A CRITIQUE OF
COMPARTMENTAL ANALYSIS

Kenneth Zierler
Department of Physiology, The Johns Hopkins University School of Medicine,
Baltimore, Maryland 21205

INTRODUCTION TO THE GENERAL EQUATIONS

Purpose of This Review

In 10 years, from 1969 through 1978, 341 articles were indexed under compartmental analysis. There must be many times that number in which data were analyzed by compartmental analysis but are not retrieved by searching only under that rubric. This article is not a review of those reports; it is a critique of uses and underlying concepts of compartmental analysis. Most of the appraisal of compartmental analysis in this critique has been made previously by others; there have been conferences and symposia (4, 49) on the subject, as well as many thoughtful individual studies. Despite the fact that caveats were published—some repeatedly—long enough ago to be well-known, even a casual thumbing through pages of current journals shows that misconceptions and misapplications persist.

This critique is meant to be helpful to authors, editors, and referees of manuscripts, but experience suggests that the aim will not be entirely fulfilled. There are two recent examples in related areas. In 1965, Dowd & Riggs (11), after analysis of the three common methods for graphic estimate of parameters of enzyme kinetics, concluded that the Lineweaver-Burk plot was most likely to mislead. Despite this publication in a widely distributed journal of biochemistry, an informal count among recent publications showed that the Lineweaver-Burk plot was still used most often. The second example is a report by Nørby et al (32) that more than 50 papers concerning nonlinear Scatchard plots, published in 5 well-regarded journals during a 4-year period, misinterpreted the data, usually owing to incorrect graphic analysis, despite the fact that descriptions and examples of correct procedure have appeared in print repeatedly since at least 1942.
Assumptions Underlying General Kinetic Models

Rescigno (37) gave the following development illustrating that compartmental analysis is a special case of a more general model.

If some substance $Q$ results from some precursor $P$, and if $P$ is labeled so that the concentration of labeled $P$ is represented by the function $g(P, t)$ and the concentration of label in its successor $Q$ is $f(Q, t)$, then, if

$$f(Q, t) = \int_0^t g(P, \tau) h(P, Q, t, \tau) \, d\tau,$$

the system is linear for the functions $g(P, t)$, provided that the function $h(P, Q, t, \tau)$ does not depend on them.

Further, if there is some time interval, $0 < \theta < T$, over which

$$h(P, Q, t, \tau) = h(P, Q, t+\theta, \tau+\theta),$$

Equation 1 becomes the convolution

$$f(Q, t) = \int_0^t g(P, \tau) h(P, Q, t-\tau) \, d\tau,$$

where $h(P, Q, t)$ is the characteristic of the system. The condition defined by Equation 2 is stationarity; the characteristic of the system is independent of the arbitrary time scale of the experiment. This is sometimes referred to as a steady state, but stationarity, as defined by Equation 2, is a more specific statement: The tracer or label is not in a steady state; only the characteristic response of the system is independent of experimental zero time.

Additional Assumptions in Compartmental Analysis

In many applications it is sufficient to determine $h(P, Q, t)$. This alone can be a powerful tool, examples of which will appear at appropriate spots in this article. However, if investigators strive to get more detail about the structure of the biological system in which $P$ and $Q$ are distributed, they must find some ways to take the system apart or to enter all the elements of the distributing system and determine the characteristics of all the elements. Often this dissection is beyond the state of the art. Rather than abandoning aspiration, investigators may make further assumptions about the system in hopes that those assumptions may lead to numerical evaluation of pertinent properties of the distributing system, and that these calculations may even be useful in prediction. The most popular of these assumptions are the basis for compartmental analysis.
There may be no canonical definition of compartmental analysis. With reference to living systems or to inferences about biological systems, the term has been used commonly to describe a method for analysis of data in which, in addition to the basic assumptions that the system is linear and that stationarity holds, two assumptions, and usually three, are made.

Assumption 1 The system is considered to be composed of separate entities, called compartments, into and from which material flows. Material may be exchanged between any compartments and between any compartment and a single phase, called the external or outside phase. The rate at which the quantity of the material under consideration changes in the \( i \)th compartment is the difference between the sum of all inputs to and the sum of all outputs from that compartment,

\[
\frac{dq_i}{dt} = \sum_{j=0}^{n} \rho_{ij} - \sum_{j=0}^{n} \rho_{ji}, \quad j \neq i,
\]

for a system of \( n \) compartments, where \( \rho \) is a rate, in dimensions of quantity per unit time. In the modern convention, adopted by a committee on nomenclature for tracer studies (7), the subscript \( ij \) is read into \( i \) from \( j \). Subscript 0 refers to the external phase, outside the system. Equation (4) is one of a set of \( n \) differential equations. It is simply a bookkeeping simplification of the general mass transfer equation, common in chemical engineering, in which it is assumed that the material under study is neither destroyed nor synthesized in any compartment.

If the assumptions are valid, Equation 4 is undoubtedly true, no matter what the structure of any compartment or of the system. However, experimenters are not usually able to measure the \( \rho \)'s; it is the object of experiment to discover the \( \rho \)'s or some property leading to calculation of the \( \rho \)'s or to some quality governing the \( \rho \)'s. One of the ways in which this problem has been approached is by making the second assumption.

Assumption 2 Each \( \rho \) is proportional to the quantity (or concentration) of the substance in the compartment from which the material flows,

\[
\rho_{ij} = \sum_{j=0}^{n} k_{ij} q_j, \quad j \neq i,
\]

where the \( k \)'s are coefficients of proportionality, with dimension of reciprocal time. It is usual to make two assumptions about the \( k \)'s: it is assumed that the \( k \)'s are independent of elapsed time; it is assumed that the \( k \)'s are independent of the conjugate \( q \), although the \( q \)'s are, of course, functions of time.
Combination of Equations 4 and 5 gives the set of \( n \) differential equations

\[
\frac{dq_i}{dt} = \sum_{j=0}^{n} k_{ij} q_j - k_{ii} q_i, \quad j \neq i,
\]

where

\[
k_{ii} = \sum_{j=0}^{n} k_{ji}, \quad j \neq i.
\]

When the \( k \)'s are independent of the conjugate \( q \)'s, Equation 6 is a \textit{first-order linear differential equation with constant coefficients}. This is the commonly made third additional assumption.

It is important to state explicitly an implication of Assumption 2. Assumption 2 implies that within any given compartment there are no gradients in whatever potentials are appropriate for movement of the material (or energy) under consideration: no chemical potential gradients, no electrochemical potential gradients, no thermal gradients. Each compartment is homogeneous, or well-stirred.

The general three-compartment model is illustrated in Figure 1. In the general \( n \)-compartment system there are \( n(n+1) \) possible \( k \)'s and \( n \) \( q \)'s. Since there are \( n \) equations in the set of Equation 6, if the \( k \)'s are known, the \( q \)'s can be found. If there is access to the compartments so that the concentration, \( c_i \), of the substance in every compartment can be measured, the volume of each compartment, \( V_i \), can be determined by first determining the \( q_i \)'s and by then exploiting the assumed relationship

\[
q_i = V_i c_i.
\]

The volume, \( V_i \), is sometimes called the \textit{space} occupied by the substance, on the assumption that its concentration is everywhere the same in that volume and that it is not sequestered or metabolized. As a hedge against the possibility that these assumptions may be invalid, the space is sometimes referred to as a \textit{virtual volume} (an etymological horror).

\textit{The General Equations of Compartmental Analysis}

Analytical solutions of the set of \( n \) equations (6) for the set of \( q_i \)'s have been given in a number of articles. Indeed, there are many articles in which the authors develop solutions for specific cases, apparently unaware that the solutions had already been published. A useful reference for solutions is the monograph by Jacquez (24).
Figure 1  The general three-compartment model.

Solutions to the set of $n$ equations (3) are each a sum of $n$ exponential terms of the form

$$q_i(t) = (a_i)_0 + \sum_{j=1}^{n} (a_i)_j e^{-\lambda_j t},$$

where for each $q_i$ there is the same set of $n$ exponents, the $\lambda_j$'s and, in general, a different set of coefficients, the $(a_i)_j$'s. The $\lambda_j$'s are functions only of the set of $k$'s. The $(a_i)_j$'s are functions of the $k$'s and the initial conditions and of the inputs to the system; the $(a_i)_0$'s are functions only of $k$'s and inputs from the outside. The initial conditions are $q_i(0)$; the inputs are functions of the $k_{i0}$'s and external concentration. For the general case, illustrated in Figure 1, each $(a_i)_j$ contains initial conditions for all three compartments.

