Dissertation

Modelling, Identification and Error Analysis of Compartment Models for the Determination of Renal Clearance and Plasma Flow

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines
Doktors der technischen Wissenschaften
unter der Leitung von

Univ. Prof. Dipl.-Ing. Dr. Felix Breitenecker
E 101
Institut für Analysis and Scientific Computing

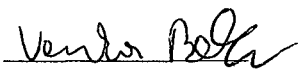
eingereicht an der Technischen Universität Wien
Technisch-naturwissenschaftliche Falkultät

durch

VERONIKA BOLTZ

Dornbacherstr. 71-73/2/1
A-1170 Wien
Matrikel Nummer: 9625931

Wien, 2004-10-27



Unterschrift

First of all, I'd like to express my thanks to Dr. Felix Breitenecker for giving me the opportunity and the support that made this thesis possible. I am also very grateful to Dr. Willibald Estelberger who was always there for me and my questions and whose knowledge and patience were of great assistance to me.

Furthermore, I want to deeply thank my parents who always believed in me and who made all of this possible for me by giving me their backing and support throughout all the years of my studies.
Special thanks are regarded to my boyfriend Michael who not only enriched my life but whose love and understanding have also carried me through all the good and tough times of the past few years.

Zusammenfassung

In der Modellierung von Ausscheidungsprozessen von exogen verabreichten Markern gibt es zwei grundsätzliche Fehlerquellen. Einerseits das sogenannte Datenrauschen, das durch zufällige und systemische Schwankungen um den idealen Modellverlauf ensteht, und andererseits, die Inhomogenität des injizierten Markers. Durch all diese Unsicherheiten ist eine ausreichend lange Protokolldauer von entscheidender Wichtigkeit, da ansonsten die Gefahr einer Überschätzung der Nierenfunktion besteht.

In den Nieren ensteht der Harn durch Filterung des Blutplasmas im Glomerulum und durch Resorption und Sekretion von Stoffen im Verlauf des anschliessenden Tubulus. Die Bestimmung dieser Filtrationsleistung (GFR) ist von grosser Bedeutung um den Nierengesundheitszustand beurteilen zu können, aber auch der renale Plasmafluss (RPF) spielt eine nicht unwesentliche Rolle in der klinischen Diagnostik.
Ein Zweikompartmentmodell, das sogenannte Grundmodell der Pharmakokinetik, kann problemlos zur Bestimmung der GFR verwendet werden. Ist man jedoch ebenfalls an dem RPF interessiert, so muss das Modell zu einer nichtlinearen Version erweitert werden. Diese Nichtlinearität ist notwendig, da die tubuläre Sekretion aufgrund der nur begrenzten Transportfähigkeit der dafür benötigten Carrier einer Sättigung unterliegt.

Fr die Bestimmung der GFR ist es jedoch essentiell die Werte der zugrunde liegenden Parameter zu kennen. Um die Parameter dieser Modelle zu bestimmen muss man ein Abweichungsmass minimieren, das als die Summe der Quadrate der Differenzen zwischen den gemessenen und errechneten Daten beschrieben ist (Methode der kleinsten Fehlerquadrate). Da das System jedoch nichtlinear in den Parametern ist bedeutet das auch die Suche nach einem Minimum einer nichtlinearen Funktion, dadurch muss die Identifikation numerisch erfolgen. Ein möglicher Algorithmus hierfür ist der Marquardt-Levenberg-Algorithmus, der eine Kombination aus der beliebten Linearisationsmethode (Gauss - Newton) und der Methode des steilsten Abstiegs darstellt.
Nun machen es die eingangs erwähnten Fehlerquellen jedoch notwendig, über die Fehlerbreiten der gefundenen Parameter bescheid zu wissen. Diese Fehlerabschätzung wird mit der sogenannten Monte Carlo Methode durchgeführt. Dabei werden Kunstprotokolle erstellt indem man einem "perfekten" Systemoutput Zufallszahlen überlagert, die aus einer Normalverteilung mit Mittelwert 0 gewonnen werden. Ungefähr 100 solche Kunstprotokolle sind ausreichend um die gewünschten Parametervarianzen auszurechenen. Diese wiederum sollten mit den Standardfehlern, die durch die sogenannte Fischer-Informationsma-trix - Methode gewonnen werden, übereinstimmen.

Abstract

In modelling elimination kinetics of exogenously applied markers there are always two kinds of error sources: first, there is always noise in the experimental data and second, the injected marker is not a homogenous substance but consists of different masses. Those uncertainties make a sufficiently long protocol length indispensable since a too short protocol length could result in an overestimation of the clearance value.

In the kidneys, a fluid that resembles plasma is filtered through the glomerular capillaries into the renal tubules (glomerular filtration). The assessment of this GFR as well as the assessment of the total renal plasma flow is of great importance in clinical decision making in the follow-up of renal patients. For the solution of this quantification problem computer-based modelling and identification techniques can be applied to the investigation of the elimination kinetics from venous plasma after infusion of suitable markers (such as inulin or sinistrin).
A two-compartment model, the so-called basic model of pharmacokinetics, is suitable for the problem of determining the renal clearance. For the assessment of the renal plasma flow this model has to be expanded into a nonlinear version. The nonlinearity results form the superposition of excretion processes obeying a saturation process (Michaelis-Menten law) in tubular secretion, the elimination process of the marker (e.g. p-amino-hippuric acid) being simply proportional to its plasma level.

In order to be able to calculate this renal clearance, the according parameter values need to be known. Parameter estimation involves searching the minimum of the objective function which is defined as the sum of the squares of the differences between each experimental data and the model response (method of least squares).
As the system is nonlinear in its parameters the problem becomes one of locating the minimum of a non-linear function and identification has to be done numerically. One of the most generally useful and widely applied techniques for doing so is the Marquardt-Levenberg algorithm which combines steepest descent and linearization methods.
Since there is the problem of different masses and as there is always noise in the experimental data consisting of random and systematic fluctuations around the ideal behaviour, it is essential to know how precise each of the found parameter estimates is. The error estimation done with a Monte Carlo method, is processed with statistics of a number of adapted artificial protocols. About 100 created protocols are sufficient for estimating the parameter variance. The standard deviations thus found are equivalent to the standard errors derived by means of the so-called Fishers information matrix method.

Contents

Chapter 1

Physiological Background

In the kidneys, a fluid that resembles plasma is filtered through the glomerular capillaries into the renal tubules (**glomerular filtration**). As this glomerular filtrate passes down the tubules, its volume is reduced and its composition altered by the processes of **tubular reabsorption** (removal of water and solutes from the tubular fluid) and **tubular secretion** (secretion of solutes into the tubular fluid) to form the urine that enters the renal pelvis. Wastes are eliminated while water and important electrolytes and metabolites are conserved. Furthermore, the composition of the urine can be varied, and many homeostatic regulatory mechanisms minimize or prevent changes in the composition of the extracellularfluid by changing the amount of water and various specific solutes in the urine.

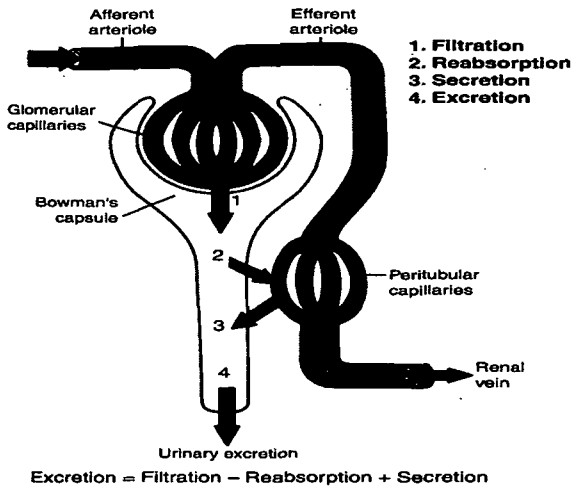


Figure 1.1: Formation of urine

From the renal pelvis, the urine passes to the bladder and is expelled to the exterior by the process of urination.

1.1 Anatomy of the Kidney

Each individual renal tubule and its glomerulus is a unit (**nephron**) of which there are about 1.3 million in each human kidney. Each nephron is capable of forming urine by itself and therefore it is not necessary to discuss the entire kidney but merely the function of a single nephron to explain the function of the kidney.
Blood enters the glomerulus through the afferent arteriole and then leaves through the efferent arteriole. The glomerulus is a network of up to 50 parallel branching and anastomosing capillaries covered by epithelial cells and encased in Bowman's capsule. Pressure of the blood in the glomerulus causes fluid to filter into Bowman's capsule, and from here the fluid flows into the **proximal tubule** that lies in the cortex of the kidney along with the glomerulus.

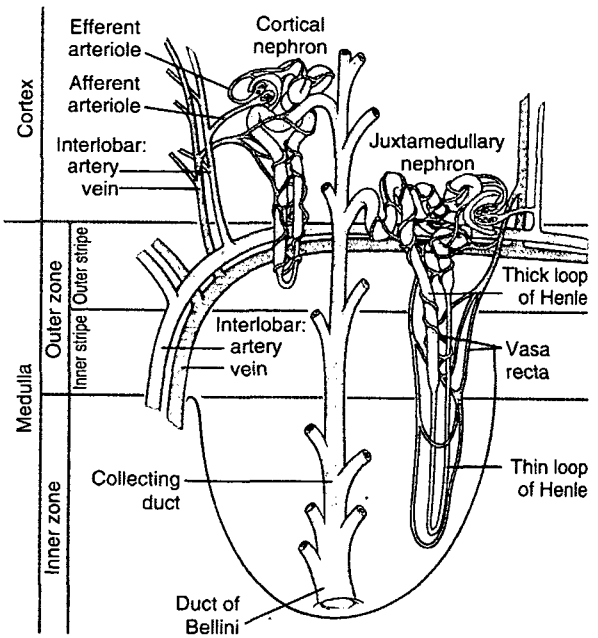
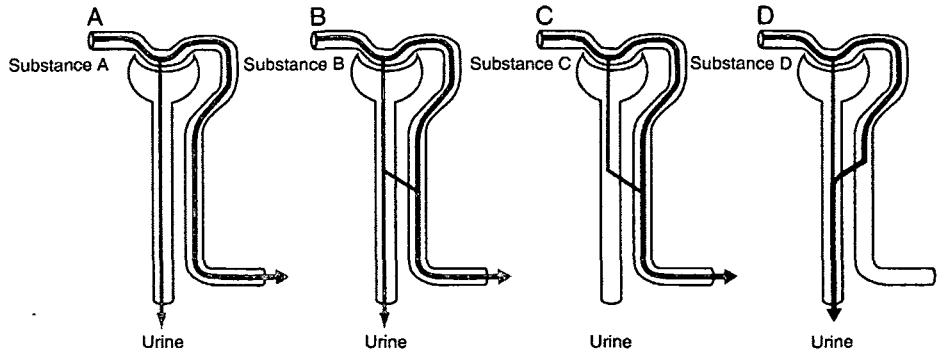


Figure 1.2: The nephron

From the proximal tubule the fluid passes into the **loop of Henle** that dips deeply into the kidney mass, some of the loops passing all the way to the bottom of the renal medulla. Each loop is divided into the **descending limb** and the **ascending limb**. The wall of the descending limb and the lower end of the ascending limb is very thin and therefore is called the thin segment of the loop of Henle. However, after the ascending limb of the loop has returned part way back in the cortical direction, its wall once again becomes thick like that of the other portion of the tubular system; this portion of the loop of Henle is called the thick segment of the ascending limb.
After passing through the loop of Henle, the fluid enters the **distal tubule**, which, like the proximal tubule, lies in the renal cortex. Then, still in the cortex, as many as eight of the distal tubules coalesce to form the cortical collecting duct (also called collecting tubule), the end of which turns once again away from the cortex and passes downward into the medulla, where it becomes the medullary collecting duct but usually called simply the **collecting duct**. Successive generations of collecting ducts coalesce to form progressively larger collecting ducts that penetrate all the way through the medulla, parallel to the loops of Henle. The largest collecting ducts empty into the renal pelvis through the tips of the renal papillae; these are conical projections of the medulla that protrude inte the renal calyces, which are themselves recesses of these very large collecting ducts, each of which transmits the urine from about 4 000 nephrons. [1, 41]

The basic function of the nephron is to clean the blood plasma of unwanted substances as it passes through the kidney. The substances that must be cleared include particularly the end products of metabolism, such as urea, creatinine, uric acid, and urates. In addition, many other substances, such as sodium ions, tend to accumulate in the body in excess quantities; it is the function of the nephron also to clear the plasma of these excesses.
The principal mechanism by which the nephron clears the plasma of unwanted substances is as follows: It filters a large proportion of the plasma in the flowing glomerular blood, usually about one fifth of it, through the glomerular membrane into the tubular system of the nephron. Then, as this filtered fluid flows through the tubules, the unwanted substances fail to be reabsorbed while the wanted substances, especially almost all of the water and many of the electrolytes, are reabsorbed back into the plasma of the peritubular capillaries.
A second mechanism by which the nephron clears the plasma of other unwanted substances is by secretion. That is, substances are secreted from the plasma directly through the epithelial cells lining the tubules into the tubular fluid. Thus, the urine that is eventually formed is composed mainly of filtered substances but also of a small amount of secreted substances.



Renal handling of four hypothetical substances. The substance in Panel A is freely filtered but not reabsorbed. The substance in Panel B is freely filtered, but part of the filtered load is reabsorbed back in the blood. The substance in Panel C is freely filtered but is not excreted in the urine because all the filtered substance is reabsorbed from the tubules into the blood. The substance in Panel D is freely filtered and is not reabsorbed but is secreted from the peritubular capillary blood into the renal tubules.

Figure 1.3: Reabsorption and secretion

1.1.1 The Nephron

As mentioned before, the glomerulus, which is about $200\mu m$ in diameter, is formed by the invagination of a tuft of capillaries into the dilated blind end of the nephron (**Bowman's capsule**). The capillaries are supplied by an **afferent arteriole** and drained by a slightly smaller **efferent arteriole**. There are two cellular layers separating the blood from the glomerular filtrate in Bowman's capsule: the capillary endothelium and the specialized epithelium that lies on top of the glomerular capillaries. These layers are separated by a basal lamina. Stellate cells called **mesangial cells** are located between the basal lamina and the endothelium. These mesangial cells are contractile and play an important role in the regulation of glomerular filtration[1].

The tremendous permeability of the glomerular membrane is caused by its special structure. The capillary **endothelial cells** lining the glomerulus are perforated by literally thousands of small holes of about $50 - 100nm$ diameter called **fenestrae**. Then, outside the endothelial cells, the **basement membrane** does not contain visible gaps or pores but is composed mainly of a meshwork of collagen and proteoglycan fibrillae that also have large spaces through which fluid can filter. The final layer of the glomerular membrane, the layer of **epithelial cells**, lines the outer surfaces of the glomerulus. However, these cells are not continuous but instead consist mainly of fingerlike

[1] see "Autoregulation of the Kidney"

projections, the **podocytes**, that cover the basal lamina. These "fingers" form **filtration slits** that are approximately $25nm$ wide and each is closed by a thin membrane.

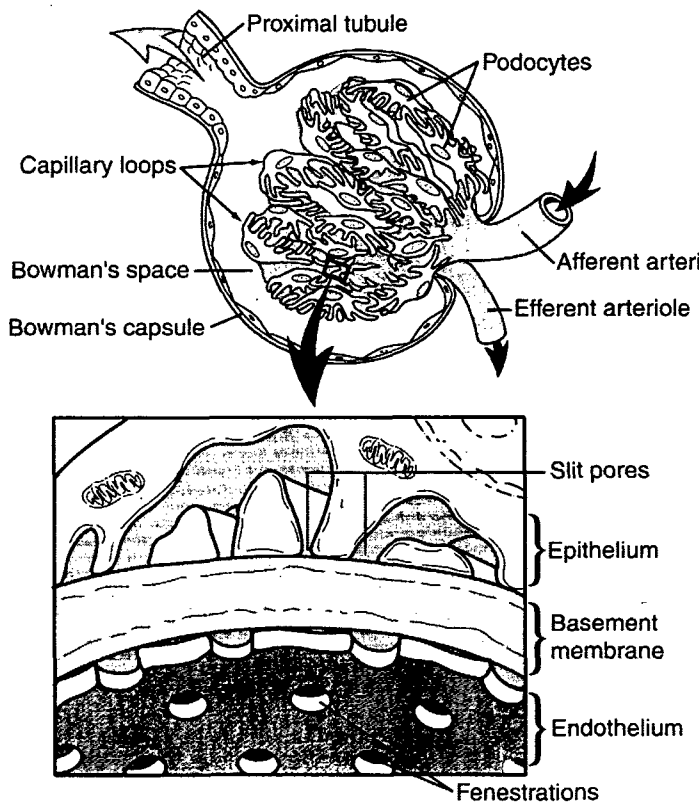


Figure 1.4: Filtration in the glomerulus

Thus, the glomerular filtrate passes through three different layers before entering Bowman's capsule, but each of these layers is several hundred times as porous as the usual capillary membrane, which accounts for the tremendous volume of glomerular filtrate that can be formed each minute.
Yet, despite its permeability, it has an extremely high degree of selectivity for the sizes of molecules that it allows to pass. Functionally, the glomerular membrane permits the free passage of neutral substances up to $4nm$ in diameter and almost totally excludes those with diameters greater than $8nm$. However, the charges on molecules as well as their diameters affect

their passage into Bowman's capsule. The reason for this is that the basement membrane portions of the glomerular pores are lined with a complex of proteoglycans that have very strong negative electrical charges. Plasma proteins, for example, also have strong negative charges. Therefore, electrostatic repulsion of the molecules by the pore walls keeps virtually all protein molecules larger than 69,000 molecular weight from passing through. [2, 41, 1]

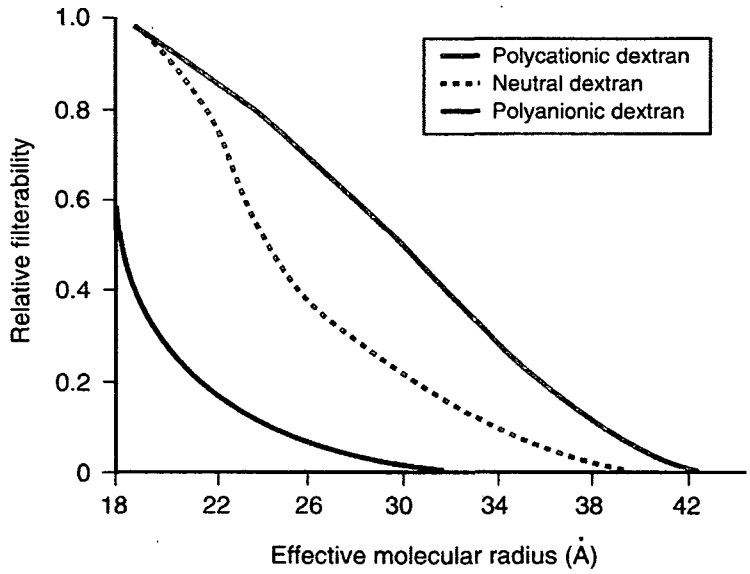


Figure 1.5: Filtration depending on the particle size

1.2 Formation of Urine: Renal Blood Flow and Glomerular Filtration

In a resting adult the rate of blood flow through both kidneys is about $1200ml/min$. The portion of the total cardiac output that passes through the kidneys is called the **renal fraction**. Since the normal cardiac output of a $70kg$ man is about $5600ml/min$, one can calculate easily that the normal renal fraction is about 21 per cent, although their weight is only about 0.5% of the total body weight. But the high blood flow is necassary for the high blood pressure of $60mmHg$ that is the driving force that causes the plasma to filter through the membrane into Bowman's capsule. [1, 41]

1.2.1 The Glomerular Filtration Rate

The quantity of glomerular filtrate formed each minute in all nephrons of both kidneys is called the **glomerular filtration rate** or short **GFR**. The GFR in an average-sized man is approximately $125ml/min$, values in women being 10 per cent lower than those in men. It should be noted that $125ml/min$ is $7.5L/h$ or $180L/d$, whereas the normal urine volume is about $0.5 - 2.0L/d$. Thus, 99 per cent or more of the filtrate is reabsorbed, with the remaining small portion passing to the urine.

