Models for protein deficiency

P. V. Sukhatme,[1] *Ph.D., D.Sc. and Sheldon Margen,*[2] *M.D.*

ABSTRACT Analysis of existing N balance data in adult man has shown that when body weight is constant protein requirements are regulated, much like energy. Data relating to daily N balance on fixed intakes have been examined for the nature and degree of intra-individual variation. It has been shown that for intakes in the range of 3.5 to 12 g N/day, the day-to-day fluctuations in N balance are not random but are serially correlated in an autoregression process. This implies that the daily N balance, like energy balance, is regulated. This regulation is produced by a probabilistic generating mechanism which remains constant through time. At very high or negligible N intakes this regulation is shown to break down, i.e., homeostasis can no longer be maintained. At high levels of protein, metabolism is altered, becoming more rapid than before, body weight increases, and the frequency and amplitude of oscillations become larger and irregular. At low levels of protein, body weight decreases, and the frequency and amplitude of the oscillations increase and decrease, respectively. In either case, the organism is under stress. The interpretation of the autoregression model is that the daily requirements for man in health will be distributed around a constant mean with stationary variance. It has been shown that the magnitude of this variance is comparable with the variation between individuals. The result is found to hold even when the daily requirement is averaged over several days. We conclude that protein deficiency must be defined as a failure of the process to be in statistical control, and not defined in the manner that assumes requirements to be fixed whereby if an individual consumes protein below this level, he suffers from protein deficiency. Based on the autoregression model, a method has been indicated for estimating the incidence of protein deficiency in the population. *Am. J. Clin. Nutr.* 31: 1237-1256, 1978.

During the past number of years two serious questions have plagued nutritionists, planners, and governmental decision makers. There is no doubt that in certain populations what has popularly become named protein-calorie malnutrition (PCM) has a fairly high prevalence and incidence. Shortly after the condition was recognized, its cause was attributed to an inadequate supply of protein in the diet of most individuals suffering from this syndrome—and it was postulated that a worldwide "protein-gap" was upon us or imminent. However, careful analysis, particularly by one of us (PVS) suggested that the etiology of this condition in most of the world was not a lack of protein, but a protein deficiency secondary to inadequate calories—of the food regularly eaten—a view which now more generally prevails. Obviously the implications for policy and implementation of those conflicting views is enormous. The resolution of this question rests in large measure upon the methodology and especially the concepts in interpreting protein "requirements" of individuals and extending these concepts to populations. In fact, a further even more important question should be asked: What do we mean by "protein requirements" and can we model the concept?

Discussion of the relative importance of protein and energy in the etiology of mal-

[1]From the Department of Biometry, Maharashtra Association for Cultivation of Science, Law College Road, Pune 4411004, India.

[2]From the Department of Nutritional Sciences, University of California, Berkeley, California 94720.

nutrition led the FAO/WHO Committee to express doubts about previous methods followed in evaluating adequacy of protein in the diets (1). Specifically, the Committee observed that the customary approach of comparing average intake with average requirement followed in the case of energy cannot be applied in the case of protein since, unlike energy, there was no evidence to suggest that the human possesses a physiological regulatory mechanism which can serve to maintain fixed body weight and N-balance over extended periods. Consequently, if an individual ate at the level of "average requirement" representing the requirements of the reference individual of his age-sex group, as he would aim to do in the case of energy (adjusted for energy expenditure or mean protein requirements) he would risk being protein malnourished since the average requirements would cover only half the individuals. The implication is that individuals vary in their true needs and that these "genetically" based needs are essentially fixed needs specific to each individual. The Committee felt that an individual should, therefore, aim to eat at the upper range of distribution of requirement. If each individual were to eat at the upper range of requirements, the needs of almost all individuals would be covered and the risk to an individual of developing protein deficiency would be negligible. By common agreement, this higher level has come to be defined as the average requirement plus twice the standard deviation. Previously known by the name of recommended intake, this level has been renamed as the "safe level" of protein intake. In explaining the meaning of the term, a principle was implied and stated, namely that an individual eating below the safe level, while not necessarily malnourished, runs the risk of developing protein deficiency and that the risk increases as the intake falls below the safe level.

As a result of their deliberations, the Committee recommended that everyone should consume protein above the safe level and workers were encouraged to calculate the rates of prevalence of those at risk, that is those with intakes below the safe level. Thus Lorstad (2) postulated as the model a normal bivariate distribution of requirement and intake and used it to calculate the proportion of the population with intakes below the safe level.

Beaton and Swiss (3) have suggested the use of protein-calorie concentration in the diet instead of protein intake as the variable in the bivariate distribution. This helps the authors to take into consideration the interrelationship between protein and energy intake. Basically, however, their model remains the same as that proposed by Lorstad (2) and assumes that requirements remain constant in an individual.

One of the perplexing problems which we will address in this paper has to do with the observed phenomenon in the most carefully controlled experiments of a large degree of intraindividual variability in requirement, which has been "ignored". It has been ignored because the variability has been assumed to be random noise, due to measurement error. Far from being random, as will be shown later, intraindividual variability is generally at least equal to and at times greater than interindividual variability. Further, as we will show, there is no reason to support the assumption of a "genetically" fixed requirement that remains constant (at least in the adult) as long as weight is constant (i.e., energy balance is maintained).

In expressing the view that requirement is essentially constant in an individual, the FAO/WHO Ad Hoc Expert Committee on Nutrition was mainly guided by the data on measurements of obligatory nitrogen losses of adult men, notably those carried out by Scrimshaw et al. (4) and the Berkeley group (5). However, as will be shown later, a reexamination of these data suggest that the intraindividual variation remains comparable with variation due to individual differences and in fact accounts for a larger part of the total variation, even when requirements are averaged over several days. Evidence on N balance in adults reported by Calloway and Margen (5) confirms this inference. Further, the nitrogen output in subjects fed on constant N levels close to the "average needs" and maintaining body weight, is seen to fluctuate in a manner which suggests that there exists a physiolog-

ical mechanism regulating N balance in man similar to that said to hold for energy balance.

In what follows, we shall present an analysis of N balance data that we believe allows us to describe the features of intraindividual variation in man's output at fixed levels of intake and how the pattern of variation changes with N intake. We shall demonstrate the biological significance of this change and go on to postulate the implications of this method of analysis for defining deficiency and for estimating its incidence in the population. Throughout the paper, it will be assumed that energy intake is not a limiting factor in the diet.

Material used

The data used in this paper are taken partly from studies described in (5) and partly from experiments recently completed at the Department of Nutritional Sciences, University of California, Berkeley. The data relate to daily N balance on fixed intake in adult subjects "maintaining" their body weight. Those previously reported (5) are shown in Figures 1 to 4 and the new series is reported in Figure 5. Specifically, the plan of experiments on subjects 1 and 2 provided for a diet containing 12 g of N from dried egg white for a continuous period of 84 days and that on subjects 3 and 4 provided for 12 g of egg N during the first 12 days, 0.64 g of N (mostly nonprotein) during the next 18 days, 3.50 g of N over the next 36 days, followed by 12 g of N during the last 18 days. The calorie allowance given to these subjects was adjusted in the course of the experiments to ensure that they maintained essentially fixed body weight.

The experimental plan on subject 5 provided for a diet containing 12 g of N from 4 to 48 days and a much higher intake of 36 g of N from days 49 to 82. Unlike in experiments on subjects 1 to 4, the caloric allowance was kept constant at 40 kcal/kg of body weight. The resulting series on body weight is shown in Figures 6 and 7. Altogether, six series of N balance are analyzed in this paper: those covering the entire period of 84

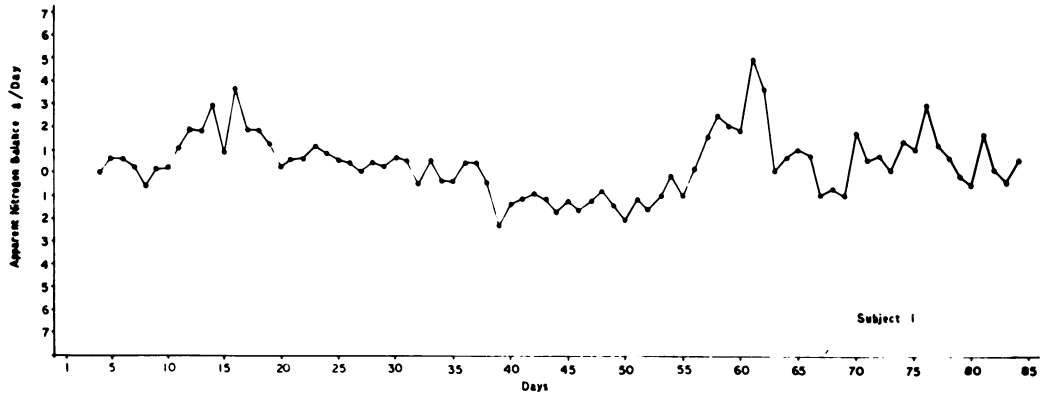


FIG. 1. Apparent nitrogen balance in grams per day: subject 1.

