78	0.83		0.81	0.60	0.60
4	0.76	0.81	0.80		0.69	0.61
5	0.63	0.63	0.61	0.68		0.71
6	0.54	0.54	0.54	0.61	0.70	

* Above the diagonal are shown the Spearman correlation coefficients, while below the diagonal are shown the Pearson correlation coefficients.

TABLE 6. Absolute values and rescaled* values of tracking coefficients using all the available longitudinal data for body height and serum cholesterol

Coefficient	Body height		Cholesterol	
	Absolute	Rescaled	Absolute	Rescaled
Cohen's κ†	0.59	0.59	0.37	0.37
Kendall's W	0.90	0.88	0.75	0.70
Guo's ρ	0.62	0.62	0.45	0.45
Foulkes and Davis's γ	0.72	0.72	0.67	0.67
McMahan's τ	0.89	0.89	0.83	0.83

* Each coefficient is rescaled so that, calculated for random numbers, the coefficient equals zero.

† Values of the coefficient are based on the division of the population into quartiles.

The value of κ is based on the division of the population into quartiles and therefore is not comparable with the other coefficients. For all the coefficients based on the division of the population into subgroups, the magnitude of the tracking coefficient is highly influenced by the choice of the arbitrary cut point. In table 8, tracking coefficients are calculated based on different subgroups (dichotomization, quartiles, quintiles, and deciles). The results show the dramatic influence of an arbitrary decision on the magnitude of the tracking coefficient.

TABLE 7. Estimated correlation coefficients (Guo's ρ) for four different correlation structures for body height and serum cholesterol

Assumed correlation structure	Body height ρ	Cholesterol ρ
Unstructured	0.62	0.52
Banded	0.70	0.51
Compound symmetry	0.88	0.64
Autoregressive	0.87	0.24

One reason for the disagreement in the magnitude of tracking coefficients is the fact that coefficients are measured on different scales. To illustrate this, the average correlation coefficients are compared with some related measures. First, for both Spearman's ρ and Pearson's ρ, the average $1/2N(N-1)$ correlation coefficients are calculated. In table 9, these values are compared with the rescaled value of Kendall's W, with the ICC, and with the amount of variance explained by the first principal component (R_{1pc}). As expected, Kendall's W is for both body height and serum cholesterol almost equal to the average Pearson's ρ and to the average Spearman's ρ. Although the values are a bit lower than the ICC and the R_{1pc}, all coefficients are comparable.

Prediction

Nonparametric parameters. Table 10 shows the "predictive value" as calculated by Palti et al. (6), which is nothing more than the percent of subjects who stayed in a certain high-risk group at the follow-up measurement and the "relative probability" as calculated by Kemper et al. (25). Both coefficients are rescaled so that when applied to random numbers the coefficients equal zero. As with the tracking coefficients, the prediction measures are highly influenced by the choice of an arbitrary cut point.

Parametric parameters. Assuming a bivariate normal distribution, Beckett et al. (35) and Guo et al. (36) provide a method to predict the probability of a certain high risk value at the follow-up measurement given the initial value. With logistic regression analysis (48), it is also possible to calculate the probability of being above a certain high risk value from the initial measurement. According to the initial measurement in 1977, the subjects are divided into deciles. For each decile, the probability of obtaining a value above the 80th percentile (P80) at the follow-up measurement in

TABLE 8. Rescaled* nonparametric tracking coefficients for body height and serum cholesterol: comparison between analysis with different group separations

Rescaled coefficient	No. of groups							
	Body height				Cholesterol			
	2	4	5	10	2	4	5	10
% at "high risk"	0.51	0.33	0.48	0.51	0.47	0.19	0.14	0
Cohen's κ	0.72	0.59	0.51	0.29	0.50	0.37	0.24	0.12
Nishio's TI	113.3	2.49	3.82	−61.7	113	2.28	4.17	113

* Each coefficient is rescaled so that, calculated for random numbers, the coefficient equals zero.