Measuring the GFR: The glomerular filtration rate can be measured in intact animals and humans by measuring the excretion and plasma level of a substance that is freely filtered through the glomeruli and neither secreted nor reabsorbed by the tubules. If the substance is designated by the letter X, the GFR is equal to the concentration of X in urine (U_X) times the urine flow per unit of time ($\dot{V}$) divided by the arterial plasma level of $X(P_X)$, or $U_X \dot{V}/P_X$. This value is called the **clearence** of X (C_X). [41, 40]
In addition to the requirement that it be freely filtered and neither reabsorbed nor secreted in the tubules, a substance suitable for measuring the GFR should not be subject to metabolismInulin, a polymer of fructose with a molecular weight of 5200 that is found in dahlia tubers, meets these criteria in humans and most animals and is extensively used to measure GFR. Another such substance used is Sinistrin.
In practice, a loading dose of Inulin (or another marker) is administered intravenously, followed by a sustaining infusion to keep the arterial plasma level constant. After the inulin has equilibrated with the body fluids, an accurately timed urine specimen is collected and a plasma sample obtained halfway through the collection. Plasma and urinary concentrations are determined and the clearance calculated (Inulinclearance = GFR).

The factors governing filtration across the glomerulus capillaries are the same as those governing filtration across all other capillaries, ie, the size of the capillary bed, the permeability of the capillaries, and the hydrostatic and osmotic pressure gradients across the capillary wall. For each nephron:

$$GFR = K_f [(P_G - P_B - (\pi_G - \pi_B)]$$

The net filtration pressure represents the sum of the hydrostatic and colloid osmotic forces that either favor or oppose filtration across the glomerular capillaries (fig.1.6). These forces include (1) the hydrostatic pressure inside the glomerular capillaries (glomerular hydrostatic pressure, P_G), which promotes filtration, (2) the hydrostatic pressure in Bowman's capsule (P_B) outside the capillaries which opposes filtration, (3) the colloid osmotic pressure of the glomerular capillary plasma protein (π_G), which also opposes filtration and (4) the colloid osmotic pressure of the proteins in Bowman's

capsule (π_B), which promotes filtration. (Under normal conditions, the concentration of protein in the glomerular filtrate is so low that the colloid osmotic pressure of the Bowman's capsule fluid is considered to be zero.) [2, 1]

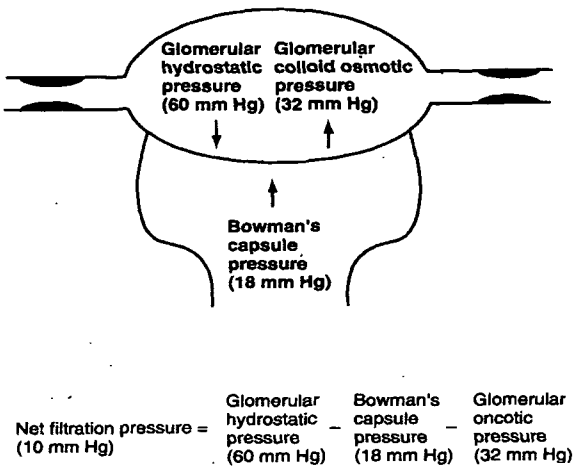


Figure 1.6: Pressures causing and opposing filtration

1.2.2 Control of the GFR and the Renal Blood Flow - Autoregulation of the Kidney

The glomerular filtration rate normally remains nearly constant hour after hour, usually varying very little either above or below the normal value of about $125\,ml/min$ for the two kidneys. Even a change in arterial pressure from as little as $75\,mn\ HG$ to as high as $160\,mn\ Hg$ hardly changes the GFR. This effect is called **autoregulation of glomerular filtration rate.**[2]

To understand the importance of maintaining a constant glomerular filtration rate, let us consider what would happen if the glomerular filtration rate should become, first, very slight or, second, very great.
At a very slight glomerular filtration rate, the tubular fluid would pass through the tubules so slowly that essentially all of it would be reabsorbed. Therefore, the kidney would fail to eliminate the necessary waste products. At the other extreme, with a much too high GFR, the fluid would pass so rapidly through the tubules that they would be unable to reabsorb those

[2] see fig. 1.7

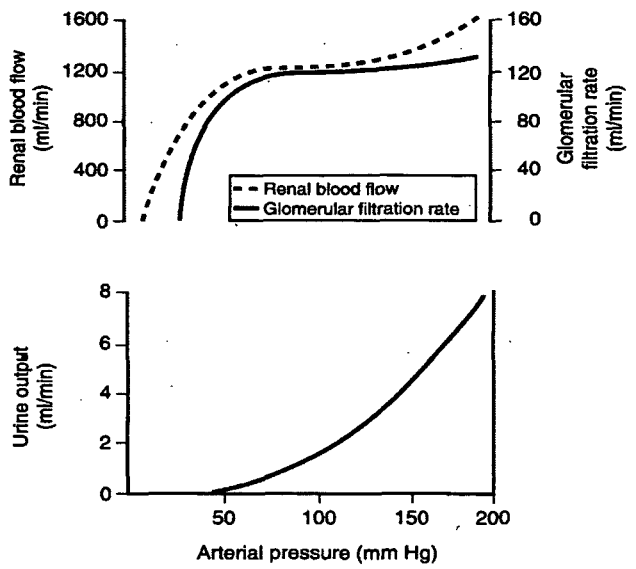


Figure 1.7: Autoregulation of the kidney

substances that need to be conserved in the body.
Thus, one can readily see that the glomerular filtrate must flow into the tubular system at an appropriate rate to allow the unwanted substances to pass on into the urine while reabsorbing the wanted substances. However, it is often not appreciated how narrow the range of glomerular filtration rate must be if optimal function of the tubular system is to occur. Analyses of tubular function have shown that even a 5 per cent too great or too little rate of glomerular filtration can have considerable effects in causing either excess loss of solutes and water into the urine or, at the other extreme, too little of the necessary excretion of waste products.

The precision with which the GFR must be autoregulated demands that there is a highly efficient system for controlling this filtration rate. Each nephron is provided with two special feedbackmechanisms that add together to provide the degree of glomerular filtration autoregulation that is required. These two mechanisms are an afferent arteriolar vasodilator feedback mechanism and an efferent arteriolar vasoconstrictor mechanism. The combination of these two is called **tubuloglomerular feedback**. The feedback mechanism occurs at the **juxtaglomerular apparatus** that consists of the macula densa and the juxtaglomerular cells:

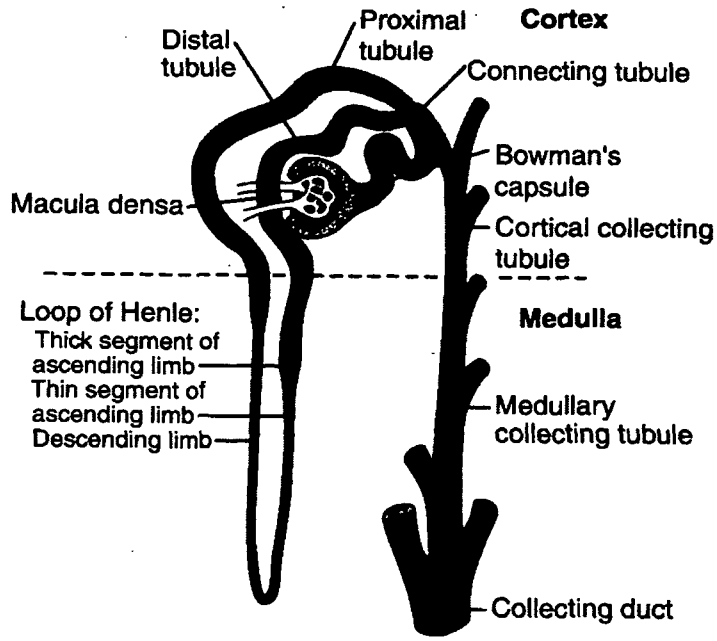


Figure 1.8: The tubulus returns to the glomerulum

As can be seen in Figure 1.8, the initial portion of the distal tubule (immediately after the upper end of the thick segment of the ascending loop of Henle) returns to the glomerulus, passing in the angle between the afferent and efferent arterioles, actually abutting each of these two arterioles. Those epithelial cells of the distal tubule that come in contact with the arterioles are more dense than the other tubular cells and are collectively called the **macula densa**.

Note also in Figure 1.9 that the smooth muscle cells of both the afferent and efferent arterioles are swollen and contain dark granules where they come in contact with the macula densa. These cells are called **juxtaglomerular cells**, and the granules are composed mainly of inactive renin.
The anatomical structure of the juxtaglomerular apparatus strongly suggests that the fluid in the distal tubule in some ways plays an important role in helping to control nephron function by providing feedback signals to both the afferent and the efferent arterioles. [2, 41]

The Afferent Arteriolar Vasodilator Feedback Mechanism: A low rate of glomerular filtration causes overreabsorption of sodium and chlo-

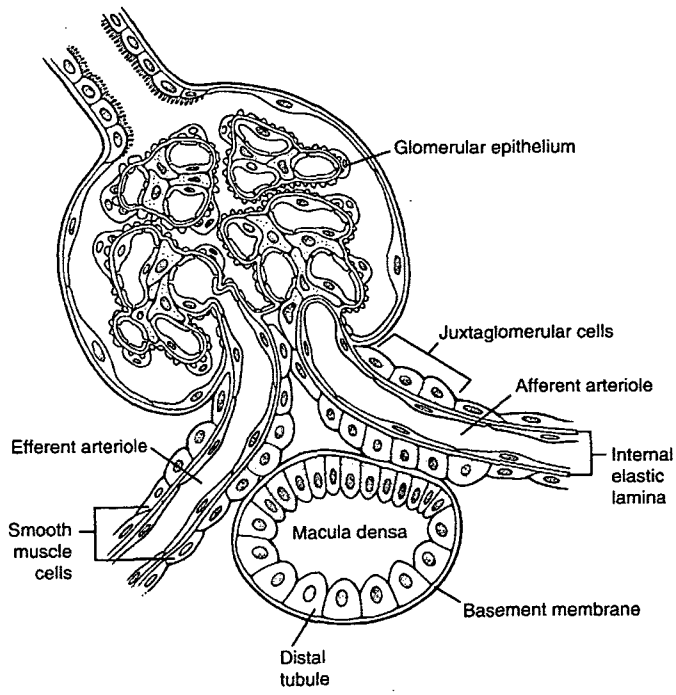


Figure 1.9: The juxtaglomerular apparatus

ride ions in the ascending limb of the loop of Henle and therefore decreases the ionic concentration at the macula densa. This decrease in ions initiates a signal from the macla densa to dilate the afferent arteriole. This in turn allows increased blood flow into the glomerulus, which increases the glomerular pressure. The increased glomerular pressure as well as the increased glomerular blood flow increases the glomerular filtration rate back toward the required level.[3]

Thus, this is a typical negative feedback mechanism for controlling the glomerular filtration rate at a steady rate. This mechanism also helps autoregulate the renal blood flow at the same time, as will be discussed subsequently.

The Efferent Arteriolar Vasoconstrictor Feedback Mechanism: Too few sodium and chloride ions at the macula densa are also believed to cause the juxtaglomerular cells to release active renin, and this in turn causes

[3]see fig 1.10

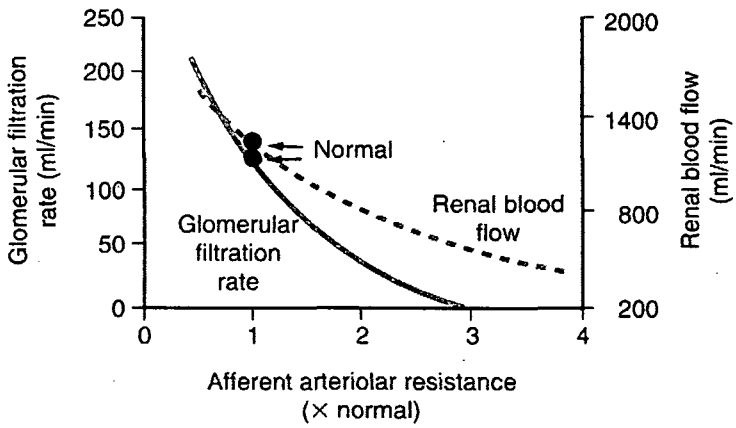


Figure 1.10: The effect of afferent resistance on the GFR

formation of angiotensin II[4]. Angiotensin II mainly constricts the efferent arteriole which is much more sensitive to it than is the afferent arteriole.

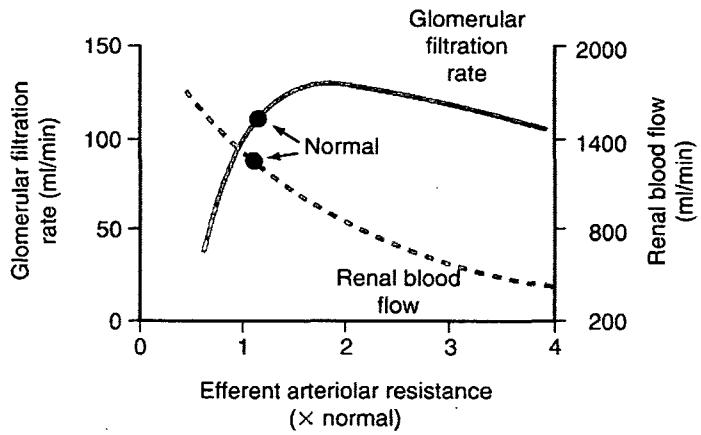


Figure 1.11: The effect of efferent resistance on the GFR

So the efferent arteriolar vasoconstrictor mechanism works as the following:

1. A too low filtration rate causes excess reabsorption of sodium and

[4]see Hormones of the kidney for more information

chloride ions in the ascending limb of the loop of Henle, therefore reducing the ionic concentration at the macula densa.

2. The low concentration of ions then causes the juxtaglomerular cells to release renin from their granules.
3. The renin causes formation of angiotensin II.
4. The angiotensin II constricts the efferent arterioles, which causes the pressure in the glomerulus to rise.
5. The increased pressure then causes the glomerular filtration rate to return back toward normal.

Thus, there is still another negative feedback mechanism that helps to maintain a very constant GFR. It does so by constricting the efferent arterioles at the same time that the afferent vasodilator mechanism dilates the afferent arterioles. When both of these mechanisms function together, the glomerular filtration rate increases only slightly even though the arterial pressure changes between the limits of 75 and $160 mmHg$. [1]

Myogenic Autoregulation of Renal Blood Flow and GFR

A second mechanism that contributes to the maintainance of a relatively constant renal blood flow and GFR is the ability of individual blood vessels to resist stretching during increased arterial pressure, a phenomenon referred to as the **myogenic mechanism**. Studies of individual blood vessels (especially small arterioles) throughout the body have shown that they respond to increased wall tension or wall stretch by contraction of the vascular smooth muscle. This contraction prevents overdistension of the vessel and at the same time, by raising vascular resistance, helps to prevent excessive increase in renal blood flow and GFR when arterial pressure increases.

1.3 Formation of Urine: Tubular Processing

As the glomerular filtrate enters the renal tubules, it flows sequentially through the successive parts of the tubule - the proximal tubule, the loop of Henle, the distal tubule, the collecting tubule and, finally, the collecting duct - before it is excreted as urine. Along this course, some substances are selectively reabsorbed from the tubules back into the blood, whereas others are secreted from the blood into the tubular lumen.

Eventually, the urine that is formed and all the substances in the urine represent the sum of three basic renal processes - glomerular filtration, tubular reabsorption, and tubular secretion - as follows:

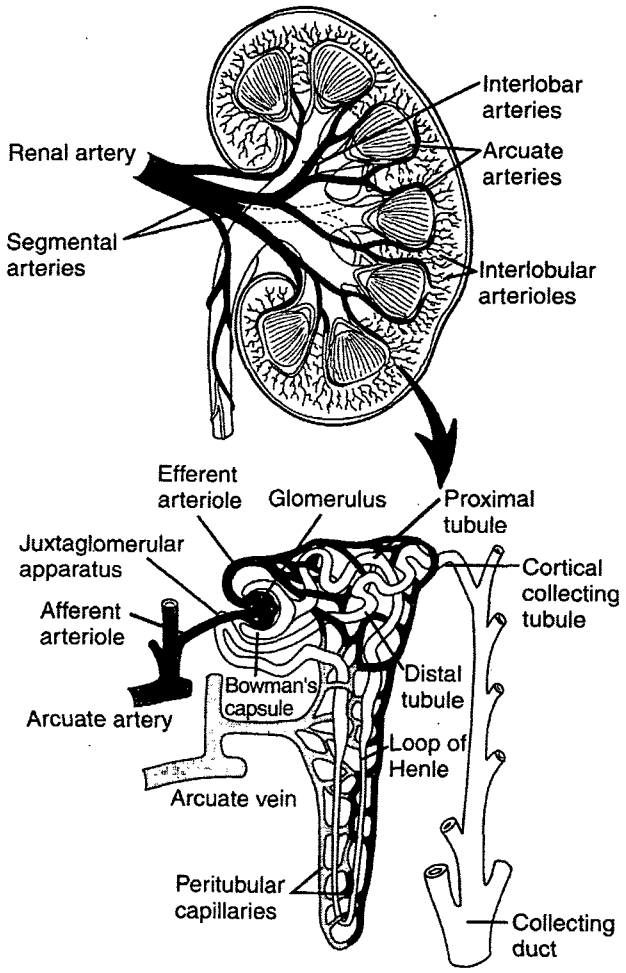


Figure 1.12: Tubular processing of the filtrate

$$\text{Urinary excretion} = \text{Glomerular filtration} - \text{Tubular reabsorption} + \text{Tubular secretion}$$

Unlike glomerular filtration, that is relatively nonselective, tubular reabsorption is highly selective. Some substances, such as glucose and amino acids, are almost completely reabsorbed from the tubules, so that the urinary excretion rate is essentially zero. Many of the ions in the plasma, such as sodium, chloride and bicarbonate, are also highly reabsorbed, but their rates of reabsorption and urinary excretion are variable, depending on the

needs of the body. Certain wast products, such as urea and creatinine, conversely, are poorly reabsorbed from the tubules and excreted in relatively large amounts.

