A convenient six-point blood sampling schedule for determining whole body cholesterol kinetics in humans

Ralph B. Dell,[1] Rajasekhar Ramakrishnan, Robert H. Palmer, and DeWitt S. Goodman

Departments of Pediatrics[2] and Medicine[3] and the Arteriosclerosis Research Center, Columbia University College of Physicians and Surgeons, New York, NY 10032

Abstract Parameters of total body cholesterol metabolism in humans can be determined by using a three-pool model to analyze the turnover of plasma cholesterol following the injection of radiolabeled cholesterol. In the past this required a rigorous schedule of approximately 36 blood samples over a 10-month period. We have developed a convenient sampling schedule involving only six large samples, each analyzed in sextuplicate. Such a reduction in the frequency of samples is possible only when considerable confidence in the model is available. In general, the simplified sampling strategy depends upon considerable prior experience with the model, only moderate biological error, and estimatable subject to subject variation in model parameters. Because the timing of the samples is critical and because the optimal times will differ for different subjects, the six-point strategy involves using the first three samples (drawn at days 1, 7, and 24 or, for hypercholesterolemic subjects, at days 1, 8, and 28) in conjunction with results from previous studies to set the time for the next sample; the process is reiterated for the last two points. In this study, we have compared parameter estimates obtained by the new six-point schedule with those obtained simultaneously (in the same, single turnover study) by the old 36-point schedule in the same 26 subjects. Both schedules gave comparable values. In particular, the coefficients of variation between values obtained by the two methods for each of the four parameters for which we have developed predictive equations were quite low: PR 1.5%, M_1 4.1%, M_3min 13%, M_{tot}min 4.3%. The simplified six-point schedule makes it feasible to study long-term cholesterol turnover in substantial numbers of patients. —**Dell, R. B., R. Ramakrishnan, R. H. Palmer, and D. S. Goodman.** A convenient six-point blood sampling schedule for determining whole body cholesterol kinetics in humans. *J. Lipid Res.* 1985. **26:** 575–582.

Supplementary key words cholesterol turnover • kinetic analysis • three-pool model

Whole body cholesterol metabolism in humans can be studied by analyzing the turnover of plasma cholesterol following injection of radiolabeled cholesterol complexed with lipoproteins (1). In previous studies we found that a three-pool compartmental model fitted the long-term plasma cholesterol specific activity-time curve in 56 subjects who were either normal or had a wide variety of abnormal serum lipid values (2, 3), and that a four-pool model did not improve the fit. Given such extensive experience with the three-pool model, we are confident that the model is a valid description of whole body cholesterol turnover in humans in vivo. The three-term exponential equation used to fit the data provides unique values for six of the eight parameters of the three-pool model, and upper and lower limits can be placed on the remaining two parameters (1, 2). Thus, from a long-term cholesterol turnover study, one can obtain estimates of the production rate, the mass of rapidly turning over cholesterol (pool 1), and of the mass and turnover rates of cholesterol in the more slowly turning over compartments, pools 2 and 3.

The blood sampling schedule used in our previous 56 studies (which involved drawing 35–40 samples over the course of 10 months) is taxing for both patient and physician, and decreases both the number of subjects who can be studied as well as the number of subjects who volunteer for the study. It would be highly advantageous to have a simpler and more convenient blood sampling schedule for obtaining this information.

In principle, the six unique coefficients of the three-term exponential equation that describes the turnover data can be determined from any six points on the specific activity-time curve. However, the accuracy of the final parameter estimates depends upon both the accuracy of the six points and when they are drawn (Ramakrishnan, R. Optimal design of experiments in nonlinear estimation. Submitted for publication). Therefore, the six points should be replicated so that they are accurately determined, and should be drawn at optimal times to permit the most accurate estimation of the six parameters of the

Abbreviations: PR, production rate; k, rate constant; M_1, M_2, M_3, M_{tot}, pool sizes.

[1]To whom reprint requests should be addressed.

[2]R. B. Dell and R. Ramakrishnan.

[3]R. H. Palmer and D. S. Goodman.

equations. In the case of a straight line y = a + bt, the accuracy of the estimates of the two parameters is highest when the two t's are determined as far apart as is feasible (4). The situation is more complex when dealing with nonlinear models (5), as discussed further under Methods. It should be noted that determining only six points precludes testing for alternate models. Thus, use of a simplified six-point sampling schedule depends upon prior experience with the model, only moderate deviations from the model (biological error, see Discussion) and estimatable subject-to-subject variation in model parameters.