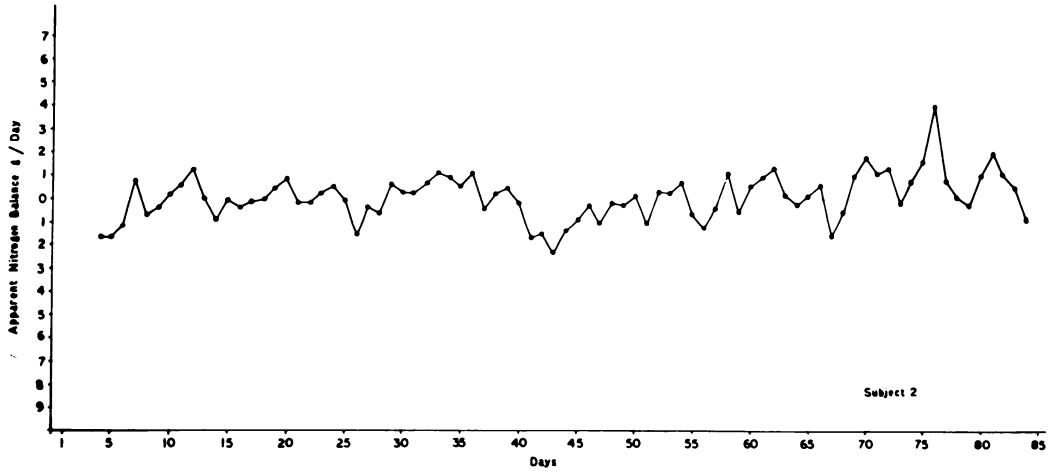


FIG. 2. Apparent nitrogen balance in grams per day: subject 2.

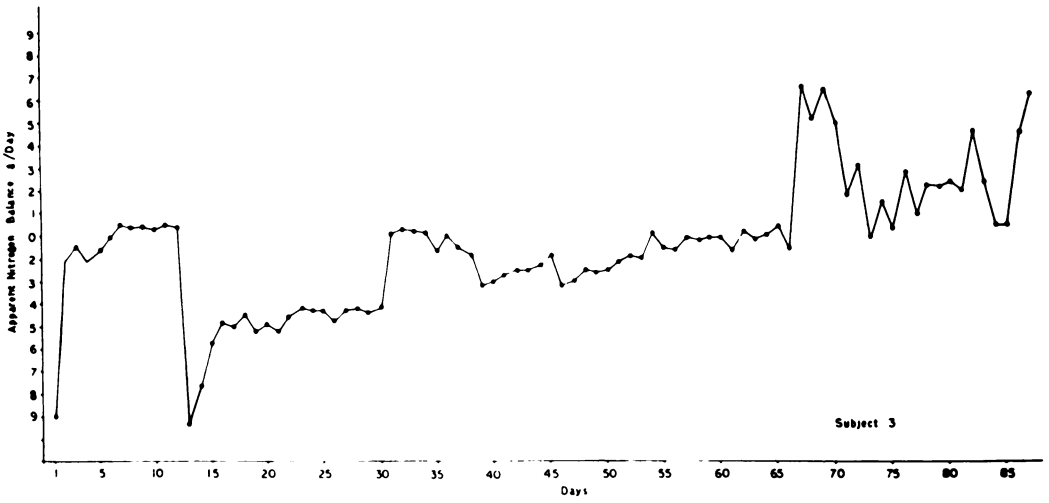


FIG. 3. Apparent nitrogen balance in grams per day: subject 3.

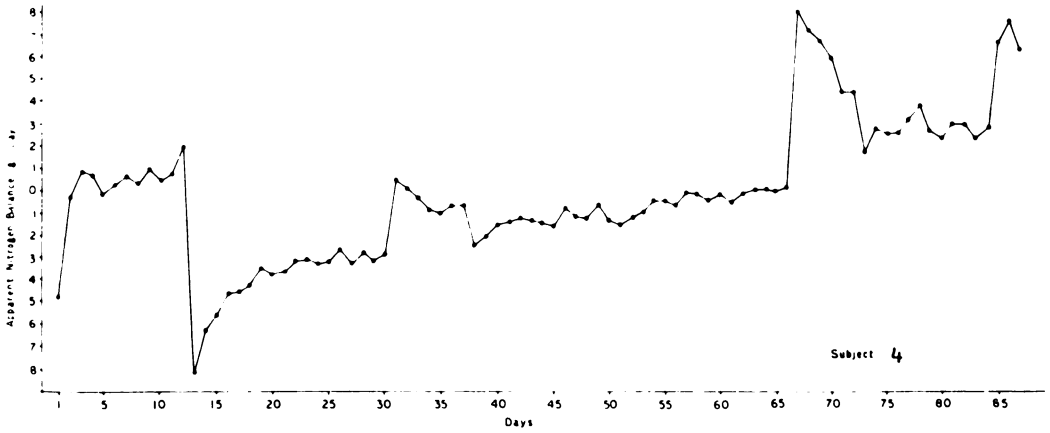


FIG. 4. Apparent nitrogen balance in grams per day: subject 4.

days in subjects 1 and 2, those covering the period from day 31 to 66 in subjects 3 and 4 and those covering the period from day 9 to 48 in the case of subject 5.

Strictly speaking, these series do not represent the daily N balance because while urinary N was determined daily, the fecal N was determined from a pooled average of 3 days. However, the limitation is not considered serious since firstly, the relative contribution of fecal N to total N output is small and secondly, the autocorrelations in the daily N balance appear to be primarily explained by those in the daily urinary N (Table 12).

In any so-called experiment involving quantitation, measurement errors occur. In metabolic studies the chances of error are greater. The studies herein reported probably are among those with the lowest measurement error possible in metabolic studies. However, even so, they are bound to contribute to the inate intraindividual variation. At the same time, as we shall show later, evidence presented in the paper suggests that there exists a regulatory message running all along the daily series of N balance. Separating out the message (pattern) if any, from the underlying errors (noise) is a major problem of communication and calls for the application of what are known as stochastic stationary processes. Errors are called by the term white noise by the engineers. The term stands for a purely random process constituting an uncorrelated sequence of identically distributed observations over time. One purpose of this paper is to illustrate the use of the stochastic processes in separating out the message regulating N balance in man in steady state from the white noise surrounding it. This can be exemplified by a simple biological analogy. The process is similar to the one that allows one to see and analyze the

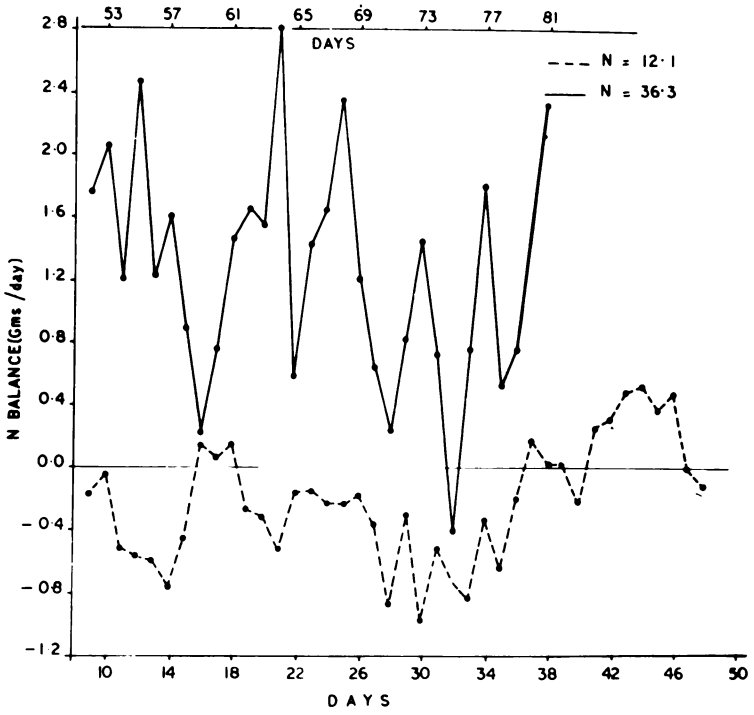


FIG. 5. Apparent nitrogen balance in grams per day: subject 5.