The investigator may measure quantity (or concentration) in one (or more) compartment of the system, but often the investigator is able to measure only the response of the system as a whole. In this case, the quantity in the whole system, as a function of time, is

$$q_s(t) = \sum_{i=1}^{n} q_i(t) = A_0 + \sum_{j=1}^{n} A_j e^{-\lambda_j t},$$

where $A_j = \sum_{i=1}^{n} (a_i)_j$, and $A_0 = \sum_{i=1}^{n} (a_i)_0$. There is, therefore, the same problem for $q_s(t)$ as for any measured $q_i(t)$. The $n \lambda$'s are functions of $n^2$ $k$'s. The $n A$'s are functions of the set of $k$'s, the $n q_i(0)$'s, and the inputs to the system. Formal solution is impossible unless a sufficient number of the $k$'s, inputs, and $q_i(0)$'s are null, in order to reduce the number of parameters to at least $n$. Later, we consider ways in which this has been attempted. We also consider alternatives to formal solutions, in which the number of parameters exceeds $n$.

It will suffice for illustration to specify the coefficients and exponents of the general two-component system. Begin with the general case, in
which
\[
\frac{dq_1(t)}{dt} = \rho_{10} + k_{12}q_2 - k_{11}q_1,
\]
\[
\frac{dq_2(t)}{dt} = \rho_{20} + k_{21}q_1 - k_{22}q_2,
\]
where \(\rho_{10}\) and \(\rho_{20}\), the input functions, may be time dependent.

A common method for solution makes Laplace transforms of the pair of differential equations, to obtain
\[
(s + k_{11})\tilde{q}_1(s) - k_{12}\tilde{q}_2(s) = \tilde{\rho}_{10}(s) + q_1(0),
\]
\[
-k_{21}\tilde{q}_1(s) + (s + k_{22})\tilde{q}_2(s) = \tilde{\rho}_{20}(s) + q_2(0),
\]
where the tilde indicates the function transformed from the time domain, \(t\), to the frequency domain, \(s\). The pair of simultaneous linear equations in \(\tilde{q}_1\) and \(\tilde{q}_2\) are solved by any conventional method. The exponents, the \(\lambda\)'s are the latent roots, or Eigenvalues, of the matrix of rate constants,
\[
\begin{bmatrix}
  k_{11} & -k_{12} \\
  -k_{21} & k_{22}
\end{bmatrix},
\]
and are always real and nonnegative. For this case, the roots are conjugate,
\[
\lambda_1 = \frac{k_{11} + k_{22}}{2} - \frac{1}{2} \sqrt{(k_{11} + k_{22})^2 - 4(k_{11}k_{22} - k_{12}k_{21})},
\]
\[
\lambda_2 = \frac{k_{11} + k_{22}}{2} + \frac{1}{2} \sqrt{(k_{11} + k_{22})^2 - 4(k_{11}k_{22} - k_{12}k_{21})}.
\]

The solutions for \(\tilde{q}_1\) and \(\tilde{q}_2\) are then inverted to the time domain to obtain
\[
q_1(t) = \left(\frac{1}{\lambda_2 - \lambda_1}\right) \left\{ \left[ (k_{22} - \lambda_1)\rho_{10} + k_{12}\rho_{20} \right] e^{-\lambda_1 t} + \left[ (\lambda_2 - k_{22})\rho_{10} - k_{12}\rho_{20} \right] e^{-\lambda_2 t} + \left[ (k_{22} - \lambda_1)q_1(0) + k_{12}q_2(0) \right] e^{-\lambda_1 t} + \left[ (\lambda_2 - k_{22})q_1(0) - k_{12}q_2(0) \right] e^{-\lambda_2 t} \right\},
\]
\[
q_2(t) = \left(\frac{1}{\lambda_2 - \lambda_1}\right) \left\{ \left[ (k_{22} - \lambda_1)\rho_{10} + k_{12}\rho_{20} \right] e^{-\lambda_1 t} + \left[ (\lambda_2 - k_{22})\rho_{10} - k_{12}\rho_{20} \right] e^{-\lambda_2 t} + \left[ (k_{22} - \lambda_1)q_1(0) + k_{12}q_2(0) \right] e^{-\lambda_1 t} + \left[ (\lambda_2 - k_{22})q_1(0) - k_{12}q_2(0) \right] e^{-\lambda_2 t} \right\},
\]
where \(\ast\) is the operator for convolution. The expression for \(q_2(t)\) is symmetrical with that for \(q_1(t)\). The sum of \(q_1\) and \(q_2\) is
\[
q_s(t) = (B_1\rho_{10} + B_2\rho_{20}) e^{-\lambda_1 t} + (B_1q_1(0) + B_2q_2(0)) e^{-\lambda_2 t} + (C_1\rho_{10} + C_2\rho_{20}) e^{-\lambda_1 t} + (C_1q_1(0) + C_2q_2(0)) e^{-\lambda_2 t}.
\]
COMPARTMENTAL ANALYSIS

537

where \( B_1 = (k_{21} + k_{22} - \lambda_1)/(\lambda_2 - \lambda_1), \quad B_2 = (k_{11} + k_{12} - \lambda_1)/(\lambda_2 - \lambda_1), \)
\( C_1 = 1 - B_1, \quad C_2 = 1 - B_2. \)

We cannot segregate terms in \( e^{-\lambda_1 t} \) and \( e^{-\lambda_2 t} \) unless we can specify the input functions, \( \rho_1(0) \) and \( \rho_2(0) \). Although the equation can be used no matter what form the input functions take, it is common practice to design experiments so that the input functions are either constant and non-zero or zero. In metabolic experiments or in studies of material flux when the input functions are constant, for example, by maintaining constant concentration in a bathing solution for organs or tissues in vitro, the tracer is loaded into the system. After the system has been loaded, then the concentration of tracer in the bathing solution is kept at zero, and the tracer is washed out.

In loading experiments, the initial conditions are \( q_1(0) = q_2(0) = 0 \), and \( \rho_1(0) \) and \( \rho_2(0) \) equal \( \bar{\rho}_1 \) and \( \bar{\rho}_2 \), respectively, both constant. Equation (12) simplifies to
\[
q_3(t) = A_0 + A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t},
\]

where \( A_0 = \frac{(k_{21} + k_{22})\bar{\rho}_{10} + (k_{11} + k_{12})\bar{\rho}_{20}}{\lambda_1 \lambda_2}, \quad A_1 = -(B_1\bar{\rho}_{10} + B_2\bar{\rho}_{20})/\lambda_1, \) and \( A_2 = -(C_1\bar{\rho}_{10} + C_2\bar{\rho}_{20})/\lambda_2. \)

An obvious independent measurement is the initial disappearance of tracer from the external phase, during an early period when return of tracer from the system to the external phase can be neglected:
\[
dq_0/dt = -(\bar{\rho}_{10} + \bar{\rho}_{20}), \quad t \to 0_+.
\]

In washout experiments, \( q_1(0) \) and \( q_2(0) \) are greater than zero and the input functions \( \rho_1(0) \) and \( \rho_2(0) \) are null. Equation (12) simplifies to
\[
q_3(t) = A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t},
\]

where \( A_1 = B_1 q_1(0) + B_2 q_2(0), \) and \( A_2 = (1 - B_1)q_1(0) + (1 - B_2)q_2(0), \) \( B_1 \) and \( B_2 \) being defined under Equation 11.

We will return to these equations when we consider methods for estimating values of the coefficients and exponents.