TABLE 9. Rescaled* tracking coefficients (average Pearson correlation coefficient, average Spearman correlation coefficient, Kendall's *W*, intraclass correlation coefficient (ICC), and the square root of the amount of variance accounted for by the first principal component (R_{1pc})) based on all six longitudinal measurements for body height and serum cholesterol

Rescaled coefficient	Body height	Cholesterol
Average Pearson's ρ	0.88	0.70
Average Spearman's ρ	0.89	0.69
Kendall's *W*	0.88	0.70
ICC	0.96	0.87
R_{1pc}	0.95	0.86

* Each coefficient is rescaled so that, calculated for random numbers, the coefficient equals zero.

1991 is calculated. In figure 8, these probabilities are shown calculated according to the method of Beckett et al. (35) and Guo et al. (36). In figure 9, the probabilities are shown calculated by logistic regression analysis.

For body height, the logistic regression approach was not possible. Calculation of probabilities for a stable (relative high tracking coefficients) variable with a moderate number of subjects is not possible, due to the lack of subjects in the lower deciles at the initial measurement and because there are no subjects who move from the lower deciles to the two highest deciles at the follow-up measurement.

For body height, subjects at P80 at the measurement in 1977 had approximately a four times higher risk of being above P80 at the measurement in 1991 than subjects at the 50th percentile (P50) in 1977.

For serum cholesterol, the results calculated by logistic regression are comparable with the results calculated according to the method of Beckett et al. (35) and Guo et al. (36). In both calculations, subjects at P80 at the measurement in 1977 had approximately three times higher risk of being above P80 at the follow-up measurement in 1991 than the subjects at P50 in 1977.

DISCUSSION

The choice between a parametric and a nonparametric approach depends in the first place on the scale in which the variable is measured. Parametric approaches can only be used for continuous variables, while nonparametric approaches can also be used for categorical variables. This choice also depends on the distribution of the observed values. The nonparametric approach makes less assumptions about the distribution of the values. Furthermore, the choice can be based on the relative computational simplicity of the nonparametric approaches. If the data are not normally distributed, it is possible to transform the data into normalized values (1).

If the population is divided into subgroups according to some arbitrary cut point (quartiles, quintiles, deciles), the magnitude of the tracking coefficient depends highly on an arbitrary decision in dividing the population (tables 8 and 10), which makes these approaches very troublesome in assessing the tracking phenomenon. Furthermore, the division into subgroups is based on sample percentiles and not on population percentiles. When the sample size is small or the measurement is unreliable, sample percentiles are very unstable. Then, the choice of the subgroups itself is subject to sampling error and can be the reason for finding low tracking coefficients.

TABLE 10. Rescaled* nonparametric prediction measures for body height and serum cholesterol: comparison between analysis with different group separations

Rescaled coefficient	No. of groups							
	Body height				Cholesterol			
	2	4	5	10	2	4	5	10
Predictive value	0.51	0.33	0.48	0.51	0.47	0.19	0.14	0
Relative probability	3.1	3.33	5.79	12.4	2.4	2.06	1.84	1

* Each coefficient is rescaled so that, calculated for random numbers, the coefficient equals zero.

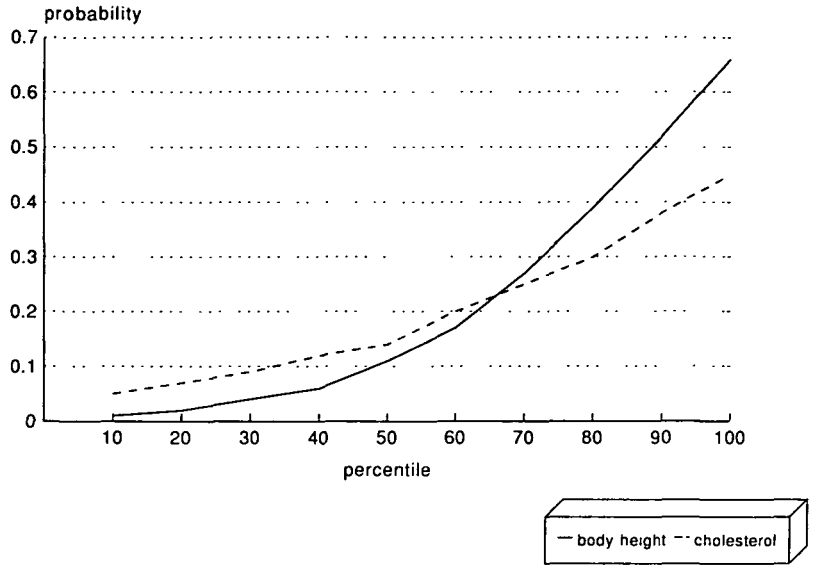

FIGURE 8. Probabilities of different deciles at the initial measurements in 1977 to reach values above P80 at the follow-up measurements in 1991, calculated according to the method of Beckett et al. (35) and Guo et al. (36), for body height and serum cholesterol.