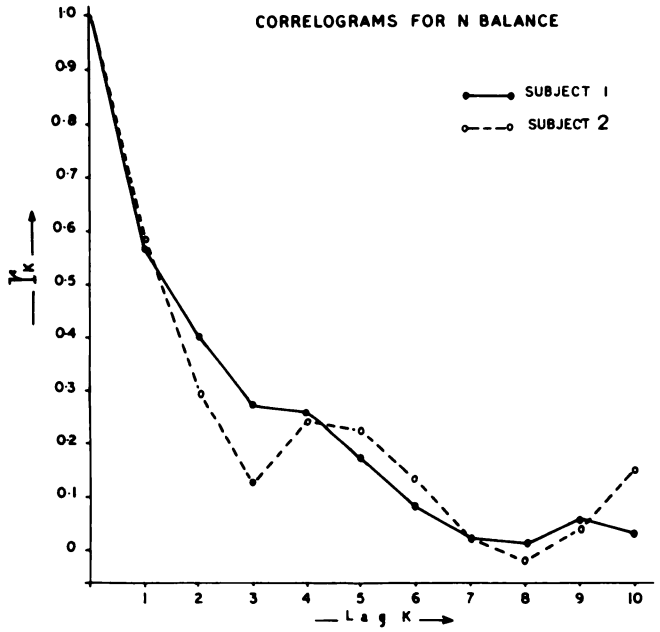


FIG. 6. Correlograms for nitrogen balance: subjects 1 and 2.

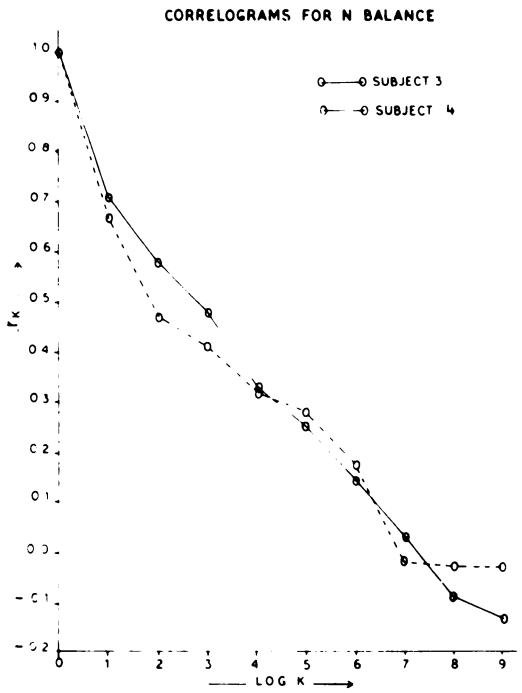


FIG. 7. Correlograms for nitrogen balance: subjects 3 and 4.

evoked potential in an EEG, that is otherwise totally obscured by the totally random electrical "noise" recorded from the entire brain with the EEG.

Daily variation in N balance

Nonrandom character

Statistical techniques for analyzing variation over time range from the straightforward to the very sophisticated. We shall mainly use simple descriptive techniques to bring out the main properties of the data plotted in Figures 1 to 5.

Figures 1 to 5 show that the N intake was fixed for various periods. The N output during these times fluctuated rather considerable from day-to-day. There is no suggestion that the output has changed systematically in a manner indicative of a trend in the mean value. When the number of observations is small, as it is in the present case, running into 84 days in the case of subjects 1 and 2 and only 30 to 40 days in the case of other subjects, a meaningful trend applicable to the age group as a whole cannot be visible. There is, however, a suggestion that the day-to-day fluctuations are not random or independent. If we draw a line through the mean of the daily series, it would appear (with the exception of series on N = 36 for subject 5) that an observation above the mean tends to be followed by one or further observations above the mean and the same tendency appears for observations below the mean.

The best indication of whether the successive observations are correlated is to calculate the correlation coefficients between observations at different distances apart in the same way as we calculate the ordinary correlation coefficient from pairs of observations on two variates and plot the values of the correlation coefficients so calculated against the differences between the observations. The coefficients of correlation so calculated are called auto or serial correlations and are denoted as r_k, the subscript k to symbol r signifying that the correlation is between values k days apart. r_k is often simply called serial correlation of order k or autocorrelation coefficient at lag k. When r_k is plotted against k, the graph is called a correlogram or the autocorrelation function. These correlograms are shown in Figure 6 for the series observed on subjects 1 and 2 and in Figure 7 for the daily data from days 31 to 66 observed on subjects 3 and 4 and in Figure 8 for the daily data shown in Figure 5 for subject 5. A correlogram shows the internal structure of the series and provides the basis for model building.

If the daily N balances were purely random, then for large values of n, i.e., long series, r_k will be approximately equal to 0 with variance approximating 1/n. That is to say, we would find most values of r_k to be within $\pm$ 0.22 for subjects 1 and 2 and within $\pm$ 0.33 for subjects 3 to 5. Instead we find that r_1 has a fairly large value in all the first 5 series on intakes from 3.5 to 12 g of N and is followed by r_2, r_3 or more coefficients, which while significantly greater than 0, tend to get successively smaller almost in an exponentially decaying manner, confirming the non-random character of the time series. The only series where the value of r_1 is not significant is series 6 on subject 5 on level of intake of 36 g N. The possible meaning of this difference will be explained later.

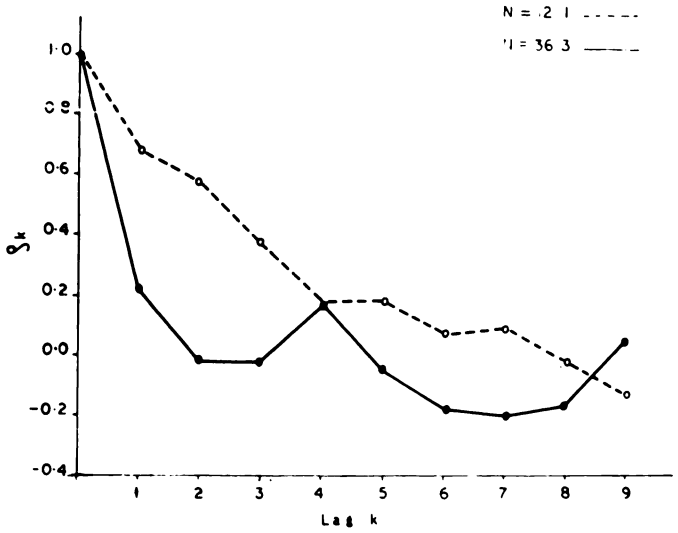


FIG. 8. Correlograms for nitrogen balance: subject 5.

Autoregressive (AR) process

When the correlograms take the form shown in Figures 6 to 8, they indicate that the process of regulating N balance on any day has two components—one arising from the current values of the process at the previous time point t − 1, t − 2, etc. and the other a random term arising from errors of measurement. The falling values of the serial correlation show that the first component is most influenced by the value of the process immediately preceding it and that the influence of the preceding values becomes less and less as the lag increases. We can, therefore, regard the N balance on the t^{th} day, w_t, as regressant on past values w_{t-1}, w_{t-2} with a random residual e_t as in a multiple regression model except that w_t is regressed not on independent variables, but on past values of w_t. Such models are known as the autoregressive models.

The simplest AR model is the first order process given by

$$w_t = \alpha_1 w_{t-1} + e_t \qquad (1)$$

Since w_t and w_{t-1} have the same variance, the expression (*1*) can also be written as

$$w_t = \rho w_{t-1} + e_t \qquad (2)$$

with the serial correlation *r* calculated from the sample providing the best estimator of ρ and hence of α_1. The first term in equation (*2*) represents the change that one expects in N-pool between the t^{th} and $t - 1^{th}$ day and e_t represents the error term distributed around zero with constant variance. The first order AR process is also often called the Markoff process. By successive substitution in equation (*2*), we obtain

$$w_t = e_t + \rho e_{t-1} + \rho^2 e_{t-2} + \cdots \qquad (3)$$

This is an important expression. It shows that errors under AR process get incorporated into the motion of the process to determine the balance on any given day and are not cancelled out as they would be in a purely random system with $\rho = 0$. Further, it is easy to see from equation (*3*) that the mean value of w_t is 0 i.e.,

$$E(w_t) = 0 \qquad (4)$$

the variance is given by

$$\sigma^2_{w_t} = \sigma^2_e (1 + \rho^2 + \rho^4 + \cdots) = \frac{\delta^2_e}{1 - \rho^2} \qquad (5)$$

σ_e^2 representing the residual variance and the autocorrelation between values separated by k days is given by

$$\rho_k = \frac{E(w_t, w_{t-k})}{E(w_t^2)} = \rho^k \qquad (6)$$

The mean value, the variance and the covariance are all seen to be independent of t,

