6 Tracer Kinetic Modeling in PET*

Richard E Carson

Introduction

The use of radiopharmaceuticals and the imaging of their biodistribution and kinetics with modern instrumentation are key components to successful developments in PET. Clever design and synthesis of sensitive and specific radiopharmaceuticals is the necessary first step. Each tracer must be targeted to measure a physiological parameter of interest such as blood flow, metabolism, receptor content, etc., in one or more organs or regions. State-of-the-art PET instrumentation produces high-quality 3-dimensional images after injection of tracer into a patient, normal volunteer, or research animal. With an appropriate reconstruction algorithm and with proper corrections for the physical effects such as attenuation and scatter, quantitatively accurate measurements of regional radioactivity concentration can be obtained. These images of tracer distribution can be usefully applied to answer clinical and scientific questions.

With the additional use of tracer kinetic modeling techniques, however, there is the potential for a substantial improvement in the kind and quality of information that can be extracted from these biological data. The purpose of a mathematical model is to define the relationship between the measurable data and the parameters of interest. In other words, an appropriate tracer kinetic model can account for all the biological factors that contribute to the tissue radioactivity signal.

In this chapter, the concepts of mathematical modeling as applied to PET are presented. Many of these concepts can be applied to radioactivity measurements from small animals made by tissue sampling or quantitative autoradiography. The primary focus in this chapter will be on methods applicable to data that can be acquired with PET imaging technology. The advantages and disadvantages of various modeling approaches are presented. Then, classes of models are introduced, followed by a detailed description of compartment modeling and of the process of model development and application. Finally, the factors to be considered in choosing and using various model-based methods are presented.

Overview of Modeling

PET imaging produces quantitative radioactivity measurements throughout a target structure or organ. A single static image may be collected at a single specific time post-injection or the full time-course of radioactivity can be measured. Data from multiple studies under different biological conditions may also be obtained. If the appropriate tracer is selected and suitable imaging conditions are used, the activity values measured in a region of interest (ROI) in the image should be most heavily influenced by the physiological characteristic of interest, be it blood flow, receptor concentration, etc. A model attempts to describe in an exact fashion this relationship between the measurements and the parameters of interest. In other words, an appropriate tracer kinetic model can account for all the biological factors that contribute to the tissue radioactivity signal.

The concentration of radioactivity in a given tissue region at a particular time post-injection primarily depends upon two factors. First, and of most interest, is the local tissue physiology, for example, the blood flow or metabolism in that region. Second is the input function, i.e., the time-course of tracer radioactivity concentration in the blood or plasma, which defines the

availability of tracer to the target organ. A model is a mathematical description (i.e., one or more equations) of the relationship between tissue concentration and these controlling factors. A full model can predict the time-course of radioactivity concentration in a tissue region from knowledge of the local physiological variables and the input function. A simpler model might predict only certain aspects of the tissue concentration curve, such as the initial slope, the area under the curve, or the relative activity concentration between the target organ and a reference region.

The development of a model is not a simple task. The studies that are necessary to develop and validate a model can be quite complex. There are no absolute rules defining the essential components of a model. A successful model-based method must account for the limitations imposed by instrumentation, statistics, and patient logistics. To determine the ultimate form of a useful model, many factors must be considered and compromises must be made. The complexity of a “100%-accurate” model will usually make it impractical to use or may produce statistically unreliable results. A simpler, “less accurate” model tends to be more useful.

A model can predict the tissue radioactivity measurements given knowledge of the underlying physiology. At first, this does not appear to be useful, since it requires knowledge of exactly the information that we seek to determine. However, the model can be made useful by inverting its equations. In this way, measurements of tissue and blood concentration can be used to estimate regional physiological parameters on a regional or even pixel-by-pixel basis. There are many ways to invert the model equations and solve for these parameters. Such techniques are called model-based methods. They may be very complex, requiring multiple scans and blood samples and using iterative parameter-estimation techniques. Alternatively, a model-based method may be a simple clinically oriented procedure. With the knowledge of the behavior of the tracer provided by the model, straight-forward study conditions (tracer administration scheme, scanning and blood data collection, and data processing) can be defined to measure one or more physiological parameters.

This chapter provides an overview of the wide assortment of ways to develop a useful model and to use the models to obtain absolute or relative values of physiological parameters.

**The Modeling Process**

Once a radioactive tracer has been selected for evaluation, there are a number of steps involved in developing a useful model and a model-based method. Figure 6.1 gives an overview of this process. Based on prior information of the expected in vivo behavior of the tracer, a “complete” model can be specified. Such a model is usually overly complex and will have many more parameters than can be determined from PET data due to the presence of statistical noise. Based on initial modeling studies, a simpler model whose parameters can be determined (identified) can be developed. Then, validation studies can be performed to refine the model and verify that its assumptions are correct and that the estimates of physiological parameters are accurate. Finally, based on the understanding of the tracer provided by these modeling studies, a simpler protocol can be defined and applied for routine patient use. This method may involve limited or no blood measurements and simpler data analysis procedures. Under many conditions, such a protocol may produce physiological estimates of comparable precision and accuracy as those determined from the more complex modeling studies.

![Figure 6.1](image-url)

**Figure 6.1.** Steps in developing a model. A priori information concerning the expected biochemical behavior of the tracer is used to specify a complete model. Initial modeling studies will define an identifiable model, i.e., a model with parameters that can be determined from the measurable data. Validation studies are used to refine the model, verify its assumptions, and test the accuracy of its estimates. After optimization procedures and error analysis and accounting for patient logistical considerations, a model-based method can be developed that is both practical and produces reliable, accurate physiological measurements.
Many factors will affect the ultimate form of a useful model. In addition to the biological characteristics of the tracer, the characteristics of the instrumentation are important. It is essential to understand the accuracy of the reconstruction algorithm and its corrections, as well as the noise level in the measurements, which depends on the injected dose, camera sensitivity, reconstruction parameters, scan time, and ROI size. It may be of little use to develop a sophisticated model if there are significant inaccuracies in the radioactivity measurements due to improper corrections for attenuation or scatter. The noise level in the data also affects the number of parameters that may be estimated. It also is the primary determinant of the precision (variability) in the estimated parameters.

**Tracers and Models**

In this chapter, the labeled compounds will be referred to as tracer, radiotracer, or radiopharmaceutical. The term tracer implies that the injected compound, including both labeled and unlabelled molecules, is present in the tissue at negligible mass concentrations, so that little or no change in the saturation of relevant enzymes or receptors occurs. For this discussion, we assume that tracer levels are appropriate, except where explicitly noted.

Figure 6.2 provides an overview of the various paths that a tracer X may follow after delivery by intravenous injection. Arterial inflow delivers X to the region of interest and venous outflow carries it away. The tracer may cross the capillary membrane and enter the tissue. From the tissue, it may be bound irreversibly or reversibly to intra- or extracellular sites, or may be metabolized (XP) into one or more chemical forms. The original labeled tracer or the metabolites may exit the tissue to the blood.

**Characteristics of Radiotracers**

Before discussing models, it is important to consider the basic characteristics of radioactive tracers. A tracer is designed to provide information about a particular physiological function of interest, such as blood flow, blood volume, a metabolic process, a transport step, a binding process, etc. However, since any given tracer will likely have many biochemical fates following injection, great care and judgment are required to choose an appropriate compound. Ideally, the only factor controlling the uptake and distribution of the tracer will be the physiological process under study. Realistically, other factors will always affect a tracer’s distribution and kinetics. For example, for a receptor-binding radiotracer, regional radioactivity concentration data are affected by regional blood flow, plasma protein binding, capillary permeability, nonspecific tissue binding, receptor association and dissociation rates, free receptor concentration, tracer clearance from blood (controlled by whole-body uptake), tracer metabolism (throughout the body), and regional uptake of any radioactive metabolites. For a well-designed tracer, the net effect of these extraneous factors is minor.

A tracer may either be a direct radiolabeled version of a naturally occurring compound, an analog of a natural compound, or a unique compound, perhaps a radiolabeled drug. An analog is a compound whose chemical properties are slightly different from the natural compound to which it is related. For example, [11C]glucose is identical to glucose except for the replacement of a 12C atom with 11C. Analogs of glucose are deoxyglucose [1] and fluorodeoxyglucose (FDG) [2-4], which are chemically different from glucose. Often, because the naturally occurring compound has a very complex biochemical fate, a model describing the tissue radioactivity data of a directly labeled compound may need to be quite complex. A carefully designed analog can dramatically simplify the modeling and improve the sensitivity of the model to the parameter of interest. Deoxyglucose and FDG are good examples. Deoxyglucose and glucose enter cells by the same transport enzyme and are both phosphorylated by the enzyme hexokinase. However, deoxyglucose is not a substrate for the next enzyme in the
glycolytic pathway, so deoxyglucose-6-phosphate accumulates in tissue. In this way, the tissue signal directly reflects the rate of metabolism, since there is little clearance of metabolized tracer. One important disadvantage of using an analog is that the measured kinetic parameters are those of the analog itself, not of the natural compound of interest. To correct for this, the relationship between the native compound and the radioactive analog must be determined. For deoxyglucose and FDG, this relationship is summarized by the lumped constant [1, 5]. To make the analog approach widely applicable, it is necessary to test if this constant changes over a wide range of pathological conditions [5-9].

Ideally, the parameter of interest is the primary determinant of the uptake and retention of a tracer, i.e., the tissue uptake after an appropriate period is directly (i.e., linearly) proportional to this parameter. This is the case for radioactive microspheres [10]. Many other compounds are substantially trapped in tissue shortly after uptake and are called chemical microspheres [11, 12]. For this class of compounds, a single scan at an appropriate time post-injection can give sufficient information about the parameter of interest. For other tracers, which both enter and exit tissue, scanning at multiple time points post-injection may be necessary to extract useful physiological information.

It is obvious that another important attribute of a tracer is that there be sufficient uptake in the organ of interest, i.e., the radioactivity concentration must provide sufficient counting statistics in a scan of reasonable length after injection of an allowable dose. Thus, the size of the structure of interest and the characteristics of the imaging equipment can also affect the choice of an appropriate tracer.

Types of Models

There are a wide variety of approaches to extract meaningful physiological data from PET tissue radioactivity measurements. All modeling approaches share some basic assumptions, in particular the principle of conservation of mass. A number of sources provide a comprehensive presentation of modeling alternatives [13–18]. Some approaches are termed stochastic or non-compartmental, and require minimal assumptions concerning the underlying physiology of the tracer's uptake and metabolism [19]. These methods permit the measurement of certain physiological parameters, such as mean transit time and volume of distribution, without an explicit description of all of the specific pools or compartments that a tracer molecule may enter.

Alternatively, there are distributed models that try to achieve a precise description of the fate of the radiotracer. These models not only specify the possible physical locations and biochemical forms of the tracer, but also include the concentration gradients that exist within different physiological domains. In particular, distributed models for capillary–tissue exchange of tracer have been extensively developed [20–26]. Since this is the first step in the uptake of any tracer into tissue, a precise model for delivery of tracer at the capillary is important. Distributed models are also used to account for processes, such as diffusion, where concentration gradients are present [27].