Furthermore, by dividing the population into subgroups, a lot of information about the data is lost. For instance, subjects can change within their original percentile group without influencing the tracking coefficient or predictability, while a minor shift at the borders of two percentile groups actually will influence tracking coefficients.

The same problem arises by comparing Foulkes and Davis's tracking coefficient γ with Cohen's κ or any other tracking coefficient based on the division of the population into subgroups. If subjects remain in the same percentile group (not influencing κ), their curves can cross many times and so influence the Foulkes and Davis tracking index γ. Similarly, vice versa, two randomly chosen individuals can have parallel lines but both can move from one percentile group to another (27).

Interpretation of coefficients

The above brings us to an important issue in discussing the different methods of describing tracking, namely the interpretation of the tracking coefficients.

One important thing is the different ranges in which the tracking coefficients can appear. Most indices (Cohen's κ, Kendall's W, and Foulkes and Davis's γ) range between 0.0 and 1.0, but McMahan's tracking index τ as well as simple correlation coefficients can also take negative values.

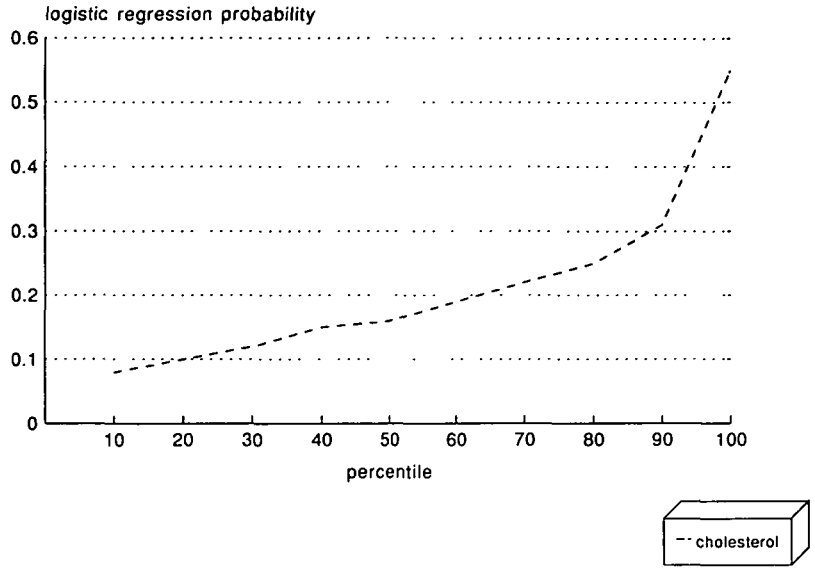

FIGURE 9. Probabilities of different deciles at the initial measurement in 1977 to reach values above P80 at the follow-up measurement in 1991, calculated with logistic regression for serum cholesterol.

Another example is the tracking index TI proposed by Nishio et al. (26). TI, calculated if the population is divided into quintiles, can take values between 1.0 and 8.3. However, if the population is divided into quartiles, $T(h)$ takes the value of 0.5 and the index can range between 1.0 and 4.0. In fact, all the coefficients should be rescaled so that when applied to random numbers the tracking coefficient equals zero.

Hibbert et al. (18) refer to the possibility of testing the significance of their tracking coefficient, Kendall's W, as an important feature in interpreting the tracking coefficient. The importance of a significance test for the tracking coefficient is, however, doubtful. This is due to the fact that the magnitude of every tracking coefficient very much depends on the length of the interperiod. A significant tracking coefficient calculated over an interperiod of one year does not have to be "better" than a nonsignificant tracking coefficient calculated over a much longer interperiod.

Michels et al. (49) demonstrated, by calculating 95 percent confidence intervals, that a significant Pearson's ρ does not imply a high level of predictability. They described the longitudinal development of systolic blood pressure and calculated on the basis of a significant correlation coefficient a 95 percent confidence interval with a width of 44.6 mmHg, which is very large.