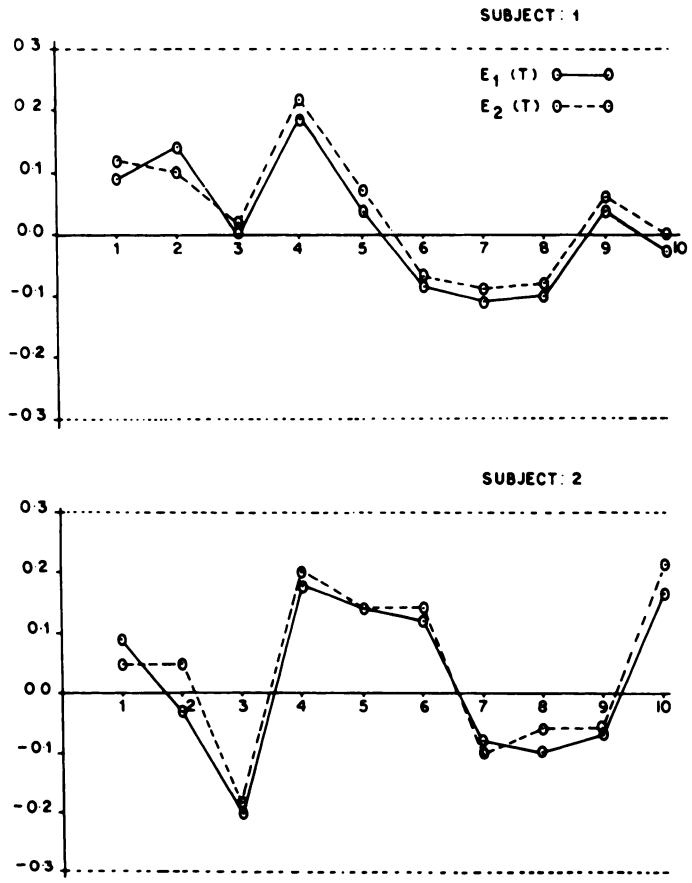


FIG. 9. Correlograms for subjects 1 and 2 after fitting AR series of order 1 and 2, respectively.

showing that the auto-regressive process of the first order is a stochastic stationary process with its correlogram given by the exponential function ρ^k. It follows that if the data should have for its correlogram a function decaying as quickly as ρ^k, which appears to be the case from inspection of Figures 6 to 8, this may be interpreted to indicate that the generating mechanism of the series albeit of a probabilistic kind, can be represented by AR process of the first order remaining constant through time.[3]

Theoretical autocorrelation functions have been worked out for most of the stochastic stationary low order AR, moving average, and mixed models. If the observed autocorrelation function is close to any of the known functions, it means that we can guess the process that is likely to provide a suitable model for the series under consideration. Further discussion of significance and interpretations of these mathematical

[3] The properties of AR process of higher order are similarly studied. Thus we may write

$$w_t = \alpha_1 w_{t-1} + \alpha_2 w_{t-2} + e_t \qquad (7)$$

with estimates of α_1 and α_2 given by

$$\hat{\alpha}_1 = \frac{r_1(1 - r_2)}{1 - r_1^2} \text{ and } \hat{\alpha}_2 = \frac{r_2 - r_1^2}{1 - r_1^2} \qquad (8)$$

and that of the residual variance given by

$$\sigma_e^2 = \sigma_w^2 \{1 - r_1\hat{\alpha}_1 - r_2\hat{\alpha}_2\} \qquad (9)$$

α_1 in equation (7) represents the partial regression coefficient of w_t on w_{t-1} when w_{t-2} is kept constant and α_2 represents the partial regression coefficient of w_t on w_{t-2} when w_{t-1} is kept constant. If the generating mechanism of the series can be satisfactorily represented by the autoregression process of first order, then clearly α_2 will be negligible relative to its standard error implying that we can depend upon w_{t-1} to act on behalf of w_{t-2}.

treatments is beyond the needs and scope of the paper.

Residual analysis

How do we know that AR process of the first order really does provide an adequate description of the data? The usual way of checking this is to examine the correlograms of residuals, which are differences between the observed N balances and the fitted (predicted) values. With a first order AR model in equation (*2*), the fitted value is clearly $r\ w_{t-1}$ and the residual error is given by

$$e_t = w_t - r\ w_{t-1} \qquad (10)$$

If an AR model of the first order does in fact give a satisfactory fit, then the correlograms of the residuals will be such that r_k is approximately normally distributed with mean 0 and variance 1/n. For reasonably large values of n, we may, therefore, expect most of the r_k values for residuals to lie between ±0.22 for subjects 1 and 2 and between ±0.33 for subjects 3 and 5. The confidence limits may be somewhat narrower for correlations at lag 1 and 2, but there is no need to enter into these refinements here.

Figures 9 and 10 show the correlograms of the residual errors for series 1 to 4. It will be seen that all the *r* values lie within the confidence limits quoted above. The correlograms can, therefore, be taken to confirm the expectation that r_k is approximately normally distributed around 0 with variance 1/n. We may accordingly conclude that an AR model of order one can be taken to adequately describe the data on daily N balance. The examination of the correlograms for the residuals obtained on fitting the AR process of the second order further reinforces this conclusion.

Another way of checking will be to fit AR series of successively higher orders and calculate the residual variance. By way of example, we have presented these calculations in Table 1 for subject 5 when fed an

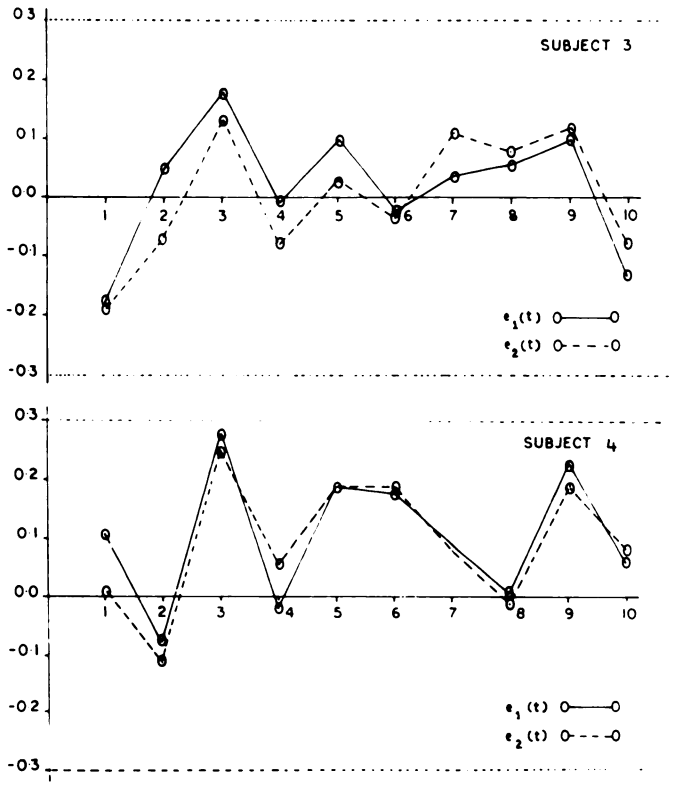


FIG. 10. Correlograms for subjects 3 and 4 after fitting AR series of order 1 and 2, respectively.

TABLE 1
Residual error obtained on fitting AR process of the first and second order to the observed N balance (in grams) in subject 5 on fixed intake of N equal to 12 g/day

t	w_t	$e_1(t)$	$e_2(t)$
1	−0.17		
2	−0.03	0.09	
3	−0.51	−0.49	−0.45
4	−0.56	−0.21	−0.28
5	−0.59	−0.21	−0.18
6	−0.75	−0.35	−0.31
7	−0.45	0.06	0.08
8	0.16	0.47	0.56
9	0.08	−0.03	0.10
10	0.16	0.11	0.08
11	−0.27	−0.38	−0.37
12	−0.32	−0.14	−0.22
13	−0.50	−0.28	−0.27
14	−0.14	0.20	0.19
15	−0.13	−0.03	0.06
16	−0.22	−0.13	−0.12
17	−0.23	−0.08	−0.09
18	−0.17	−0.01	0.00
19	−0.37	−0.25	−0.23
20	−0.87	−0.62	−0.64
21	−0.27	0.32	0.27
22	−0.96	−0.78	−0.62
23	−0.51	0.14	0.05
24	−0.73	−0.38	−0.18
25	−0.81	−0.31	−0.31
26	−0.32	0.23	0.27
27	−0.64	−0.42	−0.29
28	−0.19	0.25	0.21
29	0.19	0.32	0.44
30	0.05	−0.08	−0.01
31	0.04	0.01	−0.03
32	−0.22	−0.25	−0.25
33	0.28	0.43	0.40
34	0.32	0.13	0.22
35	0.51	0.29	0.29
36	0.54	0.19	0.20
37	0.39	0.02	0.00
38	0.49	0.22	0.17
39	0.01	−0.32	−0.33
40	−0.10	−0.11	−0.22
Mean value (g)	−0.195	−0.059	−0.048
Variance	0.149	0.083	0.081

intake of 12 g N. It will be seen that the greatest reduction in variance occurs after the AR series of the first order is fitted. Thereafter, there is hardly any reduction in the variance, the residual variance σ_e^2 remaining almost constant. The calculations for other subjects are summarized in Table 2. They show that w_t may be regarded as distributed around a constant mean with stationary variance as shown in equation (5).