ORIGIN AND USES OF COMPARTMENTAL ANALYSIS

Purposes of Compartmental Analysis

What do investigators hope to learn from compartmental analysis? There are several purposes. Mones Berman (5), whose pioneering studies made it possible for investigators to analyze complicated systems, has defined several major objectives.
One purpose is to simulate aspects of the biological system that seem to the investigator to be important properties. This is often referred to as a model, but it is worth noting that the word model is ambiguous. In mathematics, as defined for example by Suppes (43), a model is a set of instructions for exact reproduction of all properties of a system. In biology we don't know all properties of a system; we can simulate only some properties. The least we ask of a model is that it recall those properties with some acceptable verisimilitude. The number of models that meet such minimum requirements is apt to be large, perhaps infinite. We ought to challenge such models by demanding predictive value. When we alter experimental conditions, how well are results predicted? This is apt to reduce the number of satisfactory models, but, rarely, if ever, can a model be demonstrably unique. (Pioneer investigators of compartmental analysis did speak of tests for uniqueness of a model, but they meant only some index of uniqueness, such as the only one giving arbitrarily satisfactory fit, among possible compartmental models, not among all possible models.) A corollary is that mechanisms in biology cannot be proven by kinetic analysis alone. Nevertheless, the high purpose of simulation is to aid in developing an understanding of biological processes. Therefore, the model should have predictive value when experimental conditions are altered.

A second purpose is to mimic empirically the response of a system without attention to underlying mechanism. In this case it is just one of a large number of possible approximations to an observed unit impulse response, or characteristic, or density function of residence times or of transit times.

A third purpose is to calculate one or more of the parameters of the system. If there are compartments, the investigator wants to find the quantity or concentration of material in, or the volume of, one or more compartments. If there are rate constants, the investigator wants to determine one or more of them. Even if the fluxes or flows are not first-order, so that there are no rate constants to determine, the investigator wants, at least, to determine one or more of the fluxes or flows. Sometimes the purpose can be served simply by calculating the mean transit time or residence time, characteristic of the system as a whole. In this case one needs only the observed $q_x(t)$, either in response to loading or washout. For the case of number of moles of endogenous substance in the system in the steady state, or for the case of fluid volume of a system through which there is constant and equal fluid inflow and outflow, one exploits the relationship $Q = \phi \tau$, where $Q$ is either the number of moles or volume, $\phi$ is total steady input in the appropriate
dimension, and $\bar{t}$ is the mean residence or transit time, determined from analysis of the tracer curve (3, 28).

**Origins of Compartmental Analysis**

It is not clear who first used the expression, compartmental analysis, or multicompartmental analysis. The set of $n$ differential equations describing an $n$th order system of linear components has had wide application in the physical sciences. The first proposal may have been by Fourier (14), who conjectured that heat flow may be proportional to a temperature gradient. Biologists may have been influenced by Fick (13), who adopted Fourier's conjecture to attribute diffusive flow to a concentration gradient, with constant proportionality. These are limiting laws, which state that flows are proportional to forces and, apply over restricted ranges. First order kinetics were familiar to students of chemical reactions. It is not known who was the first to extend this approach to other problems in life sciences, but the approach seems to have arisen independently over a span of several decades in different subdivisions of physiological sciences. Hamilton and his colleagues (18), measured cardiac output in dogs by dilution in arterial blood of a dye injected as a bolus into the venous circulation. Desiring the first-passage concentration-time curve, which usually eluded them because it was obscured by recirculation, they plotted the data semilogarithmically, were satisfied that the downslope ultimately was linear before recirculation occurred, and so reconstructed the tail of the first passage curve. In a recent symposium on pharmacodynamics (36), credit is given to Teorell (44, 45) for two papers in 1937, which founded that discipline. For the first time, distribution of drugs in the body was modeled by multicompartments and compartments were identified with anatomic features. Although theirs was a purely empirical endeavor, Newman and colleagues (30) modeled the central circulation (right heart-pulmonary-left heart) by three mixing chambers in series. Earlier, in 1935, Dominguez et al (10) observed that urinary excretion of creatinine appeared to be proportional to plasma concentration (a relation suggested by semilogarithmic plots) and wrote appropriate linear equations.

In 1943, Zilversmit et al (52) published a seminal paper on the use of radioisotopes in estimating exchange rates in a multicompartment system. In a different field, Smith & Morales (40, 41), in 1944, published two important papers dealing with the distribution of inert gases in the body, in which a limb of an animal was treated as consisting of $n$ compartments. The second paper includes, as far as I have been able to determine, the first description of the graphic method now called
curve-peeling or curve-stripping, although Smith & Morales did not use either term. Curve-peeling is the most widely used method for empirical estimate of the number of exponential terms and the coefficient and exponent of each term. The simplicity of the method has had a great deal to do with the popularity of compartmental analysis, because one is guaranteed to get numbers. Yet, I have never seen proper attribution made by users of the technique; it is as though it is something humans have always known, like counting. The most frequently cited paper for method of analysis is by Matthews (27), who, in a review, dealt specifically with mammillary systems, those in which only one compartment is open to the outside and with which all others communicated. Even earlier, Greville (16) described a graphic method to confirm existence of a monoexponential function purporting to describe disappearance of administered glucose from blood and to obtain its rate constant by plotting the time-derivative against the function—what we recognize as a phase-plane plot. Greville’s method was rediscovered and extended 19 years later by Perl (33), who used it for an n-compartment system to determine the number of compartments, in combination with curve-peeling. Greville’s method is not used widely, although it is probably less likely to misinform than simply peeling off semilogarithmic plots of \( q_s(t) \). Although the latter is the most popular method, little attention is paid to its uncertainties, which are considered later.

Widespread use of radioisotopes in metabolic studies led to probes of complex biological systems. Interpretation of the data required a model. Multicompartmental analysis was an appealing model because it was so simple in concept and because it seemed to give the investigator more information about the system. It seemed to promise that resolution of the observed overall function \( q_s(t) \) gave insight into the size of each compartment and exchange rates between compartments and between compartments and the outside, without the need to isolate each compartment and obtain this information by a battery of separate experiments, which were often not technically possible.

Multicompartmental analysis was greatly facilitated by introduction of digital computation. Berman and colleagues [for bibliography see (6)] wrote a program that, by iterative adjustment to a least squares fit, reports all the parameters of the system, and, therefore, all non-zero connections among compartments, even in a system with more than 20 compartments. Other programs have been published since then; some deal with less general cases—for example, one designed to solve the two-compartmental open model (35). An interesting program, to be used prior to attempts to estimate values of parameters, is designed to
check the possibility that all unknown systems parameters can be estimated by a multi-input multi-output tracer experiment (9).

**Definition of a Compartment**

What is a compartment? It is, first of all, an ideational construct in which is assembled some quantity of a given species distributed randomly so that there are no potential gradients within it; the only forces driving the substance are between that compartment and the world outside it. As such, a compartment may not correspond to any anatomic feature. However, it may correspond to some anatomic feature, depending on the kind of question being asked. Even though there may be potential gradients within a cell, or even in an organ, they may have only negligible influence on overall transfers between, say, a cell or population of cells and the external phase. Even more broadly, compartmental analysis is just a set of $n$ first-order linear differential equations. Therefore, any process so described can be considered in its domain. Hearon (21) pointed out that chemical species can be treated formally as compartments and that chemical conversion can be treated formally as flow from one compartment to another. An irreversible step in a reaction is the same as loss from a compartment to the external phase. The simplest compartmental system has just one compartment. Therefore, any process described by a single first-order linear differential equation is pertinent, and some lessons from such systems are helpful in prudent handling of problems dealing with more complicated cases.

Once we admit this broad definition of compartmental analysis, it is evident that its use has been varied. In its more restricted sense it has been used to study distribution of drugs, inert gases, and so on, in the whole animal, to study intermediary metabolism in whole animals and in whole organs, to study material flux across cell membranes, blood flow and transcapillary exchange, and more. It has been used to calculate pool size and volume of distribution. It has been used to approximate diffusive flow, radioactive decay, fluorescence decay, etc.

**WHEN IS COMPARTMENTAL ANALYSIS APPROPRIATE?**

How does one decide that compartmental analysis is appropriate? If it is appropriate, how does one decide on the number of compartments and on the manner in which they are connected? How does one find the parameters? How does one convince oneself that the model is reliable?
What is the physical (real world) meaning of the parameters? Of the compartments? If compartmental analysis is not appropriate, what are the alternatives?