Therefore, by controlling the rate at which they reabsorb different substances, the kidneys regulate the excretion of solutes independently of one another, a capability that is essential for precise control of the composition of body fluids.

1.3.1 Na^+-Reabsorption

The reabsorption of Na^+ and Cl^- plays a major role in body electrolyte and water metabolism. In addition, Na^+-transport is coupled to the movement of H^+, other electrolytes, glucose, amino acids, organic acids, phosphate, and other substances across the tubule walls.

In the proximal tubules, the thick portion of the ascending limb of the loop of Henle, the distal tubules and in the collecting ducts, Na^+ moves by cotransport or exchange from the tubular lumen into the tubular epithlelial cells down its concentration and electrical gradients and is actively pumped from these cells into the interstitial space. Thus, Na^+ is actively transported out of all parts of the renal tubule except the thin portions of the loop of Henle. [41, 40]
Active transport can move a solute against an electrochemical gradient and requires energy derived from metabolism.
On the basolateral sides of the tubular epithelial cell, the cell membrane has an extensive sodium-potassium ATPase system that hydrolyzes ATP and uses the released energy to transport sodium ions out of the cell into the interstitium. This primary active transport is the driving force for the secondary active transport of Na^+ and other ions, that is different in all tubule parts.
In secondary active transport, two or more substances interact with a specific membrane protein (a carrier molecule) and are transported together across the membrane. As one of the substances (for instance, sodium) diffuses down its electrochemical gradient, the energy released is used to drive another substance (for instance, glucose) against its electrochemical gradient. Thus, secondary active transport does not require energy directly from ATP or from other high-energy phosphate sources. Rather, the direct source of the energy is that liberated by the simultanous facilitated diffusion of another transported substance down its own electrochemical gradient.

About 65% of the filtered electrolytes are reabsorbed in the proximal tubule. However, the tubular membranes are highly permeable to water, so that whenever solutes are reabsorbed, water also diffuses through the tubular

membrane by osmosis. Therefore, the osmolarity of the fluid remains about the same as the glomerular filtrate, $300mOsml/l$.
Another 25% of the sodium is reabsorbed while passing through the loop of Henle. Also, sodium is reabsorbed in the distal convoluted tubule and in the collecting duct, the latter being the place where the final Na^+-concentration is determined.

Reabsorption in the Proximal Tubule

The high capacity of the proximal tubule for reabsorption results from its special cellular characteristics: the proximal tubule epithelial cells are highly metabolic and have large numbers of mitochondria to support potent active transport processes.
The extensive membrane surface of the epithelial brush border is also loaded with protein carrier molecules that transport a large fraction of the sodium ions across the luminal membrane linked by way of the co-transport mechanism with multiple organic nutrients such as amino acids and glucose. The remainder of the sodium is transported from the tubular lumen into the cell by counter-transport mechanisms, which reabsorb sodium while secreting other substances into the tubular lumen, especially hydrogen ions. As will be discussed later on, the secretion of hydrogen ions into the tubular lumen is an important step in the removal of bicarbonate ions from the tubule (by combining H^+ with the HCO_3^- to form H_2CO_3, which then dissociates into H_2O and CO_2).
Although, the sodium-potassium ATPase pump provides the major force for reabsorption of sodium, chloride, and water throughout the proximal tubule, there are some differences in the mechanisms by which sodium and chloride are transported through the luminal side of the early and late portions of the proximal tubular membrane.
In the first half of the proximal tubule, sodium is reabsorbed by co-transport along with glucose, amino acids, and other solutes. But in the second half of the proximal tubule, little glucose and amino acids remain to be reabsorbed. Instead, sodium is now reabsorbed mainly with chloride ions. The second half of the proximal tubule has a relatively high concentration of chloride (around $140mEq/l$) compared with the early proximal tubule (about $105mEq/l$) because when sodium is reabsorbed, it preferentially carries with it glucose, bicarbonate and organic ions in the early proximal tubule, leaving behind a solution that has a higher concentration of chloride. In the second half, the higher chloride concentration favors the diffusion of this ion from the tubular lumen through the intercellular junctions into the renal interstitial fluid.
The reabsorption of all these electrolytes causes the diffusion of water and along with it the reabsorption of some other dissolved solutes (Solvent drag).

When sodium is reabsorbed through the tubular epithelial cell, negative ions such as chlorid are transported along with Na^+ because of electrical potentials. That is, transport of positively charged sodium ions out of the lumen leaves the inside of the lumen negatively charged, compared to the interstitial fluid. This causes chloride ions to diffuse *passively* through the paracellular pathway. Additional reabsorption of chloride ions occurs because of a chloride concentration gradient that develops when water is reabsorbed from the tubule by osmosis, thereby concentrating the chloride ions in the tubular lumen. Thus, the active reabsorption of sodium is closely coupled to the passive reabsorption of chloride by way of an electrical potential and a chloride concentration gradient.

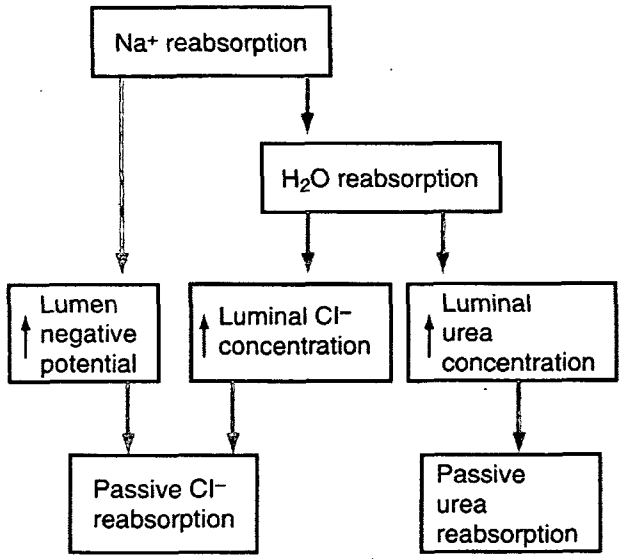


Figure 1.13: Na^+ reabsorption and its effects

The diffusion of Cl^- causes the cellmembrane to depolarise and a lumen-positive transepithelial potential is developed that causes the paracellular reabsorption of cations such as Na^+, K^+, Mg^+ and Ca^{2+}. [40]

Reabsorption in the Loop of Henle

The loop of Henle consists of three functionally distinct segments: the thin descending segment, the thin ascending segment and the thick ascending segment. The thin descending and the thin ascending segments, as their names imply, have thin epithelial membranes with no brush borders, few

mitochondria and minimal levels of metabolic activity.
The descending part of the thin segment is highly permeable to water and moderately permeable to most solutes, including sodium. The function of this nephron segment is mainly to allow simple diffusion of substances through its walls. About 20% of the filtered water is reabsorbed in the loop of Henle, and almost all of this occurs in the thin descending limb because the ascending limb, including both the thin and the thick portions, is virtually impermeable to water, a characteristic that is important for concentrating the urine.[5]
The thick segment, which begins about halfway up the ascending limb, has thick epithelial cells that have high metabolic activity and are capable of active reabsorption of sodium, chloride and potassium. About 25% of the filtered loads of sodium, chloride and potassium are reabsorbed in the loop of Henle, mostly in the thick ascending part. Considerable amounts of other ions, such as calcium, bicarbonate and magnesium are also reabsorbed in the thick ascending limb. The thin segment of the ascending limb has a much lower reabsorptive capacity than the thick segment, and the thin descending limb does not reabsorb significant amounts of any of these solutes.
An important component of solute reabsorption in the thick ascending limb is the sodium-potassium ATPase pump in the epithelial cell basolateral membranes. As in the proximal tubule,the reabsorption of other solutes is closely linked to the reabsorptive capability of the Na^+ K^+ - ATPase, which maintains a low intracellular sodium concentration. The low intracellular sodium concentration in turn provides a favorable gradient for movement of sodium from the tubular fluid into the cell. In the thick ascending loop, movement of sodium across the luminal membrane is mediated primarily by a $Na^+ - K^+ - 2Cl^-$ co-transporter. This co-transport protein carrier in the luminal membrane uses the potential energy released by downhill diffusion of sodium into the cell to drive the reabsorption of potassium into the cell against a concentration gradient.
There is also significant paracellular reabsorption of cations, such as Mg^{++}, Ca^{++}, Na^+ and K^+ in the thick ascending limb owing to the slight positive charge of the tubular lumen relative to the interstitial fluid. Although the 1-sodium, 1-potassium, 2-chloride co-transporter moves equal amounts of cations and anions into the cell, there is a slight backleak of potassium ions into the lumen, creating a positive charge in the tubular lumen. This positive charge forces cations such as Mg^{++} and Ca^{++} to diffuse from the tubular lumen through the paracellular space and into the interstitial fluid.
The thick ascending limb also has a sodium-hydrogen counter-transport mechanism in its luminal cell membrane that mediates sodium reabsorption and hydrogen secretion in this segment.
The thick segment of the ascending loop of Henle is virtually impermeable

[5]see: the countercurrentmechanism of the loop of Henle

to water. Therefore, most of the water delivered to this segment remains in the tubule, despite reabsorption of large amounts of solute. The tubular fluid in the ascending limb becomes very dilute as it flows towards the distal tubule, a feature that is important in allowing the kidney to dilute or concentrate the urine under different conditions. [2, 1, 40]

Reabsorption in the distal tubule

The thick segment of the ascending limb of the loop of Henle empties into the distal tubule. The very first portion of the distal tubule forms part of the juxtaglomerular complex that provides feedback control of the GFR and blood flow in this same nephron. The next early part of the distal tubule is highly convoluted and has many of the same reabsorptive characteristics of the thick segment of the ascending limb of the loop of Henle. It is relatively impermeable to water, and continued removal of the solute in excess of solvent further dilutes the tubular fluid. About 5% of the filtered water is removed in this segment.

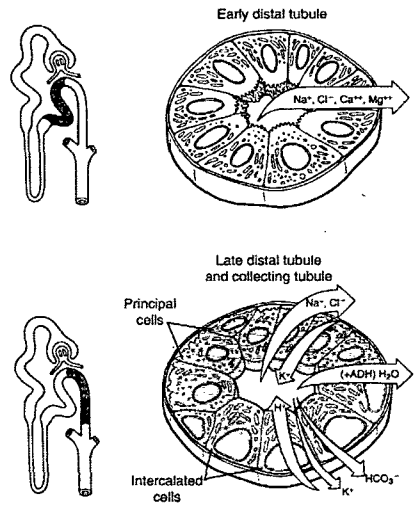


Figure 1.14: Different kind of cells in the distal tubulus

The second half of the distal tubule and the subsequent cortical collecting tubule have similar functional characteristics. Anatomically, they are composed of two distinct cell types, the *principal cells* and the *intercalated cells.* The principal cells reabsorb sodium and water from the lumen and secrete potassium ions into the lumen. The intercalated cells reabsorb potassium ions and secrete hydrogen ions into the tubular lumen. [1]

Reabsorption in the collecting duct

The collecting ducts have two portions: a cortical portion and a medullary portion through which the filtrate flows from the cortex to the renal pelvis. The changes in osmolarity and volume in the collecting ducts depend on the amount of vasopressin (ADH) acting on the ducts. This antidiuretic hormone increases the permeability of the collecting ducts to water by causing the rapid intersection of aquaporin-2 water channels into the luminal membrane of principal cells. In the presence of enough vasopressin to produce maximal antidiuresis, water moves out of the hypotonic fluid entering the cortical collecting duxts into the interstitium of the cortex, and the tubular fluid becomes isotonic. In this fashion, as much as 10% of the filtered water is removed. The isotonic fluid then enters the medullary collecting ducts. An additional 4.7% or more of the filtrate is reabsorbed into the hypertonic interstitium of the medulla, producing a concentrated urine. In humans, the osmolarity of urine reaches $1400\,mOsm/L$, almost 5 times the osmolarity of plasma, with a total of 99.7% of the filtered water being reabsorbed and only 0.3% appearing in the urine. In other species, the ability of concentrating urine is even greater. Maximal urine osmolarity is about $2500\,mOsm/L$ in dogs, about $3200\,mOsm/L$ in laboratory rats, and as high as $5000\,mOsm/L$ in certain desert rodents. [41]

When vasopressin is absent, the collecting duct epithelium is relatively impermeable to water. The fluid therefore remains hypotonic, and large amounts flow into the renal pelvis. In humans, the urine osmolality may be as low as $30mosm/L$. The impermeability of the distal portions of the nephron is not absolute; along with the salt that is pumped out of the collecting duct fluid, about 2% of the filtered water is reabsorbed in the absence of ADH. However, as much as 13% of the filtered water may be excreted, and urine flow may reach $15mL/min$ or more.[6]

Na^+ leaves the lumen through sodium-channels that are activiated by Aldosterone and ADH.[7]

[6] see fig.1.15 for more detail

[7] see Hormones of the kidney

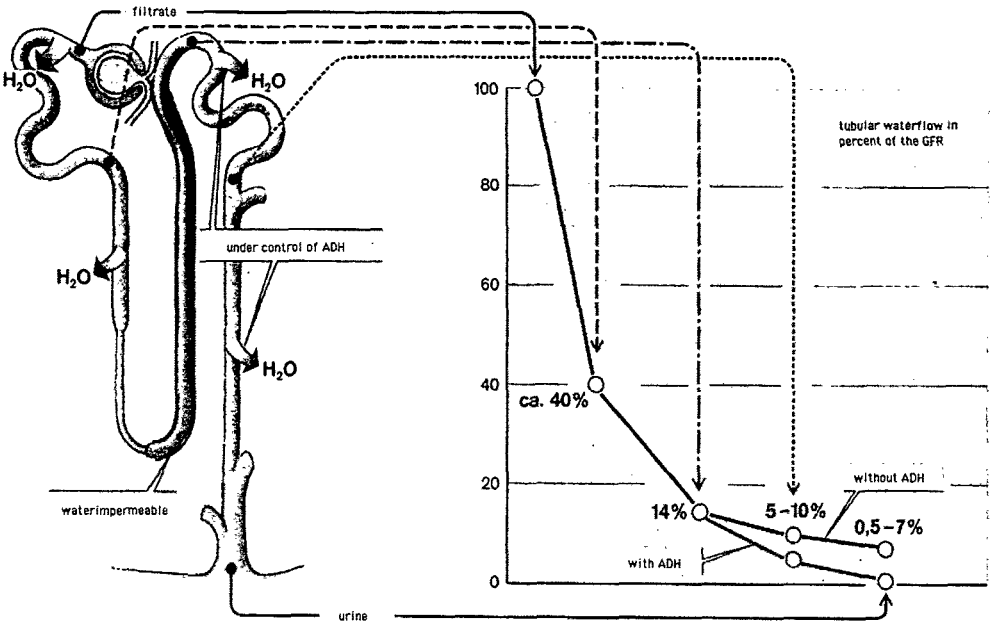


Figure 1.15: The effect of ADH on the reabsorption of water

Hormone	Site of Action and Effects
Aldosterone	Collecting tubule $\uparrow NaCl, H_2O$ reabsorption, $\uparrow K^+$ secretion
Angiotensin II	Proximal tubule, thick ascending loop of Henle, distal tubule $\uparrow NaCl, H_2O$ reabsorption, $\uparrow H^+$ secretion
Antidiuretic hormone	Distal tubule, collecting tubule and duct $\uparrow H_2O$ reabsorption
Atrial natriuretic peptide	Distal tubule, collecting tubule and duct $\downarrow NaCl$ reabsorption
Parathyroid hormone	Proximal tubule, thick ascending loop of Henle, distal tubule $\downarrow PO_4^{---}$ reabsorption, $\uparrow Ca^{++}$ reabsorption

1.3.2 The Countercurrent Mechanism

The osmolarity of interstitial fluid in almost all parts of the body is about $300\,mOsm/L$, which is similar to the plasma osmolarity. The osmolarity of the interstitial fluid in the medulla of the kidney is much higher, increasing progressively to about 1200 to $1400\,mOsml/L$ in the pelvic tip of the medulla. This means that the renal medullary interstitium has accumulated solute in great excess of water. Once the high solute concentration in the medulla is achieved, it is maintained by a balanced inflow and outflow of solutes and water in the medulla.
The major factors that contribute to the buildup of solute concentration into the renal medulla are as follows:

1. Active transport of sodium ions and co-transport of potassium, chloride and other ions out of the thick ascending limb of the loop of Henle into the medullary interstitium

2. Passive diffusion of large amounts of urea from the innermedullary collecting ducts into the medullary interstitium

3. Diffusion of only small amounts of water from the medullary tubules into the medullary interstitium, far less than the reabsorption of solutes into the medullary interstitium

The process is best understood in terms of hypothetical steps leading to the normal equilibrium condition, although of course the steps do not occur in vivo, and equilibrium is maintained unless the osmotic gradient is washed out.