Biological significance

The significance of the results in the previous sections is that N balance, like energy balance, appears to be regulated in man maintaining body weight. Not only is the autocorrelation of the first order not zero, it is large and positive when N intake is fixed within the range from 3.5 to 12 g of N. There is no published data on long-term energy balance to test the "exact" pattern of regulation from day to day in man maintaining body weight, but the autocorrelation on daily body weight shown in Figures 6 and 7 can serve to give an indication. These are shown for subject 5 in Table 3 along with the autocorrelation values for N balance and those for N balance expressed on per kilogram of body weight basis. It will be seen that when body weight is essentially constant, the daily weight is regulated in much the same manner as the daily N balance. In other words, when the body weight tends to remain constant, as we found in the case of subject 5 (see Fig. 11), it can be considered to be distributed in a stationary distribution with a constant variance in much the same manner as his daily N balance. It would therefore appear that there is no basis to adopt a different approach on

TABLE 2
Values of serial correlation and residual variance in daily apparent N balance (g/day)

	Values of r					Residual variance			
Lag	Subject 1	Subject 2	Subject 3	Subject 4	AR order	Subject 1	Subject 2	Subject 3	Subject 4
0	1.00	1.00	1.00	1.00	0	2.73	2.13	0.58	0.45
1	0.57	0.45	0.75	0.69	1	1.85	1.40	0.26	0.24
2	0.40	0.19	0.57	0.49	2	1.82	1.40	0.26	0.24
3	0.27	0.04	0.53	0.36	3	1.82	1.39	0.25	0.24

TABLE 3
Values of serial correlation for daily body weight (kg), daily apparent N balance (g/day), and daily apparent N balance (g/kg/day) for subject 5

Lag	N = 12			N = 36		
	NB	NB/kg	BW	NB	NB/kg	BW[a]
1	0.678	0.681	0.472	0.213	0.216	0.906
2	0.581	0.616	0.355	−0.021	−0.012	0.839
3	0.368	0.416	0.354	−0.022	−0.017	0.734
4	0.181	0.216	0.173	0.171	0.170	0.668
5	0.180	0.216	0.314	−0.050	−0.046	0.571
6	0.068	0.011	0.242	−0.180	−0.176	0.521
7	0.091	0.001	0.222	−0.202	−0.199	0.432
8	−0.024	−0.094	0.158	−0.174	−0.171	0.347
9	−0.138	−0.229	0.103	−0.051	0.054	0.247

[a] The high values are in part a reflection of the upward trend in body weight.

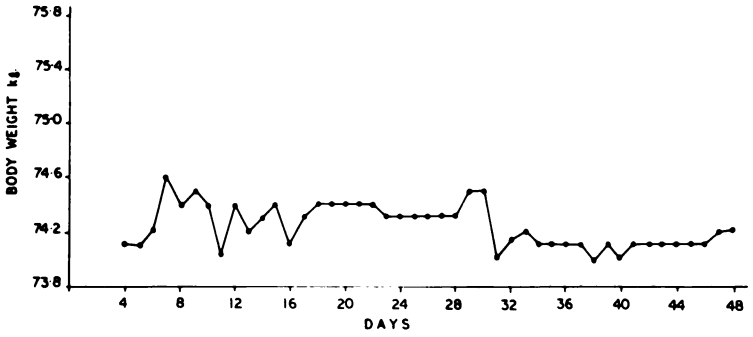


FIG. 11. Body weight, days 4 to 48: subject 5.

evaluating adequacy of protein in the diet from that followed in the case of energy.

There is a suggestion in our data that as N intake increases, the autocorrelation r decreases (see Table 2) but the number of our subjects is too small to bring out the significance of this change. However, our subject 5 leaves little doubt that when N intake is very high the body protein metabolism responds to alteration by becoming more rapid. The probable explanation is that when N intake is very high in excess of needs, the body tries to get rid of the excess N as quickly as possible by converting N containing amino acid molecules into harmless products like urea. We have also noted that when a man is on a very high N intake, the body weight increases (Fig. 12). In fact we find this to be the case in all our other experimental subjects who were given 36 g of N. By way of example, we show in Figure 13 results for a different subject, whose data was not included in the analysis. This figure clearly brings out the difference in the frequency of oscillation (and hence in the rate of N metabolism) when the intake is increased from 12 to 36 g of N. It was also observed that the body weight of this subject, as that of subject 5 increased when the intake was increased. We are not able to explain the reason for the gain in body weight. But it does tell us that the relatively large value for the average positive balance that we notice in subject 5 on 36 g N cannot be dismissed as measurement error and in fact is consistent with the observed gain in body weight. We plan to report on these results at a later date.

In a sense, the above inference flows from the exponential nature of the autocorrelation function given by $r_k = r^k$. The phenomenon represented by it can be interpreted to mean that as r decreases, that is as the power regulating N balance decreases, the

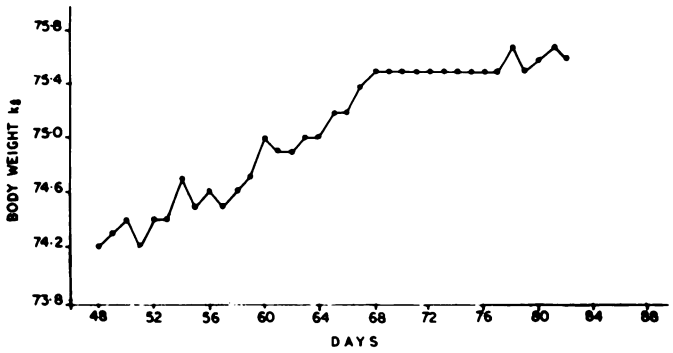


FIG. 12. Body weight, days 48 to 84: subject 5.

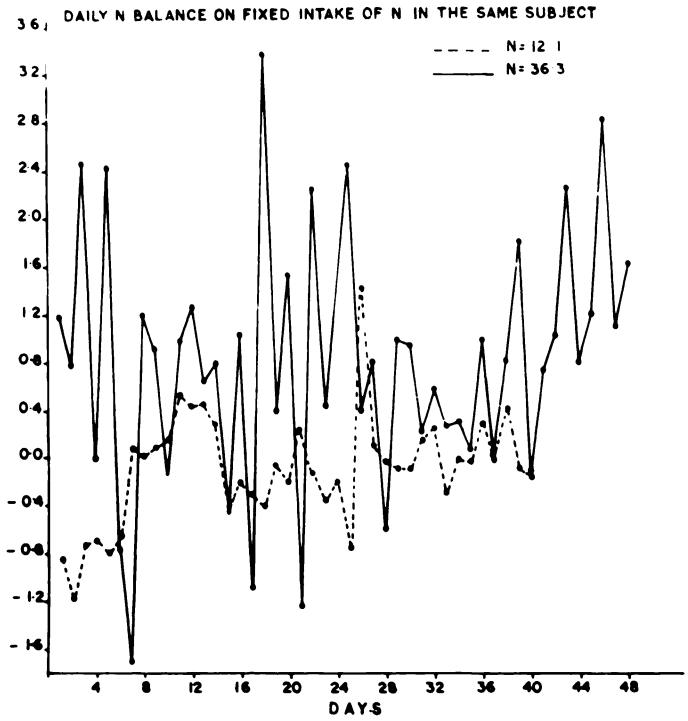
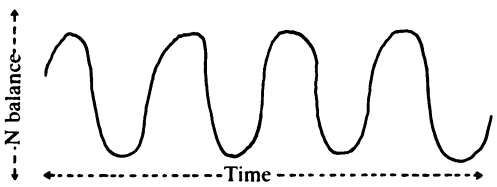


FIG. 13. Daily nitrogen balance at two levels of nitrogen intake in one adult subject.

tendency for the daily N balance to oscillate increases. This is well brought out in Figures 5 and 8 examined together. It will be seen that the traces in the series shown in the dotted line (Fig. 5, N = 12) and for which $r = 0.7$ are more smooth than those when r is close to zero (as seen when N = 36). The frequency of oscillations per unit of time is thus seen to increase as r decreases towards r_0. Apparently the subject would seem to be under "stress" when he is on high intake and that this stress increases as the power to regulate balance decreases. When r is equal to zero and the regulatory mechanism ceases to exist, this stress, measured in terms of the oscillations per day will equal 2/3 on average (2 oscillations every 3 days). When r becomes negative, the stress increases further. In other words deviations in either direction from r_0 indicate increasing stress. It may, therefore, be concluded that when the regulatory mechanism ceases to

exist, a man reaches a limit as it were in his capacity to maintain homeostasis and hence N balance. The precise value of the N intake when this limit is reached will of course vary from individual to individual, depending upon his ability to adapt requirement to high intake. Outside this range, an individual is unlikely to preserve his body mass and "maintain" N balance.