The purpose of the present study was to evaluate a convenient blood sampling schedule for the study of whole body cholesterol turnover in humans. The convenient approach adopted involves drawing only six blood samples (large enough for six replicate specific activity analyses to be carried out on each sample) at times specifically selected to be optimal or near optimal. In the present work, a single turnover study was performed on each subject, with blood being drawn during this single study according to both the usual sampling schedule and the convenient six-point schedule simultaneously. In each study, parameter estimates obtained by this method were compared with parameter estimates obtained simultaneously by drawing 36 separate samples in the conventional manner (2). The results show that parameters estimated from the specific activity-time curves determined by the two methods agree quite closely.

METHODS

Patient population

Twenty-six subjects were studied. Five of the subjects were normal, 3 were hypercholesterolemic, 15 were hypertriglyceridemic, and 3 had a combined disturbance. Results of the turnover studies in 19 of the subjects (numbers 27–30, 34–36, 40–42, 44–49, and 52–54) have been reported previously (2), and the numbers used to designate these subjects correspond to the numbers in that study. Seven subjects (numbers 64–70) will be reported in another publication (Blum, C. B., R. B. Dell, R. H. Palmer, et al. Relationship of the parameters of body cholesterol metabolism with plasma levels of HDL cholesterol and the major HDL apoproteins. Submitted for publication).

Turnover study

Details of the labeling technique have been published elsewhere (1, 6). Briefly, each subject was injected intravenously with approximately 25 μCi of [4-^{14}C]cholesterol complexed to the subject's own lipoproteins. Samples of blood were then drawn from each subject according to each of the sampling schedules described below and analyzed for the specific radioactivity of serum total cholesterol by methods described previously (1, 2, 6).

The model

The three-pool model used to characterize whole body cholesterol turnover is shown in **Fig. 1**. The model has eight unknown parameters, which are: production rate (PR), which is the total input (or output) of cholesterol into the body from both synthesis and absorption; mass of pool 1 (M_1), which is the mass of most rapidly turning over cholesterol in the body; four exchange rates, k_{12}, k_{21}, k_{13}, k_{31}; and synthetic input into the side pools, pools 2 and 3. All other parameters of the model are calculable from these basic eight (1, 6). The two parameters relating to side-pool synthesis (R_{20} and R_{30}) cannot be estimated directly from the plasma cholesterol specific activity-time curve. However, limits can be put on the possible values for R_{20} and R_{30} (1, 6). We usually allow 0.2 g/day for nonsynthetic cholesterol input; hence, R_{20} and R_{30} are constrained to be between 0 and PR–0.2. This allows calculation of a range of values for M_2, M_3, and total body exchangeable cholesterol (M_{tot}) (1, 2); when peripheral synthesis is assumed to be zero, then minimal estimates for pool sizes are obtained, and when R_{30} is assumed to be PR–0.2, maximal estimates for M_3 and for M_{tot} are obtained.

Rationale for an optimal sampling schedule

Since the plasma cholesterol specific activity-time curve can be described by a sum of three exponentials containing six parameters, it is possible to estimate these six parameters with observations at six time points. A schedule for sampling at only six times during the study would be convenient for both subject and physician, requiring only one-sixth the number of clinic visits that the usual schedule requires. The question then arises as to how these six time points should be chosen.

The answer to this question involves considering the effect of sampling times on the accuracy with which the unknown parameters of the equation can be estimated. We will first discuss the sources of imprecision in the parameter estimates and then how the timing of the samples affects the precision of the parameter estimates.

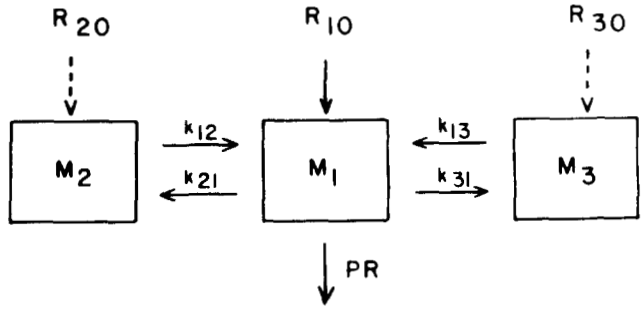

Fig. 1. Three-pool model of cholesterol turnover in humans (see text for definition and discussion of symbols).