A class of models whose complexity lies between stochastic and distributed is the compartmental models. These models define some of the details of the underlying physiology, but do not include concentration gradients present in distributed models. The development and application of these models is the principal focus of this chapter. The most common application of compartmental modeling is the mathematical description of the distribution of a tracer throughout the body [28, 29]. Here, different body organs or groups of organs are assigned to individual compartments, and the model defines the kinetics into and out of each compartment. This type of model is useful when the primary measurable data is the time-concentration curve of the tracer in blood and urine. If there are many measurements with good accuracy, fairly complex models with many compartments and parameters can be used.

In PET, compartmental modeling is applied in a different manner. Here, scanners provide one or more measurements of radioactivity levels in a specific organ, region, or even pixel. If the tracer enters and leaves the organ via the blood, the tracer kinetics in other body regions need not be considered to evaluate the physiological traits of the organ of interest. In this way, each region or pixel can be analyzed independently. Generally, there must be some knowledge of the time-course of blood radioactivity. Since each region can be evaluated separately, the models can be relatively simple, and can therefore be usefully applied to determine regional physiological parameters from PET data.

Compartmental Modeling

Compartmental modeling is the most commonly used method for describing the uptake and clearance of ra-
dioactive tracers in tissue [28, 30, 31]. These models specify that all molecules of tracer delivered to the system (i.e., injected) will at any given time exist in one of many compartments. Each compartment defines one possible state of the tracer, specifically its physical location (for example, intravascular space, extracellular space, intracellular space, synapse) and its chemical state (i.e., its current metabolic form or its binding state to different tissue elements, such as plasma proteins, receptors, etc.). Often, a single compartment represents a number of these states lumped together. Compartments are typically numbered for mathematical notation.

The compartmental model also describes the possible transformations that can occur to the tracer, allowing it to “move” between compartments. For example, a molecule of tracer in the vascular space may enter the extracellular space, or a molecule of receptor-binding tracer that is free in the synapse may become bound to its receptor. The model defines the fraction or proportion of tracer molecules that will “move” to a different compartment within a specified time. This fractional rate of change of the tracer concentration in one compartment is called a rate constant, usually expressed as “$k$”, and has units of inverse time, e.g., min$^{-1}$. The inverse minute unit reflects the fraction per minute, i.e., the proportion of tracer molecules in a given compartment that will “move” to another compartment in one minute. To distinguish the various rate constants in a given model, subscripts are used to define the source and destination compartment numbers. In much of the compartmental modeling literature, $k_{12}$, for example, reflects the rate of tracer movement to compartment 1 from 2. This nomenclature is especially convenient for large models and is motivated by the nature of matrix algebra notation. In PET applications, the number of compartments is small (1–3), as is the number of rate constants (1–6), so it is typical to use a notation with one subscript (e.g., $k_3$) where the source and destination compartments associated with each constant are explicitly defined.

The physiological interpretation of the source and destination compartments defines the meaning of the rate constants for movement of tracer between them. For example, the rate constant describing tracer movement from a receptor-bound compartment to the unbound compartment will reflect the receptor dissociation rate. For a freely diffusible inert tracer, the rate constant of transfer from arterial blood to the tissue compartment will define local blood flow. By determining these rate constants (or some algebraic combination of them), quantitative estimates or indices of local physiological parameters can be obtained. The underlying goal of all modeling methods is the estimation of one or more of these rate constants from tissue radioactivity measurements.

Examples of Compartmental Models

Figure 6.3 shows examples of compartmental model configurations. In many depictions of models, a rectangular box is drawn for each compartment, with arrows labeled with the rate constants placed between the boxes. In most whole-body compartmental models, the blood is usually counted as a compartment. Measurements from blood are often the primary set of data used to estimate the model rate constants. In the PET applications described here, we are most interested in the model constants associated with the tissue regions that are being imaged. Typically, measurements will be made from the blood to define the “input function” to the first tissue compartment (see Input Functions and Convolution, below). In this presentation, we will treat these blood input measurements as known values, not as concentration values to be pre-

![Figure 6.3](image-url)
dicted by the model. Thus, blood will not be counted as a compartment.

Figure 6.3A shows the simplest model having one tissue compartment with irreversible uptake of tracer. This irreversible uptake is shown by the presence of a rate constant $k_1$ for tracer moving from the blood to compartment 1, but with no rate constant for exit of tracer back to blood. Such a model is appropriate for radioactive microspheres [10] or for a tracer that is irreversibly trapped in tissue. This model is often used as an approximation when tissue trapping is nearly irreversible [11]. Figure 6.3B shows a one-tissue-compartment model, appropriate for a tracer that exhibits reversible tissue uptake. This is a common model for inert tracers used to measure local blood flow [13]. Here, the rate at which the tracer exits the tissue compartment and returns to the blood is denoted $k_2$. Figure 6.3C shows a model with two tissue compartments. This model may be appropriate for a tracer that enters tissue from blood, and then is either metabolized to a form that is trapped in the tissue (at a rate defined by $k_3$) or returns to blood (at a rate defined by $k_4$), such as deoxyglucose [1]. Compartment 1 represents the unmetabolized tracer and compartment 2 the metabolized tracer. Figure 6.3D shows a three-tissue-compartment model for a receptor-binding ligand where the three compartments represent free tracer, tracer specifically bound to receptor, and tracer nonspecifically bound to other tissue elements [32].

**Compartmental Modeling Assumptions**

The successful application of simple compartmental models to a complex biological system requires that many assumptions be true. These assumptions are typically not completely valid, so that successful use of these models depends upon whether errors in these assumptions produce acceptable errors in model measurements (see Error Analysis, below). Compartmental models, by their nature, assume that each compartment is well mixed, i.e., there are no concentration gradients within a single compartment. Therefore, all tracer molecules in a given compartment have equal probability of exchange into other compartments. This well-mixed assumption has the great advantage of producing relatively simple mathematical relationships. However, it limits the ability of compartmental models to provide an accurate description of some biological structures. For example, a compartmental model cannot include the change of activity concentration in a capillary from arterial to venous ends, or the heterogeneous distribution of receptors in a patch of tissue. Often, in PET applications, the “well-mixed” assumption is also violated by the nature of the imaging process. Due to low resolution, even single-pixel data from reconstructed images represent a mixture of underlying tissues. When larger ROIs are used to improve the statistical precision of the measurements, heterogeneity in the measurements increases.

A primary assumption of most compartmental models is that the underlying physiological processes are in steady state. Mathematically, this means that the rate constants of the system do not change with time during a study, and causes the mathematics of the model to be linear differential equations (see Model Implementation). If these rate constants reflect local blood flow or the rate of a metabolic or binding process, then the rate at which these processes occur should remain constant during a study. Since the rates of many biological processes are regulated by substrate and product concentrations, maintaining processes in steady state usually requires constant concentrations of these regulating molecules. In practice, this requirement is never precisely met. However, these assumptions are adequately met so long as any changes in the underlying rates of flow, metabolism, receptor binding, etc., are slow with respect to the time scale of the data being analyzed. Note that the concentrations of the injected radiopharmaceuticals may change dramatically during a study; however, this does not violate the steady-state assumption so long as the radioactive species exists at a negligible (tracer) concentration with respect to the non-radioactive natural biological substrates (see Biochemical Reactions and Receptor–Ligand Binding). For studies using injections of radiopharmaceuticals with low specific activity, saturation of receptors or enzymes can be significant, and nonlinear modeling techniques are required.

To generate the equations of a model, the magnitude of tracer movement from compartment A to compartment B per unit of time must be defined. This is called the flux ($J_{AB}$). If tracer concentration is expressed in units of kBq per mL, then flux has units of kBq per mL per min (or another appropriate time unit). The assumptions of well-mixed compartments and physiological processes in steady state lead to the mathematical relationship that the flux $J_{AB}$ is a linear multiple of the amount, or concentration, of tracer in the source compartment A ($C_A$), i.e.,

$$J_{AB} = k C_A$$

(1)

where $k$ is a rate constant with units of inverse minute and which is independent of the concentration in any compartment. This simple equation is the basis of the differential equations that describe compartmental models (see Model Implementation).
Interpretation of Model Rate Constants

The physiological interpretations of the rate constants (such as \( k \) in Eq. 1) depend upon the definition of the source and destination compartments. A single compartment of a model may often lump a number of physiological entities together, for example, tracer in extracellular and intracellular spaces or tracer that is free in tissue and nonspecifically bound. This section discusses the physiological meaning of model rate constants.

Blood Flow and Extraction

The first step in most in vivo models is the delivery of tracer to the target region from the blood. The flux of tracer into the first tissue compartment from the blood is governed by the local blood flow and the rate of extraction of the tracer from the capillary into the tissue. Conventional fluid flow describes the volume of liquid passing a given point per unit of time and has units of mL per min. A more useful physiological measure is perfusion flow, the volume of blood passing in and out of a given volume (or weight) of tissue per unit of time, which has units of mL per min per mL of tissue or mL per min per gram of tissue. In the physiological literature, the term blood flow usually means perfusion flow.

Determining blood flow and extraction information from model parameters begins with the Fick Principle (see, for example, Lassen and Perl [15]). The net flux \( (J) \) of tracer into or out of a tissue element equals the difference between the influx \( (J_{in}) \) and outflux \( (J_{out}) \), i.e.,

\[
J = J_{in} - J_{out} = F C_a - F C_v
\]

where the influx is the product of the blood flow \( (F) \) and the arterial concentration \( (C_a) \), and the outflux is the product of the blood flow and the venous concentration \( (C_v) \). The unidirectional (or first-pass) extraction fraction \( E \) is the fraction of tracer that exits the blood and enters the tissue on one capillary pass, or

\[
E = \frac{C_a - C_v}{C_a}
\]

A tracer with low extraction has a small arterial–venous difference on first pass. Equation 2 can then be rewritten as

\[
J = (F \cdot E) C_a = k C_a
\]

Equation 4 describes the unidirectional delivery of tracer from blood to tissue. The rate constant \( k \) defining this uptake process is the product of blood flow and unidirectional extraction fraction. The interpretation of the extraction fraction was further developed by Kety [13], Renkin [33], and Crone [34] by considering the capillary as a cylinder to produce the following relationship:

\[
E = 1 - e^{-\frac{PS}{F}}
\]

where \( P \) is the permeability of the tracer across the capillary surface (cm per min), \( S \) is the capillary surface area per gram of tissue (cm² per gram), and \( F \) is the blood flow (mL per min per gram). For highly permeable tracers, the product \( PS \) is much greater than the flow \( F \), so the exponential term in Eq. 5 is small, and the extraction fraction is nearly 1.0. In this case, the rate constant for delivery is approximately equal to flow. Such tracers are therefore useful to measure regional blood flow and not useful to measure permeability, i.e., they are flow-limited. For tracers with permeability much lower than flow, the relationship in Eq. 5 can be approximated as

\[
E \approx \frac{PS}{F}
\]

and the rate constant \( k \) \((F \cdot E)\) becomes \( PS \). Such tracers are useful to measure permeability and not useful to measure flow. Most tracers lie between these two extremes, so that the rate constant for delivery from arterial blood to tissue is affected by both blood flow and permeability. These relationships are directly applicable to tracers that enter and leave tissue by passive diffusion. For tracers transported into and out of tissue by facilitated or active transport, the \( PS \) product is mathematically equivalent to the transport rate, which depends upon the concentration and reaction rate of the transport enzymes (See Biochemical Reactions).