Assume first a condition in which osmolarity is $300\,mOsm/L$ throughout the descending and ascending limb and the medullary interstitium (1)[8]. Assume in addition that the pumps in the thick ascending limb can pump $100\,mOsm/kg$ of Na^+ and Cl^- from the tubular fluid to the interstitium, increasing interstitial osmolarity to $400mOsm/L$ (2). Water then moves out of the thin descending limb, and its contents equilibrate with the interstitium (3). However, fluid containing $300\,mOsm/L$ is continuously entering this limb from the proximal tubule (4), so the gradient against which the Na^+ and Cl^- are pumped is reduced and more enters the interstitium. Meanwhile, hypotonic fluid flows into the distal tubule, and isotonic and subsequently hypertonic fluid flows into the ascending thick limb. The process keeps repeating, and the final result is a gradient of osmolarity from the top to the bottom of the loop (7).
In juxtaglomerular nephrons with longer loops and thin ascending limbs, the osmotic gradient is spread over a greater distance and the osmolarity at the tip of the loop is greater. This is because the thin ascending limb

[8] see pict. 1.16

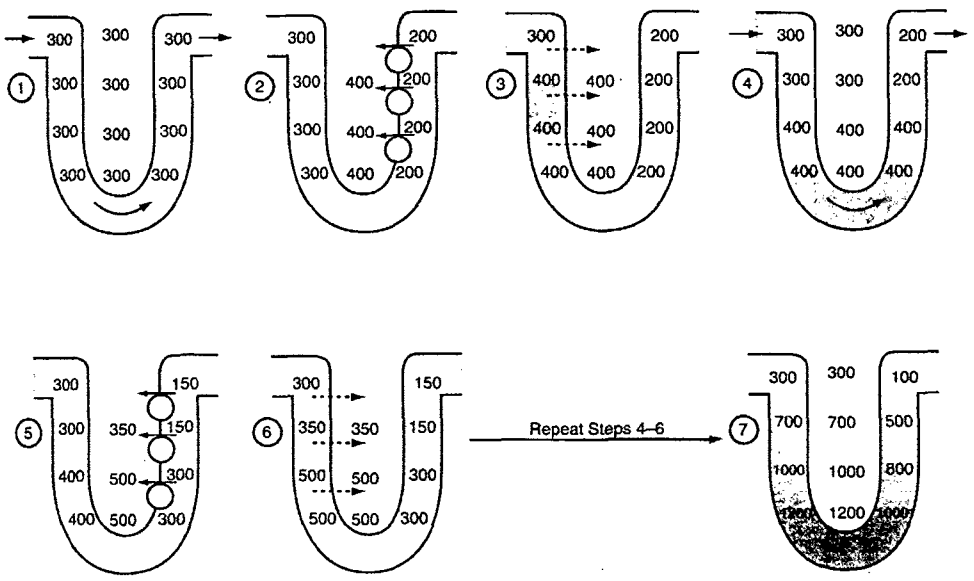


Figure 1.16: The renal countercurrent mechanism

is relatively impermeable to water but permeable to Na^+ and Cl^-. Therefore, sodium and chloride move down their concentration gradients into the interstitium, and there is additional passive countercurrent multiplication. The greater the length of the loop of Henle, the greater the osmolarity that can be reached at the tip of the pyramid.

The osmotic gradient in the medullary pyramids would not last long if the sodium and urea in the interstitial spaces were removed by the circulation. These solutes remain in the pyramids primarily because the vasa recta operate as countercurrent exchangers. The solutes diffuse out of the descending vessels and into the fenestrated ascending vessels. Therefore, the solutes tend to recirculate in the medulla and water tends to bypass it, so that hypertonicity is maintained. The water removed from the collecting ducts in the pyramids is also removed by the vasa recta and enters the general circulation.
Countercurrent exchange is a passive process; it depends upon the movement of water and could not maintain the osmotic gradient along the pyramids if the process of countercurrent multiplication in the loops of Henle were to cease. [41,2,1]

1.3.3 The Role of Urea

Urea contributes to the establishment of the osmotic gradient in the medullary pyramids and to the ability to form a concentrated urine in the collecting ducts. Urea moves passively out of the proximal tubule, but except for the inner portion of the collecting duct, the rest of the tubular epithelium is relatively impermeable to this compound. Consequently, urea is increasingly concentrated in the tubular fluid as water is removed in the loop and distal tubule.

However, when the inner medullary portion of the collecting duct is reached, urea moves into the interstitium of the pyramids, adding to the hyperosmolarity. The movement of urea in this portion of the collecting duct is facilitated by vasopressin. Conversely, when the tubular fluid reaching the inner medullary collecting duct is dilute in the absence of vasopressin, urea moves from the interstitium into the tubular lumen and the osmotic gradient in the pyramides is reduced. The amount of urea in the medullary interstitium and, consequently, in the urine varies with the dietary intake of protein. Therefore, a high-protein diet increases the ability of the kidneys to concentrate the urine. [1]

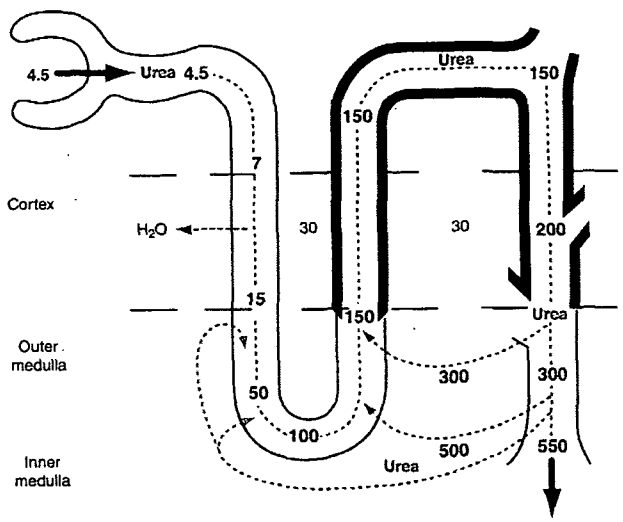


Figure 1.17: The urea cycle

1.3.4 Reabsorption of other Solutes than Sodium

1.4 Extracellular Fluid Osmolarity

For the cells to function properly, they must be bathed in extracellular fluid (ECF) with a relatively constant concentration of electrolytes and other solutes. The total concentration of solutes in the extracellular fluid - and therefore the osmolarity - is determined by the amount of solute divided by the volume of the extracellular fluid. Thus, to a large extent, extracellular fluid sodium concentration and osmolarity are regulated by the amount of extracellular water.
The body water in turn is controlled by the water intake and water excretion.

Daily Intake of Water

Water is added to the body by two major sources: it is ingested in the form of liquids or water in the food, which together normally add about $2100\,ml/day$ to the body fluids, and it is synthesized in the body as a result of oxidation of carbohydrates, adding about $200\,ml/day$. Intake of water is highly variable among different people and even within the same person on different days, depending on climate, habits and level of physical activity. But one can say that the daily water intake for a healthy adult should be approximately $1L/30\,kg$, and for babies it should be $1L/10\,kg$ bodyweight.

Daily Loss of Body Water

Some of the water losses cannot be precisely regulated. For example, there is a continuous loss of water by evaporation from the respiratory tract and diffusion through the skin, which together account for about $700\,ml/day$ of water loss under normal conditions. This is termed *insensible water loss* because we are not consciously aware of it, even though it occurs continually in all living humans.
The amount of water lost by sweating is highly variable, depending on physical activity and environmental temperature. The volume of sweat normally is about $100\,ml/day$, but in very hot weather or during heavy excercise, water loss in sweat occasionally increases to 1 to $2\,L/hour$.
Only a small amount of water ($100\,ml/day$) normally is lost in the feces, although this can increase to several liters a day in people with severe diarrhea.
The remaining water loss from the body occurs in the urine excreted by the kidneys.

Daily Intake and Output of Water		(in ml/day)
	Normal	Prolonged Heavy Exercise
Intake		
Fluids ingested	2100	?
From metabolism	200	200
Total intake	2300	?
Output		
Insensible - Skin	350	350
Insensible - Lungs	350	650
Sweat	100	5000
Feces	100	100
Urine	1400	500
Total output	2300	6600

Water makes about 46 % to 75 % of the bodyweight of an adult depending on the age and sex of a person (women have a lower percentage due to their higher bodyfat). [2, 40, 1]

About 3/5 of the bodywater is found in the intracellular space and 2/5 of it forms the extracellular fluid, the extracellular fluid consisting of the interstitium, plasmawater and transcellular fluids. With the main portion of the body Na^+ in the extracellular fluid and sodium being the most abundant ion in that compartment, the regulation of the extracellular fluid osmolarity is closely linked to the regulation of the sodium concentration.
A rise of the ECF osmolarity due to sodium intake or water loss causes water to flow from the intracellular space into the extracellular fluid, because osmotic equilibrium between the ECF and the ICF has to be maintained. To save the cells such unwanted volume changes the osmolarity of the extracellular fluid has to be highly controlled. Osmoreceptors (mainly in the hypothalamus), Adiuretin (ADH) as a hormone and the kidneys as its target play a major role in this.

Water Deficit

If H_2O loss is not - or not sufficiently - substituted, the ECF gets hyperton. Even a rise in osmolarity as low as 1 % is enough to cause or increase the excretion of ADH in the back of the hypophyse. ADH decreases the renal water excretion, water is held back and therefore the body water concentration is increased.
Simularily, many of the factors that stimulate ADH secretion also increase thirst, which is defined as the conscious desire for water.

Control of thirst	
Increase Thirst	Decrease Thirst
↑ Osmolarity	↓ Osmolarity
↓ Blood volume	↑ Blood volume
↓ Blood pressure	↑ Blood pressure
↑ Angiotensin II	↓ Angiotensin II
Dryness of mouth	Gastric distention

Water Overflow

The intake of hypotonic fluid decreases the extracellular space osmolarity. This, on the other hand, decreases the excretion of ADH causing a water diuresis, which brings the plasmaosmolarity back to normal.

The daily $NaCl$ intake is about 8 to 15 g/d. The very same amount per time should be excreted by the kidneys to make sure that the Na^+ in the body and related to that, the extracellular fluid volume, is held constant. Several mechanisms are involved in this:

- Atriopeptin (ANP): is excreted by the cells in the atrium when the extracellular space volume is increased. ANP increases the renal Na^+ excretion by rising the filtration fraction and decreasing the $NaCl$ reabsorption in the collecting duct.
- The renin-angiotensin-system (resulting in Na^+ retention)
- ADH: secretion in the hypothalamus is stimulated by an increase in osmolarity measured in the atrium through a decreased (more than 10%) ECF volume and by angiotensin II.
- Pressure diuresis: an overage amount of water and sodium is excreted. It is caused by an increased blood pressure due to an increase in the extracellular space volume.

Lack of Sodium

A too low blood osmolarity decreases the secretion of ADH, therefore increasing the H_2O secretion. This causes a reduction of the ECF volume and along with that the plasma volume and the blood pressure are also decreased. This activates the renin-angiotensin-system.
Because of the sodium retention water is held back, simularily water is consumed (angiotensin II causes thirst), so the volume of the extracellular space is normalised again.

Increased Salt Intake

With no significant change in the body water amount the plasma osmolarity rises (causing thirst) and more ADH is excreted. Because of this, the ECF volume increases and this brings the renin-angiotensin-system to slow down or end. Also, ANP increases the $NaCl$ and H_2O excretion so that the ECF volume is brought back to normal. [40]

1.5 The Kidney Hormones

1.5.1 Aldosterone

Aldosterone is *the* major mineralcorticoid secreted by the adrenals. As already mentioned before, aldosterone increases the absorption of sodium and the secretion of potassium. Therefore, aldosterone causes sodium to be conserved in the extracellular fluid while increasing potassium excretion in the urine.
The regulation of aldosterone secretion (by the zone glomerulosa cells) is so deeply intertwined with the regulation of extracellular fluid electrolyte concentrations, extracellular fluid volume, blood volume, arterial pressure, and many special aspects of renal functions that it is difficult to discuss its regulation independently of all these other factors.
The following factors are known to play essential roles in the regulation of aldosterone. In the probable order of their importance, they are as follows:

1. Increased potassium ion concentration in the extracellular fluid greatly *increases* aldosterone secretion.
2. Increased activity of the renin-angiotensin system also greatly *increases* aldosterone secretion.
3. Increased sodium ion concentration in the extracellular fluid *very slightly decreases* aldosterone secretion.

Aldosterone is an important regulator of sodium reabsorption and potassium secretion by the renal tubules. The primary site of aldosterone action is on the principal cells of the cortical collecting tubule. The mechanism by which aldosterone increases sodium reabsorption while at the same time increasing potassium secretion is by stimulating the sodium-potassium ATPase pump on the basolateral side of the cortical collecting tubule membrane. Aldosterone also increases the sodium permeability of the luminal side of the membrane.
In the absence of aldosterone, as occurs with adrenal destruction or malfunction, there is marked loss of sodium from the body and accumulation of potassium. Conversely, excess aldosterone secretion, as occurs in patients with adrenal tumors is associated with sodium retention and potassium depletion.

1.5.2 The Renin-Angiotensin System

Renin itself is not a hormone but a small protein enzyme released by the kidneys when the arterial pressure falls too low. In turn, it raises the arterial pressure in several ways, thus helping to correct the initial fall in pressure.
Renin is synthesized and stored in an inactive form called *prorenin* in the juxtaglomerular cells of the kidneys. When the arterial pressure falls, many of the prorenin molecules in the juxtaglomerular cells are caused to split and release renin.
Renin itself is an enzyme, not a vasoactive substance. Instead, it acts enzymatically on another plasma protein, a globulin called renin substrate (or *angiotensinogen*), to release a 10-amino acid peptide, *angiotensin I*. Angiotensin I has mild vasoconstrictor properties, but not enough to cause significant changes in circulatory function. The renin persists in the blood for 30 to 60 minutes and therefore, continues to cause formation of still more angiotensin I during this entire time.
Within a few seconds after formation of the angiotensin I, two additional amino acids are split from the angiotensin I to form the 8-amino acid peptide *angiotensin II* . This conversion occurs almost entirely in the lungs during the few seconds while the blood flows through the small vessels of the lungs, catalyzed by an enzyme called *converting enzyme* that is present in the endothelium of the lung vessels.
Angiotensin II is an extremely powerful vasoconstrictor, but it persists in the blood only for 1 or 2 minutes because it is rapidely inactivated by enzymes called angiotensinase.

During its persistence in the blood, angiotensin II has two principal effects that can elevate arterial pressure:

- *Vasoconstriction* occurs rapidely, intensively in the arterioles and much less so in the veins. Constriction of the arterioles increases the total peripheral resistance, thereby raising the arterial pressure. Also, the mild constriction of the veins promotes increased venous return of blood to the heart, thereby helping the heart pump against the increasing pressure.

- The second principal means by which angiotensin increases the arterial pressure is to act directly on the kidneys themselves to *decrease the excretion of both, salt and water.* This slowly increases the arterial pressure over a period of hours and days. This long-term effect is even more powerful than the acute vasoconstrictor mechansim in eventually returning the arterial pressure all the way back to normal after a low blood pressure event.
  Angiotensin II causes the kidneys to retain salt and water in three ways:

1. Angiotensin II stimulates adlosterone secretion which in turn increases sodium reabsorption
2. Angiotensin II constricts the efferent arterioles. This reduces peritubular capillary hydrostatic pressure, which increases net tubular reabsorption, especially from the proximal tubules. Also, by reducing renal blood flow, efferent arteriolar constriction raises the filtration fraction in the glomerulus and increases the concentration of proteins and the colloid osmotic pressure in the peritubular capillaries; this increases the reabsorptive force at the peritubular capillaries and raises tubular reabsorption of sodium and water.
3. Angiotensin II directly stimulates sodium reabsorption in the proximal tubules, the loops of Henle, and the distal tubules. One of the direct effects of angiotensin II is to stimulate the sodium-potassium ATPase pump on the tubular epithelial cell basolateral membrane. A second effect is to stimulate sodium-hydrogen exchange in the luminal membrane, especially in the proximal tublule.

For all this reasons, angiotensin II is perhaps the body's most powerful sodium-retaining hormone.

1.5.3 Antidiuretic Hormone

Adiuretin or vasopressin or antidiuretic hormon or short ADH is a polypeptide, synthesized and released in the hypothalamus and the pituitary gland from where it is transported to the hypophyse, where it is stored in the secretory granules. Increased osmolarity, decreased arterial pressure, decreased blood volume or other stimuli cause ADH to be released, which then enters the systemic circulation.
Secretion of ADH in response to an osmotic stimulus is rapid, so that plasma ADH levels can increase severalfold within minutes, thereby providing a rapid means for altering renal excretion of water.
Extremely minute quantities of ADH can cause decreased excretion of water by the kidneys (antidiuresis). Briefly, in the absence of ADH, the collecting tubules and ducts become almost impermeable to water, which prevents significant reabsorption of water and therefore allows extreme dilution of the urine. Conversely, in the presence of ADH, the permeability of the collecting ducts and tubules to water increases greatly and allows most of the water to be reabsorbed as the tubular fluid passes through these ducts, thereby conserving water in the body and producing very concentrated urine. [1, 41]
Aside from the effect of minute concentrations of ADH in causing increased water conservation by the kidneys, higher concentrations of adiuretin have

potent effect of constricting the arterioles throughout the body and therefore of increasing the arterial pressure. For this reason, ADH has another name, *vasopressin.*

Chapter 2

The LADMER - System

In the relationship between dose and effectiveness or dose response, not only the amount of drug administered and the pharmacological effect of the drug are of importance but many other factors are responsible for the entrance of a drug into the body. These factors are based on the physical and chemical properties of the drug substance and of the drug product. What happens to the active ingredient in the body after administration of a drug product in its various dosage forms? This entire cycle of processes is termed **fate of drugs**.
Whether a blood level curve will reach its peak rapidly or slowly depends on the route of administration, the dosage form, the liberation rate of the drug from the dosage form, diffusion, penetration and permeation of the drug, its distribution within the body fluids and tissues, the type, amount and rate of biotransformation, recycling processes and elimination. In addition to these factors there are also others, depending on the individual disposition, diseases, etc.
The fate of drugs is described by the so-called LADMER-system (often also termed LADME-system) showing that **l**iberation, **a**bsorption, **d**istribution, **m**etabolism and **e**limination are involved to elicit the **r**esponse. [4, 8]

2.1 Liberation

Liberation (drug release) is the first step which determines onset of action, rate of absorption, availability, etc., which is true for all drug products by all routes of administration, except intravenous and the preoral use of true solutions.
After the drug has been administered it disintegrates into small particles (Disintegration) and releases the active agent, which now dissolves in the intestinal fluids (Dissolution). The dissolution of the drug is the necessary condition for the consecutive reabsorption - for a drug can only be absorbed if it is present in the form of solution. Speed and completeness of dissolution

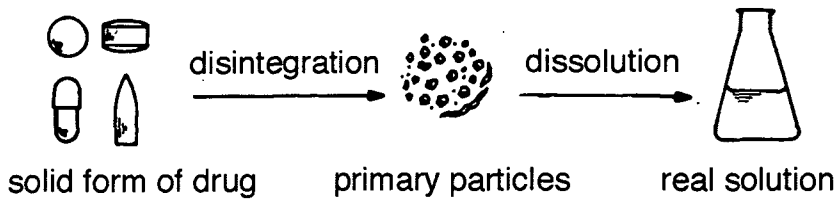


Figure 2.1: Liberation of drugs

and reabsorption together determine the drug's bioavailability. [4, 26]

2.2 Absorption

After dissolution, the drug diffuses to the site of absorption. Only drugs administered intravenously in solution enter the circulatory system immediately. With all other routes of administration, the drugs must pass membranes which act as lipid barriers. Different transport mechanisms are employed to penetrate into and to permeate through these membranes.
Most of the drugs are absorbed or transported by passive diffusion, which depends on the pK_a values of the drug, the pH of the solution and the lipid solubility of the unionized form. Drugs passing through the lipid barrier may directly enter the central compartment.