If the observations can be made with no error whatsoever, the parameters can be determined exactly, with no uncertainty, from observations at any six distinct times. In fact, of course, experimentally derived data deviate from the true specific activity curve that is characteristic of a given subject. There are two reasons why a point can deviate from the true curve: *a*) measurement error associated with the experimental determination of the specific activity of the sample, and *b*) a transient fluctuation in cholesterol metabolism in the patient yielding a change in plasma specific activity. This latter reason can be called biological error, and is discussed further below. Measurement error can be reduced by performing replicate assays on a given sample, but cannot be eliminated. Therefore, parameters calculated from the observations have some uncertainty associated with them, usually expressed as confidence limits on the parameter estimates or as standard deviations of the parameter estimates.

Expression for precision of multiple parameters

When the model has more than one parameter, there are many ways (ref. 7, pp 51–53) to combine the precision of the estimates of the several model parameters into a single quantity. In this study, we used the sum of squared coefficients of variation (defining the coefficient of variation as the standard deviation of a parameter estimate divided by the parameter estimate itself) of the parameter estimates. We made this choice for two reasons: first, it removes the dimensionality of each parameter; secondly, minimizing this quantity tends to make the coefficients of variation of the individual parameters as close to one another as possible. Thus, this approach gives similar weight to each of the model parameters (i.e., parameters with small and large values will be equally important in determining the final sampling schedule).

A finite imprecision in determining the data points leads to imprecision in parameter estimates, and the extent of this imprecision depends on the sensitivities of the values predicted from the model at the times of observations to the model parameters. If the values predicted from the model change considerably with small changes in the parameters (i.e., are very sensitive to changes in the parameters), then the unknown parameters can be estimated quite precisely from the values determined experimentally and the parameter estimates will have narrow confidence limits. Thus, the accuracy of parameter estimates depends not only upon the accuracy of the observations but also on how sensitive the values computed from the model are to small changes in the parameters. Furthermore, the sensitivities of the values to the parameters depends on sampling times. This dependence can be illustrated by discussing the simple case of a linear model and the more complicated situation of nonlinear models.

Linear model

In fitting a straight line $y = a + bt$ to a set of data, observations at two points, t_1 and t_2, suffice to estimate values for the unknown parameters a and b. The standard error of these estimates is inversely proportional to the squared difference between t_1 and t_2 (4):

$$SE\ (a,b) \propto \frac{1}{(t_1 - t_2)^2}.$$

The variance of the parameter estimates diminishes as the two time points are chosen to be further and further apart. In other words, drawing two samples as far apart as possible allows the most precise parameter estimation. The only limitations to this are the feasibility of experimentally determining y and the fact that the phenomenon under study must be linear over the whole range of t. Therefore, one must have techniques for measuring, precisely, low levels of y and must be assured that a linear model is appropriate because testing of linearity is impossible with two data points.

Nonlinear models

The situation becomes more complicated with nonlinear models (such as sums of exponentials) because the optimal sampling times now become dependent on the unknown parameters. A simple monoexponential equation with one unknown parameter will illustrate this point. For

$$y = e^{-\alpha t}$$

an observation at a single time is sufficient to estimate α. However, if the observation is made at time zero, y equals 1 regardless of the values of α. Thus at time zero the model-predicted value (y) is totally insensitive to the unknown parameter (α). At large times, y approaches zero and the value is again highly insensitive to α. Hence, making observations at either time zero or at infinite time provides no information about α, and such observations cannot be used to estimate α. At intermediate times, the sensitivity varies with time. Therefore, observation times should be chosen that maximize the sensitivity and hence the accuracy of the parameter estimates.

As discussed above, the accuracy of each of the parameter estimates depends upon both the accuracy of the observations and the sensitivity of the value predicted from the model to the parameter. In the monoexponential equation, the standard error of α can be stated as:

$$SE_\alpha = \frac{s_y}{\left|\frac{\delta y}{\delta \alpha}\right|} = \frac{s_y}{te^{-\alpha t}}$$

where s_y is the error in measuring y. Just as in the linear case, the standard error of the parameter estimate de-

pends upon the sample time, t. However, in contrast to the linear case, the standard error of the parameter estimate also depends upon the value of the unknown parameter, α. In fact, if s_y is constant, then SE_α is at a minimum when $t = 1/\alpha$. Thus, the most precise estimate of α (minimum SE_α) is obtained when $t = 1/\alpha$. If s_y is not constant but depends on y in some fashion (we typically find that it is a constant percentage of y unless y is very small), the precise result is different but the principle remains the same, which is that there is a single optimum sampling time, which is a function of α.