The interpretation of a delivery rate constant \( k \) as the product of flow and extraction fraction may depend upon whether the blood activity concentration \( C_a \) is measured in whole blood or in plasma. If there is very rapid equilibration between plasma and red blood cells, then the whole blood and plasma concentrations will be identical. However, if equilibrium is slow with respect to tracer uptake rates into tissue, or if there is trapping or metabolism of the tracer in red blood cells, the plasma concentration should be used. In the extreme of no uptake of tracer into red cells, then the delivery rate constant \( k \) is the product of extraction fraction and plasma flow, where plasma flow is related to whole blood flow based on the hematocrit. If binding of tracer to plasma proteins is significant, similar changes in interpretation of the rate constants may also be required.
Diffusible Tracers and Volume of Distribution

One of the simplest classes of tracers is those that enter tissue from blood and then later return to blood. The net flux of tracer into a tissue compartment can be expressed as follows:

\[ J = K_1 C_a - k_2 C \]  

(7)

\( K_1 \) is the rate of entry of tracer from blood to tissue and is equal to the product of extraction fraction and blood flow, and \( C_a \) is the concentration of tracer in arterial blood. The rate constant \( k_2 \) describes the rate of return of tracer from tissue to blood, where \( C \) is the concentration of tracer in tissue. The physiological interpretation of \( k_2 \) can best be defined by introducing the concept of the volume of distribution. Suppose the concentration of tracer in the blood remained constant. Ultimately, the concentration of the diffusible tracer in the tissue compartment would also become constant and equilibrium would be achieved. The ratio of the tissue concentration to the blood concentration at equilibrium is called the volume of distribution (or alternatively the partition coefficient). It is termed a volume because it can be thought of as the volume of blood that contains the same quantity of radioactivity as 1 mL (or 1 gram) of tissue. Once the blood and tissue tracer concentrations have reached constant levels, i.e., equilibrium, the net flux \( J \) into the tissue compartment is 0, so the volume of distribution \( V_D \) can be expressed as

\[ V_D = \frac{C}{C_a} = \frac{K_1}{k_2} \]  

(8)

where the last equality is derived by setting the flux \( J \) in Eq. 7 to 0. Therefore, the physiological definition for the rate constant \( k_2 \) is the ratio of \( K_1 \) to \( V_D \). Thus, \( k_2 \) has information concerning flow, tracer extraction, and partition coefficient.

Biochemical Reactions

Often, two compartments of a model represent the substrate and product of a chemical reaction. In that case, the rate constant describing the "exchange" between these compartments is indicative of the reaction rate. For enzyme-catalyzed reactions [35], the flux from substrate to product compartments is the reaction velocity \( v \):

\[ v = \frac{V_m C}{K_m + C} \]  

(9)

\( V_m \) is the maximal rate of the reaction, \( C \) is the concentration of substrate, and \( K_m \) is the concentration of substrate that produces half-maximum velocity. This is the classic Michaelis–Menten relationship. It shows that the velocity is not a linear function of the substrate concentration, as in Eq. 1. However, when using tracer concentrations of a radioactive species and if the concentrations of the native substrates are in steady state (see Compartmental Modeling Assumptions), the linear form of Eq. 1 still holds. In the presence of a native substrate with concentration \( C \) and the radioactive analog with concentration \( C^* \), the reaction rate for the generation of radioactive product \( v^* \) is as follows:

\[ v^* = \frac{V_m^* C^*}{K_m^* \left( 1 + \frac{C}{K_m} + \frac{C^*}{K_m} \right)} \]  

(10)

\( V_m^* \) and \( K_m^* \) are the maximal velocity and half-maximal substrate concentration for the radioactive analog. If the radioactive species has high specific activity (the concentration ratio of labeled to unlabelled compound in the injectate) so that its total concentration (labeled and unlabelled) is small compared to the native substrate, i.e., \( C/K_m \ll C/K_m^* \), then Eq. (6.10) reduces to

\[ v^* = \left( \frac{V_m^*}{K_m^* \left( 1 + \frac{C}{K_m} \right)} \right) C^* = k C^* \]  

(11)

The term in large brackets in Eq. 11 is composed of terms that are assumed to be constant throughout a tracer experiment. Therefore, when using radiopharmaceuticals at tracer concentrations, enzyme-catalyzed reactions can be described with a linear relationship as the product of a rate constant \( k \) and the radioactive substrate concentration \( C^* \). The rate constant \( k \) includes information about the transport enzyme and the concentration of unlabelled substrate.

Receptor–ligand Binding

For radiotracers that bind to receptors in the tissue (see, for example, Eckelman [36]), the rate of binding, i.e., the rate of passage of tracer from the free compartment to the bound compartment, can also be described by the linear form of Eq. 1 under tracer concentration assumptions. For many receptor systems, the binding rate is proportional to the product of the concentrations of free ligand and free receptor. This classical bi-molecular association can be described mathematically as

\[ v = k_{on} \left( B_{\text{max}} - B \right) F \]  

(12)
where \( k_{on}^* \) is the bi-molecular association rate (nM\(^{-1}\)min\(^{-1}\)), \( B_{max} \) is the total concentration of receptors (nM), \( B \) is the concentration of receptors currently bound (either by the injected ligand or by endogenous molecules), and \( F \) is the concentration of free ligand. When a radioactive species is added and competes with endogenous compound for receptor binding, the radiopharmaceutical binding velocity is

\[
\dot{v}^* = k_{on}^* \left( B_{max} - B - B^* \right) F^*
\]

(13)

where \( k_{on}^* \) is the association rate of the radiopharmaceutical and \( B \) is the mass concentration of the bound radiopharmaceutical. If the radioactive compound has high specific activity, then \( B \ll B \) and Eq. 13 becomes

\[
\dot{v}^* = k_{on}^* B_{max}^* F^* = kF^*
\]

(14)

where \( B_{max}^* \) is the free receptor concentration \( (B_{max} - B) \). Thus, using a high specific activity receptor-binding ligand, measurement of the reaction rate constant \( k \) provides information about the product of \( k_{on}^* \) and \( B_{max}^* \), but cannot separate these parameters. Since \( B_{max}^* \) is sensitive to change in total receptor and occupancy by endogenous or exogenous drugs, receptor-binding ligands can be extremely useful to measure receptor occupancy or dynamic changes in neurotransmitter levels [37]. Note that the description of receptor-binding radioligands is mathematically identical to that for enzyme-catalyzed reactions, although the conventional nomenclature is different.

**Model Implementation**

This section presents an overview of the mathematics associated with compartmental modeling. This includes the mathematical formulation of these models into differential equations, the solution equations to a few simple models, and a summary of parameter estimation techniques used to determine model rate constants from measured data. Here, we concentrate on applications where we have made measurements in an organ or region of interest which we wish to use to ascertain estimates of the underlying physiological rates of this region.

**Mathematics of Compartmental Models**

This section describes the process of converting a compartmental model into its mathematical form and determining its solution. For a more complete discussion of these topics, consult basic texts on differential equations [38] as well as a number of specialized texts on mathematical modeling of biological systems [16, 28].

First, we start with a particular model configuration like those in Fig. 6.3. The compartments are numbered 1, 2, …, and the radioactivity concentration in each compartment is designated \( C_1, C_2, \ldots \). Radioactivity measurements in tissue are typically of a form such as counts per mL or kBq per gram. The volume or weight unit in the denominator reflects the full tissue volume. However, the tracer may exist only in portions of the tissue; for example, just the extracellular space. In this case, the concentration of the tracer within its distribution space will be higher than its apparent concentration per gram of tissue. When these concentration values are used to define reaction rates, instead of the true local concentration, the interpretation of the relevant rate constant should include a correction for the fraction of total tissue volume in which the tracer distributes.

**Differential Equations**

The net flux into each compartment can be defined as the sum of all the inflows minus the sum of all the outflows. Each of these components is symbolized by an arrow into or out of the compartment, and the magnitude of each flux is the product of the rate constant and the concentration in the source compartment. The net flux into a compartment has units of concentration \( (C) \) per unit time and is equal to the rate of change \( (dC/dt) \) of the compartment concentration, or \( dC/dt \). Consider the simple one-tissue-compartment model in Fig. 6.3B. The differential equation describing the rate of change of the tissue concentration \( C_i \) is

\[
\frac{dC_i}{dt} = K_i C_a(t) - k_2 C_i(t)
\]

(15)

Here, \( C_a(t) \) is the time course of tracer in the arterial blood, also called the input function. \( K_i \) is the rate constant for entry of tracer from blood to tissue, and \( k_2 \) is the rate constant for return of tracer to blood. The capitalization of the rate constant \( K_i \) is not a typographical error. \( K_i \) is capitalized because it has different units than other rate constants. The blood radiotracer measurements are typically made per mL of blood or plasma. In non-imaging studies in animals, tissue concentration measurements are made per gram of tissue. Thus, \( C_i \) had units of kBq per gram, and \( C_a \) had units of kBq per mL. Therefore, \( K_i \) must have units of mL blood per min per gram tissue (usually written as mL/min/g).
The other rate constants have units of inverse minute. PET scanners actually acquire tissue radioactivity measurements per mL of tissue. Thus, to present results in comparable units to earlier work, corrections for the density of tissue must be applied to convert kBq per mL tissue to kBq per gram tissue.

Before solving Eq. 15 for a general input function \( C_a(t) \), first consider the case of an ideal bolus input, i.e., the tracer passes through the tissue capillaries in one brief instant at time \( t = 0 \), and there is no recirculation. If \( C_a \) is the magnitude of this bolus, the model solution for the time-concentration curve for compartment 1 is as follows:

\[
C_1(t) = C_a K_1 \exp(-k_2 t)
\]

Thus, at time zero, the tissue activity jumps from 0 to the tracer density of tissue must be applied to convert kBq per mL tissue to kBq per gram tissue.

Note that in most cases, the measured tissue activity will be the total in both compartments, so that the model prediction will be the sum \( C_1(t) + C_2(t) \).

These solutions for tissue concentration are linearly proportional to the magnitude of the input, \( C_a \). Doubling the magnitude of the input (injecting more) will double the resultant tissue concentration. The equations are non-linear with respect to many of the model rate constants (those that appear in the exponents) but is linear in \( K_1 \).