As the marker used for the determination of the clearence (Sinistrin) is administered intravenously absorption as well as liberation is not of great interest to us. [8, 35]

2.3 Distribution

After their reabsorption drugs are quickly distributed throughout the circulatory system. From the blood they pass over into adjacent tissues.
However, most drugs are at least partially bound to protein in the bloodstream and it is only the free, unbound form of the drug that is available for action. A protein-bound drug cannot migrate from blood into tissue and can therefore not reach its destination, it remains ineffective. On the other hand, plasma protein-bound drugs cannot be excreted by the kidneys either as they can't be filtered in the glomerulum. The protein-bound drug therefore longer remains in the body.

But the plasma protein-bound fraction is not permanently trapped but is

in equilibrium and will be released from the protein as the free drug is eliminated from the plasma. The drug then may enter the peripheral compartment by again passing a lipid barrier until it finally reaches the biophase. This is a cell, or even a cell component, where the final interaction between drug and receptor takes place.
So, plasma proteins may exert a buffer and transport function in the distribution process. As only the free, non-protein-bound fraction of a drug can leave the circulatory system and diffuse into tissue, the equilibrium between free and unbound drug acts as a buffer system, since a relatively constant concentration of free drug can be maintained over a relatively long period of time due to the dissociation of the drug-protein complex.

Drug binding to protein is usually rather nonspecific, i.e., many drugs bind to the same binding sites (receptors) on the protein molecule. The drug with the higher affinity will, therefore, displace a drug of lower affinity from its binding site by competition. This can lead to some dangerous situations, since only the non-protein-bound fraction is free for pharmacological action; intensity of pharmacological response, side effects, and toxicity increase upon displacement from protein binding. This is of importance only in drugs highly bound to protein. Imagine, for example, a drug already bound to a large amount ($\approx 99\,\%$) being displaced by a drug administered later than the first one. This second drug now supplants a small fraction, say $3\,\%$, of the bound drug increasing the free and active part to $4\,\%$, 4 times the value it was before.! Such a raise could already result in severe toxic symptoms.

The extent of protein binding is determined in vitro by dialysis, ultracentrifugation, ultrafiltration, molecular filtration, Sephadex-gel filtration, electrophoresis or by agar plate test.
The extent of protein binding is usually given in percentage. However, one must be aware that percent bound is a function of the capacity of the protein and the concentration of the drug bound in the environment. [4, 8, 27]

2.4 Metabolism

After releasing the drug from its receptor binding, the drug again passes through a lipid barrier and reenters the central compartment, from which the drug, by again passing a lipid barrier, is metabolized in the liver, kidney or in the tissue of the plasma.
Metabolism generally serves three principal purposes: (1) to supply energy for body functions and maintenance, (2) to break down ingested compounds, i.e., catabolism, to simpler structures, and biosynthesis of more complex

molecules, i.e., anabolism, usually requiring energy; and (3) for the conversion or biotransformation of foreign compounds to more polar, water-soluble and ionized structures which can be eliminated more easily.
Drug metabolism refers solely to the chemical biotransformation of a drug by the biological environment. The principle site of drug metabolism is the liver, less important are the kidneys, muscle tissues and gut wall.

Drug metabolism is very complex. Often, it is also termed detoxication indicating that one of the main functions of metabolism is the formation of more polar and water-soluble compounds resulting in reduction of their pharmacological activity and more rapid excretion from the body.
Drug metabolism reactions are divided into Phase 1, or non-synthetic reactions, and Phase 2, or synthetic reactions. From the point of view of pharmacokinetics, the synthetic reactions of metabolism are more interesting as they are responsible for the formation of the final metabolic product of the drug to be excreted.
The most common drug metabolism reactions comprise oxidation, reduction, hydrolysis, and conjugation. Also, a drug can be metabolized simultaneously by competing reactions. The extent of formation of the different metabolites thus depends on the relative rates of reaction. Additionally, many drugs are subjected to different metabolic reactions sequentially, where oxidation, reduction or hydrolysis reactions are followed by conjugation.

2.4.1 First Pass Effect:

After peroral and also after deep rectal administration, drugs may be metabolized in the gastrointestinal epithelium during absorption, or by the liver before they reach systemic circulation. This latter process is called the *first-pass-effect* and reduces the systhemic availability of the drug. This first-pass-effect is also one of the explanations for differences in the elimination half-life of some drugs when administered I.V (intravenous) and P.O. (peroral).
The importance of the first-pass-effect depends on the metabolic capacity for the particular drug, the rate of metabolism and the rate of absorption. If the given amount of drug is small. but capacity and rate of metabolism are high, a large fraction of the drug may be metabolized, therefore reducing the extent of bioavailability. With increasing dose sizes, on the other hand, first-pass metabolism may become saturated, thus increasing the extent of bioavailability. [8]

2.5 Excretion

Exrection of drugs is the final elimination (or loss of drug) from the body. Excretion can happen by various pathways (urine, bile, intestines, salvia,

alveolar air, sweat and milk), but the two most important ways of excretion are via the kidney into the urine and via the liver into feces.
For the markers used in this paper, sinistrin and p-amino-hippuric acid, the kidney remains the pathway of interest. So, let's shortly recall the most important facts about the renal clearance:

2.5.1 Renal Clearance:

The renal clearance for a drug is that volume of blood that is cleared of the drug during one minute via the kidneys. When a compound is filtered only by the glomerulus and is not otherwise acted on by the kidney (like it is the case for inulin, sinistrin and mostly for creatinine), its clearance equals the glomerular filtration rate. Such substances are, therefore, used as test substances for the determination of the glomerular filtration rate (GFR): [2, 4, 8]

$$Cl_{ren.} = \frac{C_u * V}{C} \; [ml/min]$$

$Cl_{ren.}$	corrected renal clearance [ml/min]
C_u	concentration of drug in urine $[mg/ml]$
V	volume of urine excreted $[ml/min]$
C	concentration of drug in plasma (no protein binding present) $[mg/ml]$

In the case a drug is bound to protein the corrected renal clearance is calculated in the following way:

$$Cl_{ren.corr.} = \frac{C_u * V}{(1-p) * C} \; [ml/min]$$

$Cl_{ren.corr.}$	corrected renal clearance $[ml/min]$
p	fraction of drugs bound to protein $[fraction\ of\ 1]$
C_u	concentration of drug in urine $[mg/ml]$
V	volume of urine excreted $[ml/min]$
C	concentration of drug in plasma $[mg/ml]$

Chapter 3

Compartmental Models

Basically, compartmental analysis is used to evaluate a system by measuring the input and output of a tracer introduced into that system.
This method can be used to access systems assumed to consist of one to several compartments and has a large number of applications not only in medicine and biology, but also in pharmacy, social science, behaviourmetrics and ecology.
Models are used to describe and interpret a set of data obtained by experimentation. A model in pharmacokinetics is a hypothetical structure which can be used to characterize with reproducibility the behaviour and the "fate" of a drug in biological systems when given by a certain route of administration.

The system to be evaluated is divided into a number of compartments and the transport of substances from one compartment to another is described as a flux.
Such a compartment is an entity which can be described by a definite volume and a concentration (of drug contained in the volume). Meaning that at a certain time, the concentration of the administered drug is the same throughout that compartment. [16]
Multiplying that concentration with the volume yields the total amount of drug in that compartment:

$$M = C * V$$

M amount of drug in the compartment $[mcg]$
C concentration of the drug $[mcg/ml]$
V volume of the compartment $[ml]$

Speaking in a greatly simplified manner, a compartment can be seen as a vessel (e.g. a bath tub) of unknown volume filled with water. Now, if you

add a substance to that vessel, stir up real well so that this substance is equally distributed throughout that vessel, and then take a probe, you can easily calculate the volume of the tub:

$$V = \frac{M}{C}$$

The same principle works for the human body: one injects some (known) amount of a drug (dose D) into the left arm-vein, waits until the blood has circulated once through the body (about 1 to 3 minutes), takes a blood sample from the right arm-vein which gives the drug concentration (blood level C). As in the beginning all of the drug is still in the circulatory system (the vessel of unknown volume), this value equals the "fictive initial concentration" (C_0). Knowing the dose and this initial concentration the **volume of distribution** can easily be calculated:

$$V_d = \frac{D}{C_0}$$

V_d volume of distribution $[ml]$
D dose administered $[mcg]$
C_0 fictive initial concentration $[mcd/ml]$

We now have assumed that the drug is evenly distributed in the body like it was in the tub, but the human body is not a simple vessel and there is no substance that is totally homogeneous in the entire organism. Actually, the human body is a multi-million compartment model considering drug concentration in different organelles, cells or tissues.

Usually, the behaviour of a drug in a biological system can be described by a one-compartment model or a two-compartment model. Sometimes it is necessary to employ multicompartmental models. In order to keep the model as simple as possible one should begin by determining whether experimental data can be fitted to a one-compartment model. And only if no fitting is obtained, one continues trying more sophisticated models.

However, in the human body we have access to only two types of body fluid - blood (or plasma or serum) and urine. Compartment models in pharmacokinetics are, therefore, used to fit experimental data from blood level versus time curves or urinary cumulative excretion versus time curves to models. A certain type of model is not necessarily specific for a particular drug. Often a blood level versus time curve upon extravascular administration can be fitted to a simple one-compartment model, whereas the blood level versus time curve upon intravascular administration is best fitted to a

two-compartment model. [4]
Two or more compartments can be linked together because a drug may move from one compartment into another and back. The movement occurs at different rates (speeds) and is described by distribution rate constants.

3.1 Open One-Compartment Model

We are talking about an open one-compartment model if the drug entering the body (input) distributes (equilibrates) instantly between the blood and other body fluids or tissues. In an open one-compartment model the drug is not necessarily (and indeed is rarely) confined to the circulatory system. The drug may occupy the entire extracellular fluid, "soft" tissue or the entire body. However, distribution occurs instantly and is not "pooled" in a specific area.

3.2 Open Two-Compartment Model

We are talking about an open two-compartment model if the drug entering the body (input) does **not** instantly distribute (equilibrate) between the blood and those other body fluids or tissues which it eventually reaches. The distribution of the drug in blood and other "soft" tissues, on the one hand, and into other "deep" tissues, on the other hand, occurs at different rates (speeds). Eventually a steady state will be reached which terminates the "distribution" phase.
The body fluids or tissues which are in equilibrium with the circulatory system comprise the **central compartment** which is accessible through blood sampling. Those body fluids or tissues into which the drug distributes slowly comprise the **peripheral compartment** which is not accessible by blood sampling.

The term "open" in conjunction with a compartment model refers to the fact that we do not have a closed system, but have a unidirectional input and output into and out of the system. [8]

3.3 Mathematical Representation

The development of the mathematical realization of strictly compartmental models involves, essentially, two stages. The first consists of writing mass balance equations for each compartment in terms of material flowing into that compartment and the flux of material from that compartment to other sites. This flux of material from one compartment to another can be assumed to depend, linearly or nonlinearly, on the mass or concentration of material in the source compartment only. [7, 42]

The general form of equation defining the dynamics of the i-th compartment in such a model is

$$\dot{Q}_i(t) = R_{i0} + \sum_{j=1,j\neq i}^{n} R_{ij}(Q_j) - \sum_{j=1,j\neq i}^{n} R_{ji}(Q_i) - R_{0i}(Q_i), \quad i = 1, 2, \ldots, n$$

where

Q_i	quantity of material in compartment i
R_{ij}	flux of material into compartment i from compartment j, depending on Q_j only
R_{ji}	flux of material from compartment i to compartment j, depending on Q_i only
R_{i0}	flux of material into compartment i from external environment
R_{0i}	flux of material from compartment i into external environment

It should be noted that all fluxes and masses cannot be negative.
If the compartment is accessible to measurement, a desired output is represented as a dashed line with a bullet. Measurements are typically made in units of concentration, $c_i(t)$. The concentration is the mass within the volume of the compartment, $v_i(t)$, where

$$c_i(t) = \frac{Q_i(t)}{v_i(t)}$$

The second stage requires specifying the functional dependences of each flux. The nature of these dependences, which may be linear or nonlinear, and their values are obtained either from *a priori* knowledge or from parameter estimation. Two commonly occurring types of functional dependencies are the linear dependence and the threshold/ saturation dependence, which includes the Michaelis-Menten form.
The linear and Michaelis-Menten dependences can be described mathematically in the form

Linear Dependence

$$R_{ij}(Q_j) = k_{ij}\, Q_j$$

For a linear time-invariant system, each flux is described as the product of a constant, k_{ij}, and the mass of the source compartment, resulting only from diffusion. So this linear compartment model can be described as

$$\dot{Q}_i(t) = R_{i0} + \sum_{j=1,j\neq i}^{N} k_{ij}\, Q_j(t) + \sum_{j=1,j\neq i}^{N} -k_{ji}\, Q_i(t) - k_{0i} * Q_i$$

where k_{ij} is a constant defining the fractional rate of transfer of material into compartment i from compartment j. Each of this rate constants has the unit of inverse time.

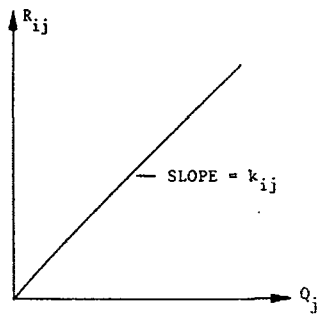


Figure 3.1: Graphical representation of linear dependence of material flux R_{ij} on the quantity of material Q_j in the source compartment j

Michaelis-Menten Dynamics

$$R_{ij}(Q_j) = \frac{\alpha_{ij} Q_j}{\beta_{ij} + Q_j}$$

where α_{ij} is the saturation value of flux R_{ij} and β_{ij} is the value of Q_j at which R_{ij} is equal to half its maximal value.

3.4 Example

The following example shows the distribution of a test substance (sinistrin) in a two-compartment model. After administration of the drug in the central compartment, it rapidly distributes in this compartment while distribution of the drug in other tissues (peripheral compartment) occurs at a measurable speed.

As can be seen in the graphic, the whole process can be divided into two phases: in phase 1, distribution between the compartments takes place, while phase 2 consists of the elimination from the central volume.

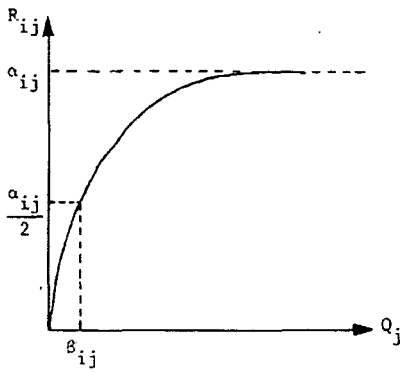


Figure 3.2: Graphical representation of a Michaelis-Menten dependence of material flux R_{ij} on the quantity of material Q_j in the source compartment

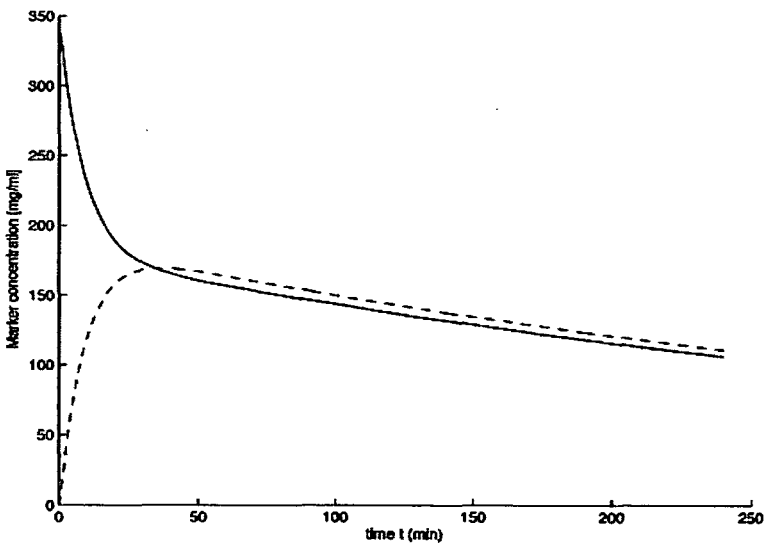


Figure 3.3: Marker concentration in the central and peripheral compartment

Chapter 4

The Markers

4.1 Inulin

Inulin, $(C_6H_{10}O_5)_4$, is an indigestible polysaccharide occurring in the rhizome of certain plants (Compositae). It is a polymer of fructofuranose and yields fructose on hydrolysis.
A long time Inulin was used as *the* standard substance used in the test for determining the renal clearance. But it has the big disadvantage of not being water soluble at room temperature what makes infusion rather difficult. That's why today, other substances (as Sinistrin) are also used as markers for the determination of the renal clearance.

4.2 Sinistrin

Sinistrin is a mucilaginous carbohydrate, resembling achrodextrin, extracted from squill as a colorless amorphous substance; - so called because it is levorotatory.
Sinistrin unlike Inulin is water soluble at room temperature what makes it ideal for intravenous applications in renal diagnostics.

4.3 Para-Amino Hippuric Acid

PAH (Para-Amino Hippuric Acid) was isolated from horse urine (hippos: horse, uron: urine) by Liebig in 1829.
The N-acetic acid, $C_9H_{10}N_2O_3$, of para-aminobenzoic acid, is a white crystalline powder which is used as a pharmaceutic aid. As PAH - unlike Inulin and Sinistrin - is secreted in the renal tubules resulting in an excretion that can be up to 5 times higher than the filtered amount, its sodium salt is used for the measurement of the effective renal plasma flow and for determining the functional capacity of the tubular excretory mechanism.