These questions have often been contemplated and debated. It is agreed that the more one knows about the system under study, the better. One uses all reliable information to begin model formulation. One must have a clear idea of the question asked and design experiments to answer that question. The practice of data collection and post hoc model-making is not likely to yield incisive results. There are clear examples of appropriate conditions for compartmental analysis and clear examples of inappropriate conditions, but most of the areas are gray. It is obvious that radioactive or fluorescence decay, no matter in what other processes it is embedded, is described by first-order differential equations, as are chemical reactions known to be first-order. On the other hand, there are continuous processes for which other formulations are known. For example, streamline or parabolic fluid flow in tubes is properly described by nonexponential equations. A solution for diffusion equations is a Bessel function. Compartmental analysis has been applied to diffusion, and there are circumstances in which this is useful, but the parameters of the functions do not have the physical meaning usually given them in compartmental analysis. The mimicry works because the Bessel function can be expressed as an infinite series of exponential terms. If the series is weighted heavily by the first one or two terms, the result could be fitted by compartmental analysis, but the coefficients and exponents would not give the information usually sought by compartmental analysis. An example of a useful circumstance is an experiment in which a whole organ or tissue is incubated in a well-stirred bath. The tissue contains interstitial fluid through which the tracer must pass to reach the bath after it leaves cells. Transfer through interstitial fluid is diffusive. However, the cell membrane barrier may be such a bottleneck, and it usually is for water-soluble substances, that transfer through interstitial fluid is an order of magnitude more rapid than any other transfer in the system; that is, the first term of the infinite series may so dominate that the treatment of the process as a single exponential may be an acceptable representation of transfer out of cells. Whether or not this is the case depends on an understanding of the relative contributions made by the several processes in the given system and on the resolution required. For example, if the second term of the series expressing the contribution by diffusion is close to one of the membrane transport terms, then the analysis will give misleading results.
There are occasions in which the likelihood of first-order kinetics is inescapable, even though there may be no compelling a priori argument. For example, Larsen & Lassen (25) reported that washout of an inert gas from adipose tissue in situ in man, a phenomenon presumably dependent only upon blood flow, was monoexponential over a 24-hour period, during which all but a small percent was removed from the tissue. This is not generally true for inert gas washout from other tissues, but in this case it would be pointless to avoid formal analysis by a first-order equation. A similar but less dramatic experience by Rogus & Zierler (39) occurred in studies of $^{24}$Na efflux from mammalian skeletal muscle. Washout could be followed until the radioactivity in the whole tissue, including interstitial space, was no greater than 1 part in $10^4$ of initial activity. It was inescapable that the curves, after correction for the contribution from interstitial fluid, were resolvable into sums of two exponential terms; the coefficient of the slow component was only about 5% of that of the fast component. The slow component could not have been detected had it not been possible to continue the observations for such a long period.

There are cases in which a monoexponential response occurs unexpectedly. For example, it is not surprising that $^{42}$K efflux from a single muscle fiber might be described by a monoexponential, at least until only 10% of $^{42}$K remains in the fiber. However, because the rate constant is, in theory, inversely proportional to fiber radius and because in a whole muscle there is a distribution of fiber radii usually over at least a twofold range, it is expected that washout of $^{42}$K from a whole muscle would be described by a sum of exponential terms, reflecting the distribution of radii. In fact, however, $^{42}$K washout from a whole muscle gives a good fit to a monoexponential. The explanation lies in the insensitivity of semilogarithmic plots. Even if the two rate constants differ by a factor of two, it is difficult to distinguish a biexponential from a monoexponential, particularly in the presence of biological and analytical noise. At least a factor of three is required (50). The question is, What is the meaning of the single coefficient and the single rate constant? The coefficient, $A$, is the sum of all the $A_i$ of the $n$-exponentials contributed by the population of cells, and the single $\lambda$ is some weighted mean of all the $\lambda$'s. If the initial slopes are indistinguishable, $\lambda A = \Sigma \lambda_i A_i$; hence, $\lambda = (\Sigma \lambda_i A_i)/(\Sigma A_i) = \bar{\lambda}$; that is, the zero time intercept should be the quantity of $^{42}$K in the muscle fibers at zero time, and the rate constant is a mean of rate constants for the fiber population.

In general, however, with the exception of clear-cut physical phenomena known on independent grounds to be described by first-order linear equations, there is usually no a priori case for compartmental analysis.
This means that when compartmental analysis is used, it must be used only as an exercise in curve-fitting, in which no real-world meaning is attributed to the coefficients and exponents or to the number of terms; or, if it is the investigator's aim to associate these with properties of the biological system, there must be tests of the validity of these assignments. Some tests are described later.

HOW TO CARRY OUT THE ANALYSIS

Starting Out (Including a Critique of Curve-Peeling)

How does one begin compartmental analysis? Berman (5) has given a full description. First, it is best if the investigator knows in advance, on independent grounds, how many compartments there are and how they are connected. He may know one and not the other. For example, in studies of fluorescence lifetimes, it may be accepted that the lifetime of one fluorophor is independent of another. Therefore, if the decay curve from which the time course of the excitation light component is deconvolved is a sum of \( n \) exponentials, then one is justified in associating each slope with the decay of a separate fluorophor. In metabolic studies in whole organs or animals, the situation is less clear, but all knowledge of the system may help in assigning numbers of components. If each compartment is a separate organ and if the connections are only vascular, knowledge of the circulatory system will suggest the hookup between compartments and between compartments and the external phase. Knowledge of the means of excretion, whether only urinary or only in bile, for example, will locate the coefficient of loss from the system, the \( k_{0i} \).

If one knows the number of components but not all possible interconnections, it is sometimes possible to design experiments to give specific information on this matter, which is just the problem of determining which \( k_{ij} \)'s are zero and which are non-zero.

In the absence of at least a tentative sense that there are exactly \( n \) compartments, and, usually, even with that sense, a custom is to begin with curve-peeling to find the number of compartments. How reliable is curve-peeling?

First, the raw data must be impeccable. Even under the best circumstances, for a monoexponential the uncertainty of the estimate of slope is several times the standard deviation of an experimental point. For example, Figure 2 is a photograph of a semilogarithmic fluorescence decay curve, which is the convolution of a monoexponential on the time distribution of a brief photic excitation. Biological data are rarely as
Photographic record of fluorescence decay to illustrate uncertainty of determination of slope of a monoexponential function under conditions that are far more reliable than is usually possible in biological experiments.

Reliable. There are nearly $10^6$ total counts. The peak is $2 \times 10^5$. The last point is 50. The slope, between amplitudes of $10^5$ and $10^3$, is $0.0532 \pm 0.0056 \text{ nsec}^{-1}$, an uncertainty of 11%. This occurs despite the fact that at $10^5$ the uncertainty is only 0.3% and at $10^3$ the uncertainty is only 3%. A coefficient of variation of about 5%, which is very good for biological data, usually leads to an uncertainty of about 15% in estimate of slope of a semilogarithmic plot.

Second, the data must cover a broad range, the extent of which depends on what is sought. In almost all cases, that range ought to extend down to a few percent of peak for washout experiments and possibly further if contributions from slow components with small intercepts are sought. In many experiments in biological preparations, such a range is not attainable, either because the baseline noise is large or the steady state cannot be held sufficiently long, or for other reasons. Although this is an unfortunate fact of life, it greatly reduces confidence in the resulting curve-peeling. One of the first things a reader should do in looking at semilogarithmic plots is to identify the range. There are many reports in which the last point is as much as 70% of the peak in washout experiments, and it is quite common to find it not less than 40%.
How reliably does curve-peeling demonstrate the number of exponential terms? It is rare that one can by graphic analysis decompose the function into more than three exponential terms. Errors introduced by curve-peeling are not random, because errors in estimate of the first peel make the residue erroneous, so that error in estimate of the second peel is compounded. Error in estimate of the slope leads to error in the same direction in estimate of the intercept. Too large an estimate of the first peel leads to too small an estimate of the second. But the key question is, If one, by curve-peeling, finds \( n \) exponentials, how does one know there are not \( n + 1 \)? Unless one can pursue the data until the function is complete, one cannot know; there may be more terms. Furthermore, even if there are in fact only \( n \) terms, unless one has a sufficiently complete curve one may indeed peel off \( n \) terms, but the coefficients and exponents may all be wrong. An example appears in Figure 3. This is washout of \( ^{24}\text{Na} \) from a two-compartment system. When followed for

![Graph showing washout of \( ^{24}\text{Na} \)](https://www.annualreviews.org/aronline)

**Figure 3** The two curves are from the same set of data. Curve A is truncated after only 70 min, yet is resolvable into a sum of 2 exponential terms, neither of which matches either of the 2 exponential terms into which the more complete curve is resolved, despite the fact that curve A decayed to only 1% of initial value. (Reprinted from *Computer Processing of Dynamic Images*, ed. K. B. Larson, J. R. Cox, Jr., Soc. Nucl. Med., NY)
only 70 min, the curve is resolvable into a sum of two exponentials. When observations are continued for two hours, the curve is still resolvable into two exponentials, but neither exponential is the same as estimated from shorter washout (51).