### Input Functions and Convolution

In the previous section, mathematical solutions were presented for simple models under the condition of an ideal bolus, i.e., the tracer appears for one capillary transit with no recirculation. In reality, the input to the tissue is the continuous blood time–activity curve. The equations above are linear with respect to the input function \( C_a \). This permits a direct extension of these bolus equations to be applied to solve the case of a continuous input function. Fig. 6.4 illustrates this concept. Figure 6.4a and 6.4c show ideal bolus input functions of different magnitudes at different times. Figure 6.4b and 6.4d show the corresponding tissue responses for the model with one tissue compartment (Fig. 6.3b). Suppose, as in Fig. 6.4e, the combination of the two inputs is given, i.e., there is a bolus input of magnitude \( A \) at time \( t = T_1 \), and a second bolus of magnitude \( B \) at \( t = T_2 \). The resulting tissue activity curve is:

\[
C_1(t) = K_1 A \exp(-k_2(t - T_1)) \quad \text{for} \quad T_1 \leq t < T_2
\]

\[
C_2(t) = K_1 A \exp(-k_2(t - T_1)) + K_1 B \exp(-k_2(t - T_2)) \quad \text{for} \quad t \geq T_2
\]

In other words, the tissue response is a sum of the individual responses to each bolus input. The responses are scaled in magnitude and shifted in time to match each bolus input.

Suppose now there is a series of bolus administrations at times \( T_i \), \( i = 1, \ldots \), each of magnitude \( C_a(T) \) as depicted by the square waves in Fig. 6.4g. The total tissue response (Fig. 6.4h) can be written as the summation:

\[
C_1(t) = \sum_i C_a(T_i) K_1 \exp(-k_2(t - T_i))
\]

where the exponentials are defined to have zero value for negative arguments (i.e., \( t < T_i \)). If we now consider the continuous input function \( C_a(t) \) (bold line in Fig. 6.4g) as an infinite summation of individual boluses,
Figure 6.4. Convolution. a, c, e, and g show the input functions that produce the tissue responses in b, d, f, and h, respectively. 

- a: an ideal bolus input at time $T_1$.
- b: the corresponding tissue response to the one-compartment model of Fig. 6.3b, i.e., exponential clearance following Eq. (16).
- c: a single bolus of half the magnitude in a at time $T_2$.
- d: the tissue response to c, which is altered in time and magnitude in a corresponding manner from that in b.
- e: combination of inputs in a and c.
- f: the tissue response to e, which is the sum of the tissue responses from each bolus administered separately (b+d). This demonstrates the linearity of the model equations.
- g: a continuous input function (bold line) which can be interpreted as a series of bolus injections of varying magnitudes (fine lines).
- h: the tissue activity response to g (bold line) which can be interpreted as the sum of the responses to each bolus (fine lines).
the summation in Eq. 25 becomes an integral, and the tissue response (bold line in Fig. 6.4h) becomes:

$$C(t) = \int C_a(s)K_1\exp(-k_2(t - s))ds$$

(26)

Here, $s$ is the integration variable. This is called a convolution integral, and is often written as

$$C(t) = C_a(t) \otimes K_1\exp(-k_2t)$$

(27)

with the symbol $\otimes$ denoting convolution. This presentation corresponds to the one-compartment model of Eq. 25. However, convolution applies to any compartmental model whose solution has a linear relationship to its input function. Let $h_i(t)$ be the impulse response function for compartment $i$, i.e., the time course of tissue response from a bolus input of magnitude 1 [$K_1\exp(-k_2t)$ for the one-compartment model]. Then the tissue activity resulting from the general input function $C_a(t)$ is written as

$$C_i(t) = C_a(t) \otimes h_i(t)$$

(28)

Thus, for linear compartmental models, the tissue time–activity curve is the convolution of the input function with the impulse response function. For compartmental models, the latter is a sum of exponentials, typically one exponential per compartment. A number of approaches have been used to implement and solve Eq. 28 on computers if the arterial input function is determined from serial samples. One approach is to fit the measured input function data to a suitable model [39] and then solve the convolution integral by standard mathematical methods. Alternatively, a continuous input function can be approximated by linear interpolation between the sample data values, and then Eq. 28 can be solved by analytical integration over each time period between blood samples.

Figure 6.5 shows the effects that variations in the input function can produce on the resulting tissue response. Figure 6.5a shows three input functions. The solid line is a measured arterial input function. The other two input curves were calculated based on the measured data so that the area under all curves is a constant. The tissue concentration curves produced in response to each input function are shown as the corresponding curves in Fig. 6.5b. In all cases, the tissue response is calculated from the one-compartment model, Eq. 27, using the same parameters ($K_1 = 0.1$ mL/min/mL and $k_2 = 0.1$ min$^{-1}$). The difference in shape between the input functions produces comparable differences in the tissue concentration curves. These differences in shape do not reflect differences in the local physiological parameters of the tissue, since the rate constants were the same in all cases. Thus, the main point of Fig. 6.5 is that a time–activity curve in a tissue region cannot be interpreted without knowledge of the input function.

The linear compartmental models discussed to this point have the tremendous advantage of providing exact mathematical solutions, predicting the tissue response in the form of Eq. 28. In some cases, the flux between compartments cannot be described mathematically as the product of a rate constant times the concentration of tracer in the source compartment ($J = kC$). For example, in modeling receptor-binding ligands, the linear flux assumption holds if the radiopharmaceutical is administered at tracer levels and does not produce detectable saturation of the receptor sites (Eq. 14). If such a ligand is administered in low specific activity so that it produces a change in receptor occupancy during the data collection period, the

![Figure 6.5](image-url)
differential equations governing the model are no longer linear. This can be seen in Eq. 13 where the flux from the free to the bound compartments $v'$ cannot be described as a constant multiplied by $F'$. Since the bound concentration $B'$ also appears in the relationship. To solve the model in this and most other non-linear cases requires techniques of numerical integration of differential equations [40, 41]. The basic idea to numerically estimate the activity in each compartment is to take small steps in time and use the differential equations to determine how much each compartment's concentration should change over each time step. The most commonly used method for numerical integration is called Runge–Kutta, which provides increased accuracy with longer time steps by averaging multiple estimates of the derivative $dC/dt$.

In implementing models, it is important that the model formulation matches the nature of PET scan data. The models presented above predict the tissue concentration at an instant in time. Image values represent the average tissue activity collected over each scan interval. The instantaneous model value can be used to determine the integrated scan value. For example, for the one-compartment model (Eq. 15 and Eq. 27), the integrated scan value from time $T_1$ to $T_2$ is as follows:

$$\int_{T_1}^{T_2} C_i(t)dt = \frac{K_i \int_{T_1}^{T_2} C_a(t)dt - \left(C_i(T_2) - C_i(T_1)\right)}{k_2}$$

(29)

Another practical issue is radioactive decay. This can be handled either by explicit decay correction of both the tissue and blood data or by incorporating decay into the model formulation. The latter approach can be accomplished by adding an additional rate constant corresponding to the decay rate (0.693/half life) to each compartment. This method is slightly more accurate than explicit decay correction for short-lived tracers, since decay correction does not account for biological change in tracer concentration within one scan interval.

### Parameter Estimation

The previous sections presented the mathematical techniques necessary to solve the model equations. Thus, with knowledge of the input function $C_a(t)$, the model configuration, and its rate constants, the tissue concentration curve can be predicted mathematically. This section provides an overview of the inverse problem, i.e., given measurements of the tissue activity and the input function and a proposed model configuration, one can produce estimates of the underlying rate constants. Many references are available on the topic of parameter estimation [42–44].

There are many ways to accomplish the estimation of model parameters. The choices available and the success of any given method depend upon the form of the model and the sampling and statistical quality of the measured data. If only a single measurement of tissue radioactivity is made, obviously only a single parameter can be determined. Collection of multiple time points permits the estimation of some or all of the parameters of a model. Since measured data always have some associated noise, the estimates of model parameters from such data will also be noisy. It is often the goal of a statistical estimation method to minimize the variability of the resulting parameter estimates. Note also that the values of the parameters will affect the statistical quality of the results. For example, blood flow estimates produced by a particular method may be reliable for high-flow regions but unreliable for low-flow regions.

When many tissue measurements are collected after radionuclide administration, the most commonly used method of parameter estimation is called least-squares estimation. Qualitatively, the goal of this technique is to find values for the model rate constants that, when inserted into the model equations, produce the “best” fit to the tissue measurements. Quantitatively, the goal is to minimize an optimization function, specifically the sum of the squared differences between the measured tissue concentration data and the model prediction, i.e.,

$$\sum_{i=1}^{N} \left(C_i - C(T_i)\right)^2$$

(30)

where there are N tissue measurements, $C_i$, $i=1,\ldots,N$, at times $T_i$, and $C(T_i)$ is the model prediction of tissue activity at each of these times. This particular form is used because of the nature of the noise in the measured data. Parameter estimates produced by minimizing the sum of squared differences have minimum variability if the noise in each scan measurement is statistically independent, additive, Gaussian, and of equal magnitude. Additive and independent statistical noise is usually a good assumption for PET image data, however, often the variance of the measurements will not be constant across different scans in one multiplex scan acquisition, particularly for short-lived isotopes such as $^{15}$O or $^{11}$C. In this case, the least-squares function can be modified to accommodate variable noise levels as follows:

$$\sum_{i=1}^{N} w_i \left(C_i - C(T_i)\right)^2$$

(31)
where $w_i$ is a weight assigned to data point $i$. This method is called \textit{weighted least-squares estimation}, and the optimal weight for each sample is the inverse of the variance of the data [43]. For simple count data, the variance of the data can be estimated from the count data itself based on its Poisson distribution [45]. For reconstructed data, many algorithms have been proposed to calculate or approximate the noise in pixel or region-of-interest data [46–52].

It is important to recognize that there are many non-random or deterministic error sources in the modeling process that cause inconsistencies between the model and the measured data (see section on random and deterministic errors). When fitting data to a model, the parameter estimation procedure is naïve in that it believes that the specified model is absolutely correct. The algorithm will do its best to minimize the optimization function. Therefore, if there are deterministic errors in the model or the input function, the estimation algorithm can produce unsuitable results. It may be appropriate in some situations to adjust the weights of some data points (e.g., early time points where errors in the model due to intravascular activity are most significant) to reduce the sensitivity of the model to the presence of deterministic errors.

Once an optimization function (Eq. 30 and Eq. 31) has been defined, there are many algorithms available to determine the values of the model parameters that minimize it [41, 43]. Unfortunately, in most cases with compartmental models, there are no direct solutions for the parameters. This is true because, although the models themselves are linear (i.e., all fluxes between compartments are linear multiples of the concentration in the source compartment), the solutions to these models are functions that are non-linear in at least one of the model parameters. For example, Eq. 27, the solution to the one-compartment model (Fig. 6.3B) is linear with respect to the parameter $K_1$, but is non-linear with respect to the parameter $k_2$. To solve for the parameters, iterative algorithms are required. First, an initial guess is made for the parameter values. Then the algorithm repeatedly modifies the parameters, at each step reducing the value of the optimization function. Convergence is reached when changes to the parameters from one iteration to the next become exceedingly small. Great care is required in the use of iterative algorithms, because incorrect solutions can be obtained, particularly if the initial guess is not appropriate.