Choosing optimal sampling times from parameter estimates based on prior studies

The dependence of the best sampling times on the values of the unknown parameters poses a dilemma inasmuch as the purpose of the study is to estimate the unknown parameters. This problem has been considered in the statistical literature (5, 7) and the recommendation is to calculate the optimal times using the mean parameter values from previous studies, assuming that these means will be close to the parameter estimates that would result from the study being planned.

That approach assumes that there is a single universal true value for each parameter, which is approached by the mean values from previous studies, and that each study brings the parameter estimate closer to this single value. However, in biological studies, each subject has a different true value of the parameter; past studies, however many, cannot precisely predict the true value for a subject not yet studied. This problem has been analyzed by us (Ramakrishnan, op. cit.) and an approach has been worked out that determines the best sampling times for a range of possible parameter values rather than for a single value.

This approach requires a knowledge of the distribution of parameter values in the population. We estimated this distribution from our past experience by treating the parameter values from all previous turnover studies as equally likely to occur. Since hypercholesterolemic subjects were found to have different parameters and hence different optimal sampling times, two separate distributions were used—one for hypercholesterolemic subjects and one for others.

It may be noted here that the sampling times obtained in this manner give the optimal times if one plans to draw only six blood samples. Optimal schedules for drawing more samples will be different from and probably better than a six-point schedule, especially when the true parameter values have a broad distribution. Thus, it would be better in a statistical sense to draw 12 or 18 samples clustered in some fashion around the six times, but in this study we were willing to sacrifice some precision for the convenience of a six- or nine-point sampling schedule. Examples are given (Ramakrishnan, op. cit.) for optimal schedules when the number of samples is greater than the number of model parameters.

Sequential design

While the prior distributions of parameters can be used to choose sampling times that are optimal for the population as a whole, there is considerable variability in the parameters, and hence in the optimal sampling times, within the population. This variability is illustrated for nonhypercholesterolemic subjects in **Table 1**.

In order to refine the choice of sampling dates (which otherwise would be based solely on the prior distribution of parameter values) to be better suited for a given individual, we adopted a sequential design approach that incorporates early information from a given study in the calculations of successive sampling times. The approach is described in detail elsewhere (8), and outlined in **Fig. 2**.

Samples were obtained at days 1, 7, and 24 (or days 1, 8, and 28 for hypercholesterolemic subjects) and analyzed rapidly. The data were then used in conjunction with data from previous studies in order to obtain the most likely first approximation of the six parameters. This was done by noting that the previous studies had provided a family of specific activity decay curves with a distribution of parameter values, some of which would be more likely than others, given the first three data points. We chose the most likely parameter estimates that would still fit the first three data points exactly. Using these preliminary estimates, the most likely best time for the fourth sample was chosen. After each successive sample, the process was repeated to choose the next sampling time. After all six samples had been obtained, the final parameter estimates were made without reference to the prior distribution.

The sequential design was implemented after the first three samples had been obtained. The first sample was drawn on day 1, the earliest feasible time consistent with complete mixing of the label with the mass of readily exchangeable cholesterol. The second and third sampling times were chosen to correspond with the most likely best times based on the prior distributions for nonhypercholes-

TABLE 1. Subject-to-subject variation in optimal sampling times[a]

Sample No.	Minimum	10th Percentile	90th Percentile	Maximum
2	4	5	9	10
3	12	16	31	34
4	35	40	87	102
5	87	110	197	214
6	253	267	300	330

[a]This table illustrates the distribution of optimal sampling times (in days after the administration of [^{14}C]cholesterol) for samples 2–6 of a six-sample schedule. The optimal sampling times were determined from cholesterol turnover curves from 38 subjects without hypercholesterolemia (2); times for hypercholesterolemic subjects were slightly longer.