The same interpretation problem arises when authors want to evaluate their tracking coefficient by saying that if their tracking coefficient is above a certain value the population tracks for that variable. A Foulkes and Davis tracking index of $\gamma > 0.5$ indicates tracking (30) but, if the index is calculated over a very short interperiod, the conclusion that a population tracks does not indicate anything.

The problem with the interpretation of coefficients was also noticed by Rogossa and Willett (50), who calculated both Foulkes and Davis's tracking coefficient γ and McMahan's tracking coefficient τ for the same data. The indices not only revealed different results, but they led to different conclusions. The McMahan index τ indicated a tracking variable, while the Foulkes and Davis coefficient γ did not.

In our example, however, both indices reveal tracking for serum cholesterol as well as for body height, although the absolute values for McMahan's index τ are higher than the values for Foulkes and

Davis's index γ. The fact that the values for τ are higher than γ is, however, not surprising. For instance, if two longitudinal curves cross at the end or at the beginning of a time interval, γ will indicate no tracking, while τ will indicate tracking, because the relative deviation of the population mean will be unchanged over a considerable portion of the time interval.

Statistical modeling

Which model to use to describe tracking is a very important question. Choosing a priori a straight line model is perhaps too simple, because a lot of longitudinal data are not fully described by such linear models. Of course, a simpler model is preferable to a more complex one, but only if the simple model provides a precise description of the data (36). If not, a more complex model has to be used.

On the other hand, it is striking that tracking models almost never contain confounding variables although they can highly influence the magnitude of the tracking coefficients or the predictability of future values by early measurements. Lauer et al. (5) showed, for instance, a relation between the longitudinal development of blood pressure and the longitudinal development of such anthropometric measures as body weight and body height. Calculation of a tracking index based only on the longitudinal development of blood pressure will underestimate the "real" tracking coefficient of that variable. Armstrong et al. (51) pointed out that levels of serum cholesterol and blood pressure are influenced by sexual maturation. Therefore, when interested in tracking of serum cholesterol and blood pressure, a variable indicating sexual maturation has to be added to the tracking model, because, otherwise, the tracking coefficient and predictability of early measurements for future values will be underestimated. So, in modeling tracking, one has to account for certain confounders. Which variables act as confounders depend on the variable for which tracking is described and on the questions to be answered by the tracking analysis.

Another problem that sometimes arises in longitudinal studies is a change of behavior of the subjects because of their knowledge of results. For instance, if a high blood pressure is measured at a certain age, there is a good possibility that subjects will change their behavior in relation to their knowledge of this high blood pressure (e.g., they will change their dietary practices, change their physical activity patterns, take drugs, etc.). This selective change of behavior can influence the tracking coefficients as well as the predictability. To illustrate this, the female subjects of the Amsterdam Growth and Health Study (the population used in the example) were divided into a moderate/high risk group and a low risk group according to their serum cholesterol levels at the measurement in 1985 at age 21 years. The moderate/high risk group had serum cholesterol levels above 5.2 mmol/liter and the low risk group had levels under 5.2 mmol/liter. The moderate risk criterion value of 5.2 mmol/liter was chosen according to the Report of the National Cholesterol Education Program (52).

For both groups, we analyzed the percent change in absolute fat intake, fat intake as percent of energy intake, cholesterol intake, and sum of skinfolds between the measurements in 1985 and the measurements in 1991 (at age 27 years). The differences between the two groups are shown in figure 10. The moderate/high risk group decreased significantly ($p < 0.05$) their fat intake as a percent of energy intake and their sum of skinfolds. Therefore, in modeling tracking, possible confounding by changes of behavior must be taken into consideration.

Another problem that often occurs in longitudinal studies is a test or learning effect. This means that differences between repeated measurements are only caused by a changing attitude toward the measurement itself. These test/learning effects can influence the prediction of future values. Only if the test/learning effects are different for each individual can they lead to an underestimation of the tracking coefficient.