PAH is metabolized in the liver in different extents varying on the patient, its metabolite being excreted by the kidneys as well. That of course makes it necessary not only to measure the amount of p-amino hippuric acid, but also the concentration of its metabolites.
Otherwise, this would lead to a false identification of the model parameters (because of wrong underlying measurements) resulting in a too low plasma-concentration calculation. [3]

Chapter 5

The Basic Model of Pharmacokinetics

The so-called basic model of pharmacokinetics is an open two-compartment model, which means that the extracellular space is considered to be composed of two functionally separated spaces: one, that is perfused rapidly (comprising the blood, the brain, the liver, the kidneys etc.), and the other one which is perfused slowly (like the muscles at rest, the adipose tissue, etc.).
The marker is infused into the vein, it is distributed in the extracellular body water, and finally it is eliminated by the kidneys via the process of urination.

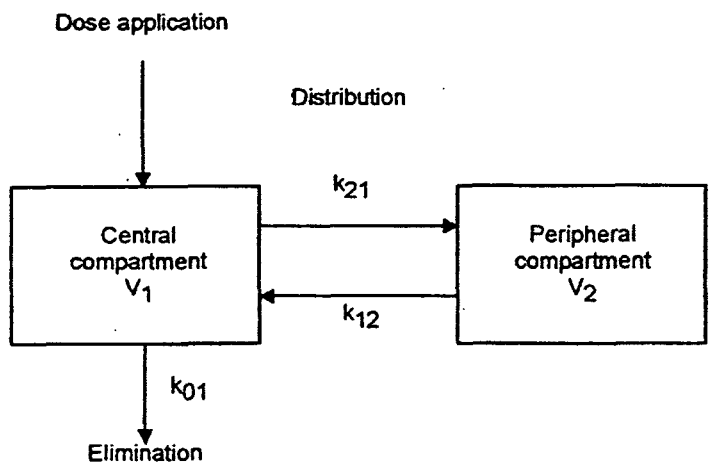


Figure 5.1: Schematic Diagram of the Basic Model of Pharmacokinetics

The marker kinetics, represented by the temporal courses of the marker amounts in the compartments, therefore, is the result of the chosen infusion strategy, the exchange transport between the two compartments, and finally the renal elimination process.

This system can be described by a set of two simultaneous differential equations describing the rate of change of the marker amounts in the respective compartments:

$$\frac{dx_1}{dt} = f(t) - (k_{01} + k_{21})\,x_1 + k_{12}\,x_2 \tag{5.1}$$

$$\frac{dx_2}{dt} = k_{21}\,x_1 + k_{12}\,x_2 \tag{5.2}$$

Equations (2.1) and (2.2) can be stated verbally in the following way:
Firstly, the rate of change of the marker amount in the central compartment, dx_1/dt, is defined by the input strategy chosen, the loss of marker from the central to the peripheral compartment on the one hand, and its gain by the central from the peripheral volume on the other hand, and its elimination via the renal excretion process. Secondly, the rate of change of the marker amount in the peripheral compartment, dx_2/dt, is due to gain from and loss to the central volume.
The input function $f(t)$ is described by the following equation:

$$f(t) = D/\tau, \quad if \;\; 0 \leq t < \tau \tag{5.3}$$

The initial marker amounts are given by:

$$x_1(0) = c_1(0)\,V_1 = x_{10} \tag{5.4}$$

$$x_2(0) = c_2(0)\,V_2 = c_2(0)\,V_1\,(k_{21}/k_{12}) = x_{20} \tag{5.5}$$

with $c_1(t) = x_1(t)/V_1$ and $c_2(t) = x_2(t)/V_2$.

The symbols in the expressions have the following meaning:

$f(t)$	the input strategy as a function of time t
x_1	the amount of the marker in the central compartment
x_2	the amount of the marker in the peripheral compartment
k_{21}	the relative rate of transport from compartment 1 to 2
k_{12}	the relative transport rate from compartment 2 to 1
k_{01}	the relative rate of elimination
D	the priming dose
τ	the infection duration
V_1	the volume of the central compartment
V_2	the volume of the peripheral compartment
c_1	the concentration of the marker in the central compartment
c_2	the concentration of the marker in the peripheral compartment

This linear two-compartment model is suitable for the problem of determining the renal clearance.
For the assessment of the renal plasma flow this model has to be expanded into a nonlinear version [1]. The nonlinearity results from the superposition of excretion processes obeying the so-called Michaelis-Menten law as in tubular secretion, the elimination process of the marker (e.g. p-amino-hippuric acid) being simply proportional to its plasma level.

5.1 Solving the System

Equations (5.1) and (5.2) can also be written in the following form:

$$\dot{\mathbf{x}} = \begin{pmatrix} -(k_{01} + k_{21}) & k_{12} \\ k_{21} & -k_{12} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} f(t) \\ 0 \end{pmatrix}$$

which is of the form

$$\dot{\mathbf{x}} = A\mathbf{x} + b$$

If we want to solve that system we must first find a solution of the corresponding homogenous system:

$$\dot{\mathbf{x}} = A\mathbf{x}, \quad \mathbf{x}(0) = 0$$

As the solution of a linear homogenous system of differential equations is a superposition of exponential functions, one gets a solution of the following

[1] see chapter 9

form:

$$x_1(t) = v_{11}\, e^{\lambda_1 t} + v_{12}\, e^{\lambda_2 t}$$
$$x_2(t) = v_{21}\, e^{\lambda_1 t} + v_{22}\, e^{\lambda_2 t}$$

λ_i being the eigenvalues and v_{ij} the corresponding eigenvectors.

For eigenvalues it counts that:

$$det\,(A - \lambda E) = 0$$

So, in our case that's

$$\begin{vmatrix} -(k_{01} + k_{21} + \lambda) & k_{12} \\ k_{21} & -(k_{12} + \lambda) \end{vmatrix} =$$

$$k_{01}\, k_{12} + \lambda\, k_{01} + k_{12}\, k_{21} + \lambda\, k_{21} + \lambda\, k_{12} + \lambda^2 - k_{12}\, k_{21} =$$

$$\lambda^2 + \lambda\,(k_{01} + k_{12} + k_{21}) + k_{01}\, k_{12} = 0$$

resulting in the following solution:

$$\lambda_{1,2} = \frac{-\,(k_{01} + k_{21} + k_{12}) \pm \sqrt{(k_{01} + k_{21} + k_{12})^2 - 4\, k_{01}\, k_{12}}}{2}$$

The eigenvectors can now be determined from the following equations system:

$$\begin{pmatrix} -(k_{01} + k_{21} + \lambda_i) & k_{12} \\ k_{21} & -(k_{12} + \lambda_i) \end{pmatrix} \begin{pmatrix} v_{i(1)} \\ v_{1(2)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

which leads to

$$v_1 = c \begin{pmatrix} 1 \\ \frac{k_{01}+k_{21}+\lambda_1}{k_{12}} \end{pmatrix}, \quad v_2 = d \begin{pmatrix} 1 \\ \frac{k_{21}}{k_{12}-\lambda_2} \end{pmatrix}$$

So we arrive at the following solution for the homogenous system:

$$\mathbf{x}(t) = c \begin{pmatrix} 1 \\ \frac{k_{01}+k_{21}+\lambda_1}{k_{12}} \end{pmatrix} e^{\lambda_1 t} + d \begin{pmatrix} 1 \\ \frac{k_{21}}{k_{12}-\lambda_2} \end{pmatrix} e^{\lambda_2 t}$$

with c and d arbitrary.

Next, we need to determine a particulary solution: $A\,\mathbf{x}_p + b = \dot{\mathbf{x}}_p$.

If $\mathbf{x}_p = const$, then - clearly - $\dot{\mathbf{x}}_p = 0$ which results in $A\mathbf{x}_p + b = 0$ with

$$b = \begin{pmatrix} f(t) \\ 0 \end{pmatrix} = \begin{cases} \begin{pmatrix} \frac{D}{\tau} \\ 0 \end{pmatrix} & \text{for } 0 < t \leq \tau \\ \\ \begin{pmatrix} 0 \\ 0 \end{pmatrix} & \text{for } \tau < t < \infty \end{cases}$$

One arrives at the following system:

$$\begin{pmatrix} k_{01} + k_{21} & -k_{12} \\ \\ -k_{21} & k_{12} \end{pmatrix} \begin{pmatrix} x_{p_1} \\ \\ x_{p_2} \end{pmatrix} = \begin{pmatrix} \frac{D}{\tau} \\ \\ 0 \end{pmatrix}$$

which is equal to solving:

$$\begin{pmatrix} k_{01} + k_{21} & -k_{12} \\ \\ k_{01} & 0 \end{pmatrix} \begin{pmatrix} x_{p_1} \\ \\ x_{p_2} \end{pmatrix} = \begin{pmatrix} \frac{D}{\tau} \\ \\ \frac{D}{\tau} \end{pmatrix}$$

from which follows that $x_{p_1}\,k_{01} = D/\tau \longrightarrow x_{p_1} = D/(k_{01}\,\tau)$, leading to

$$\begin{aligned} (k_{01} + k_{21})\,\frac{D}{k_{01}\,\tau} - k_{12}\,x_{p_2} &= \frac{D}{\tau} \\ \frac{D}{\tau} + \frac{k_{21}\,D}{\tau\,k_{01}} - k_{12}\,x_{p_2} &= \frac{D}{\tau} \end{aligned}$$

resulting in

$$x_{p_2} = \frac{D}{\tau\,k_{01}} * \frac{k_{21}}{k_{12}} = \frac{k_{21}}{k_{12}} * x_{p_1}$$

So, the solution looks like this:

$$\mathbf{x}(t) = c \begin{pmatrix} 1 \\ \frac{k_{01}+k_{21}+\lambda_1}{k_{12}} \end{pmatrix} e^{\lambda_1 t} + d \begin{pmatrix} 1 \\ \frac{k_{21}}{k_{12}-\lambda_2} \end{pmatrix} e^{\lambda_2 t} + \begin{pmatrix} \frac{D}{\tau k_{01}} \\ \frac{k_{21}}{k_{12}} \frac{D}{\tau k_{01}} \end{pmatrix}$$

c and d can be determined from the initial condition:

$$\mathbf{x}(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = c \begin{pmatrix} 1 \\ a \end{pmatrix} e^{\lambda_1 0} + d \begin{pmatrix} 1 \\ b \end{pmatrix} e^{\lambda_2 0} + \begin{pmatrix} x_{p_1} \\ x_{p_2} \end{pmatrix}$$

with

$$a = \frac{k_{01}+k_{21}+\lambda_1}{k_{12}} \quad \text{and} \quad b = \frac{k_{21}}{k_{12}-\lambda_2}$$

So:

$$\begin{aligned} c + d + x_{p_1} &= 0 \quad |*(-a) \\ da + db + x_{p_2} &= 0 \end{aligned}$$

$$-ad + db - ax_{p_1} + x_{p_2} = 0$$

It isn't hard to see that

$$d(b-a) = ax_{p_1} - x_{p_2} \quad \longrightarrow \quad d = \frac{ax_{p_1} - x_{p_2}}{b-a}$$

Having obtained the value for d one can easily calculate the one for c:

$$\begin{aligned} c + \frac{ax_{p_1} - x_{p_2}}{b-a} + x_{p_1} &= 0 \\ c + \frac{ax_{p_1} - x_{p_2} + bx_{p_1} - ax_{p_1}}{b-a} &= 0 \\ \longrightarrow \quad \frac{-bx_{p_1} + x_{p_2}}{b-a} &= c \end{aligned}$$

which results in the following vector:

$$\begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} \frac{x_{p_2} - b\,x_{p_1}}{b-a} \\ \\ \frac{a\,x_{p_1} - x_{p_2}}{b-a} \end{pmatrix}$$

For $\tau < t < \infty$ the solution is the following:

$$x'(0) \;=\; x(\tau) \;=\; c' \begin{pmatrix} 1 \\ a \end{pmatrix} e^{\lambda_1 (t-\tau)} \;+\; d' \begin{pmatrix} 1 \\ b \end{pmatrix} e^{\lambda_2 (t-\tau)}$$

whereas

$$\begin{pmatrix} 1 & 1 \\ a & b \end{pmatrix} \begin{pmatrix} c' \\ d' \end{pmatrix} = \begin{pmatrix} x_1(\tau) \\ x_2(\tau) \end{pmatrix}$$

$$\Longrightarrow \begin{pmatrix} c' \\ d' \end{pmatrix} = \begin{pmatrix} \frac{b\,x_1(\tau) - x_2(\tau)}{b-a} \\ \frac{x_2(\tau) - a\,x_1(\tau)}{b-a} \end{pmatrix}$$

So we finally arrive at the following complete solution:

For $0 \leq t \leq \tau$:

$$\mathbf{x}(t) \;=\; \frac{b\,x_1(\tau) - x_2(\tau)}{b-a} \begin{pmatrix} 1 \\ \frac{k_{01}+k_{21}+\lambda_1}{k_{12}} \end{pmatrix} e^{\lambda_1 t} \;+\; \frac{x_2(\tau) - a\,x_1(\tau)}{b-a} \begin{pmatrix} 1 \\ \frac{k_{21}}{k_{12}-\lambda_2} \end{pmatrix} e^{\lambda_2 t}$$

$$+ \begin{pmatrix} \frac{D}{\tau\,k_{01}} \\ \frac{k_{21}}{k_{12}} \frac{D}{\tau\,k_{01}} \end{pmatrix} + \begin{pmatrix} \frac{D}{\tau\,k_{01}} \\ \frac{k_{21}}{k_{12}} \frac{D}{\tau\,k_{01}} \end{pmatrix}$$

And for $\tau < t < \infty$:

$$\frac{b\,x_1(\tau) - x_2(\tau)}{b-a}\begin{pmatrix} 1 \\ \frac{k_{01}+k_{21}+\lambda_1}{k_{12}} \end{pmatrix} e^{\lambda_1(t-\tau)} + \frac{x_2(\tau) - a\,x_1(\tau)}{b-a}\begin{pmatrix} 1 \\ \frac{k_{21}}{k_{12}-\lambda_2} \end{pmatrix} e^{\lambda_2(t-\tau)}$$

Chapter 6

Identification

The most immediate goal of scientific or industrial experimentation is to find relationships among manipulated and observed variables, or to validate such relationships coming from some underlying theory. A mathematical description almost invariably involves estimating the values of some unknown parameters to best match the available body of experimental observations.

The simplest mathematical description or model of a system is the function

$$y = f(\mathbf{x}, \mathbf{p})$$

assumed to predict the dependent variable y in terms of the independent variables $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ and unknown parameters $\mathbf{p} = (p_1, p_2, \ldots \ldots, p_m)^T$.
Assuming that the independent variables can be manipulated or observed error-free, and only the dependent variable y is corrupted by measurement errors, the outcome of the i-th experiment is given by the vector $(x_{i1}, x_{i2}, \ldots, x_{in}, \tilde{y}_i)$, where $\tilde{y}_i = f(\mathbf{x_i}, \mathbf{p}) + \epsilon_i$.

Our basic assumption is that the response function $f(x, p)$ is a correct one and the random quantity ϵ_i represents the measurement error. The question of interest then is what the true value $\mathbf{p}$ of the parameters is, though by the imprecise nature of measurements we can never hope to determine it with absolute certainty.
However, having a set of observations and assuming some statistical properties of the errors, it is reasonable to seek parameter estimates that yield not only a good fit to the data, but on the average come firmly close to the true values, and do not vary excessively from one set of experiments to the next. [9, 17]

6.1 Linearity and Non-Linearity in Parameter Estimation

In parameter estimation we vary the parameter values, having already fixed the independent and dependent variables at values determined by the results of the experiment. Whether an equation is linear or non-linear in its parameters can be decided by inspection. It is linear if the dependent variable y is a linear function of the parameters, the value of x being regarded as constant.
An alternative definition is that a function is linear in its parameters if none of the sensitivity coefficients are functions of any parameter(s). The **sensitivity coefficients** are the first derivatives of the model equation with respect to each parameter and they can be very useful measures.

It is very much easier to estimate parameter values producing the best fit of a model to data if the parameters are related linearly. If a non-linear relationship exists between them, simple calculation is no longer possible. Instead, it is necessary to search for the required values by trial and error methods.

6.2 Defining the "Best" Fit of a Model to Data

The most obvious criterion of the goodness of fit of a model to data is that the differences between them should be as small as possible. A simple and well-established technique for ensuring this, and which is usually approptiate, is the method of **least squares**. It is a special case of a more general approach known as the method of maximum likelihood.
In the method of least squares, the **objective function** is defined as the sum of the squares of the differences between each experimental determination and the response of the model $f(x)$ at the same value of x. That is

$$\text{objective function} = \sum_{i-1}^{N}[y_i - f(x_i, p_1, p_2, \ldots, p_m)]^2$$

where the nomenclature means that the contents of the square brackets are to be squared and summed over all values of i, from 1 to the number of data points N. The measured response of the system is y_i, and $f(x_i, p_1, p_2, \ldots, p_m)$ represents the value of the equation of the model at point x_i.
Parameter optimization is the adjustment of the parameter values until the objective function is minimized. The remaining differences between model and data are known as the **residuals**. [7, 9]

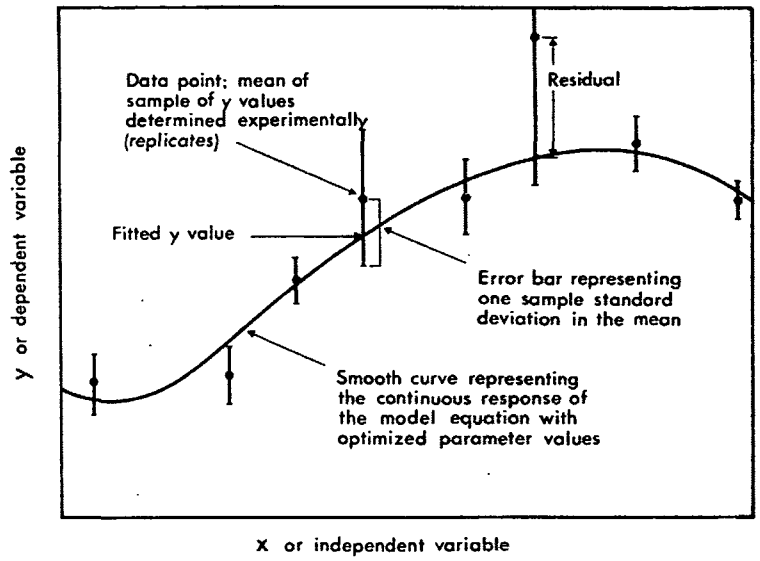


Figure 6.1: Pictorial representation of residuals and error bars

Several important assumptions underly the method of least squares. These are

1. the correct form of the model has been chosen
2. the data are typical
3. the values of y are uncorrelated in the statistical sense
4. there is no error in the values of x

The most difficult of this to satisfy is (1); (4) can be allowed for if grossly untrue, and minor deviations from (2) and (3) are tolerable.
An example of a common occurence of correlation is when measurements have been made over an extended period of time during which an uncontrolled, progressive change has taken place in the experimental conditions, or the properties of the system. If these requirements are not met, the statistical basis of the method is weakened and correct results will not be obtained.