One can, therefore, visualize a model for defining nitrogen requirements for an individual in terms of well used biological concepts. We have shown that protein requirements of man are regulated. We do not (as with most biological processes) understand the causes or nature of the regulation. The fact that we are dealing with regulation obviously implies that man also has a regulating system, involving some kind of N or protein "stat" which operates within certain limits which can only be crudely defined at the present time. However, within limits that we call homeostasis there is a pattern of regulation that can be described—namely a series of oscillations which when measurement errors are eliminated becomes clear. These oscillations have a band of frequencies and amplitudes which although varying some with protein intake and protein composition (and generally being of lesser magnitude when protein intake is close to the "requirements" of the individual at the time of observation) still are always present. This can be illustrated in Scheme 1 where the regulated state indicates that the individual is in homeostasis with regard to N intake. However, it also means that there is no "absolute" or "true" requirement. Depending on the time that the N balance is determined, it may be different but at any point on the curve the individual is in balance. This is the concept of N balance—a biologically regulated phenomenon we wish to emphasize in our paper.



Scheme 1.

Now what is the effect of altering N intake when considering our homeostatic model? When we exceed the homeostatic mechanism a state of "stress" is said to exist. The first manifestation of this would be an alteration amounting to a breakdown in the oscillation pattern: (*1*) the timing of the oscillations would change, becoming more rapid than before; (*2*) the amplitude will generate becoming sharper and larger; and (*3*) the fluctuations in general will look more irregular than before indicating lack of periodicity. Depending on protein level the organism might or might not adapt. Obviously in the case of zero protein, adaptation is impossible, and death would ensue. In the case of "high" protein—at least at 36 g/day over the period of observation we have not yet seen any adaptation. The organism has remained in a state of "stress" as defined above. This phenomenon of stress may also be observed in other biological parameters at high levels of N intake—as marked increase in calciuria, and in experiments in our laboratory at even higher levels of protein, failure of renal glucose regulating mechanisms.

However, it would be expected that with less severe perturbations of N intake, after a period of stress, as with any regulating biological system, adaptation (a characteristic of regulated biological systems) would become operative and the individual would again come into homeostasis albeit at a new level of N intake and body mass.

This approach therefore denies the concept of a fixed nitrogen requirement for man and redefines protein requirements in a dynamic fashion, fully compatible with that of most other biological systems.

We now return to a more mathematical treatment of this concept. This representation of the daily variation in N balance in terms of the number of oscillations per day, also called the frequency per unit of time is known as the spectral representation of the stochastic process regulating N balance in man. The variance of the time series σ_w^2 represents the total power of the spectral function and describes how the total power is distributed with respect to frequency. When r is positive, the power is seen to be concentrated at lower frequencies, as r de-

creases the power is seen to shift evenly over all the frequencies and when r is negative, the power is concentrated at high frequencies. The adjective "power" derives from engineer's use of this word in connection with the passage of an electric current through a resistance. For a periodical input, the power is directly proportional to the square of the amplitude of the oscillation, but when, as in the experiments reported in this paper, the input is fixed or is of stochastic nature as in a free-living population in steady state and the distribution of output is stationary, the power is distributed over several frequencies. This is probably the reason why Edholm et al. (6) have searched in vain to conclude that energy intake and expenditure in individuals maintaining body weight do not balance even when averaged over several days or weeks. Although further discussion is outside the scope of this paper, we are tempted to add a figure to illustrate how change in r influences the spectra. Figure 14 gives examples of the four spectra corresponding to four values of (*1*) $r = 0.7$; (*2*) $r = 0.3$; (*3*) $r = 0$; and (*4*) $r = -0.3$. It will be seen that as the intake increases and r decreases, the frequency per unit of time is increased.

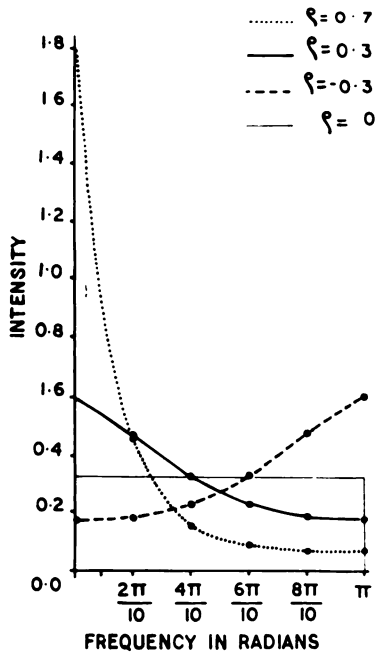


FIG. 14. Spectral representation of the autoregressive series.

Implications for day-to-day variation in requirement

The results described in the previous sections can be directly used in analyzing daily variation in requirement within individuals. For when the daily intake is fixed, the daily variation in output is identical with that in N balance. In particular, when the daily intake is methodologically defined as equal to the estimated true requirement of the individual, say $\hat{Y}$, the output y_t on any day t can be taken as the observed requirement for that day, being the sum of $\hat{Y}$ and w_t given by

$$y_t = \hat{Y} + w_t \qquad (11)$$

Y, the true requirement will vary from individual to individual but not in the same individual, w_t representing the deviation from Y arising from the fact that the determination refers to the chosen day or period of time and not to the whole timespan represented by the age-sex group to which the individual belongs. It follows that if the series on daily N balance, w_t, within an individual can be represented by AR process of the first order as shown in equation (2), so can also be the series on output y_t.

Now Y is commonly estimated from the regression equation relating N balance to intake. For adults, it represents the lowest intake at which N balance is zero. Thus, if an individual is fed for about 1 week at each of several levels of N near the point of balance, the regression equation can be written as

$$w = kx + \alpha \qquad (12)$$

where w stands for the balance, x for intake and α for the intercept on w axis when x is zero representing the loss of N on no protein diet and k for utilization of the N fed. Putting $w = 0$, we have for the true adult requirement an estimate given by

$$\hat{Y} = -\frac{\alpha}{k} \qquad (13)$$

Alternative estimates of Y can also be suggested but these will not be considered here.

It will suffice to say that the estimate in the form given in equation (*13*) above is the one generally used in nutrition literature and is in line with the procedure used in the report on protein requirements (1) for estimating "true" protein needs of individuals.

It is different when we come to consider the component w_t. Ordinarily an individual will be rarely in zero balance on any day. As the data presented in the previous sections show, most of the time w_t will be found to wander around a stationary mean with varying intervals between peaks and troughs and varying amplitudes suggestive of a generating mechanism of a probabilistic kind remaining constant through time. The meaning of such a mechanism is that if it were possible to repeat the circumstances which gave rise to the observed value of w_t on the t^{th} day, then w_t would be expected to have the same frequency distribution for any t and w_t and w_{t-k} to have the same multivariate distribution for any t and k. Specifically, it was shown that this generating mechanism can be reasonably adequately described by AR process of the first order, implying that the requirement y_t for any day t will be distributed around Y with variance σ_w^2 given by equation (5) and that the correlation ρ_k between requirements on the t^{th} and $(t + k)^{th}$ day will be given by equation (6).

In practice it will not be possible to repeat the circumstances which gave rise to the observed y_t because it is impossible to make more than one observation at a given point of time. Nevertheless, we can regard the observed series of output as just one example of the infinite set of series which might have been observed. Every member of this set is, therefore, a possible realization of the stationary stochastic process represented by AR series of the first order. Because the population involved is only a notional population, the idea is difficult to comprehend but once it is grasped it becomes clear that the most we can say about the N requirement of man on any day is that it will vary around the true requirement Y most of the time within the range $\pm 2\sigma_w$,σ_w being estimated from the range of values actually observed over time within individuals. The statement is similar to the one we make for body weight, which we saw is also distributed in a stationary distribution. The only difference is that whereas the coefficient of variation in the case of an adult maintaining body weight is small, of the order of 1%, it is larger in the case of N balance because of larger errors that inevitably arise in the measurement of N balance compared to the error of measurement of body weight.