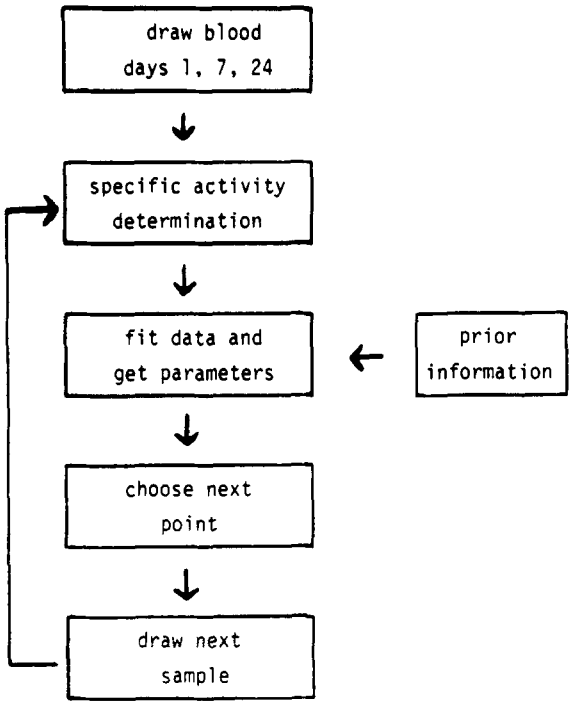

Fig. 2. Sequential design approach for determining a six-point sampling schedule.

terolemic subjects (days 7 and 24) and hypercholesterolemic subjects (days 8 and 28) separately, as these distributions had slightly different mean values. Theoretically, the sequential design could be implemented earlier (e.g., after the second sample), but the ranges of optimal times for the third sample were reasonably narrow, and completing the analyses in time to refine the optimal third time would present logistical difficulties. Therefore, we chose to implement the procedure after the third sample.

Computer programs

Optimal sampling times, for an individual or for a population, are determined by minimizing a nonlinear function of several variables, requiring the use of a computer program. Details of the algorithm used by the program are given elsewhere (Ramakrishnan, op. cit.).

Sampling schedules

Two sampling schedules were used simultaneously in each subject. In one, the usual sampling schedule, 36 samples were drawn over 40 weeks. They were drawn 1, 2, 3, 4, 7, 10, and 12 days after injection of tracer, twice a week for 6 weeks, weekly for 4 weeks, every 10 days for 6 weeks, every 2 weeks for 10 weeks, and then every 3 weeks for the remaining 12 weeks of the study. In the second schedule, samples were drawn at only six time points, at times which coincided with 6 of the 36 times in the usual schedule. The six time points were chosen as described above (Fig. 2). In order to increase the accuracy with which each point was determined, six analyses were performed on each sample, and results falling outside statistical limits were excluded (9).

Residual error

The residual error about the fitted curve provides an indication of the extent to which an observation may deviate from the model-predicted value.

$$\text{Residual error (coefficient of variation)} = \sqrt{\frac{\sum_{i=1}^{36} (Y_{i\,obs} - Y_{i\,calc})^2 / Y_{i\,calc}^2}{36 - 6}}$$

where $Y_{i\,obs}$ is the observed plasma cholesterol specific activity at the i-th time, $Y_{i\,calc}$ is the specific activity calculated from the model at the i-th time, and division by $Y_{i\,calc}$ is done because measurement error was found to be a constant percent of the value. As discussed above, both measurement error and biological error cause deviations that contribute to the residual error in the 36-point strategy. In contrast, in the 6-point strategy, biological error does not contribute to the residual error. This is because the mean values of the replicates at the six time points define one and only one six-parameter curve, with no residual error by definition. (Of course, biological variation can alter the value of any given point, and hence the value of the parameters, but there is still no residual error in the curve defined by six points.) Thus, the only contribution to the calculated residual error in the 6-point strategy is the measurement error arising from variability in measuring the replicate samples.

Since measurement error is the same with either sampling schedule, the difference between the residual errors with the two methods should reflect biological variation. We have chosen to call this difference biological error, and computed it as:

$$BE = \sqrt{RE_{36}^2 - RE_6^2}$$

where BE is biological error and RE_{36} and RE_6 are the residual errors found with the 36- and 6-point strategies, respectively.

RESULTS

Parameter estimates

Model parameters estimated from the 36 points drawn by the usual sampling schedule and from the 6 points of the new schedule are given in **Table 2** and **Table 3**. Table 2 gives values for each patient as well as means for PR, M_1, M_3min and M_{tot}min, parameters for which we have

TABLE 2. Model parameters for each subject estimated from both the usual 36-time point sampling schedule and the new 6-time point schedule