Figure 6.6 provides an example of the process of parameter estimation applied to time-activity data collected after a bolus injection of fluorodeoxyglucose (FDG) [53]. Figure 6.6a shows a plot of region-of-interest values (occipital cortex) taken from reconstructed PET images. The solid line through the data points is the best fit obtained by minimizing the weighted sum of squared differences between the data and the two-compartment model (Figure 6.3c). Figure 6.6b shows a plot of the weighted residuals versus time. The residual is the difference between each data point and the model prediction. When weighted least squares is used, the residuals are scaled by the square root of each weight, $w_i$, so that the sum-of-squares optimization function equals the sum of squared residuals. Ideally the residuals would be random, have zero mean, and uniform variance. If a good estimate of the noise level in the data is known, the weighted residuals should have a standard deviation of approximately 1. Thus, when plotting the residuals versus time or versus concentration, the residuals would appear as a uniform band centered on zero. The residuals in Fig. 6.6b reasonably satisfy these expectations.

Many parameter estimation algorithms provide estimates of the uncertainties of the parameter estimates...
called standard errors. These values can be used as estimates of the minimum random uncertainty in the estimate. The algorithm determines these standard errors based on the structure of the model, the parameter estimates, and the magnitude of the residual sum of squares. However, this measure is an underestimate of the true uncertainty of the parameters, since there are usually many sources of “real-world” errors that are not explicitly included.

It is often useful to calculate functions of the rate constants which provide different physiological information. For example, in the one-compartment model (Fig. 6.3B), a parameter estimation problem may be posed to estimate the rate constants $K_1$ and $k_2$. From these parameters, the distribution volume ($V = K_i/k_2$) can be calculated. To determine the uncertainty in the distribution volume estimate, information about the individual standard errors in $K_i$ and $k_2$ is required along with the correlation between them. The parameter estimates will be correlated since both values are determined simultaneously from the same noisy data. The coefficient of variation ($CV$, the ratio of the standard error to the parameter value) of the distribution volume can be calculated by propagation of errors calculations:

$$CV^2(V) = CV^2(K_i) + CV^2(k_2) - 2\rho_{12}CV(K_i)CV(k_2)$$

where $\rho_{12}$ is the estimated correlation coefficient between the parameter estimates $K_i$ and $k_2$.

Least squares is the best optimization criterion for estimating parameters when a large number of assumptions are met. If any of these assumptions are not true, better estimates may be obtained by other methods (see section on error analysis). A better estimate is one that may be more accurate (less biased) or more precise (less variable). In addition, iterative least-squares algorithms may be very computationally intensive, particularly if it must be carried out individually for every pixel in an imaging volume. Often, iterative least-squares procedures are used only for a small number of regions of interest. However, it is often more useful if the data analysis procedure produces functional images where each pixel represents a physiological parameter of interest. To do this, rapid computation schemes are required. Rapid implementations of iterative least-squares procedures have been developed for the simplest non-linear models with just one non-linear parameter, e.g., the one-compartment model with solution in Eq. 27. These techniques have been applied to the measurement of cerebral blood flow [54, 55] and total volume of distribution of receptors [56–58]. In addition, a number of methods have been derived that allow direct non-iterative calculation of the parameter estimates by reformulating the problem in terms of integrals of the tissue and blood data [59–66]. These methods do not minimize the sum-of-squares optimization function, but in many cases have been shown to have comparable statistical quality to the least-squares techniques and often have less sensitivity to deterministic errors in the model. Another interesting approach for parameter estimation from non-linear models is called spectral analysis and uses the methods of linear programming with the knowledge that all the exponential clearance terms ($\alpha_i$ in Eq. 19 and Eq. 20, for example) are positive [67].

As shown above, the measured tissue activity is the convolution of the input function with the underlying impulse-response function (Eq. 28). This impulse-response function has a much simpler mathematical form (usually a sum of exponentials) and is therefore more easily analyzed. Some investigators have used the approach of deconvolution, whereby an estimate of the impulse-response function is determined from measurements of the tissue response and the input function [68]. However, because the process of deconvolution greatly amplifies noise in the tissue measurements and is often mathematically unstable, great care is required in the application of these techniques.

### Development of Mathematical Models

The primary factor affecting the form of a model is the nature of the tracer itself. Usually, a priori information can be used to predict all of the relevant metabolic paths of the tracer in tissue, i.e., a complete model. However, technical and statistical limitations of the available data will prevent the use of such a comprehensive model, which includes all steps in the physiological uptake, metabolism, and clearance of a tracer.

Figure 6.1 shows the process of development and application of a model in PET [69, 70]. This section presents the steps starting with a complete model, then generating an identifiable model, and ultimately a practical model. An identifiable model is one which can be applied to regional kinetic data and used to extract estimates of model parameters. Such a model is a simplified version of a comprehensive description of the interactions of a radiotracer in tissue. However, this model may not be workable if its parameter estimates are too variable or inaccurate. A useful model may be derived by further simplification of the identifiable model. The
useful model provides reproducible and accurate estimates of model parameters. Validation studies are necessary to demonstrate these characteristics.

**Model Identifiability**

The first step in defining a model is to determine identifiability, meaning that the parameters of a model can be uniquely determined from measurable data. There is an extensive literature on this topic [16, 19, 71–77], including studies with particular attention to PET applications [78–80]. In some cases, the structure of the model itself does not permit the unique definition of parameter values, even with noise-free data. One example of this is the case of high specific-activity pharmaceutical are collected. Then, a number of possible model configurations are proposed. Usually, the number of compartments covers a range from very complex to very simple, and there is a range of different numbers of parameters to be estimated in these models. Parameter estimation procedures are performed with the measured data using each model. The goodness-of-fit of each model to the data is assessed from the residual sum of squares (Eq. 31) using statistical tests such as the F-test, the Akaike information criterion [85], or the Schwarz criterion [86] to determine which model is most appropriate. In general, the use of a more complex model with additional parameters will produce a better fit to the data and a smaller residual sum of squares. However, this will be the case even if the additional parameters added by the more complex model are only providing a better fit to the noise in the data and have no relationship to the underlying true tissue model. The statistical tests used for model comparison determine whether the residual sum of squares has been reduced using the more complex model by an amount that is significantly greater than what is expected by random chance.

Another very useful approach for model comparison is the examination of the pattern of residuals as in Fig. 6.6b [43]. If the residuals from a fit of one model configuration do not appear as randomly distributed around zero, then a more complex model may be appropriate. However, given all the error sources (see section on random and deterministic errors), no model will ever be perfect. It will therefore often be the case that an overly complex model will still provide a statistically significant improvement in the fit compared to a simpler model. The modeler must have a good understanding of the degree of accuracy in the data in order to avoid an unduly complicated model.

To simplify models, pairs of compartments can be combined together. Two compartments can be collapsed
into one by assuming that the rate constants “connecting” them are large enough so that the two compartments remain in continuous equilibrium. When this occurs, the physiological interpretation of the remaining rate constants in the reduced model must be changed. In this way, a set of “nested” models can be defined. Figure 6.3 shows some examples of nested models. Here, a simple model with a few rate constants can be considered to be a special case of a more complex model with more parameters. For example, the model in Fig. 6.3B is a simplified version of Fig. 6.3C which is itself a simplified version of Fig. 6.3D.

It is good practice to test a set of nested models to determine which one best characterizes a set of measured data [56]. An example of this process is shown for the opiate receptor antagonist [18F]cyclofox [84, 87–89]. Figure 6.7a shows a typical time–activity curve measured in the thalamus with PET after bolus injection. The symbols are measured data points. The solid line is the best-weighted least squares fit using a model with two tissue compartments (Fig. 6.3c). The dashed line is the best fit using a model with one tissue compartment (Fig. 6.3B). Both models also included an additional parameter to account for radioactivity present in the tissue vascular space, so five and three parameters were estimated, respectively. The plots of weighted residuals versus time from these fits are shown in Fig. 6.7b (one compartment) and Fig. 6.7c (two compartment). The one-compartment results show a deterministic pattern of residuals. The residual points are not randomly distributed about zero, but instead show runs of sequential values that are all positive or all negative. The residual pattern is more random when using the two-compartment model. In this case, the more complex model was found to have produced a statistically significant reduction in the residual sum of squares. However, this improvement was not large and was not found uniformly for all patients or for all brain regions.

The absolute magnitude of the residual noise can also be useful in determining if a particular model configuration is appropriate. If the model is exactly correct and the magnitude of data noise is known,
the weighted sum of squares (Eq. 31) will be approximately equal to \( N - n_p \), where \( N \) is the number of data points and \( n_p \) is the number of model parameters being estimated. If the actual sum of squares from data fits are close to this value, the modeler gains additional confidence that the chosen model configuration is appropriate.

**Model Constraints**

A typical situation in PET modeling problems is that a simple model with few parameters is often not adequate to describe the tissue concentration curve. However, a more complex model that does adequately describe the data frequently produces parameter estimates that have large uncertainties (standard errors). Specifically, a simple one-compartment, two-parameter model is often insufficient, whereas a two-compartment, four-parameter model is “better” by various statistically significant measures. A number of authors have dealt with this conflict by applying constraints. These entail specifying exact values for certain parameters or defining relationships between the parameters that must be met. In either case, the effect is to reduce the number of parameters that must be determined from the model. If the constraints are accurate (or reasonably so), then the sensitivity of the model data to the remaining parameters is increased and the uncertainty in their estimation is reduced. Often the constraint equations use a priori values for physiological constants based on the presumed interpretation of the model parameters in terms of Michaelis–Menten parameters [90–92]. Alternatively, some parameters may be constrained based on measurements made in other regions [93]. For example, a common approach for receptor-binding tracers is first to analyze a reference region known to have little or no specific binding to determine parameters associated with the magnitude of nonspecific binding. Then, regions with specific binding are analyzed with nonspecific-binding rate constants constrained to equal those estimated in the reference region [81, 94]. Alternatively, additional studies can be performed to aid the estimation process by constraining parameters to be common to the analysis of both studies. For receptor-binding tracers, a study with an inactive enantiomer can be used to determine parameters of nonspecific binding [95, 96]. In addition, paired studies with high and low specific activity injections and/or displacement can be performed and analyzed simultaneously with some parameters shared in the models for the two studies [97, 98].

**Validation of Physiological Measures**

In the process of developing and selecting a suitable model formulation and methodology, it is important to perform validation studies which prove that the parameter estimates produced by a model are correct. These studies determine the precision and accuracy of model estimates, verify the legitimacy of the model assumptions, and help choose between various approaches. Such an evaluation invariably must be done in animals because of constraints on experimental design, scan duration, and radiation dosimetry in humans. Practical limits on animal studies include limitations on total blood sampling for input function measurements and the effects of anesthesia.