Definition of deficiency

The fact that requirement varies from day to day has been well recognized. What is new in the above findings is that the fluctuations in requirement appear to be represented by a stationary stochastic process of the Markoff kind. From the practical viewpoint, the question which arises here is whether the influence of intraindividual variation described by the process can be reduced or even eliminated altogether by averaging requirement over a sufficiently long period such as 3, 5, 7 or more days to provide what is called the habitual requirement. If not, would it not be right to define deficiency as a failure of the process to be in statistical control, rather than as a situation in which the observed intake falls short of the individual's true requirement?

The variance of the mean will invariably decrease as the number of days over which the requirement is averaged will increase. However, when, as in the present case, the daily variation is not random (i.e., not arising from chance error alone), the precision of the estimated mean will be influenced not only by the size of the period but also by the distribution of requirement over time. In particular it can be shown that when the time series is represented by the Markoff process, the variance of the estimated mean of P days will vary as λ/P, where

$$\lambda = \frac{1 + r}{1 - r} - \frac{2r}{P}\frac{1 - r^P}{(1 - r)^2} \qquad (14)$$

Table 4 shows the values of λ/p for different values of r and P. It will be seen that when the serial correlation of the first order has a value around 0.6 to 0.7, the variance of the estimated requirement even when based on 7 days average (and losing in the process six pieces of information) will be three times as

TABLE 4
Variance of the mean of P values with unit variance when successive observations are correlated in the Markoff fashion

P \ r	0.0	0.50	0.66	0.80
1	1.00	1.00	1.00	1.00
3	0.33	0.61	0.73	0.83
5	0.20	0.45	0.58	0.72
7	0.14	0.35	0.48	0.58

large as when based on the assumption of independence between successive observations. This means that even when the average refers to 7 days, the standard error will be reduced only in the ratio 1 to 0.7 and not become negligible, as Beaton and Swiss (3) believe it to be, based on the assumption that successive observations are independent over time. Table 2 will show that the coefficient of day-to-day variation in requirement within an individual is of the order of 20%. It follows that it will not be possible to reduce it to much below 12 to 15% even if requirement figures refer to 7-day periods. The conclusion is that requirement of man in N equilibrium maintaining his body weight will vary not only from day to day but also from week to week rather considerably as a matter of course with stationary variance. When the observed intake for any day or period is therefore less than the requirement of the reference individual for the age sex group to which a person belongs, it cannot be taken to imply that a man is deficient in protein unless his intake is so low as to be below the lower limit of the confidence interval for the chosen level of significance. The definition of deficiency as a situation in which an individual's consumption is less than his estimated true requirement cannot, therefore, be valid unless σ_w were negligible relative to σ_y which does not appear to be so on the evidence presented above.

Balance data are the appropriate data to study the relative magnitudes of inter- and intraindividual variation in requirement. However, to our knowledge, no daily balance data beyond what are reported in this paper are available for investigation. In this situation, measurement of obligatory losses of N in adults for 5 consecutive days (7 to 11 days) reported by Scrimshaw and his associates provides the only source for studying variation (4). Table 5 presents the analysis of variance of these data. Under the model described in equation (11), which for convenience we can also write as

$$y_{it} = \mu + b_i + w_{it} \qquad (15)$$

with μ denoting the true requirement of the reference man and b_i denoting the deviation of an individual's true requirement from μ then the mean square between individuals in the analysis of variance table is an estimate of $5\sigma_b^2 + \sigma_w^2$ and the mean square within individuals is an estimate of σ_w^2. Inter- and intra-individual components of variation are thus found to have the following values:

$$\sigma_w^2 = 0.23$$

$$\sigma_b^2 = 0.16$$

$$\sigma^2 = \sigma_w^2 + \sigma_b^2 = 0.39$$

Far from being negligible, intraindividual variation is seen to account for some 60% of the total variation. However, σ_w^2 calculated from this table does not constitute the true intra-individual variation in requirement. It only reflects the variation between days within a single period from 7th to 11th day. What we need is an estimate of intraindividual variation between periods (of 5 days each) within individuals. If N balance on successive days could be assumed to be independent, we could estimate it to be 0.05 and consider it as negligible compared to the value of σ_b^2, namely 0.16 but as we saw already, the successive observations are not independent and it would seem that intraindividual variation between periods within individuals is likely to be much larger and possibly comparable to variance between periods.

There is also a further consideration. As the individual advances in time, he becomes increasingly a different individual. The specialized environment in which he is being

TABLE 5
Analysis of variance of daily N losses (in g) on no-protein diet in 83 adult males

Source	D.F.	S.S.	M.Sq
Between subjects	82	82.80	1.01
Wtihin subjects	332	75.68	0.23
Total	414	158.48	

brought up interacting with the "genetic" component of the requirement will further add to the variance. It will, therefore, not be surprising if we should find that intraindividual variation even when requirement is based on several days' average does not decrease relative to σ_b^2. Unfortunately, we have no data to go by. The only data available to examine this hypothesis are the one reported by Scrimshaw on four adult subjects at intervals of 3 years (4) and as we shall see, his data confirm that intraindividual variation does not become negligible relative to σ_b^2 but remain the most fundamental variation in requirement. Table 6 shows these data and Table 7 their analyses of variance. It will be seen that even though the data are presented as 5-day averages, the intraindividual variation comes to 0.17 compared with the estimated true interindividual variation of only 0.07, thus accounting for some 70% of the total variation. On the other hand, part of the variation in N losses at intervals of three years is evidently due to difference in body weights which occurred over the same period. These body weights are shown in Table 6. The correlation between N loss and body weight is found to be 0.75, giving for the regression of N loss on body weight z,

$$y = -2.19 + 0.83\,z$$

Adjusting N loss for body weight and analyzing the variation in the adjusted figures, we obtain the analysis of variance reported in Table 8. It will be seen that intraindividual variation now comes to 0.09 compared with the estimated true interindividual variation of only 0.01, thus accounting for even a larger share of the total variation than before. If the analysis of variance is carried out on N loss per unit of body weight basis as reported by Beaton and Swiss (3), intraindividual variation is found to account for even a larger part of the total variation (Table 9). As Table 9 shows, almost all the variation is intraindividual and not interindividual as wrongly concluded by Beaton and Swiss (3). The very small number of subjects of course precludes generalization. Nevertheless, one sees in these data evidence to question the validity of the assumption that intraindividual variation is primarily a chance variation due to measurement errors and therefore can be reduced to a negligible fraction of the total variation by referring estimates of requirement to a sufficiently long period. Elsewhere (7), we have reported analysis of variance of two other data—those of Muller and Cox (8) and by Gopalan and Narasinga Rao (9). These data confirm that intraindividual variation forms a substantial part of the total variation.

The conclusion that intraindividual variation is not negligible nor can it be made negligible relative to true interindividual variation is also substantiated by data presented in this paper. These data are summarized in Table 10 and their analysis of variance is given in Table 11. Table 11

TABLE 7
Analysis of variance of N losses (g/day) shown in Table 6

Source	D.F.	S.S.	M.Sq
Between subjects	3	0.93	0.31
Within subjects	4	0.68	0.17
Total	7	1.61	

TABLE 8
Analysis of variance of N loss (g/day) adjusted for body weight

Source	D.F.	S.S.	M.Sq
Between subjects	3	0.33	0.11
Within subjects	4	0.36	0.09
Total	7	0.69	

TABLE 6
Comparative data on urinary (U) and faecal (F) excretion on no-protein diet for four adult male subjects studied at 3-year intervals

Subject	14		54		57		68	
Year	1966	1969	1966	1969	1966	1969	1966	1969
(U + F) N[a]	2.98	3.96	2.89	2.88	3.64	3.78	2.66	3.28
Body wt (kg)	63.8	70.7	57.7	62.5	69.7	68.3	65.4	66.9

[a] Nitrogen (in grams) – 5-day average.

TABLE 9
Analysis of variance of N loss (g/day) per kilogram of body weight

Source	D.F.	S.S.	M.Sq		
Between subjects	3	68.52	22.84	→	$2\sigma_b^2 + \sigma_w^2$
Within subjects	4	90.98	22.74	→	σ_w^2
Total	7	159.50	22.79	→	$\sigma^2 = \sigma_b^2 + \sigma_w^2$
		$\sigma_b^2 = 0.05$			
		$\sigma_w^2 = 22.74$			

TABLE 10
Comparative data on daily N output (in grams) on diet containing 12.38 g of N for five male adults

	A	B	C	D	E	Weighted average
Period 1	11.61	12.31	12.45	13.90	13.07	12.29
	(35)[a]	(35)	(9)	(9)	(9)	(97)
Period 2	11.50	11.62	14.41	15.61	16.64	13.45
	(25)	(25)	(15)	(15)	(15)	(95)
Weighted average	11.57	12.02	13.67	14.97	15.30	
	(60)	(60)	(24)	(24)	(24)	

[a] Figures in parentheses show the number of days on which the average values are based.