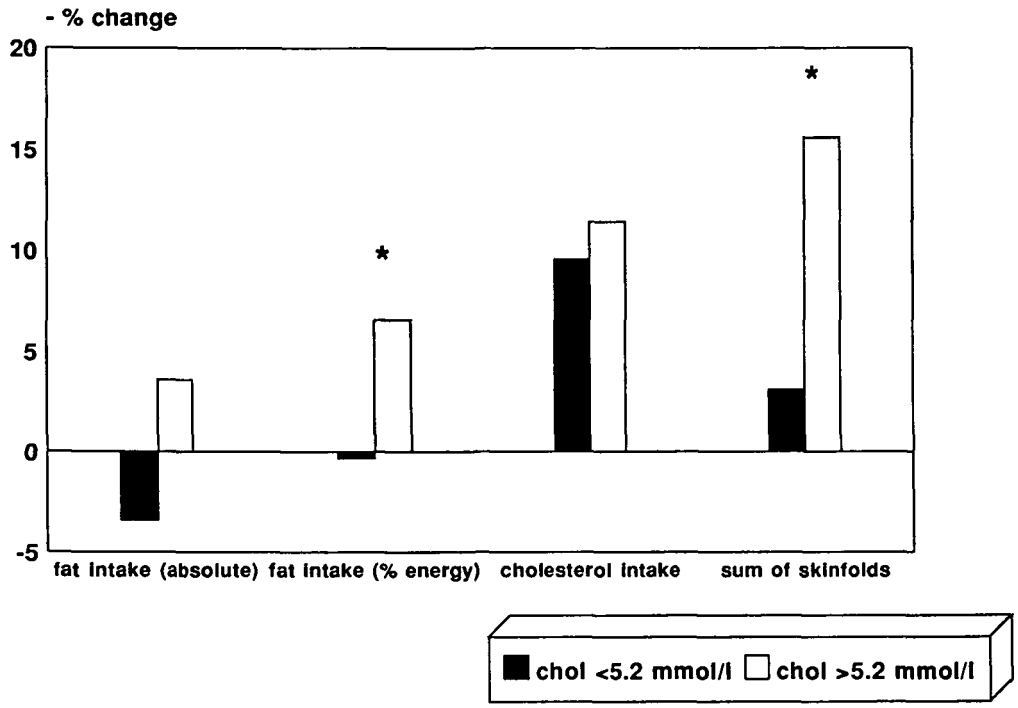

FIGURE 10. Percent change between ages 21 and 27 years (1985 to 1991) in absolute fat intake, fat intake as percent of energy intake, cholesterol intake, and sum of skinfolds for a moderate/high risk group and a low risk group for serum cholesterol according to the measurements at age 21 years. * $p < 0.05$.

Within-person variability

Because tracking models do not consider random measurement error, within-person variability can also lead to underestimation of the tracking coefficient.

Gillman et al. (53) used average values of multiple measurements with a very short time period as the initial value in calculating tracking correlations. They compared these corrected tracking coefficients with coefficients based on the correlation between just two measurements. The corrected correlations were substantially higher. Taking the average of multiple measurements will partially account for the within-person variability.

Rosner and Willett (54) estimated a corrected correlation coefficient that could be interpreted as if an infinite number of measurements were available for each subject. Although the precise description of this estimation goes beyond the scope of this review, the within-person variability has to be taken into account by interpreting the tracking analysis.

Risk factors for chronic diseases

One of the major issues involved in tracking analysis is the longitudinal development of risk factors for chronic diseases. Before interpreting the tracking coefficient or predictive value, one has to be aware of the fact that the maintenance of a relatively high value for a risk factor through time may not be as important in the development of a disease as a certain increase in this value. That is probably the reason why some authors (5, 8) calculated not only the percent of subjects who maintained a certain rank order, but also the percent of subjects with rising rank orders.

Another point related to this is that although a subject has a high value of a certain risk factor in relation to the other subjects of the observed population it does not mean that the absolute value of the risk

factor is high. Thus, the calculation of the predictability of an early measurement for developing a certain absolute high risk value, for instance, as supposed by Beckett et al. (35) and Guo et al. (36), gives perhaps more information than the calculation of the predictability in a relative sense.

CONCLUSIONS

We do not intend here to provide a "perfect" tracking coefficient or a "perfect" model to describe tracking. We wish merely to draw attention to some important points about tracking analysis.