Subject #	PR		M_1		M_3min		M_{tot}min		k_{21}		k_{31}	
	36	6	36	6	36	6	36	6	36	6	36	6
27	0.87	0.87	18.9	18.9	17.2	21.2	58.3	57.7	.085	.085	.013	.022
28	0.92	0.92	23.6	23.7	20.2	18.1	60.5	60.5	.063	.068	.019	.014
29	0.96	0.96	20.7	20.7	27.6	26.9	66.0	66.2	.076	.079	.021	.022
30	1.36	1.36	22.6	21.8	45.3	52.0	84.5	86.9	.050	.063	.036	.046
34	1.50	1.51	31.1	32.0	32.8	30.4	92.3	97.0	.039	.036	.012	.008
35	1.53	1.56	30.2	30.8	53.8	56.3	98.9	105.0	.043	.052	.032	.030
36	1.28	1.30	23.9	23.0	35.4	38.0	80.6	79.6	.051	.054	.018	.025
40	1.09	1.07	22.5	21.8	41.6	43.5	74.5	71.3	.047	.096	.045	.055
41	1.76	1.76	34.3	34.6	69.1	69.6	111.4	114.8	.036	.043	.050	.047
42	1.82	1.82	25.7	24.6	46.5	49.2	85.4	85.6	.041	.051	.038	.044
44	1.04	1.04	25.1	23.7	39.8	42.4	91.6	86.0	.046	.054	.015	.022
45	2.42	2.30	27.0	27.7	53.6	51.6	94.9	96.4	.050	.054	.040	.031
46	1.57	1.55	22.3	21.1	44.3	44.9	82.6	81.2	.066	.078	.027	.030
47	1.87	1.89	28.4	27.1	58.2	67.4	113.3	112.5	.048	.052	.025	.036
48	2.15	2.12	27.4	26.0	47.0	47.0	94.7	94.7	.058	.059	.024	.027
49	1.49	1.44	21.8	20.2	27.1	44.6	79.8	71.9	.056	.076	.010	.045
52	1.29	1.30	25.3	25.5	47.6	50.6	83.2	89.4	.054	.059	.038	.032
53	1.95	1.93	26.9	24.5	48.2	44.0	93.2	88.6	.042	.056	.026	.024
54	1.30	1.28	20.2	20.4	25.2	22.3	69.1	60.7	.066	.066	.008	.014
64	2.09	2.10	31.6	29.8	44.0	41.5	95.9	98.8	.056	.081	.029	.024
65	1.13	1.13	21.3	21.2	81.1	86.4	119.6	123.2	.059	.073	.049	.053
66	0.82	0.82	17.0	16.8	38.8	38.5	66.6	66.2	.067	.080	.044	.043
67	1.15	1.15	30.4	29.7	23.8	27.2	87.5	84.4	.052	.052	.009	.012
68	1.29	1.24	28.8	30.5	35.6	45.9	79.8	77.9	.038	.024	.026	.044
69	0.66	0.66	19.9	20.3	35.8	39.2	65.7	66.5	.032	.032	.032	.037
70	0.90	0.91	25.0	25.0	27.8	36.6	65.0	65.1	.031	.012	.020	.033
Mean	1.385	1.385	25.08	24.71	41.02	43.67	84.42	84.15	.0520	.0592	.0271	.0315
Coef. var (%)	1.5		4.1		13		4.3		27		36	

developed predictive equations (2), and for k_{21} and k_{31}. From these six parameters all other model parameters can be derived (equations are given in (1)). Table 3 gives only mean parameter values for k_{12} and k_{13} and various mass estimates obtained from the two sampling schedules. It will be noted that there is very good agreement between the parameters derived from the two sampling strategies, suggesting that the convenient 6-point schedule does not introduce any substantial bias in the parameter estimates. Also presented in the tables are the coefficients of variation of each of the parameters estimated with the 6-point schedule compared to values obtained via the 36-point schedule. This coefficient of variation provides an estimate of how much the values determined from the 6-point schedule varied from the values obtained by the usual 36-point schedule and was computed by:

$$\text{coefficient of variation (\%)} = \frac{\sqrt{\sum_{i=1}^{26} (P_6^i - P_{36}^i)^2/26}}{\text{Mean } P_{36}} \times 100$$

when P_6^i and P_{36}^i are parameters of the i-th subject estimated from the 6- and 36-point sampling schedules, respectively. The smaller the coefficient of variation, the closer the parameters estimated from the two schedules agree with each other.

The coefficients of variation for PR, M_1, and M_{tot}min were less than 5% and for M_3min the coefficient was 13%. These four parameters are parameters for which we have developed predictive equations (based on physiological variables such as total body weight, excess body weight, age, and plasma cholesterol and triglyceride concentration (2)), and they are determined quite accurately by the new sampling schedule.