Minimization of the objective function is done using the methods of the differential calculus for determining minima, namely setting the differential of the function with respect to each parameter to zero and solving for the

parameters.

$$\frac{\partial Q(p)}{\partial p_j} = 2 \sum_{i=1}^{n} [\tilde{y}_i - f(\mathbf{x}_i, \mathbf{p})] \left(-\frac{\partial f}{\partial p_j}(\mathbf{x}_i, \mathbf{p})\right) = 0$$

The solutions of these simultaneous equations are the required parameter estimates.

When the equations are linear with respect to the parameters, the solution is easy to obtain as an algebraic equation. This is the case, for example, in the common least squares estimation of the slope and intercept (parameters) of a straight line.
However, when the model equation is non-linear in its parameters, the derivatives of the objective function are no longer linear and the problem becomes one of locating the minimum of a non-linear function. Algorithms for doing this have been known since the time of Newton, but it is only with the development of computers that they have become practicable.

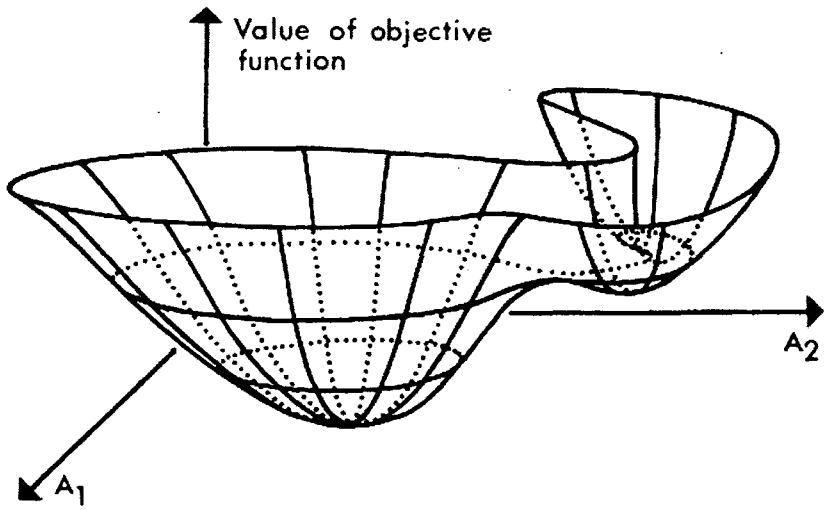


Figure 6.2: The parameter space

6.3 Methods of Parameter Search in Non-Linear Models

Finding the combination of parameter values that minimizes the objective function of a model equation non-linear in its parameters, entails a sequential search of all combinations of parameter values. These combinations define **parameter space** which may be visualized as a kind of landscape of many dimensions where the valleys represent minima in the objective function.

The search is conducted in a series of steps or iterations controlled by a search algorithm. These algorithms may be classified according to whether they are direct or gradient methods.
Direct methods are easier to implement computationally, but **gradient techniques** find the minimum more quickly.
In contrast to the linear case there is the need to provide initial parameter estimates $\mathbf{p}^0$ of the unknown parameter vector $\mathbf{p}$ to be estimated. The choice, which has to rely on independent knowledge or preliminary processing of the data, is of importance since, as there is the possibililty of the iterative scheme converging to a local minimum, or even not converging at all, rather than the desired global minimum. Local minima occur unpredictably in the parameter space of non-linear models. It is essential that they be identified and discarded in favour of the lowest minimum. Unfortunately, it is impossible to prove that any minimum found is not a local one.
One good practical test is to repeat the iterative scheme for a variety of different initial estimates of the parameter vector. If the same minimum is found, confidence increases progressively. [6, 9]

6.3.1 Direct Search Methods

The simplest form of direct search is to divide the feasible range of parameter space into a grid of values, and to evaluate the objective function at each point. The combination of parameter values yielding the minimum sum of differences is then selected.
The simplex approach, in which the objective function is evaluated at $p+1$ mutually equidistant points in the space of p parameters, is more efficient. The principle of the method requires that in a model in which there are two parameters the response of the equation is evaluated at three combinations of parameter values at the vertices of an equilateral triangle, known as the **simplex**. The vertex at which the objective function is maximum is noted, and the simplex is reflected about the other two vertices to define a new point, where a new evaluation is made. This process is repeated until a minimum is located.
The simplex method replaces "trial and error" techniques and is greatly superior to the common practice of varying one parameter at a time, especially

when the parameters are not independent of one another but interact.
On the whole, direct search methods are less efficient than gradient ones, However, the simplex is easy to understand, requires no elaborate calculations, is capable of following an optimum which moves with time and is readily applied in the laboratory to stochastic or deterministic problems.

6.3.2 Gradient Search Methods

Gradient search methods select the search direction using information about the response of the model equation to changes in its parameters. This information is contained in the values of the partial derivatives (sensitivity coefficients) which show how fast the objective function is diminishing with changes in each parameter, and also in the results of previous steps in the search. The direction of parameter space providing the most rapid diminution in the value of the objective function is known as the **direction of steepest descent**. One very important advantage of gradient search methods is that the derivatives can, at the minimum, be used to calculate the allimportant precision of each parameter estimate [1].

Methods which follow the path of steepest descent fail as the minimum is approached, tending to "hunt" inefficiently. Techniques relying on linearization of the fitting function (as the Gauss-Newton method) are more effective in the vicinity of the minimum.
This Gauss-Newton method seeks the minimum in a single step by attempting to calculate its position analytically as if the objective function were really linear, but fall short of this goal in proportion to the degree of non-linearity. However, relatively close to the minimum even a non-linear equation becomes fairly linear, so the objective function can be recast without serious distortion as a linear approximation by using a Taylor series expension and omitting all but the linear terms. An almost direct approach to the minimum can then be calculated analytically, further iterations being required only to adjust for error caused by the assumption of linearity.
Serious error may be introduced by the arbitrary linearization and convergence to the minimum prevented, even after a number of iterations. Succesfull use of the Gauss technique therefore requires guessed initial parameters which are close to the optimum values.

One of the most generally useful and widely applied techniques combines steepest descent and linearization methods, and is due to Levenberg (1944) and Marquardt (1963).
In this **Marquardt-Levenberg algorithm** , the steepest descent approach is automatically applied when it is most effective at points relatively far from

[1] see chapter blablabla: Fisher Information Matrix

the minimum, while linearization of the fitting function is made dominant as the minimum is approached.

6.4 The Marquardt-Levenberg Algorithm

As already heard before, our goal is to find parameter values such that they minimize some error norm

$$Q(\mathbf{p}) = \sum_{i=1}^{n} [\tilde{y}_i - f(\mathbf{x}_i, \mathbf{p})]^2 w_i$$

where the w's are (a priori fixed) weighting coefficients measuring the importance of particular observations in the sum. [6, 9, 22]

6.4.1 Weighting

The testing of a model is begun by collecting experimental measurements. These measurements usually consist of pairs of values, one each for the independent and dependent variables. The independent variable is often time, and the dependent one the response of the experimental system at that time. An additional extremly important piece of information is always potentially available. This is the estimate of reliability or uncertainty in each measured value of the dependent variable. The most useful estimate of uncertainty is the **standard deviation**, or its square the **variance**, because this has a precise statistical meaning.
Fortunately, it is usually possible to be much more certain of the value of the independent , or x variable, which greatly simplifies the analysis.

There are different possibilities for the weighting matrix and the choice which one to use is of significant importance. One possibility is to take $\mathbf{W} = \mathbf{I}$ which means that there is no weighting at all.
Another option is to choose $\mathbf{W} = \mathbf{R}^{-1}$ whereas there are a number of different possibilities for $\mathbf{R}$ as well:

1. If the measurement errors are white (i.e. uncorrelated so that the off-diagonal elements of $\mathbf{R}$ are zero) and stationary (so that all the diagonal elements of $\mathbf{R}$ are equal), then $\mathbf{R} = \sigma^2 \mathbf{I}$ with σ being the measurement error variance.

2. If the measurement errors are white and nonstationary, then

$$\mathbf{R} = \begin{pmatrix} \sigma^2(t_1) & & 0 \\ & \ddots & \\ 0 & & \sigma^2(t_N) \end{pmatrix}$$

Standard deviations (and therefore variances) are easily calculated from replicate measurements made under constant conditions. The estimated errors are then entirely random. The uncertainty in the mean of replicates is found by dividing the standard deviation by the square root of the number of replications, and is known as the **sample mean standard deviation**, or **standard error of the mean.** The term sample is used because the replicates form a sample of the entire population of replicates that might ever be observed.

The standard error of the mean, when squared, provides a good measure of the precision of the mean relative to other means and the weight used is the reciprocal of the variance. In this way, the weight of a mean grows as its variance decreases; the more precisely a value is known, so is its weight increased.

3. Very often, measurement errors are known apart from a scale factor:

$$\mathbf{R} = \mathbf{W}' \sigma^2$$

where $\mathbf{W}'$ is a known weighting matrix and σ^2 is unknown.

With the use of a weighting matrix, it is thus possible to associate a numerical confidence with each mean value of the dependent variable and the fitting procedure will effectively take more notice of variable values that have a high confidence. [7, 6]

As we only have one measurement series per patient a weighting matrix wouldn't make much sense. So we go for the very first case and can take an identity matrix instead.

6.4.2 The Algorithm

For simplicity reasons I will use the following notations:

$$\tilde{\mathbf{Y}} = \begin{bmatrix} \tilde{y_1} \\ . \\ . \\ . \\ \tilde{y_n} \end{bmatrix}, \quad \mathbf{F(p)} = \begin{bmatrix} \mathbf{f(x_1, p)} \\ . \\ . \\ . \\ \mathbf{f(x_n, p)} \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} \mathbf{W_1} & & & \\ & . & & \\ & & . & \\ & & & \mathbf{W_n} \end{bmatrix}$$

thereby reducing the objective function to the form

$$Q(\mathbf{p}) = [\tilde{\mathbf{Y}} - \mathbf{F(p)}]^T \mathbf{W} [\tilde{\mathbf{Y}} - \mathbf{F(p)}].$$

As already heard before the Gauss-Newton method relies on local linear approximation of the function $\mathbf{F}$ around the initial estimate $\mathbf{p}^0$ of the parameters:

$$\mathbf{F(p)} \doteq \mathbf{F(p^0)} + \mathbf{J(p^0)}\,[\mathbf{p} - \mathbf{p^0}].$$

The Jacobimatrix $\mathbf{J}$ of $\mathbf{F}$ is defined by

$$\mathbf{J(p)} = \begin{bmatrix} \frac{\partial f(x_1,p)}{\partial p_1} & . \; . \; . & \frac{\partial f(x_1,p)}{\partial p_m} \\ \frac{\partial f(x_2,p)}{\partial p_1} & . \; . \; . & \frac{\partial f(x_2,p)}{\partial p_m} \\ . & & \\ . & & \\ . & & \\ \frac{\partial f(x_n,p)}{\partial p_1} & . \; . \; . & \frac{\partial f(x_n,p)}{\partial p_m} \end{bmatrix} \tag{6.1}$$

With this one gets the following quadratic approximation of $Q(\mathbf{p})$:

$$\tilde{Q}(\mathbf{p}) = [\tilde{\mathbf{Y}} - \mathbf{F} - \mathbf{J}\,(\mathbf{p} - \mathbf{p}^0)]^T \mathbf{W} [\tilde{\mathbf{Y}} - \mathbf{F} - \mathbf{J}\,(\mathbf{p} - \mathbf{p}^0)]$$

where the argument $\mathbf{p}^0$ of $\mathbf{F}$ and $\mathbf{J}$ is dropped for notational simplicity. The next estimate $\mathbf{p}^1$ is then the minimum point of the quadratic function $\tilde{Q}(\mathbf{p})$, which can easily be found. If we regard $\mathbf{\Delta p} = \mathbf{p} - \mathbf{p}^0$ as the unknown parameter vector the problem is equivalent to solving a linear regression problem with the vector of dependent variables $\tilde{\mathbf{Y}} - \mathbf{F}$ and the

matrix of independent variables $\mathbf{J}$.

$$\frac{\partial}{\partial(\mathbf{p}-\mathbf{p}^0)}(\tilde{\mathbf{Y}} - \mathbf{F} - \mathbf{J}(\mathbf{p}-\mathbf{p}^0))^T \mathbf{W} (\tilde{\mathbf{Y}} - \mathbf{F} - \mathbf{J}(\mathbf{p}-\mathbf{p}^0)) =$$
$$(-\mathbf{J})^T \mathbf{W} (\tilde{\mathbf{Y}} - \mathbf{F} - \mathbf{J}(\mathbf{p}-\mathbf{p}^0) + (\tilde{\mathbf{Y}} - \mathbf{F} - \mathbf{J}(\mathbf{p}-\mathbf{p}^0))^T \mathbf{W} (-\mathbf{J})$$

As $\mathbf{a}^T\mathbf{c}\,\mathbf{b} = \mathbf{b}^T\mathbf{c}\,\mathbf{a}$, this equals:

$$(-\mathbf{J})^T \mathbf{W} (\tilde{\mathbf{Y}} - \mathbf{F} - \mathbf{J}(\mathbf{p}-\mathbf{p}^0) + (-\mathbf{J})^T \mathbf{W} (\tilde{\mathbf{Y}} - \mathbf{F} - \mathbf{J}(\mathbf{p}-\mathbf{p}^0)$$
$$= -2\,\mathbf{J}^T \mathbf{W} (\tilde{\mathbf{Y}} - \mathbf{F} - \mathbf{J}(\mathbf{p}-\mathbf{p}^0)) = 0$$

$$\longrightarrow \quad \mathbf{J}^T \mathbf{W} (\tilde{\mathbf{Y}} - \mathbf{F}) = \mathbf{J}^T \mathbf{W}\,\mathbf{J}(\mathbf{p}-\mathbf{p}^0)$$

After differentiation of $Q(\mathbf{p})$ with respect to the parameters and simple transformation of the resulting equations one arrives at the following solution:

$$\Delta\mathbf{p} = [\mathbf{J}^T \mathbf{W}\,\mathbf{J}]^{-1}\,\mathbf{J}^T \mathbf{W}\,[\tilde{\mathbf{Y}} - \mathbf{F}]$$

Repeated application of this idea would yield the Gauss-Newton iteration

$$\mathbf{p}^{k+1} = \mathbf{p}^k + [\mathbf{J}^T \mathbf{W}\,\mathbf{J}]^{-1}\,\mathbf{J}^T \mathbf{W}\,[\tilde{\mathbf{Y}} - \mathbf{F}], \tag{6.2}$$

where $\mathbf{J}$ and $\mathbf{F}$ are computed at $\mathbf{p}^k$.

This Gauss-Newton algorithm offers quadratic convergence close to the minimum, further apart, however, the step size is frequently inflated, particularly when $[\mathbf{J}^T \mathbf{W}\,\mathbf{J}]$ is nearly singular. Then $\mathbf{p}^{k+1}$ might be a worse approximation of the minimum than $\mathbf{p}^k$ itself.

Gradient Search Methods which follow the iteration

$$\mathbf{p}^{k+1} = \mathbf{p}^k + [diag(\mathbf{J}^T \mathbf{W}\,\mathbf{J})]^{-1}\,\mathbf{J}^T \mathbf{W}\,[\tilde{\mathbf{Y}} - \mathbf{F}]$$

may become ineffective when they approach the minimum.

The **Marquardt-Levenberg modification** seeks to overcome this problem through the iteration

$$\mathbf{p}^{k+1} = \mathbf{p}^k + [\mathbf{J}^T \mathbf{W}\,\mathbf{J} + \lambda^{k+1}\,diag(\mathbf{J}^T \mathbf{W}\,\mathbf{J})]^{-1}\,\mathbf{J}^T \mathbf{W}\,[\tilde{\mathbf{Y}} - \mathbf{F}]$$

with $\mathbf{I}$ being the $n \times n$ unit matrix and λ^{k+1} being a nonnegative scalar, the so-called Marquardt parameter.
With λ sufficiently large, the additional term moderates the length of the step and forces its direction toward the negative gradient of the objective function. A variety of rules has been proposed for selecting this Marquardt parameter in subsequent iterations, In a convergent iteration most of the the methods decrease its value, so that near the minimum we practically return to the Gauss-Newton method. [4, 5, 6]

One simple rule for selecting the Marquardt parameter would be the following:

$$\begin{array}{lrcl} \textit{Initially} & \lambda^{(0)} & = & 0.01, \textit{ whereas in subsequent iterations} \\ & \lambda^{(k+1)} & = & 0.1\,\lambda^{(k)} \textit{ if } Q(p^{(k+1)}) < Q(p^{(k)}), \textit{ and} \\ & \lambda^{(k+1)} & = & 10\lambda^{(k)} \textit{ otherwise} \end{array}$$

6.4.3 Solution of the Model Differential Equations

The solution of the model differential equations within the estimation algorithm must be carried out by numerical integration. Algorithms for numerical integration need to be both stable and efficient as numerical integration is required within each iteration of the estimation process.
For the implementation of this algorithm I chose Matlab's ODE solver45, a Runge Kutta solver.

6.4.4 Computation of Derivatives and Sensitivity Equations

If a gradient type of estimation algorithm is adopted, there is obviously the need to evaluate first derivatives and in some cases (not in ours) there might as well be second derivatives of the model response with respect to the parameters.

One good approach for this is the use of sensitivity equations. This allows 'exact' numerical computation of the derivatives even if it does involve complex computation.

Let's examine this sensitivity approach in more detail. Suppose we have a model specified by:

$$\dot{\mathbf{x}}(t, \mathbf{p}) = \mathbf{A}(\mathbf{p})\mathbf{x}(t, \mathbf{p}) + u(t) \tag{6.3}$$

$$y(t, \mathbf{p}) = \mathbf{C}(\mathbf{p})\mathbf{x}(t, \mathbf{p}) \tag{6.4}$$

$y(t, \mathbf{p})$ being the model response.