How does one know that the curve really is an exponential? Unless there are independent reasons, as in fluorescence decay, one doesn't. The custom is to test the straight line against observed values for goodness of fit by some least squares method. Acceptance of the result is subjective. An acceptable fit can be erroneous. Many examples exist in many fields. On at least one occasion misinterpretation had tragic results. In the example in Figure 4, the decay curve is acceptably small. But note that the early points and the late points tend strongly to be above the line, and the middle points tend to be below the line; differences are not random. Extrapolation of the line to negligible quantities occurs before the real curve reaches negligible quantities. If

![Figure 4](image_url)

*Figure 4* A curve treated mistakenly as a monoexponential. Abscissa: arbitrary time; ticks at intervals of 10 arbitrary units. Ordinate: log scale. Experimental values, x. Continuous line, best fit monoexponential with time constant 0.052 per time unit and regression coefficient < 0.999. In fact, the experimental points were generated by imposing slight noise on the function \( y = 80e^{-0.069t} + 20e^{-0.034t} \). Tip-off that the plotted values were not generated by a monoexponential is that early and late points are above the line, middle points are below it; deviations are not random.
this were a virus-kill curve in response to formalin, the error in estimate of time required to destroy live virus would be serious. There is a way to avoid this error. In the example illustrated, the error is obvious on inspection, but it can be subtler. To avoid the error, the reconstructed curve should be tested against the observed curve by autocorrelation. In the example in Figure 4, the autocorrelation function would have large-amplitude oscillations. The practice of testing by autocorrelation is routine in some laboratories, for example, in testing reconvolved functions against observed fluorescence lifetime curves, a practice introduced by Grinvald & Steinberg (17) whose paper gives references to the method.

One must always remember that all sorts of nonexponential functions can be decomposed into sums of exponentials, the number of which depends on the range of observation. Some examples of exponential masquerade have been published (51). Goodness of fit provides no assurance that the decomposition into exponentials is warranted. René Thom (46) gives the delightful example shown in Figure 5. Two models are tested against real data. Over the range indicated, model 1 fits by statistical tests far better than model 2. But, as Thom points out, anyone looking at the extended range would know that model 1 did not describe the system that gave rise to the data, whereas model 2 was on

![Figure 5](image-url)  
*Figure 5* Curves generated by two tentative models, presented with plot of observed data. The model that gives the better fit to observed data over a limited range does not describe the system, and the model that gives the poorer fit appears to deserve further study. The best fit by statistical tests does not guarantee the correct model. (Reprinted from René Thom, *Structural Stability and Morphogenesis*, Benjamin-Cummings, Reading, Mass.)
the right track, differing only by a simple transformation of coordinates. There are examples in the literature in which the best fit was to a model shown subsequently to be wrong, a result that has occurred due to systematic experimental error.

**Analytical Solutions**

How does one determine hookup among compartments? Analysis of a two-compartment system illustrates one method. The general two-compartment system has six rate constants. How many are zero? Possible arrangements of two-compartment systems open to the outside are diagrammed in Figure 6. Table 1, developed from Equations 13 and 14, lists the $A$'s and $\lambda$'s for each arrangement for loading and washout experiments. The $A$'s refer to coefficients of equations describing the time course of the quantity of material in the entire system, $q_s(t)$. In Table 1, $A_0$ applies to loading experiments. In washout experiments, $A_0 = 0$. $A_1$ in Table 1 applies to washout experiments. $A_2$ for washout experiments is obtained from the relationships, $A_1 + A_2 = q_1(0) + q_2(0)$. For loading experiments, the $A_1$ is obtained from the tabulated washout $A_1$ by changing the sign and by substituting $\bar{\rho}_{10}/\lambda_1$ and $\bar{\rho}_{20}/\lambda_1$ for $q_1(0)$ and $q_2(0)$, respectively. In configurations 2, 4, 5, 7, and 8 there is no input from the outside into compartment 2, and terms in $\bar{\rho}_{20}$ disappear. $A_2$ in loading experiments is obtained from the loading relationship $A_0 + A_1 + A_2 = 0$ or by changing sign of the washout $A_2$ and substituting $\bar{\rho}_{10}/\lambda_2$ and $\bar{\rho}_{20}/\lambda_2$ for $q_1(0)$ and $q_2(0)$, respectively. For loading experiments for substances not destroyed in the system, $q_s(t)$ can be determined either by measuring the quantity in the system or by measuring the loss of the substance from the external phase to the system. In washout experiments, $q_s(t)$ is obtained either as the quantity remaining in the system or as the quantity recovered from the system in the external phase.

Often it is possible to measure the quantity in one of the compartments, particularly one into which there is an external input, say $q_1(t)$. The equation for $q_1(t)$ in the general two-compartment system (configuration 1 of Figure 6) is, for loading,

$$ q_1(t) = \frac{k_{22}\bar{\rho}_{10} + k_{12}\bar{\rho}_{20}}{\lambda_1\lambda_2} - \frac{(\lambda_2 - k_{11})(\bar{\rho}_{10}/\lambda_1) + k_{12}(\bar{\rho}_{20}/\lambda_1)}{\lambda_2 - \lambda_1} e^{-\lambda_1 t} $$

$$ - \frac{(k_{11} - \lambda_1)(\bar{\rho}_{10}/\lambda_2) - k_{12}(\bar{\rho}_{20}/\lambda_2)}{\lambda_2 - \lambda_1} e^{-\lambda_2 t}, \quad 16. $$
Figure 6 The 10 possible configurations of two-compartment systems open to the outside, with no blind compartments. See Table 1 for corresponding equations.

and for washout,

\[ q_1(t) = \frac{(\lambda_2 - k_{11})q_1(0) + k_{12}q_2(0)}{\lambda_2 - \lambda_1} e^{-\lambda_1 t} \]

\[ + \frac{(k_{11} - \lambda_1)q_1(0) - k_{12}q_2(0)}{\lambda_2 - \lambda_1} e^{-\lambda_2 t}. \]

The exponents, \( \lambda_1 \) and \( \lambda_2 \), are the same for \( q_1(t) \) as given in Table 1 for \( q_s(t) \). From these general equations for \( q_1(t) \), equations for the other nine configurations are obtained by nulling terms containing null rate constants and substituting simplified expressions for the \( \lambda \)'s. From this the reader can construct a table of coefficients for \( q_1(t) \) similar to that given for \( q_s(t) \).

Inspection of Figure 6 and Table 1 shows the following. In loading experiments, each configuration leads to a formally unique equation. This is not true of washout experiments in which there are three pairs of identical equations and a fourth pair that are equivalent if compartments 1 and 2 are reversed (configurations 8 and 10). Therefore, from washout equations alone, there are six configurations that lead to no unique solution without independent experiments to get at the components of the system. Moreover, even in loading experiments there are systems that will lead to indistinguishable solutions in the absence of additional information. For example, configurations 2 and 5 differ only in that \( k_{02} \) is null in 5. It is unlikely that if, in configuration 2, \( k_{12} \) were large compared to \( k_{02} \), a best-fit computer program could distinguish between the two configurations.