Residual error

The residual error for the 36-point strategy (RE_{36}) varied from 2.23 to 8.61%, with a mean of 4.45%. The residual error for the 6-point strategy (RE_6) varied from 1.41 to 6.53%, with a mean of 2.91%. In 23 of the 26 subjects, the residual error was larger for the 36-point strategy

TABLE 3. Mean parameter values for k_{12}, k_{13}, and various mass estimates obtained from both the 36-point and the 6-point sampling schedule

Parameter	36-Point Schedule	6-Point Schedule	Coefficient of Variation (%)
k_{12}	0.077	0.119	133
k_{13}	0.016	0.018	25
M_2min	18.3	15.8	38
M_2int	26.8	22.6	39
M_2max	35.3	29.4	39
M_3int	81.6	80.4	14
M_3max	122.2	117.2	20

than for the 6-point strategy. As discussed in Methods, this is interpreted to mean that there is a significant deviation from the model due to biological variation as well as measurement error.

Assuming that mean RE_6 is an estimate of measurement error, then the measurement error of plasma cholesterol specific activity has a coefficient of variation of 2.9%. The estimated error due to biological variation is 3.1%, roughly comparable to measurement error. The mean overall error rate is 4.4%. None of the residual errors exceeded 10% in any of the 26 subjects studied. Approximately half the residual error appears to be due to measurement error and half due to biological error.

DISCUSSION

It is clearly inconvenient and onerous to require a subject to return to the clinic 35 to 40 times during a 10-month turnover study. In fact, such frequent visits dissuade many subjects from participating in such studies. In theory, six parameters can be estimated from six points, an observation that leads to a convenient sampling schedule. However, several conditions must be met before one can use such a simplified or convenient schedule. First, one must be confident that the three-pool model is a good description of the data, since the convenient strategy does not permit exploration of alternative or more complex models. After studying nearly 60 subjects, with a wide variety of plasma lipid values, in every one of whom a three-pool model was found to fit the data as well as a four-pool model and better than a two-pool model, we now have confidence in the three-pool model as a generally valid description of whole body cholesterol turnover in humans.

Secondly, there should not be serious biological error, that is, large systematic deviations from the model in the plasma specific activity data. This requirement arises because the 6-point strategy will not control for deviations from the model, i.e., for biological error. If one of the samples happens to be drawn during a major deviation from the model then the parameter estimates would be seriously affected by the perturbation. Fortunately, comparison of the residual errors, each expressed as a coefficient of variation, from the two sampling strategies shows that biological variation is comparable in magnitude to measurement error; both are roughly 3%.

Thirdly, parameter estimates produced by the convenient 6-point strategy should not deviate systematically from parameters estimated in the same patient at the same time using the usual sampling strategy. Results presented in Tables 2 and 3 show that mean parameter estimates from the two strategies agree quite closely.

Fourth, the parameter estimates from the 6-point schedule should not vary significantly around those from the usual method. Again, the data presented in Table 2 show that the accuracy of the parameter estimates given by the 6-point strategy (as measured by the coefficients of variation) is quite high for PR, M_1, M_3min, and M_{tot}min. These are the parameters that seem of particular physiological relevance, and they are parameters for which we have developed predictive equations (2). The coefficients of variation for some of the other parameters (e.g., k_{12}, Table 3) are, however, substantial.

Finally, in any given instance, one must be certain not only that a three-pool model will be an adequate description of the plasma cholesterol specific activity-time curve, but also that the dynamics of the curve are not too dissimilar from our prior experience. In our studies of cholesterol turnover in abetalipoproteinemia (3), we had to use a more frequent sampling schedule, and the sampling strategy will similarly have to be validated in other unusual disorders and in patients who are being treated with plasma cholesterol-lowering drugs.

While it may be possible to estimate six parameters with the minimum number of samples (5), from a practical standpoint it is probably unwise to use only the minimum number of samples if it is feasible to obtain a slightly larger number (e.g., nine or ten samples). In any case, times at which samples are drawn are quite critical for accurate parameter estimation. There are several ways of choosing the optimal sampling times (5, 7), all of which depend on knowing the unknown parameter values, i.e., the sampling times depend upon the parameter values. The optimal times were computed for our previously studied subjects and the first three were found to be quite close to days 1, 7, and 24 (or days 1, 8, and 28 for hypercholesterolemic subjects) but after that the times varied considerably. Hence, we have adopted the scheme described in Fig. 2 to select the times for collecting the final three samples.

There is a certain amount of statistical literature on optimal design of experiments with some application to modeling of biological data (10, 11). While Landaw and DiStefano (11) present formulas for computing optimal sampling times, the effect of subject-to-subject variability on optimal times was not discussed and has received surprisingly little attention. Also there has been little work on sequential design within one subject.