On the one hand, as explained earlier, it is preferable to use a tracking coefficient that is as simple as possible. In that sense, a parametric coefficient is less suitable than a nonparametric coefficient, because the latter makes no assumptions about the distribution of values. Furthermore, if the data are multivariate normally distributed, the parametric coefficients are comparable with the nonparametric coefficients.

However, on the other hand, tracking is part of the description of the longitudinal development of certain variables and the possible prediction of future values by early measurements. The latter is especially important for epidemiologic purposes, because, if some degree of tracking is observed, it can lead to early detection of risk factors related to chronic diseases and therefore to early treatment.

Both topics, describing the longitudinal development and quantifying the predictability follow naturally from each other. Because of this, it is recommended to use a statistical approach that can deal with both questions, so that the results can be related to each other. Besides that, the approach has to have the possibility to use all the available longitudinal data and to control for "a selective change of behavior" and other confounding variables. Therefore, a regression modeling technique, which can be used for the calculation of a tracking coefficient as well as a predictive value, seems the most appropriate.

Early detection of risk factors for chronic diseases makes it possible to develop preventive strategies. Therefore, it is worthwhile to assess the influence of certain factors on the longitudinal development of the risk factor under consideration. Consequently, the choice of a certain regression model has to depend on the possibility to quantify these influences.

Classic regression analysis, however, is not suitable for longitudinal data, because the different observations for each individual are not independent, which is the assumption in classic regression analysis. This nonindependence of observations for a given individual can be characterized by choosing a priori a certain correlation structure for the longitudinal data. The method used by Beckett et al. (35) and Guo et al. (36) deals with this problem by allowing the longitudinal observations to be correlated. The only assumption in this method is the multivariate normality of the observations. So, this method is only suitable for continuous variables. In epidemiology, however, one is often interested in the development of a certain "high risk" group and not in the total spectrum of values. In that situation, the population is divided into two groups to form a dichotomous variable. A solution for this problem is given by Zeger and Liang (55) who developed a longitudinal data-analyzing technique (generalized estimating equations (GEE)), which is suitable for longitudinal analysis for both discrete and continuous outcomes. One has to be very careful in defining the "high risk" group, because when this is based on an arbitrary decision, the choice highly influences the magnitude of the tracking coefficient or predictive value. Thus, when a dichotomization is made, it has to be based on an actual objective "risk value" rather than on an arbitrary decision based on the sample distribution.

In choosing a tracking coefficient, the following points should be kept in mind:

1. One should take as many time points as feasible, as spread out over the time period as possible.
2. One should use a measurement as sensitive and reliable as possible, with uncorrelated measurement errors.

3. The coefficient has to be interpreted easily. A coefficient whose values can range between 0.0 for no tracking and 1.0 for perfect tracking seems to be the most appropriate. A coefficient which does not equal zero when applied to random numbers or is not 1 by perfect tracking should be rescaled.
4. In most studies, only a point estimate of a tracking coefficient is given. Conclusions based on a point estimate are questionable because no information about the reliability is taken into account. Our suggestion is to calculate a 95 percent confidence interval around the tracking coefficient, so conclusions can be made more thoroughly.
5. It is very dangerous to provide strict rules for the interpretation of tracking coefficients, because the value of a coefficient is very dependent on the period under consideration. Probably the best strategy is to give some advice about the interpretation of the value of the coefficient and let the reader decide what to do with the results.