An important use of Figure 6, Table 1, and the reader's tabulation of coefficients for \( q_1(t) \) is to suggest experiments designed to give more detailed information about the configuration and numerical values of the parameters. Possibilities are numerous. The following examples are
### Table 1 Parameters of the ten configurations of two compartment systems

<table>
<thead>
<tr>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
</tbody>
</table>
| \[
\frac{(k_{21} + k_{22}) \bar{p}_{10}}{\lambda_1 \lambda_2} + \frac{(k_{11} + k_{12}) \bar{p}_{20}}{\lambda_1 \lambda_2} = \frac{(\lambda_2 - k_{01}) q_1(0) + (\lambda_2 - k_{02}) q_2(0)}{\lambda_2 - \lambda_1} + \frac{k_{11} + k_{22}}{2} + \frac{1}{2} \sqrt{(k_{11} + k_{22})^2 - 4(k_{11}k_{22} - k_{12}k_{21})}
\] |
| 2     |
| \[
\frac{(k_{21} + k_{22}) \bar{p}_{10}}{\lambda_1 \lambda_2} = \text{same as 1}
\] |
| 3     |
| \[
\frac{(k_{21} + k_{12}) \bar{p}_{10}}{\lambda_1 \lambda_2} \quad \frac{(\lambda_2 - k_{01}) q_1(0) + (\lambda_2 - k_{02}) q_2(0)}{\lambda_2 - \lambda_1} \quad \frac{k_{11} + k_{12}}{2} + \frac{1}{2} \sqrt{(k_{11} + k_{12})^2 - 4k_{01}k_{12}}
\] |
| 4     |
| \[
\frac{(k_{21} + k_{12}) \bar{p}_{10}}{\lambda_1 \lambda_2} = \text{same as 2}
\] |
| 5     |
| \[
\frac{(k_{21} + k_{12}) \bar{p}_{10}}{\lambda_1 \lambda_2} \quad \text{same as 3}
\] |
| 6     |
| \[
\frac{(k_{21} + k_{22}) \bar{p}_{10} + k_{11} \bar{p}_{20}}{k_{11}k_{02}} = \frac{(k_{02} - k_{01}) q_1(0)}{k_{02} - k_{11}} \quad \frac{k_{11} + k_{12}}{2} \quad \text{same as 6}
\] |
| 7     |
| \[
\frac{(k_{21} + k_{22}) \bar{p}_{10}}{k_{11}k_{02}} = \text{same as 6}
\] |
| 8     |
| \[
\frac{(k_{21} + k_{22}) \bar{p}_{10}}{k_{21}k_{02}} = \frac{k_{21}q_1(0)}{k_{02} - k_{21}} \quad k_{21}, k_{02}
\] |
| 9     |
| \[
\frac{\bar{p}_{10}}{k_{01}} + \frac{\bar{p}_{20}}{k_{02}} \quad q_1(0) \quad k_{01}, k_{02}
\] |
| 10    |
| \[
\frac{\bar{p}_{10}}{k_{01}} + \frac{(k_{01} + k_{12}) \bar{p}_{20}}{k_{01}k_{12}} = \frac{q_1(0)}{k_{12} - k_{01}} + \frac{k_{12}q_2(0)}{k_{12} - k_{01}} \quad k_{01}, k_{12}
\] |

*aFormal expressions for coefficients and exponents of observed loading and washout curves for all non-equivalent possible configurations of a 2-component system open to the outside. For definitions see Equations 9 through 15 and related text. \( A_1 \) is given for washout experiments. For loading experiments, change sign and see text for guidance in substituting \( \bar{p}_{10} \) and \( \bar{p}_{20} \) for \( q_1(0) \) and \( q_2(0) \), respectively. \( A_2 \) is determined from \( A_2 = q_1(0) + q_2(0) - A_1 \), in washout experiments, and from \( -A_2 = A_0 + A_4 \), in loading experiments. It is assumed that the observed function is \( q_s(t) \), the quantity in the total system. For formulation of a similar table for the case of measurement of the quantity in one of the compartments, see text.*
intended to illustrate and to suggest how the investigator may find other approaches appropriate to his experimental conditions.

First, consider possibilities for formal analysis. Begin with the assumption that either it is known on other grounds that one is dealing with a two-compartment system open to the outside or that some curve-peeling technique has been so reliable as to make the possibility highly likely. In a loading experiment, the knowns are $q_s(t)$, $A_0$, $A_1$, $A_2$, $\lambda_1$, and $\lambda_2$. The unknowns are $\bar{\rho}_{10}$, $\bar{\rho}_{20}$, $k_{01}$, $k_{21}$, $k_{12}$ and $k_{02}$. Since the sum of the three $A$'s is zero, the $A$'s are in fact a set of only two independent values. Furthermore, when $q_s(t)$, $A_0$, $A_2$, and $\lambda_2$ are known, $\lambda_1$ and $A_1$ are available; that is just what curve-peeling is all about. There are, therefore, only four independent statements about loading experiments for the general two-compartment system: $q_s(t) = A_0 + A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t}$, the equations for $A_0$ and for $A_1$ for configuration 1, Table 1, and the equation for either $\lambda_1$ or $\lambda_2$ but not both. The statement that $A_0 + A_1 + A_2 = 0$ is implicit in the equation for $q_s(t)$ because at $t=0$, $q_s(0)=0$ in loading experiments. The unknowns are not entirely independent. Obviously, if any 3 $k$'s are known, the remaining $k$ can be found since it is just a linear function of one $\lambda$ and the other $k$'s. In addition, the following establishes an equation for determining one of the $\bar{\rho}$'s if the other is known.

$$\frac{dq_s(t)}{dt} = \bar{\rho}_{10} + \bar{\rho}_{20} - k_{01} q_1(t) - k_{02} q_2(t).$$

At $t=0$, $q_1$ and $q_2$ are zero. Therefore,

$$\frac{dq_s(0)}{dt} = \bar{\rho}_{10} + \bar{\rho}_{20} = -\lambda_1 A_1 - \lambda_2 A_2.$$  

Thus, $(\bar{\rho}_{10} + \bar{\rho}_{20})$ can be found either from the initial slope of the loading curve or from the formal statement relating the empirical $\lambda$'s and $A_1$ and $A_2$. There are, thus, three independent unknown $k$'s and an independent $\bar{\rho}$. However, in arriving at the statement for the sum of $\bar{\rho}_{10}$ and $\bar{\rho}_{20}$ we had to use one of the four independent statements, leaving only three equations for the four unknowns. There is, therefore, no possible unique solution without additional information.

Two kinds of additional information help.

One might know qualitatively how the compartments are hooked up. For example, if one had good reason to state that there is only one compartment open to the outside, say compartment 1, this is configuration 5. The only input, $\bar{\rho}_{10}$ is found from Equation 18. From Table 1, $A_0$ gives an expression for the sum $(k_{21} + k_{12})$. The only unknown in the loading $A_1$ is $k_{01}$. When the expression $\lambda_1$ is expanded, with substitution
for \((k_{12} + k_{21})\) from \(A_0\), \(k_{12}\) is obtained. Since \(k_{11} + k_{22} = \lambda_1 + \lambda_2\) and, in this configuration, \(k_{22} = k_{21}\), \(k_{11}\) is obtained. But \(k_{11} = k_{01} + k_{21}\), from which the unknown \(k_{21}\) is found. Thus, all \(k\)'s and the single \(\bar{p}\) are obtained formally. The amount of uncertainty with each solution is great because each estimate contains cumulative errors of estimate of \(\lambda\)'s, \(A\)'s and the successive inputs and rate constants.

A second possible approach is to enter one of the compartments, say compartment 1, obtaining \(q_1(t)\) as well as \(q_2(t)\). Again, formal solution for all \(k\)'s and both \(\bar{p}\)'s is possible by making use of the initial slope of \(q_1(t)\) to obtain \(\bar{p}_{10}\), from which result \(\bar{p}_{20}\) is found. From \(A_0\), \(A_1\), \((a_1)_0\), and \((a_1)_1\) and from the expressions for the \(\lambda\)'s there are now ample linear equations to obtain the four \(k\)'s. But again, formal estimates will have increasing unreliability as the errors of estimate of compartments accrue.

Other analytical approaches have been taken; one approach assumes that instead of a discontinuous system of \(n\) compartments there is a continuous distribution of time constants. These approaches have been analyzed critically by Peslin et al (34), whose paper should be consulted also for references to these alternative proposals. Peslin et al found no advantage and additional errors in these methods.