Although much of the work of model development and validation is performed using small or large animals, it is important to realize that there are considerable differences between PET image data and autoradiographic or tissue-sampling measurements, as well as the species differences among rodents, large animals, and humans, which may limit the applicability of the information obtained in the animal experiments. For example, measurement of tissue concentration data at multiple time points in rodents requires multiple animals. PET studies allow acquisition of multiple time points in a single study, avoiding inter-individual variability. However, the spatial resolution and statistical reliability of scan data are substantially worse than measurements from tissue samples in rats. Therefore, kinetic parameters that can be reliably determined from rat data may not be numerically identifiable from human scan data. Therefore, many validation studies should be repeated wherever practical with human subjects.

The simplest test of a model is reproducibility, i.e., the variability of the model parameters under identical conditions, either on the same day or different days [99]. Repeating studies on the same day will generally produce smaller differences in scan data results, since there will be less variation in subject positioning and scanner calibration. Measurements of population variability of model estimates provide information concerning the most useful model configurations. Clearly, model parameters with large coefficients of variation will not generally be useful. Also, models should provide physiologically reasonable values. Although in vivo measurements can cer-
tainly produce different results than in vitro tests, it is up to the investigator to demonstrate the accuracy of a model that produces parameter estimates inconsistent with previous results in the literature. Additionally, one can compare parameters from one tracer to another when there is reason to believe they should be similar, e.g., comparing the $K_i$ values of two tracers with high extraction, in which case both $K_i$ values should approximate flow [58].

The next steps in validation of a model are intervention studies. Here, one or more of the physiological parameters that affect tracer uptake are altered, and the model is tested to verify that the parameters change in the proper direction and by an appropriate magnitude in response to a variety of biological stimuli. For example, brain blood flow can be altered by changing arterial pCO$_2$, or free receptor concentration can be reduced by administration of a cold ligand. It is also useful to test whether the parameters of interest do not change in response to a perturbation in a different factor, e.g., does an estimate of receptor number remain unchanged when blood flow is increased [100]? Alternatively, changing the form of the input function should ideally have no effect on the model parameters [84, 101–103]. Model assumptions, e.g., parameters whose values have been constrained, should be tested. At a minimum, computer simulations of the effects of errors in various assumptions upon model results can be performed (see section on error analysis). The limitation of these simulations is that they are only as good as the models on which they are based. Therefore, experimental validation of model assumptions should be performed where possible.

Finally, the absolute accuracy of model parameters can be tested by direct comparison with a “gold standard.” To test the accuracy of regional measurements, such a validation study can only be carried out with animals. While this validation step is very appealing, it is often very difficult to achieve. There is often no gold standard available for the measurement of interest. Even if such a standard is available, the comparison will require careful matching of scan data with tissue sample data. If the regions being compared are small, the effects of inaccurate registration and scanner resolution can make evaluation of the model’s accuracy difficult at best. Even without a gold standard, other validations of the model can be performed. For example, model predictions of concentrations in separate compartments can be compared to biochemical measurements of tissue samples [104]. Also, microdialysis provides a method to assess extracellular tracer concentration directly for comparison with model predictions [105].

### Model-based Methods

To this point, the design, development, and validation of a tracer kinetic model have been presented. Ideally, the modeling effort generates a complete, validated model that describes the relationship between tissue measurements and the underlying physiological parameters. With this knowledge, we can design a method of data acquisition and processing suitable for human studies. This section concerns this final step in the modeling process shown in Fig. 6.1: the adaptation of such a useful model to produce a practical patient protocol [69, 106]. It is often the case that the original modeling studies are complex and may not be suitable for human subjects, particularly certain patient populations. For example, arterial blood sampling may not be feasible, or a long data acquisition period may not be practical, or the statistical quality of data in humans may limit the number of parameters that can be reliably estimated. From the understanding of the characteristics of the tracer and with knowledge of the limitations imposed by instrumentation and logistical considerations, a model-based method can be developed that can achieve a useful level of physiological accuracy and reliability.

Many questions must be considered in converting a model into a model-based method. To what extent are the extra complexities of a full modeling study necessary or useful? What are the best trade-offs to maintain an adequate signal-to-noise ratio in the data? Can an appropriate input function be measured less invasively than from arterial samples, e.g., from direct scan measurements, from venous samples, or from a reference region? What is a practical data collection period that is compatible with the time availability on the scanner, the statistical requirements of the collected images, and the characteristics of the patients? Which parameters are of prime importance? Can parameter estimates be calculated on a pixel-by-pixel basis to generate functional images or must time-consuming iterative nonlinear methods be applied to region-of-interest data? What reasonable assumptions can be incorporated into the model to reduce the number of parameters to a workable set that can be determined with reasonable precision? Is the method overly sensitive to measurement errors or to inaccuracies in model assumptions, particularly in patient groups? If the method is simplified too much, could differences between patients and controls be exaggerated or hidden because physiological factors properly included in the original model are now ignored?
This section presents some approaches that have been used to produce model-based methods. These methods are generally simpler than the full parameter estimation studies, use additional assumptions, and typically allow production of functional images of the physiological estimates. In addition, sources of error in model approaches are discussed along with error analysis methodology. Finally, the trade-offs between using model-based techniques and simple empirical methods are examined.

**Graphical Analysis**

One increasingly common method applied to tracer kinetic data is that of graphical analysis [90, 107–112]. The basic concept of this method is that after appropriate mathematical transformation, the measured data can be converted into a straight-line plot whose slope and/or intercept has physiological meaning. This approach has advantages, since it is simple to verify visually the linearity of the data and it is simple to determine the slope and intercept by non-iterative linear regression methods. It is also generally easy to determine these values on a pixel-by-pixel basis, thus producing a functional image of the parameter [113]. For many models, the simplified equations of graphical analysis will not apply for all times post-injection, e.g., at early times when the blood activity is changing rapidly and some tissue compartments have not yet reached equilibrium with the blood. Therefore, care must be taken in selecting the time period for determination of the slope and intercept. However, it is also true that avoiding the time periods where the kinetics are rapid also makes the method less sensitive to errors introduced by oversimplifications in the model, particularly those dealing with tracer exchange between capillary, extracellular space, and intracellular space.

The most widely used graphical analysis technique is the Patlak plot [107–109]. This approach is appropriate when there is an irreversible or nearly irreversible trapping step in the model. Conceptually, the transformations of the Patlak plot convert a bolus injection experiment to a constant infusion. A simple example of this model is the two-compartment model (Fig. 6.3C), in which the rate constant for return of tracer from compartment 2 to compartment 1, $k_2$, is zero or is small, i.e., irreversible trapping. In this case, the model solution (from Eqs. 21, 22 and 28) for the total tissue tracer concentration $C(t)$ for an arbitrary input function $C_a(t)$ is

$$C(t) = C_a(t) \left( \frac{K_1 k_2}{k_2 + k_3} \exp[-(k_2 + k_3)t] + \frac{K_1 k_3}{k_2 + k_3} \right)$$  \hspace{1cm} (33)$$

If the arterial input function were held constant ($C_a$), the solution to Eq. 33 would be

$$C(t) = C_a \left( \frac{K_1 k_2}{k_2 + k_3} (1 - \exp[-(k_2 + k_3)t]) + \frac{K_1 k_3}{k_2 + k_3} t \right)$$  \hspace{1cm} (34)$$

After an appropriate time, $t^*$, after which the exponential term in Eq. 34 becomes sufficiently small, the ratio of tissue to blood activity becomes

$$\frac{C(t)}{C_a} = \frac{K_1 k_2}{(k_2 + k_3)^2} + Kt$$  \hspace{1cm} (35)$$

which is a linear equation. The slope of this equation, $K$, is

$$K = \frac{K_1 k_3}{k_2 + k_3}$$  \hspace{1cm} (36)$$

The term $K$ is the net uptake rate of tracer into the irreversibly bound compartment 2. It is the product of two terms: $K_1$, the rate of entry into the tissue from the blood, and $k_3/(k_2 + k_3)$, the fraction of the tracer in the tissue that reaches the irreversible compartment (Fig. 6.3C).

For the case when the input function is not a constant, the Patlak transformation is as follows:

$$C(t) = V_0 + K \left( t \int C_a(s) ds \right)$$  \hspace{1cm} (37)$$

The term in brackets in Eq. 37 is often called stretched time or normalized time, since it has units of time and it distorts time based on the shape of the input function. If the ratio of tissue to blood activity, which is called the apparent volume of distribution, is plotted versus stretched time, under the appropriate conditions a linear plot is obtained with slope $K$ and intercept $V_0$ (the initial volume of distribution). Note that in the case of a constant arterial input, stretched time becomes exactly equal to true time.

In applying this graphical method, it is important to verify that the Patlak plot is in fact linear over the range of time used, an assumption that can often be evaluated in animal studies, where longer experiments can be performed [114]. For purposes of fitting data to estimate $K$, instead of fitting Eq. 37, it is equivalent to use multiple linear regression to fit the measured tissue data directly.
This approach is better if the later values of the input function are noisy (e.g., due to metabolite correction), and more easily allows regression weights based directly on image noise estimates to be added to the estimation process.

Figure 6.8 provides an example of the use of a Patlak plot as applied to brain PET data after the injection of FDG [53]. In this study, subjects were studied on two occasions, approximately one week apart. For one scan, the subjects underwent a hyperinsulinemic euglycemic clamp, whereby high levels of insulin were infused, and simultaneously blood glucose levels were maintained at a constant level, thus maintaining the steady-state assumption of the tracer kinetic model. On the second occasion, a sham clamp was performed, i.e., a control study. The high insulin levels in the clamp study caused a dramatic change in the plasma input function, i.e., the rate of FDG clearance from plasma was much higher. Figure 6.8a shows the tissue curves for an average of gray matter regions in one individual. There is clearly a dramatic difference in the two curves. The Patlak transformation of Eq. 37 was applied to these data and is shown in Fig. 6.8b with a plot of the apparent volume of distribution versus stretched time. The two plots nearly overlay each other, demonstrating that the hyperinsulinemic study covers a longer period in stretched time than the control study.

A second graphical approach is that developed for measurement of parameters for reversible neuroreceptor ligands, i.e., those that approach equilibrium during the time period of the experiment. As described above, the total volume of distribution \( V \) is the most commonly estimated parameter for these types of tracers. The Logan graphical relationship [111] allows the estimation of \( V \) from the slope of a plot produced by a transformation of the data, like the Patlak plot described above. The Logan relationship can be derived exactly from the one tissue compartment model, Eq. 15, and integrating:

\[
C(t) = V_0 C_a(t) + K \int_0^t C_a(s)ds
\]  

(38)

Dividing Eq. 39 by \( k_2 \) and \( C(t) \), and rearranging yields:

\[
\frac{\int_0^t C(s)ds}{C(t)} = V \frac{\int_0^t C_a(s)ds}{C(t)} - \frac{1}{k_2}
\]

(40)

where the slope of this relationship \( V \) is the volume of distribution for the one tissue compartment model \((K_1/k_2)\). In cases where the data are not consistent with a one-compartment model, the graph becomes linear after an appropriate time, and the linear regression is performed for those later data. In that case, the slope is the estimate of the total volume of distribution. Figure 6.9 shows an example of Logan graphical analysis [111] as applied to PET time–activity data for the 5-HT\(_{1A}\) antagonist [\(^{18}\)F]FCWAY [115] as measured in the rhesus monkey. The three curves show regions with different receptor levels, with the highest slope (frontal cortex) corresponding to a region with high specific binding.
Logan graphical analysis of regional PET data acquired from the 5-HT1A antagonist $[^{18}F]$FWAY in rhesus monkey in frontal cortex (●), thalamus (♦), and cerebellum (▲). Data are transformed as specified in Eq. 40. Following a certain time, the graphs become straight lines with slopes equal to the volumes of distribution for each region. Regions with greater specific binding have higher slopes.

and a high value of $V$. Note that the time to achieve linearity of these plots differs between regions due to different receptor levels/kinetic parameters.