REFERENCES

1. Clarke WR, Schrott HG, Leaverton PE, et al. Tracking of blood lipids and blood pressures in school age children: the Muscatine Study. Circulation 1978;58:626–34.
2. Rosner B, Hennekens CH, Kass EH, et al. Age-specific correlation analysis of longitudinal blood pressure data. Am J Epidemiol 1977;106: 306–13.
3. Lauer RM, Anderson AR, Beaglehole R, et al. Factors related to tracking of blood pressure in children. US National Center for Health Statistics Health Examination Surveys Cycles II and III. Hypertension 1984;6:307–14.
4. André JL, Deschamps JP, Petit JC, et al. Change of blood pressure over five years in childhood and adolescence. Clin Exp Hypertens A 1986;8: 539–45.
5. Lauer RM, Mahoney LT, Clarke WR. Tracking of blood pressure during childhood: the Muscatine Study. Clin Exp Hypertens A 1986;8: 515–37.
6. Palti H, Gofin R, Adler B, et al. Tracking of blood pressure over an eight year period in Jerusalem school children. J Clin Epidemiol 1988;41:731–5.
7. Freedman DS, Shear CL, Srinivasan SR, et al. Tracking of serum lipids and lipoproteins in children over an 8-year period: the Bogalusa Heart Study. Prev Med 1985;14:203–16.
8. Stuhldreher W, Donahue R, Drash A, et al. The Beaver County Lipid Study: sixteen-year cholesterol tracking. Ann NY Acad Sci 1991;623: 466–8.
9. Baumgartner RN, Guo S, Roche AF. Tracking of lipids and lipoproteins in adolescents from 12 to 22 years of age: the Fells Longitudinal Study. Ann NY Acad Sci 1991;623:406–9.
10. Rolland-Cachera MF, Deheeger M, Guilloud-Bataille M, et al. Tracking the development of adiposity from one month of age to adulthood. Ann Hum Biol 1987;14:219–29.
11. Cronk CE, Roche AF, Kent R Jr, et al. Longitudinal trends in subcutaneous fat thickness during adolescence. Am J Phys Anthropol 1983;61: 197–204.
12. Baumgartner RN, Roche AF. Tracking of fat pattern indices in childhood: the Melbourne Growth Study. Hum Biol 1988;60:549–67.
13. Casey VA, Dwyer JT, Coleman KA, et al. Body mass index from childhood to middle age: a 50-y follow-up. Am J Clin Nutr 1992;56:14–18.
14. Stein AD, Shea S, Basch CE, et al. Variability and tracking of nutrient intakes of preschool children based on multiple administrations of the 24-hour dietary recall. Am J Epidemiol 1991; 134:1427–37.
15. Garry PJ, Hunt WC, Koehler KM, et al. Longitudinal study of dietary intakes and plasma lipids in healthy elderly men and women. Am J Clin Nutr 1992;55:682–8.
16. Lee IM, Paffenbarger RS Jr, Hsieh CC. Time trends in physical activity among college alumni, 1962–1988. Am J Epidemiol 1992;135:915–25.
17. Beunen G, Lefevre J, Claessens AL, et al. Age-specific correlation analysis of longitudinal physical fitness levels in men. Eur J Appl Physiol 1992;64:538–45.
18. Hibbert ME, Hudson IL, Lanigan A, et al. Tracking of lung function in healthy children and adolescents. Pediatr Pulmonol 1990;8:172–7.
19. Mellies MJ, Laskarzewski PM, Tracy T, et al. Tracking of high- and low-density-lipoprotein cholesterol from childhood to young adulthood in a single large kindred with familial hypercholesterolemia. Metabolism 1985;34:747–53.
20. Porkka KVK, Viikari JSA, Akerblom HK. Tracking of serum HDL-cholesterol and other lipids in children and adolescents: the Cardiovascular Risk in Young Finns Study. Prev Med 1991;20:713–24.
21. Harsha DW, Smoak GS, Nicklas TA, et al. Cardiovascular risk factors from birth to 7 years of age: the Bogalusa Heart Study. Tracking of body composition variables. Pediatrics 1987;80: 779–83.
22. Freedman DS, Byers T, Sell K, et al. Tracking of serum cholesterol levels in a multiracial sample of preschool children. Pediatrics 1992;90:80–6.
23. Freedman DS, Srinivasan SR, Cresanta JL, et al. Cardiovascular risk factors from birth to 7 years of age: the Bogalusa Heart Study. Serum lipids and lipoproteins. Pediatrics 1987;80:789–96.
24. Harada Y, Mori C, Watanabe K, et al. Left ventricular diastolic function of children in high blood pressure tracking group. Acta Paediatr Jpn