Another factor to be considered is the criterion by which different optimal designs are compared. A number of criteria have been suggested (7), but the most widely used is the determinant or D-optimality criterion. In this criterion the standard errors of the parameter estimates are minimized. We have chosen to minimize the coefficients of variation of all six parameters, i.e., the estimated standard error of the parameter estimate divided by the estimated value itself. Times chosen in this way are sensitive to estimates of the slow rate constants (typically 0.005 $days^{-1}$) as well as the much faster (typically 0.02 $days^{-1}$) rate constants. If sampling times were chosen to minimize the average standard errors of the parameter estimates

alone, then the standard error of the larger estimates would predominate and times would be chosen which would be optimal for the fast parameters but not the slow. Scaling the standard errors by the parameter values means that times will be chosen which minimize the overall relative standard error and slow parameters will be estimated as accurately (in a relative sense) as the fast parameters.

We have used this convenient sampling schedule in 50 studies of cholesterol turnover since 1980. The schedule has simplified the study of cholesterol turnover inasmuch as patients need come to the clinic only 6 times for the drawing of blood samples rather than 35 to 40 times as was the case in the past. This has saved considerable professional time per study and has certainly made patients' acceptance of the protocol much easier. Further, we are collaborating with other groups using this convenient strategy in order to estimate cholesterol turnover in special study populations. The convenient sampling schedule now makes it feasible to study long-term cholesterol turnover in substantial numbers of subjects.

The approach used in this study to develop the convenient sampling schedule for cholesterol turnover can be applied to other studies involving parameter estimation from kinetic data. The requirements are that the model or mathematical equation be known, e.g., a sum of three exponentials or a logistic curve, and that there should be sufficient prior experience to permit estimation of the distribution of parameters in the population. When these requirements are met, an optimal sampling schedule may be determined by convenience/compliance considerations as here, by the total amount of blood that may be drawn, or by other practical considerations. In addition, if rapid analysis of early samples is possible, a sequential design approach can be used for optimizing the later sampling times for the individual subject under study.

We thank Ms. M. Myers for expert assistance. This work was supported by grant HL 21006 (SCOR in Arteriosclerosis) from the National Heart, Lung, and Blood Institute, Bethesda, MD.

Manuscript received 27 July 1984.

REFERENCES

1. Goodman, D. S., R. P. Noble, and R. B. Dell. 1973. Three-pool model of the long-term turnover of plasma cholesterol in man. *J. Lipid Res.* **14:** 178–188.
2. Goodman, D. S., F. R. Smith, A. H. Seplowitz, R. Ramakrishnan, and R. B. Dell. 1980. Prediction of the parameters of whole body cholesterol metabolism in humans. *J. Lipid Res.* **21:** 699–713.
3. Goodman, D. S., R. J. Deckelbaum, R. H. Palmer, R. B. Dell, R. Ramakrishnan, G. Delpre, Y. Beigel, and M. Cooper. 1983. Cholesterol turnover and metabolism in two patients with abetalipoproteinemia. *J. Lipid Res.* **24:** 1605–1611.
4. Finney, D. J. 1952. Statistical Method in Biological Assay. 2nd ed., Hafner Publishing Co., New York. 214–216.
5. Box, G. E. P., and H. L. Lucas. 1959. Design of experiments in non-linear situations. *Biometrika.* **46:** 77–90.
6. Goodman, D. S., and R. P. Noble. 1968. Turnover of plasma cholesterol in man. *J. Clin. Invest.* **47:** 231–241.
7. Fedorov, V. V. 1972. Theory of Optimal Experiments. Academic Press, New York. 120–122.
8. Dell, R. B., and R. Ramakrishnan. 1982. A mathematical model for cholesterol kinetics. *In* Lipoprotein Kinetics and Modeling. M. Berman, S. M. Grundy and B. Howard, editors. Academic Press, New York. 313–330.
9. Pearson, E. S., and H. O. Hartley. 1962. Biometrika Tables for Statisticians. Vol. 1. Cambridge. 49–51.
10. DiStefano, J. J., III. 1981. Optimized blood sampling protocols and sequential design of experiments. *Am. J. Physiol.* **240:** R259–R265.
11. Landaw, E. M., and J. J. DiStefano III. 1984. Multiexponential, multicompartmental, and noncompartmental modeling. II. Data analysis and statistical considerations. *Am. J. Physiol.* **246:** R665–R677.