These are examples of formal analytical solutions. Whether solutions are obtained this way or by a trial and error best-fit computer program, as developed first by Berman (6), there should be tests for the ranges of uncertainties of the estimates. Rockoff (38) proposed some sensible tests. The simplest and most important is based on trials with other values for the rate constants and other unknowns. If it is possible to substitute other values, differing by some arbitrary percent from those calculated, with little or no change in the synthesized function, then the system is poorly sensitive to those rate constants. Van Liew (48) has a critical and helpful study of graphic analysis of multicompartmental functions, including some special case examples and principles guiding investigators to innovation of other forms designed for special purposes. Again, as suggested earlier in this critique, the function \(q_2(t)\) should be synthesized from the calculated mean value, and the synthetic and observed functions should be tested for autocorrelation.

**Tests for Partial Solutions**

Even when analytical solutions are impossible, one can still learn a good deal about the structure of the system, and one may even be able to estimate some or all of the rate constants and pool sizes as satisfactorily as if there were an analytical solution. For example, consider a two-
compartment system in which both compartments are known to be reversibly open to the outside. It is unknown how, or whether, they are connected to each other. There are three possible configurations: 1, 6, and 9. We now perform the following experiment based on prior information about the system. Suppose we have reason to know that most or all of the property that determines $k_{01}$ is metabolically dependent; furthermore, there is a specific inhibitor of that metabolic process, which, in small concentrations, affects no other rate constant and which acts so rapidly compared to all time constants of the system that it can be considered to act instantly. This means that upon its addition there is a step change from $k_{01}$ to some substantially smaller $k'_{01}$. We carry out a paired experiment in which specimens of the biological material that form the system are incubated in the absence of the inhibitor with the loading solution until a steady state obtains. The specimens are then divided into two groups. Group C, the controls, is washed out in the absence of inhibitor. Group E is washed out in the presence of inhibitor. What happens to the $\lambda$'s and the $A$'s when $k_{01}$ is step-changed? To obtain the answer, the $\lambda$'s and $A$'s for the various possible configurations are differentiated with respect to $k_{01}$. The major points of analysis of those derivatives are tabulated in Table 2, which shows that the set of changes in the $A$'s and $\lambda$'s unequivocally distinguishes formally among the three possible configurations.

Configuration 9 is the parallel configuration. It is the only one in which the washout $A$'s are exactly the $q_i(0)$'s for any $n$-compartment system. It is, therefore, the only one for which the $A$'s measure pool size formally, within analytical error. Although there are other configurations for which the step change in one rate constant may not alter the

<table>
<thead>
<tr>
<th>Configuration</th>
<th>$dA_1/dk_{01}$</th>
<th>$d\lambda_2/dk_{01}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+, −, 0</td>
<td>+</td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Among two-compartment systems, configurations 1, 6, and 9 can be distinguished by washout in paired experiments under conditions in which one rate constant, say $k_{01}$, is changed in one member of the pair and not affected in the other. Differentiate $A$’s and $\lambda$’s (given in Table 1) with respect to $k_{01}$. $\lambda_2$ is taken as greater than $A_1$, and $k_{11} = k_{01} + k_{21}$ is less than $k_{22}$. In configuration 1, $\lambda_2$ changes, and $A_1$ may or may not change. In configuration 6, $A_1$, but not $\lambda_2$, changes, and in 9 neither changes, compared to the control.*
$A$'s, independence of the $A$'s from changes in rate constants is characteristic only of the parallel configuration. For example, for configuration 1, if a step change in $k_{01}$ does not alter the washout $A$'s, it is because the ratio

$$\frac{q_1(0)}{q_2(0)} = \frac{2k_{12}k_{21} + (k_{22} - k_{11})k_{12}}{2k_{12}k_{21} - (k_{22} - k_{11})k_{21}}.$$ 

This, of course, requires a coincidence that is apt to be unusual, and, in most cases for configuration 1, the washout intercepts will change in response to a step change in rate constant. Furthermore, even when the ratio $q_1(0)/q_2(0)$ happens to be such that the $A$'s do not change with a step change in $k_{01}$ for configuration 1, there will be a change in the $A$'s when, say, $k_{02}$ is changed, because the requirement that $dA_1/dk_{02} = 0$ for configuration 1 is different from the requirement that $dA_1/dk_{01} = 0$.

The parallel configuration is the only one for which a step change in rate constant at zero washout time never alters the $A$'s. This is a generally useful distinction.

There is a similarly useful distinction between a one-compartment system and all others. The phase-plane plot, $dq(t)/dt$ against $q(t)$, proposed by Greville (16) and Perl (33), is linear over its entire length only for a monoexponential, and this plot will pick up early and late departures from a monoexponential more reliably than a semilogarithmic plot of either alone.

**Estimates From Inequalities**

We turn to the question of incomplete estimates of rate constants, using a method illustrated by the study by Rogus & Zierler (39). For the general $n$-compartment system, there is formal symmetry in the analytical expressions for the $A$'s. The real system may be very asymmetric. There may be large differences among pool sizes and among rate constants. These can be exploited.

Begin with

$$k_{11} + k_{22} = \lambda_1 + \lambda_2$$ 

which follows at once from the expressions for the $\lambda$'s.

Recall Equation 10 for $\lambda_2$. It is recast to

$$2\lambda_2 = k_{22} + k_{11} + \sqrt{(k_{22} - k_{11})^2 + 4k_{12}k_{21}},$$

from which

$$\lambda_2 > k_{22}.$$
With this inequality and Equation 19 and for \( \lambda_2 > \lambda_1 \) and \( k_{22} \) set arbitrarily greater than \( k_{11} \), we draw the number line (Figure 7) and note the following:

\[
0 < \lambda_1 < k_{11} < k_{22} < \lambda_2, \\
k_{11} > b = \lambda_2 - k_{22} = k_{11} - \lambda_1 > 0, \\
k_{22} > a = \lambda_2 - k_{11} = k_{22} - \lambda_1 > 0.
\]

Therefore, both \( k_{02} \) and \( k_{12} \) are less than \( \lambda_2 \), but either \( k_{12} \) or \( k_{21} \) or both may be greater than \( \lambda_1 \). Note further that

\[
a + b = \lambda_2 - \lambda_1, \\
a - b = k_{22} - k_{11}, \\
ab = k_{12}k_{21}.
\]

The problem is to place more restrictive limits on \( a \) and \( b \). Detailed solutions have been given (39). In the reference cited, the authors were able to confine some of the estimates within such small limits that the uncertainties from this source were less than the analytical errors.

Other useful approaches are perturbation methods developed by Hart (19) and a method for finding parameters of a system imbedded in another system (20).

**COMMON ERRORS**

*Violation of Assumptions*

**STATIONARITY** Stationarity is crucial not only to compartmental analysis but also to all other deterministic as well as alternative stochastic analyses. Its existence can be tested by repeated experiments with different concentrations of loading solutions, by double tracer experiments with delay between addition of the two tracers, and so on. These tests also confirm that superposition holds. Thron (47) considers this subject in a critical examination of compartmental analysis of uptake and washout data.
COMPARTMENTS EXIST When there are no real compartments and compartmental analysis is used, the result is only an empirical fit to an experimental function, obtained in response to a distributed process in which pool sizes and individual rate constants have no real representation in the biological system. Examples are analysis of indicator-dilution curves through vascular trees, and diffusion.

EACH COMPARTMENT IS HOMOGENEOUS This is the assumption that \( p_{ij}(t) = k_{ij}q_j(t) \). Necessary and sufficient proof is that for every \( ij \) \( dp_{ij}/dq_j \) is constant over a broad range of \( q_j \). It is seldom possible to make these observations, except in a one compartment system. A necessary but not sufficient test is to vary \( q_j(0) \) and see if the same set of rate constants is recovered.

Inadequate Data

DATA MAY BE TOO NOISY TO MERIT ANALYSIS Atkins (1) and Myhill (29) investigated the effect of data error on the determination of systems parameters by compartmental analysis. The error of the estimates increases more rapidly than the increase in data error. In general, it is a wiser use of an investigator’s time to try to reduce noise to a minimum than to analyze data that are so noisy that they can accommodate an infinite number of models. Many investigators choose the alternative of smoothing the data. There are two kinds of smoothing. One is based on fitting the data to some analytic function, which assumes the conclusion. The other is based on some sort of time interval averaging, which is a filtering process. Smoothing procedures introduce problems. Although a smooth curve may look satisfactory and yield approximately the same integral over all its observed extent, higher moments can be increasingly erroneous. This is particularly true of filtering methods. The smoothed curves should be tested against the observed curve for autocorrelation.