Reference Region Methods

The emphasis in this chapter has been the determination of kinetic rate constants using the relationship between tissue data measured with the PET scanner and the input function, usually derived from arterial blood samples. For studies in the chest with tracers that do not metabolize, the input function can be measured from the imaging data in the left ventricle, atrium, or the aorta [116–119]. Other approaches have been used where smaller blood vessels can be imaged but corrections for partial volume effect are required [120, 121]. However, in a number of other cases, approaches have been developed to avoid the measurement of the arterial input function and still deduce kinetic parameter information by comparison of the time-activity curve in the region of interest to that in a reference region. The most significant application of this approach has been in receptor modeling where the comparison of regions with and without receptors provides a natural application [122–124], which can often be extended to pixel-by-pixel analysis [125]. The general idea of these approaches is to use the mathematics of the model to infer the shape of the arterial input function based on the time-course measured in the reference region. This permits a mathematical relationship to be developed for the region-of-interest concentration in terms of the reference region data and the kinetic parameters of both regions. Usually, the number of available parameters is reduced, e.g., in this situation the uptake constant $K_r$ for either the region of interest or the reference region cannot be determined, but the ratio between them can be estimated.

In addition, there are reference region methods adapted for graphical analysis, either for irreversible [109] or reversible [126, 127] tracer uptake. As with all the graphical methods, only a subset of the kinetic parameters can be determined, and with the use of reference regions, the estimated parameters are typically ratios of the original parameters between their values in the region of interest and that in the reference region. However, it is often the case that the most sensitive biological parameter is a normalized model value. Normalization tends to eliminate certain methodological errors which add common variance to both the region-of-interest and the reference region results [106]. Therefore, these reference-region graphical methods tend to directly estimate the parameter ratios of interest.

Single-scan Techniques

A common approach to produce simplified model-based methods is the use of single-scan techniques. Here, based on a good understanding of the relationship between tissue radioactivity and the underlying physiological parameters, tissue radioactivity information is acquired during one scan interval. This single measurement permits the estimation of a single unknown physiological parameter. Since most models have multiple rate constants, some corrections must be applied to account for these other unknowns. Careful design of a single-scan technique ensures that variation in these nuisance parameters produces only minor errors in the parameter of interest.

For the measurement of cerebral blood flow with $[^{15}O]$water or comparable diffusible tracers, two approaches have been taken to produce single-scan methods. Some of the earliest studies used continuous inhalation of $[^{15}O]CO_2$ [128], which is rapidly converted to $[^{15}O]water in the lungs. By achieving constant radioactivity levels, the derivative in the differential equation of uptake of the tracer (Eq. 15 with additional terms for radioactive decay) can be set to zero, and $K_1$ can be determined from an algebraic formula in terms of tissue and blood radioactivity. A different approach uses a bolus injection followed by a single short scan [129, 130]. This autoradiographic method uses the explicit solution of the model (Eq. 27) to determine $K_1$ from the integrated tissue radioactivity and a measured input function. Both of these methods treat the estimated $K_1$ values as equal to blood flow, assuming a large permeability-surface (PS) area product for the tracer (Eq. 5). Both methods also
require the use of an assumed value for the tracer distribution volume $V$ in order to specify $k_2$ as $K_f/V$. As only one tissue measurement is made, only one unknown parameter can be determined. The short scan of the autoradiographic method was designed in part to minimize the sensitivity of this method to errors in the assumed value of the distribution volume.

Another example of single-scan, model-based techniques is the autoradiographic method for measurement of glucose metabolism, which was developed in rats with $[14C]$deoxyglucose [1] and extended to PET using $[18F]2$-fluoro-2-deoxy-D-glucose [2–4]. These methods take advantage of the fact that most of the radioactivity in the tissue by 45 min post-injection has been phosphorylated, so that the total tissue radioactivity can be used to estimate the net flux into tissue of deoxyglucose, $K$. This is the same rate constant as determined from the slope of the Patlak plot. Effectively, these methods estimate the slope of a Patlak plot by using the measured tissue value at one data point and by using population values of the model rate constants to estimate the $y$-intercept of the straight line. A number of other formulations of this approach have been developed [131–133], each with different sensitivities to errors in the assumed rate constants. Finally, since FDG is an analog of glucose, the metabolic rate of glucose is estimated from the measured net flux of FDG using the measured plasma glucose level and an assumed scaling factor, the lumped constant [1, 5–9].

Equilibrium Methods

Another single-scan technique has been developed for quantification of receptors by using infusion to produce true equilibrium [84, 134, 135]. By administering tracer as a combination of bolus plus continuous infusion (B/I), constant radioactivity levels can be reached in blood and in all regions of interest. The total tissue volume of distribution can be determined from the ratio of tissue activity to metabolite-corrected plasma activity. This value will include free, non-specifically bound, and specifically bound tracer. Estimates of the nonspecific component, e.g., from a region with low receptor binding, from measurements with an inactive enantiomer, or from data acquired after displacement with excess cold ligand, can be subtracted to estimate the binding potential, $B_{max}/K_D$ [136] ($K_D$ is the dissociation equilibrium constant). Multiple infusions at different specific activities can be used to determine $B_{max}$ [137, 138].

This infusion approach can be extended to provide receptor-binding data in two states: at baseline and post-stimulus (e.g., drug-induced neurotransmitter changes), with a single administration of tracer. Without infusion, such data are conventionally acquired with paired studies, each with a bolus injection. In the first study, control levels of binding are measured, for example, by determining $V$ by compartment modeling [56] or graphical analysis [111]. Then, following the pharmacological intervention, a second measurement of binding is made with a second injection of tracer. This approach has been used successfully with the $D_2$ ligand $[11C]$raclopride [95, 139] as well as with a number of other tracers. For example, Dewey et al. have demonstrated the effects of changes in synaptic dopamine by direct effects on the dopamine system itself [140] and by indirect pharmacological interventions [141, 142]. In humans, this paired-study approach has been used to measure drug occupancy [143–146].

The alternative study design is to administer the tracer as a combined bolus plus continuous infusion (B/I) to measure short-term changes in free receptor concentration [101, 147, 148]. First, the B/I administration of tracer is performed to achieve constant radioactivity levels in blood and all brain regions. Once equilibrium is achieved, control binding levels can be determined. For example, the volume of distribution $V$ can be measured directly from the tissue-to-plasma concentration ratio. Then, a stimulus is administered while the infusion of radiotracer continues, and the change in specific binding of the tracer can be monitored. An example of B/I data is shown in Fig. 6.10 assessing the effects of amphetamine-induced dopamine release with $[11C]$raclopride. By comparing the pre- and post-amphetamine levels of specific binding determined directly from the tissue concentration values (Basal Ganglia/Cerebellum $–1$), the change in specific

---

**Figure 6.10.** ROI data from basal ganglia (●) and cerebellum (■) following combined bolus plus infusion administration of the $D_2$ dopamine ligand $[11C]$raclopride. At 40 min (arrow), 0.4 mg/kg of amphetamine was administered intravenously, producing displacement of raclopride due to competition with increased synaptic dopamine.
Different kinetic responses. This can have minimal to large effects on model results depending on the magnitude of heterogeneity and how the parameter of interest affects the tissue concentration measurements. This effect has been studied in great detail for a number of methods [155–161]. Since finite resolution is unavoidable in real imaging data, ideally application of modeling techniques will not introduce artifactual changes in the data. In other words, suppose a heterogeneous region was composed of two tissue types. Ideally, the final kinetic estimates from that area would be the weighted averages of the appropriate values for each tissue type, weighted by the fraction of the region occupied by each tissue type. If the parameter is estimated in a linear fashion from the data, this will be the case. For non-linear methods, heterogeneity will introduce a bias. An important approach to deal with the partial volume effect is to correct the PET data for this effect [162–166]. Recently, investigators have begun to assess the effects of these corrections on kinetic modeling [167] with tendencies toward major increases in the parameter values and the noise level of the estimates.

Another source of error in model applications is the presence of intravascular radioactivity in the tissue measurements [168–173]. Some fraction of the measured counts originates from radioactivity in the blood within the tissue. Since the radioactivity time-course in blood differs from that in tissue, errors in model measurements will occur unless this effect is properly handled. In some cases, the fraction of tissue volume occupied by blood can be measured in a separate tracer experiment. Alternatively, this vascular fraction is added as a parameter to account for this effect. Obviously, these errors are most important in regions with large blood volumes or in regions near the heart chambers or large blood vessels. Typically, errors due to vascular radioactivity are more significant when data collected immediately after injection are included in the analysis. However, these early data are often most sensitive to the parameter of interest, such as in the case of blood flow tracers where the rate constant for movement of tracer from blood into tissue ($K_i$) is of prime importance. Various strategies involving selection of time intervals for analysis or optimal region-of-interest placement have been proposed to handle these effects [171, 174].

A key to successful quantitative methods is the accurate measurement of the input function. Typically, the blood time–activity curve is measured in a peripheral blood vessel (usually radial artery) unless the heart chambers can be imaged directly [116–118]. When individual blood samples are drawn by hand, they must be taken at a sufficiently rapid rate to characterize the

**Random and Deterministic Errors**

In making the choices necessary to implement a tracer method, it is important to be aware of the many sources of error that affect the precision and accuracy of these physiological measurements [106]. A good understanding of what effects are more or less significant to a given tracer and to the biological question of interest is essential in designing a sensitive, reliable technique that is not overly complex.

One aspect to consider is random errors, i.e., the effects of random statistical noise in the data on model parameters. The duration of data acquisition, the amount of smoothing of the images, the use of pixel or region-of-interest data, the number of parameters in the model, the mathematical structure of the model, and the actual parameter values affect the statistical accuracy of the parameter estimates. Many investigators have assessed the sensitivity of PET data to model parameters and methods to optimize the statistical quality of the model estimates [149–152]. In addition, noise in measured data can directly introduce bias in parameter estimates when non-linear methods are used [66, 153].