TABLE 11
Analysis of variance of N output (g/day)

Source	D.F.	S.S.	M.Sq		
Between adults	4	408.34	102.09	→	$37.6\sigma_b^2 + 18.8\sigma_p^2 + \sigma_d^2$
Between periods within adults	5	116.90	23.38	→	$18.8\sigma_p^2 + \sigma_d^2$
Error	182	227.98	1.20	→	σ_d^2
Total	191	743.23			
$\sigma_b^2 = 2.09$		$\sigma_p^2 = 1.18$			$\sigma_d^2 = 1.20$
$\sigma_p^2 + \sigma_d^2 = 2.38$					
$\sigma^2 = \sigma_b^2 + \sigma_d^2 = 4.47$					

TABLE 12
Correlation between successive values of urinary output, urinary plus faecal N output, and N balance

Lag	Subject 2			Subject 3			Subject 5		
	$(U + F)_N$	U_N	Balance	$(U + F)_N$	U_N	Balance	$(U + F)_N$	U_N	Balance
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	0.42	0.51	0.36	0.61	0.63	0.75	0.58	0.49	0.69
2	0.14	0.29	0.10	0.48	0.52	0.57	0.35	0.30	0.49
3	0.02	0.16	−0.02	0.39	0.46	0.48	0.31	0.28	0.36

shows that intraindividual variation, even when output is averaged over as long a period as two weeks, remains comparable with the estimated true interindividual variation. The marked daily variation in the use of food protein and regulatory mechanism governing the loss and repletion of the N pool following the Markoff pattern accounts for this phenomenon. But we have no explanation to offer in the case of data on obligatory losses on no protein diet reported in Tables 6 to 8.

Another advantage of use of this method is that when protein absorption is high, urinary excretion alone can give data similar to balance data (Table 12).

Incidence of protein deficiency

Although the purpose of this paper is not directly to discuss the problems inherent in the various models of protein requirements on the estimation of incidence of protein deficiency in populations, or how these

models form the basis of program decision making, we do believe that the model which we have suggested is sufficiently different from others to make some tentative remarks regarding their implications for defining incidences of malnutrition in populations and possible implications for social policy.

Using our model, we can now evaluate intake data for estimating the incidence of protein deficiency in the population. Ordinarily in an adequately nourished population of individuals of the "reference" type, with no one protein deficient and assuming in line with the conclusion that N balance like energy balance is regulated in man maintaining body weight and that further the true requirement is the same for all individuals in health of the reference type, one would expect, given normal distribution, 95% of the individuals to have intake within the range given by $\mu \pm 2\sigma_w$. Consequently, in any observed intake distribution nutrition unit[4] basis, the proportion of the individuals below $\mu - 2\sigma_w$ may be taken as the estimate of the incidence of protein deficiency.

To assume that each adult individual in health has the same fixed requirement as the reference adult is, however, to impose too strong an assumption on the model for requirement. For it is well known that some adaptation does occur in men without implying that a man who has adapted himself to a requirement higher than μ is protein deficient relative to the reference adult and vice versa. We may, therefore, weaken the assumption to one represented by a model which may be simply written as

$$y = \mu + b + w \qquad (15a)$$

where b represents the difference between the true requirement of the i^{th} individual and that of the reference adult and w represents the deviation from $\mu + b$ arising from the fact that the determination refers to a chosen period and not to the whole time span represented by the age sex group. Every individual has its value b so that b can be considered a random variable distributed around zero with variance σ_b^2. Likewise, the term w is distributed around zero with known stationary variance σ_w^2 given by equation (5). Briefly, the model visualized is a one-way random model and represents a data situation in which the population of individuals is partitioned as it were into subpopulations with each individual constituting a subpopulation, observations over time in the same individual constituting a stochastic stationary time series to make the subpopulation. The variance σ^2 in the population is made up of two components—one represented by the true interindividual variance σ_b^2 and the other by intraindividual variance σ_w^2 in the same individual with $\sigma^2 = \sigma_w^2 + \sigma_b^2$.

Clearly most individuals in health in the framework of this model will have an intake between $\mu \pm 2\sigma$. It follows that the proportion of individuals below the lower critical limit $\mu - 2\sigma$ may be taken to represent the estimate of the incidence of protein deficiency.

In two respects the model developed about the incidence of protein deficiency differs from that developed by Lorstad (2). Firstly, it does not accept that requirement is constant in an individual i.e., $\sigma_w = 0$. To assume so is to ignore the evidence presented in the paper and the finding based thereon that the daily N balance in man maintaining his body weight is regulated by a generating mechanism of a probabilistic kind remaining constant through time. Secondly, it does not accept that knowledge of the true requirements as such of the individuals included in the sample can be of major interest (or even ascertained sufficiently accurately) in evaluation of intake data. Thus, an individual could have for his observed intake a value above $\mu - 2\sigma$ but still be below his true requirement without implying that he is deficient.

The principle of evaluating intake data laid down by the FAO/WHO Expert Committee namely that an individual eating below the safe level by definition runs the risk of developing protein deficiency and that the risk increases as the intake falls below the safe level, does not therefore apply in the case of our model, except of course at the extremes of low intake, at levels where homeostasis breaks down. On the other hand, this principle is basic to the models developed by Lorstad (2) and by Beaton

[4] A nutrition unit for protein has the same daily requirement as the reference adult, namely μ.

and Swiss (3). For this reason data used in their models must represent the habitual or usual intake and the true requirement. In other words, there is no place in their approach for day-to-day or week-to-week variability in intake and requirement. Their approach thus amounts to accepting what is usually called a fixed model in place of a one-way random model given by equation (15). There is no evidence that adult individuals have fixed needs beyond that of individuals of the reference type to adapt themselves to variable intake. Whether their adaptation is a temporary physiological adjustment to environmental change or gets transformed into population adaptation over generations we do not know. Whatever the nature of this adaptation, if we should accept the fixed model we must at least ensure that the variance of requirement approximate the true interindividual variation σ_b^2 and not σ^2 and the intake data likewise represent habitual or usual intakes which make full use of both the intra- and interindividual components of variance in their estimation and are not simple averages of the daily intakes over the period of the survey. If this were not done, the incidence will be overestimated using the previously proposed models (2, 3).

In practice, it is unlikely that there will be any significant difference because in the present state of our knowledge of protein requirement (1) it is found that almost every diet which meet man's energy needs has a protein content greater than requirement. However, the last word has not been said; protein requirements are still tentative based as they are on limited data mainly collected in the West and on approaches for which there is no general agreement. They will undoubtedly undergo a change as more and better experimental data on N balance become available, not only in regard to level but even more importantly in the evolution of the model for interpreting them.

References

1. Joint FAO/WHO Ad-Hoc Expert Committee. Report on energy and protein requirements. WHO Tech. Rep. Ser. No. 522, 1973; FAO Nutr. Meetings Rep. Ser. No. 52, 1973.
2. Lorstad, M. H. Recommmended intake and its relation to nutrient deficiency. FAO Nutr. Newslet. 9: 18, 1971.
3. Beaton, G. H., and L. D. Swiss. Evaluation of nutritional quality of food supplies: prediction of desirable or safe protein-calorie ratios. Am. J. Clin. Nutr. 27: 485, 1974.
4. Scrimshaw, N. S., M. A. Hussein, E. Murray, W. M. Rand and V. R. Young. Protein requirements of man: variations in obligatory urinary and faecal nitrogen losses in young men. J. Nutr. 102: 1595, 1972.
5. Calloway, D. H., and S. Margen. Variation in endogenous nitrogen excretion and dietary nitrogen utilisation as determinants of human protein requirement. J. Nutr. 101: 205, 1971.
6. Edholm, O. G., J. M. Adam, M. J. R. Healy, H. S. Wolff, R. Goldsmith and T. W. Best. Food intake and energy expenditure of army recruits. Brit. J. Nutr. 24: 1091, 1970.
7. Sukhatme, P. V. The protein problem, its size and nature. J. Roy. Statist. Soc. 137: 166, 1974.
8. Muller, A. G., and W. M. Cox. Nitrogen retention studies on rats, dogs and man. J. Nutr. 34: 285, 1947.
9. Gopalan, C., and B. S. Narasinga Rao. Effect of protein depletion on urinary nitrogen excretion in undernourished subjects. J. Nutr. 90: 213, 1966.