1990;32:530–9.
25. Kemper HCG, Snel J, Verschuur R, et al. Tracking of health and risk indicators of cardiovascular diseases from teenager to adult: Amsterdam *Growth and Health Study*. *Prev Med* 1990;19: 642–55.
26. Nishio T, Mori C, Haneda N, et al. Quantification of blood pressure tracking of children by tracking index: the Shimane Heart Study. Jpn Circ J 1987;51:1404–8.
27. Schneiderman ED, Kowalski CJ, Ten Have TR. A GAUSS program for computing an index of tracking from longitudinal observations. Am J Hum Biol 1990;2:475–90.
28. Schneiderman ED, Kowalski CJ. Ten Have TR, et al. Computation of Foulkes and Davis nonparametric tracking index using GAUSS. Am J Hum Biol 1992;4:417–20.
29. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968;70:213–20.
30. Fleiss JL. Statistical methods for rates and proportions. New York, NY: John Wiley and Sons, 1973.
31. Voors AW, Webber LS, Berenson GS. Time course studies of blood pressure in children—the Bogalusa Heart Study. Am J Epidemiol 1979; 109:320–34.
32. Ibsen KK, Rotne H, Hougaard P. Tracking of blood pressures in diabetic children. Acta Paediatr Scand 1983;72:495–8.
33. Török E, Gyárfás I, Csukás M. Prediction of stable high blood pressure in adolescents. Clin Exper Hypertens A 1986;8:547–55.
34. Aleksandrov AA, Shamarin VM, Petrosian KY, et al. Development of main risk factors of ischaemic heart disease in children and adolescents (eight-year prospective study). Cor Vasa 1987;29:412–20.
35. Beckett LA, Rosner B, Roche AF, et al. Serial changes in blood pressure from adolescence into adulthood. Am J Epidemiol 1992;135:1166–77.
36. Guo S, Beckett L, Chumlea WC, et al. Serial analysis of plasma lipids and lipoproteins from individuals 9–21 y of age. Am J Clin Nutr 1993; 58:61–7.
37. Jennrich RI, Schluchter MD. Unbalanced repeated-measures models with structured covariance matrices. Biometrics 1986;42:805–20.
38. Schluchter MD. Unbalanced repeated measures models with structured covariance matrices. In: Dixon WJ, Brown MB, Engelman L, et al, eds. BMDP statistical software manual. Vol 2. Berkeley, CA: University of California Press, 1990:1207–44.
39. Foulkes MA, Davis CE. An index of tracking for longitudinal data. Biometrics 1981;37:439–46.
40. Goldstein H. Measuring the stability of individual growth patterns. Ann Hum Biol 1981;8: 549–57.
41. Siegel S. Nonparametric statistics for the behavioral sciences. New York, NY: McGraw-Hill, 1956.
42. McMahan CA. An index of tracking. Biometrics 1981;37:447–55.
43. Matthews KA, Woodall KL, Stoney CM. Changes in and stability of cardiovascular responses to behavioral stress: results from a four-year longitudinal study of children. Child Dev 1990;61:1134–44.
44. Darby SC, Fearn T. The Chatham Blood Pressure Study: an application of Bayesian growth curve models to a longitudinal study of blood pressure in children. Int J Epidemiol 1979;8: 15–21.
45. Ware JH, Wu MC. Tracking: prediction of future values from serial measurements. Biometrics 1981;37:427–37.
46. Berkey CS, Laird NM, Valadian I, et al. Modelling adolescent blood pressure patterns and their prediction of adult pressures. Biometrics 1991; 47:1005–18.
47. Kemper HCG, ed. Growth, health, and fitness of teenagers: longitudinal research in international perspective. Basel, Switzerland: Karger, 1985.
48. Lauer RM, Clarke WR. Childhood risk factors for high adult blood pressure: the Muscatine Study. *Pediatrics* 1984;84:633–41.
49. Michels VV, Bergstrahl EJ, Hoverman VR, et al. Tracking and prediction of blood pressure in children. Mayo Clin Proc 1987;62:875–81.
50. Rogossa DR, Willett JB. Comparing two indices of tracking. Biometrics 1983;39:795–6.
51. Armstrong N, Balding J, Gentle P, et al. Serum lipids and blood pressure in relation to age and sexual maturity. Ann Hum Biol 1992;19: 477–87.
52. Report of the National Cholesterol Education Program Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Arch Intern Med 1988;148:36–69.
53. Gillman MW, Cook NR, Rosner B, et al. Childhood blood pressure tracking correlations corrected for within-person variability. Stat Med 1992;11:1187–94.
54. Rosner B, Willett WC. Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. Am J Epidemiol 1988; 127:377–86.
55. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986;42:121–30.