A primary source of deterministic error is the measurement of regional radioactivity from the PET scanner. Although the quantitative accuracy of PET continues to improve, there are still many sources of inaccuracies. For example, the accuracy of the scatter correction is limited, particularly for whole-body imaging and for 3D acquisition. A key effect corrupting PET imaging data is finite resolution, i.e., the partial volume effect [154]. The magnitude of bias in concentration measurements depends on the size of the underlying structure, the distribution of radiotracer within and around the structure, the resolution of the scanner, the reconstruction algorithm, and the strategy for extracting regional concentration values. Definition of the regions of interest using registered anatomical images (MR or CT) is important, as long as registration errors are minimized.

The partial volume effect produces heterogeneity, i.e., the tissue response measured from even a single pixel will represent a weighted average of the tissue in the surrounding region, and is thus a combination of different kinetic responses. This can have minimal to

**Positron Emission Tomography**
curve accurately. Careful attention is required for accurate sample timing, centrifugation, pipetting or weighing, radioactivity counting, counting corrections (background, decay, dead time, etc.), and data handling. Some investigators have developed devices for automatic withdrawal and measurement of whole-blood radioactivity [175]. These devices provide consistent data, but they may have increased statistical noise depending upon their counting geometry. Timing and dispersion differences between the brain and the peripheral artery require correction, particularly for studies of short duration with sharp bolus inputs [176–179]. A number of studies have been undertaken to assess the effects of statistical noise in the input function on estimated parameters and to develop appropriate estimation methodology [180–183]. If there are radioactive metabolites of the tracer in blood, it is important to determine the fraction of blood radioactivity that corresponds to the original tracer as well as the extent to which these metabolites pass into tissue. Since metabolite determinations are often complex, particularly for short-lived tracers, metabolite measurements are made at only a small number of samples. Appropriate interpolation or modeling schemes are necessary to generate a continuous estimate of the metabolite fraction throughout the study [101, 184]. Alternatively, other modeling approaches can be used to infer the metabolite correction [185].

**Error Analysis**

Error analysis is a useful tool in the development of an appropriate model-based method. Performance of a thorough error analysis is a critical step in the assessment of the utility of a given method. Papers dedicated solely to error analysis are common in the literature [155, 156, 171, 186–196]. These analyses usually proceed as follows. Choose a particular source of error. Select values for the model parameters and use the model equations to simulate tissue data including this error effect, usually covering a range of effect magnitudes. Then, analyze these simulated measurements with one or more methods, compare the derived parameter estimates to their original values, and determine the magnitude of error that is produced.

Figure 6.11 provides an example of the results of an error analysis. Cerebral blood flow (CBF) measurements with the tracer \(^{[15]}\)O water are altered in the presence of errors in correction for the time delay between the measured arterial input function and the actual input to the brain. Using an actual measured input function, tissue time-activity data were simulated over a 4-min period using the model of Eq. 27, with a flow value of 0.5 mL/min/g and a distribution volume of 0.8 mL/g. CBF (K

\[
K_1
\]

was then calculated by direct estimation of the two model parameters for total time intervals of 90, 120, and 240 sec. In each case, the tissue data were shifted with respect to the arterial input function by –3 to +3 sec (a positive shift means that the tissue data have been shifted forward in time with respect to the arterial input function). The three curves show the percent error in estimated flow, based on data collection periods of 90 sec, 120 sec, and 240 sec. See text for additional details.
actual errors may be larger or smaller than those predicted from independent error analyses. Therefore, it is best if the ultimate choice of a method can be made by analyzing many studies with a variety of techniques and choosing the approach that has the best reproducibility, the minimum population variability, or the maximum statistical power to extract a particular physiological signal.

**Selection of Model-based Methods**

This chapter has presented an overview of modeling methods, from the most complex dynamic data acquisition with iterative parameter estimation to simplified methods including Patlak and Logan plots or single-scan techniques. Choosing the best approach is not simple, and other options are available when selecting a tracer method. In some studies, investigators normalize the physiological measurements. Instead of using the absolute values provided by a method, the results are scaled in some manner by a reference value, such as the average value in the entire organ or in a particular reference structure. This procedure may significantly reduce inter-subject variation introduced by instrumentation, reconstruction, errors in the measurement of the input function, as well as variability due to global flow, metabolism, etc. In some cases where the model equations are linear (or nearly so) with respect to the parameter of interest, investigators can avoid the measurement of the input function. However, in other cases a ratio is used. These different formulations are more common because it can be estimated without measurement of the plasma input function. In some cases, the difference of the $V$ values is used and in other cases a ratio is used. These different formulations have different characteristics in terms of biological interpretation as well as within-subject and between-subject variability. For example, the ratio formulation is more common because it can be estimated without measurement of the plasma input function. However, in that case, the results depend upon the assumption that the level of nonspecific tracer binding is unchanged between regions and between subject groups.

An alternative to using a model-based method is to use a simple empirical approach. Such approaches make no explicit attempt to estimate the physiological parameter(s) of interest. Instead, an index based on tissue measurements is used and presumed to reflect the underlying physiology. Empirical indices include absolute radioactivity values, radioactivity values corrected for dose and/or subject weight, and ratios of radioactivity values between target and reference regions (normalized values).

How can an investigator determine the best approach when using a tracer? Many trade-offs must be considered in designing a study, and there are no simple answers [106]. As an example, consider the use of a receptor-binding radiotracer for measurements in the brain with PET. Suppose the tracer binds reversibly, i.e., its dissociation rate from the receptor is sufficiently fast to approach equilibrium during the study period. Possible model-based quantification approaches include the following: 1) complete modeling study with iterative parameter estimation; 2) use of a simplified model with estimation of the volume of distribution [56]; and 3) use of a linearization formula to derive the volume of distribution from the later portion of the data [111]. Empirical alternatives to model-based methods include the following: 1) ratio of tissue region of interest to (metabolite-corrected) blood (apparent volume of distribution); or 2) ratio of tissue region of interest to reference region with few receptors during the apparent equilibrium phase.

Although the empirical approaches are the simplest, they can provide misleading results. For tracers that can reversibly bind with receptors, indices derived from ratios of tissue concentration to reference regions or to plasma levels can be significantly distorted due to lack of true equilibrium [84]. This effect is demonstrated in Fig. 6.12 with a "bolus plus infusion" protocol using the opiate antagonist [18F]cyclofoxy (see section on single-scan techniques). Radioactivity in the tissue regions (Fig. 6.12a) reached steady levels by ~20 min. At 70 min post-injection (arrow), the infusion was discontinued, and plasma and tissue concentrations dropped. Figure 6.12b shows the apparent volume of distribution plotted against time. Discontinuing the infusion caused a dramatic increase in the values for the receptor-rich thalamus with smaller increases in frontal cortex and cerebellum. The magnitude of this effect depends upon the relative magnitudes of the rate of tracer clearance from plasma and the receptor dissociation rate. The change in the apparent distribution volume value (Fig. 6.12b) is due solely to the change in clearance of radiotracer from plasma and demonstrates that this ratio measure can be significantly affected by the plasma clearance rate.
Another choice to be made with this type of study is whether tracer administered by continuous infusion is a good idea [84, 134]. With infusions, data analysis is greatly simplified, since the volume of distribution can be obtained directly from the ratio of tissue radioactivity to metabolite-corrected blood. Scans need only be collected during the equilibrium period providing more patient comfort. Fewer measurements in blood are required. Also, the technique is model-independent and only relies on equilibrium conditions. However, if true equilibrium is not obtained, errors that could have been eliminated by a more complex modeling procedure will occur. Due to normal variation in plasma clearance rates of the tracer, deviations from equilibrium will add variability to the results, although biases here will be smaller than those following bolus injections [204]. The time interval corresponding to true equilibrium must be carefully assessed and ideally verified in each subject. There are also increased logistical requirements due to a long infusion of radioactivity compared to a simple bolus injection. It is also not at all clear whether bolus or infusion approaches provide better statistical quality in the final physiological measurements.

Does the use of model-based methods improve the signal-to-noise characteristics of data? In other words, can small biological signals be detected more easily by using modeling methodology? Use of appropriate quantification methodology can reduce intersubject variability by accounting for factors affecting the raw concentration measurements that are unrelated to the physiological measure of interest. If inter-subject variability is decreased, the power of the study to detect group differences is typically increased. However, if this extraneous variability is small, then use of a model-based method may produce little improvement in the signal-to-noise ratio. In fact, since there are a large number of potential sources of error in applying modeling techniques, errors in these corrections or in the implementation of these procedures can actually increase variability over simpler, empirical methods. The net effect of applying a model on measurement variability thus depends upon the magnitude of physiological variation in the patient groups that can be removed by the model versus the accuracy of the model and the reliability of the additional measurements that it requires.

Model-based methods have one important advantage over empirical approaches. With model-based results, it is easier to justify the conclusion that any significant findings are in fact due to real differences in the biological function of interest and not due to extraneous physiological factors. When empirical methods detect significant differences, these other physiological factors may contribute substantially to the measured differences. Thus, interpretation of the results is less straightforward. This is particularly true when there are known differences in physiology between subject groups in a study. For example, if plasma tracer clearance differs between patients and control subjects, substantial errors may be made if tissue radioactivity values are directly interpreted as reflecting the relevant physiological process. On these grounds, model-based methods, which usually require a more complicated study procedure, are superior to empirical approaches. It is important, however, to remember that model-based methods rely on many assumptions, which can produce misleading results when applied inappropriately.
Summary

This chapter discussed the use of mathematical models to extract physiological information from PET studies with radioactive tracers. Modeling methods offer a number of advantages. Application of a model can explain to what extent the tissue radioactivity measurements reflect the physiological function of interest. It can produce quantitative estimates of one or more physiological parameters. Use of a model can explain the cause of different levels of uptake between subjects. It may improve the signal-to-noise characteristics of the data by removing additional variation caused by extraneous physiological factors. The application of modeling methodology also has disadvantages. Usually, modeling procedures are more complex, often requiring longer scanning sessions, blood sampling, metabolite analyses, and complex data processing. Violations in the assumptions made by models can produce misleading results.

Validation studies can demonstrate that a model-based method accurately measures the parameter(s) of interest and is not influenced by other factors. The understanding provided by a model allows the development of study procedures that maximize sensitivity to key parameters and minimize the effects of violations in model assumptions. Ideally, the understanding provided by the model will allow the design of a simple straightforward study procedure. In that way, the radiopharmaceutical can be applied to the appropriate patient groups without a complex procedure while still generating an accurate regional physiological assay. The final configuration of a model-based method may be as simple as an empirical technique but as accurate as a more complex study procedure.

It is essential to have a good understanding of the relationship between the tissue measurements and the underlying physiology, i.e., a model. A useful model will provide a mathematical description that is sufficient to predict the tracer’s physiology and biochemistry within the limitations of available instrumentation and the logistics of a practical patient procedure. In addition, the assumptions and limitations of the technique must be clearly delineated. Without a model, it is difficult to assess how physiological differences between study populations affect an empirical method. Ideally, use of a model will significantly improve the physiological significance of the resulting data and may also improve the sensitivity of the tracer to the underlying physiological processes under study.

References