DATA DO NOT COVER SUFFICIENT RANGE Unless it is known beyond a doubt that the system is described in full by a monoexponential, it is necessary to follow loading experiments from \( q_s(t) = 0 \) to \( q_s(t) = A_0 \), and washout experiments from \( q_s(0) \) until \( q_s(t) = 0 \). How close one can come to these desiderata depends on the signal/noise ratio and on how long stationarity can be sustained. As was pointed out earlier, unless washout is followed to very low levels, uncertainties are so great as to challenge the cogency of any analysis, except for an independently proven monoexponential. A more subtle violation is the practice of introducing step changes in experimental conditions, sometimes multiple, during a single
washout. It is only for monoexponentials that this may not give misleading results. This practice is, in fact, a violation of the stationarity assumption. The information required, either for genuine multicomartment systems or for stochastic systems, can better be obtained by paired experiments under different conditions, each observed in full under stationary conditions. The possibility of erroneous interpretation is illustrated by an example for the general two-compartment system. Consider a washout experiment in which at time \( t=a > 0 \), conditions are altered to effect a step change in \( k_{01} \). \( k_{01} \) becomes \( k'_{01} \), and \( k_{11} \) therefore becomes \( k'_{11} \). At time \( t=b > a \), the first conditions are reimposed, with a step change back to \( k_{01} \) and \( k_{11} \). The pair of differential equations describing this experiment are

\[
\begin{align*}
\frac{dq_1(t)}{dt} &= k_{12}q_2(t) - \left[ U(t) - U(t-a) + U(t-b) \right] k_{11} \\
&\quad - \left[ U(t-a) - U(t-b) \right] k'_{11} q_1(t), \\
\frac{dq_2(t)}{dt} &= k_{21}q_1(t) - k_{22}q_2(t),
\end{align*}
\]

where \( k_{11} = k_{01} + k_{21}, \ k'_{11} = k'_{01} + k_{21} \). The step changes affect \( \lambda_1, \lambda_2, A_1, \) and \( A_2 \). At \( t=a \), this change from \( A_1 \) to \( A'_1 \) involves changing \( q_1(0) \) and \( q_2(0) \) to \( q_1(a) \) and \( q_2(a) \), respectively, making appropriate changes in the \( \lambda \)'s and \( k \)'s incorporated in \( A'_1 \), changing the factor \( e^{-\lambda_1 t} \) to \( e^{-\lambda_1 (t-a)} \) and similarly for the term \( A_2 e^{-\lambda_2 t} \). The reader who is contemplating experiments of this type may find it useful to write the full equation for \( q_s(t) \) under these conditions, plot it semilogarithmically, and see whether the changes in slope yield unambiguous information that \( k_{01} \) has changed to \( k'_{01} \) and back to \( k_{01} \).

INCOMPLETE LOADING A related problem in washout experiments is failure to load until a steady state is reached. In experiments with radioisotope, as Nims (31) has pointed out, load should continue until specific activity everywhere in the system is the same as specific activity of the loading solution. An exception is in the case of a one-compartment system, in which it makes no difference. The two major difficulties introduced by incomplete loading are the following: First, the washout curve is complicated by internal redistribution of the tracer, from compartments with higher initial specific activity to those with lower. This particularly distorts the early portion of washout. Second, the quantities \( q_i(0) \) do not measure pool size. Exact expression for \( q_s(t-a) \) for washout following loading from time zero to time \( a \) is obtained by
solving the set of $n$ differential equations

$$\frac{dq_i(t)}{dt} = [U(t) - U(t-a)] \tilde{p}_{i0} + \sum_{j=1}^{n} k_{ij} q_j(t) - k_{ii} q_i(t), \quad j \neq i.$$ 

When at least three compartments are connected in series the system exhibits delay; time is required for a change in output to reflect a change in input. In such systems, if washout observations immediately follow loading, the first part of the washout curve includes effects of delay. If loading is incomplete, the curve may even rise during the initial washout period.

**Errors in Analysis; Misinterpretations of Models**

It is common to find the $\lambda$'s interpreted as individual $k_{ij}$'s and the $A_i$'s interpreted as $q_i(0)$'s in washout experiments even when it is assumed that the model is not the parallel $n$-compartment system (configuration 9, Figure 6). It is only in such a system that these interpretations are formally valid, although there are many possibilities in which they may be acceptable approximations. However, when such approximations are made, the uncertainties of the estimates ought to be stated. For some reason, not at all obvious, a paper by Huxley (22) is often referred to as justification for the practice of assuming that the $A_i$'s are equivalent to the $q_i(0)$'s even though the system is known not to be parallel. This is a misreading of that paper, in which Huxley points out just the opposite; the error can be several-fold. The fractional error is given by $(q_i(0) - A_i)/A_i$. Since, in all possible models, $q_i(0) < A_i$, where equality holds only for the parallel model, $A_i$ overestimates $q_i(0)$. In general, for $\lambda_1 < \lambda_2$, for the two-compartment model, the smaller the ratio $\lambda_1/\lambda_2$, the closer $A_1$ approaches $q_1(0)$; the larger the ratio, the worse the estimate. For example, if configuration 8, Figure 6, is the correct model, the assumption that $A_1 = q_1(0)$ fractionally overestimates $q_1(0)$ by exactly $\lambda_1/\lambda_2$.

Other examples of misinterpretation were given previously in this critique; for example, misinterpretation of a curve on a semilogarithmic plot as a straight line (see Figure 4). It is also common to see curves (more or less of the shape of log-normal functions) markedly skewed, with relatively rapid upsweeps and slower downsweeps, for which an author claims that an exponential term (or terms) comprises some portion of the down-limb (usually middle or tail) because there is a linear segment on a semilogarithmic plot. This may or may not be a correct appraisal, but the evidence is insufficient for interpreting the
slope of that segment in terms of some specific physical property of the system. In this example, if compartmental analysis is valid, there is a minimum of three compartments or there must be a more appropriate model than compartmental analysis.

Ellis & Duggleby (12), in an article entitled, "What happens when data are fitted to the wrong equation?" proposed that the shape of a plot of the residual (experimental minus calculated value of dependent variable) against each experimental variable may be helpful in choosing more appropriate equations. The information in such plots is related to that in autocorrelation plots, and some such test ought to be required wherever the model is quantitatively important.

ALTERNATIVES TO COMPARTMENTAL ANALYSIS

Compartmental analysis, when each compartment and each rate constant is assumed to map identified biological elements, is one form of deterministic statement. If the process is understood on other grounds to obey some other deterministic law, such as diffusion, then this is an alternative to compartmental analysis.

There are many phenomena in biology that are either not yet well described deterministically or that, for some purposes, are more conveniently described as stochastic processes. The stochastic approach can be considered model-free. The observed function, in stationary conditions, is accepted as the result of all those factors that determine how long it takes the substance under study to get into, distribute throughout, or get out of the system. The observed function is, or is a function of, a probability density function of residence times in, or of transit times through, the system (28, 42). By judicious simultaneous use of multiple indicators, a good deal can be learned about the system (8). For more extended analysis and applications of stochastic processes in biological systems, see several monographs (2, 15, 23). A hybrid approach is used by Matis & Carter (26), dealing specifically with two- and three-compartment models, in which finite populations in any one compartment are treated probabilistically.

Investigators often want to know more than they have determined experimentally. To satisfy this yearning there are two kinds of choices. Either they can imagine that the system is constructed in a certain way, and seek estimates of parameters of the imagined model by some deterministic analysis, or they can design experiments to get information directly about subunits of the system, treating the result stochastically or using them to generate more substantial deterministic models. It
is as though the set of entropy consists of the information content of the biological system under study plus the dreams of the investigator. Partition of these factors, information and imagination, is different for different types of analysis. This is not intended to be pejorative. There are segments of the community of life scientists who regard model-making pejoratively. In fact, every quantitative statement about biological mechanisms is based on a model; results cannot be expressed quantitatively except in the framework of a model. Every investigator who plots data semilogarithmically is using a model. What is important is that the investigator thoroughly comprehend his model, design experiments to test it, and, when it has been validated, understand how to exploit it.

Literature Cited

30. Newman, E. V., Merrell, M., Genecin,