If we now build the partial derivatives:

$$\frac{\partial y(t,\mathbf{p})}{\partial p_i} = \mathbf{C(p)}\frac{\partial \mathbf{x}(t,\mathbf{p})}{\partial p_i} + \frac{\partial \mathbf{C(p)}}{\partial p_i}\mathbf{x}(t,\mathbf{p}) \tag{6.5}$$

$$\frac{\partial \dot{\mathbf{x}}(t,\mathbf{p})}{\partial p_i} = \mathbf{A(p)}\frac{\partial \mathbf{x}(t,\mathbf{p})}{\partial \mathbf{p}} + \frac{\partial \mathbf{A(p)}}{\partial \mathbf{p}}\mathbf{x(t,p)} \tag{6.6}$$

At the first iteration, for the initial set of parameter estimates $\mathbf{p^0}$, (6.3) and (6.6) can be solved in series (with $\mathbf{p} = \mathbf{p^0}$), thus giving $y(t)$ and $[\partial y(t,\mathbf{p})]/\partial p_i$. This then can be used to produce the new estimates $\mathbf{p^1}$ (by using $[\partial y(t,\mathbf{p})]/\partial p_i$ in the Jacobimatrix J) [2] This sensitivity procedure is repeated at each iteration step.
The sensitivity approach can also be applied to a model that is nonlinear in the dynamics.

To further illustrate the sensitivity approach, let's return to our model described by

$$\frac{dx_1}{dt} = f(t) - (k_{01} + k_{21})\,x_1 + k_{12}\,x_2 \tag{6.7}$$

$$\frac{dx_2}{dt} = k_{21}\,x_1 + k_{12}\,x_2 \tag{6.8}$$

The model response is given by

$$y(t) = \frac{x_1(t)}{V_1} \tag{6.9}$$

In this case the sensitivity equations are given by:

$$\frac{\partial \dot{x}_1}{\partial k_{01}} = -(k_{01}+k_{21})\frac{\partial x_1}{\partial k_{01}} - x_1 + k_{12}\frac{\partial x_2}{\partial k_{01}} \tag{6.10}$$

$$\frac{\partial \dot{x}_1}{\partial k_{12}} = -(k_{01}+k_{21})\frac{\partial x_1}{\partial k_{12}} - x_2 + k_{12}\frac{\partial x_2}{\partial k_{12}} \tag{6.11}$$

$$\frac{\partial \dot{x}_1}{\partial k_{21}} = -(k_{01}+k_{21})\frac{\partial x_1}{\partial k_{21}} - x_1 + k_{12}\frac{\partial x_2}{\partial k_{21}} \tag{6.12}$$

$$\frac{\partial \dot{x}_2}{\partial k_{01}} = k_{21}\frac{\partial x_2}{\partial k_{01}} - k_{12}\frac{\partial x_2}{\partial k_{01}} \tag{6.13}$$

[2] $\frac{\partial f(x_i,p)}{\partial \mathbf{p_j}}$ equaling $[\partial y(t,\mathbf{p})]/\partial p_i$ in (6.1)

$$\frac{\partial \dot{x}_2}{\partial k_{12}} = k_{21}\frac{\partial x_1}{\partial k_{12}} - k_{12}\frac{\partial x_2}{\partial k_{12}} - x_2 \tag{6.14}$$

$$\frac{\partial \dot{x}_2}{\partial k_{21}} = k_{21}\frac{\partial x_1}{\partial k_{21}} - k_{12}\frac{\partial x_2}{\partial k_{21}} + x_1 \tag{6.15}$$

$$\frac{\partial y}{\partial k_{01}} = \frac{1}{V_1}\frac{\partial x_1}{\partial k_{01}} \tag{6.16}$$

$$\frac{\partial y}{\partial k_{12}} = \frac{1}{V_1}\frac{\partial x_1}{\partial k_{12}} \tag{6.17}$$

$$\frac{\partial y}{\partial k_{21}} = \frac{1}{V_1}\frac{\partial x_1}{\partial k_{21}} \tag{6.18}$$

$$\frac{\partial y}{\partial (\frac{1}{V_1})} = x_1 \tag{6.19}$$

We first need to solve the equation set (6.7) - (6.9) and (6.10) - (6.15) since the sensitivity system (6.10) - (6.19) requires the variables x_1 and x_2.
It should also be noted that most of the computer cost in estimating the parameters is in the repeated numerical integration of these equations.

6.5 Example

To derive appropriate measurements a marker (Sinistrin) is injected, a short amount of time (but not too long, otherwise elimination has already occured) is waited to give the marker the chance to equilibrate throughout the bloodstream and then a number of measurements is made. At first measurements are taken in short periods that become larger as time passes by.

Not only the amount of the injected marker D can differ but also the time of the injection τ can vary between seconds and hours.
The following test data resulted from an injection with overall dose size $D = 2500$ and injection duration of $\tau = 0.5$ min, measurements were made over a period of four hours.

Time	Concentration
5 min.	276 mg/l
10 min.	227 mg/l
15 min.	203 mg/l
20 min.	190 mg/l
25 min.	184 mg/l
30 min.	174 mg/l
35 min.	176 mg/l
40 min	171 mg/l
45 min.	167 mg/l
50 min.	163 mg/l
55 min.	151 mg/l
60 min.	155 mg/l
75 min.	150 mg/l
90 min.	142 mg/l
105 min.	141 mg/l
120 min.	135 mg/l
150 min.	128 mg/l
180 min.	120 mg/l
240 min.	111 mg/l

Identification was done with the Marquardt-Levenberg algorithm described before to find the parameter values that would minimize the difference between these measured values and the function values calculated at the same time points (least squares).
Identification resulted in the following parameter estimates for the independent parameters k_{01}, k_{12}, k_{21} and V_1 and the dependent parameter V_2:

k_{01}	k_{12}	k_{21}	V_1	V_2
0.0039	0.0616	0.0466	7.8836	5.9637

The curve fitted to the experimental data can be seen in fig.6.3

The squared and summed remaining differences between the measured and the modeled data, termed *residual sum of squares*, is 261.0549 for our case.

But what about the renal clearance? The renal clearance can be calculated from the identified parameters by:

$$\text{renal clearance} = k_{01} * V_1 * 1000$$

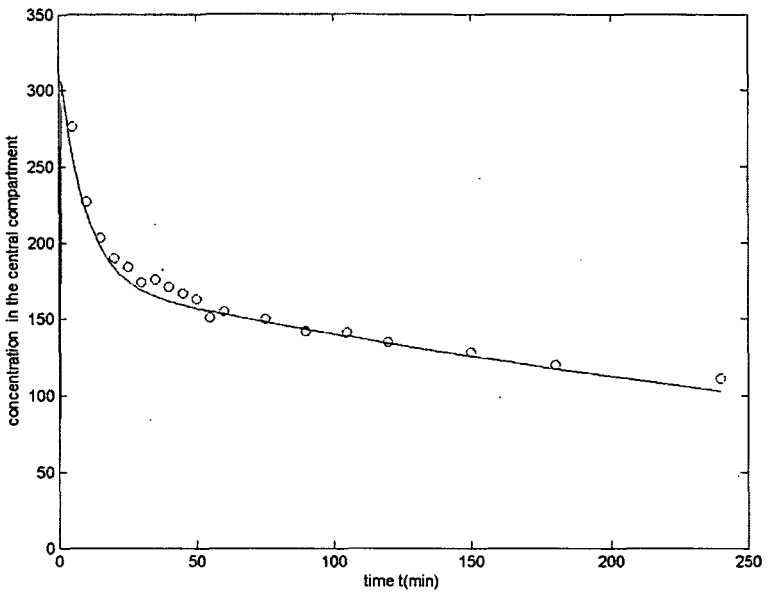


Figure 6.3: Plotting the fitted curve to the experimental data

which results in the following estimation for the kidneys filtration capacity:

$$\text{renal clearance} \approx 31.1109$$

Chapter 7

Accuracy of Parameter Estimates

It is not enough to only compute the estimates $\hat{\mathbf{p}}$ of the parameters, we must also investigate their reliability and precision.
A mathematical model is defined by a series of equations, input factors, parameters and variables aimed to characterize the process being investigated. Input is subject to many sources of uncertainty including errors of measurements, absence of information or natural intrinsic variability of the system, such as the occurrence of stochastic events. All this imposes a limit on our confidence in the response or output of the model.

In the following two approaches for estimating the effect of 'noisy' data on the parameters are introduced. The first one is the Fisher information matrix approach which makes use of the fact that the diagonal elements of the covariance matrix of the parameter estimates contain the variances of the particular parameter estimates.
Determination of this covariance matrix in order to access the parameter accuracy can be a difficult problem. A lower bound of this matrix can be obtained by computing the inverse of the Fisher information matrix. For a sufficiently large set of data, this lower bound is likely to yield a reasonable estimate for the covariance matrix. In practice with only small data sets the inverse of the Fisher information matrix can still be computed, but care must be taken in its interpretation. Moreover, the covariance matrix is usually evaluated through linearization and thus further care must be taken in its interpretation. [20, 14, 7]

Another possibility is the Monte Carlo approach that deals with the generation of artificial protocols by superposition of Gaussian random numbers. For a given number of independent system parameters the random numbers can be taken from a distribution with mean zero and a standard deviation

(sd) given by

$$sd = \sqrt{\frac{Q}{n-m}}$$

7.1 The Fisher Information Matrix

7.1.1 Maximum Likelihood Estimation

As shortly mentioned in the last chapter, the method of least squares is a special case of the maximum likelihood principle. Under the condition of normally distributed errors, the maximum likelihood estimator coincides with the least squares estimator.

As we already know, an estimator is a function that relates measured data to an estimate of an unknown quantity, which means that the estimator can be regarded as a set of rules such that from the data a numerical value (*estimate*) can be assigned to the unknown parameters. The estimate so achieved is obviously dependent on the data.
Assuming that a true value $\mathbf{p}_0$ of the parameter vector $\mathbf{p}$ exists then corresponding to this true value, an exact description of the data can be achieved from the probability density function $f_{\mathbf{p}_0}$. Clearly, the value of $\mathbf{p}_0$ is unknown.

Expressing the problem in a more theoretical mathematical way, the measured date, $z(t_1), \ldots, z(t_N)$ can be viewed as a particular realization of the random variables $Z(t_1), \ldots, Z(t_N)$ which are extracted from an unknown probability density distribution belonging to

$$\{f_{\mathbf{p}}, \mathbf{p} \in \mathbf{R}^P\}$$

where $\mathbf{R}^P$ is the Euclidean space of dimension P, the dimension of the parameter vector $\mathbf{p}$.

Let us define an estimator of $\mathbf{p}$, any function Φ (independent of the parameter)

$$\Phi : \mathbf{R}^N \rightarrow \mathbf{R}^P$$

The value taken by Φ corresponding to the particular realizations $z(t_1), \ldots, z(t_N)$ of $Z(t_1), \ldots, Z(t_N)$ is called an *estimate* $\mathbf{s}$ of $\mathbf{p}$ based on the data:

$$\mathbf{s} = \Phi(z(t_1), \ldots, z(t_N))$$

To yield good estimates from the estimation process, an estimator should possess a number of properties: unbiasedness, minimum variance, efficiency and consistency.

Unbiasedness

An estimator is said to be unbiased if the sampling distribution is evenly clustered about the true value. Mathematically this can be expressed as

$$E_{\mathbf{p}_0}\{\Phi(Z(t_1),\ldots,Z(t_N))\} \;=\; \mathbf{p}_0$$

where $E_{\mathbf{p}_0}\{.\}$ is the expectation operator with respect to the true probability density function $f_{\mathbf{p}_0}$ of the samples.
Since the true value of $\mathbf{p_0}$ is unkown, it is required that

$$E_{\mathbf{p}}\{\Phi(Z(t_1),\ldots,Z(t_n))\} \;=\; \mathbf{p} \qquad \forall \mathbf{p} \in \mathbf{R}^P$$

This is termed uniform unbiasedness.

Minimum Variance

A second desirable property of an estimator is that is should result in an estimate, the variance of which is as small as possible. So, if Γ is the class of unbiased estimators, Φ is said to be a minimum variance estimator if

$$\sigma^2_{\Phi}(\mathbf{p}_0) \leq \sigma^2_{\Psi}(\mathbf{p}_0), \qquad \forall \Psi \in \Gamma$$

where

$$\sigma^2_{\Phi} \;=\; E\{(\Phi - \mathbf{p}_0)(\Phi - \mathbf{p}_0)^T\}$$

Again it is required that

$$\sigma^2_{\Phi}(\mathbf{p}) \leq \sigma^2_{\Psi}(\mathbf{p}), \quad \forall \Psi \in \Gamma, \quad \forall \mathbf{p} \in \mathbf{R}^P$$

Efficiency

An unbiased estimator is called efficient if its covariance is equal to the Cramér-Rao lower bound, that is, the inverse of the *Fisher information matrix* [1]. This is the lowest value of covariance that is theoretically possible to achieve.

[1] for more information on the Cramér-Rao lower bound see page ??

Consistency

The final desirable property of an estimator is that it should be consistent. That means that as the sample size (number of observations) increases, the estimate tends to the true value. Expressing this concept mathematically we have

$$lim_{N\to\infty}\,\Phi[Z(t_1),\ldots,Z(t_N)] \;=\; \mathbf{p}_0$$

If the conditions of unbiasedness and efficiency are satisfied only for the case of N, the number of samples tending to infinity, the properties are termed asymptotic unbiasedness and asymptotic effiency.

Let's return to our maximum likelihood estimator whose underlying principle is simple: select the parameters such that the occurence of the observed values $y_1,\ldots,y_n$ is the most likely among all the possible outcomes of the experiment. While the least squares estimator did require comparatively little a priori knowledge, only information regarding the measurement errors, the maximum likelihood approach is also in need of a probability density function defining the generation of the experimental test data.
Let this a priori probability density function be defined as

$$f(Z(t_1),\ldots,Z(t_N),\mathbf{p}) \tag{7.1}$$

As the measurements are made a posteriori we know the particular realization of the stochastic variables

$$Z(t_1) \;=\; z(t_1), Z(t_2) \;=\; z(t_2),\ldots,Z(t_N) \;=\; z(t_N)$$

and these are the values from which we want to estimate the unknown parameter vector $\mathbf{p}$.

So, $\mathbf{z} \;=\; [z(t_1)\; z(t_2)\; \ldots\; z(t_N)]^T$ is given by the probability density function described by (7.1), but in order to emphasize the dependence of this density function on the measured data, it can be rewritten as

$$L\{z(t_1),\ldots,z(t_N),\mathbf{p}\} \;=\; L\{\mathbf{z},\mathbf{p}\} \tag{7.2}$$

where L is called the *likelihood function.* As the sample values are known, $\mathbf{p}$ is the only unknown argument of L.

As mentioned before, the principle of the maximum likelihood method is to choose as an estimate of $\mathbf{p}$ the value of the parameter vector that most probably renders the observed values $z(t_1), \ldots, z(t_N)$ what means that this chosen estimate $\hat{\mathbf{p}}$ maximizes L:

$$L\{z(t_1), \ldots, z(t_N), \hat{\mathbf{p}}\} \leq L\{z(t_1), \ldots, z(t_N), \mathbf{p}\}, \quad \forall \mathbf{p}$$

For computational reasons it is more convenient to take the logarithmic form of the likelihood function, $ln\, L$, so the estimate $\hat{\mathbf{p}}$ of $\mathbf{p}$ can be found solving

$$\frac{\partial}{\partial \mathbf{p}} lnL\{z(t_1), \ldots, z(t_N), \mathbf{p}\} = 0$$

But it is not enough to only compute the estimates $\hat{\mathbf{p}}$ of the parameters, we should also investigate their reliability and precision. The accuracy of the estimates is normally assessed through the use of the covariance matrix of the parameter estimates $\mathbf{C}$. This is a square matrix with dimension equal to the number of parameters. Having evaluated the covariance matrix corresponding to the parameter estimates obtained from a single set of experimental data, the square root of each of the diagonal elements equals the standard deviation of the corresponding parameter. Thus this matrix provides information as to the reliability of the estimates that have been achieved. [7]
The theoretical basis for evaluating the parameter accuracy relies on the following theorem:

Cramér-Rao Theorem (for an unbiased estimator)

$$\mathbf{C}(\mathbf{p}) \geq \mathbf{F}^{-1}$$

where $\mathbf{F}$ is the Fisher information matrix, which is given by

$$\mathbf{F} = E\left\{\left(\frac{\partial ln\, L}{\partial \mathbf{p}}\right)\left(\frac{\partial ln\, L}{\partial \mathbf{p}}\right)^T\right\}$$

Care must be taken in the interpretation of this measure of accuracy, however, since in practice the number of samples in the data set is restricted, whereas the lower bound is only reached theoretically when there is an infinite number of samples and maximum likelihood estimation is being employed.
As I chose to use a least squares estimator it now needs to be shown that

in our case these two are equivalent and that the Fisher information matrix approach can therefore be used to evaluate the achievable parameter accuracy.

7.1.2 Equivalence of the Maximum Likelihood and Least Squares Estimators

The only thing we need for these two estimators to coincide is the asumption that with

$$z(t_l) \;=\; y(t_l) \,+\, e(t_l)$$

the $e(t_l)$ belong to a normal distribution of zero mean and are of variance $\sigma^2(t_l)$. Moreover the measurement errors are white (i.e. uncorrelated).

Then $z(t_l)$ is normally distributed with mean given by $y(t_l)$ and variance given by $\sigma^2(t_l)$:

$$\mathbf{z} \;\sim\; \mathrm{norm}(\mathbf{y}, \mathbf{R})$$

where

$$\mathbf{R} \;=\; \begin{pmatrix} \sigma^2(t_1) & & 0 \\ & \ddots & \\ 0 & & \sigma^2(t_N) \end{pmatrix}$$

So, for the case of normally distributed errors the probability density function (7.1) (or in likelihood function form (7.2)) is given by

$$L \;=\; \frac{1}{(2\,\pi)^{N/2}\,\mathbf{R}^{1/2}}\, e^{-\frac{1}{2}\,(\mathbf{z}-\mathbf{y})^T\,\mathbf{R}^{-1}\,(\mathbf{z}-\mathbf{y})}$$

with the following logarithmic form:

$$\ln L \;=\; c \,-\, \frac{1}{2}\,(\mathbf{z}\,-\,\mathbf{y})^T\,\mathbf{R}^{-1}\,(\mathbf{z}\,-\,\mathbf{y}) \tag{7.3}$$

with c being a constant.

As can readily be seen, maximizing the likelihood function is equivalent to minimizing $(\mathbf{z}\,-\,\mathbf{y})^T\,\mathbf{R}^{-1}\,(\mathbf{z}\,-\,\mathbf{y})$:

$$\max \ln L \;\Leftrightarrow\; \min\,(\mathbf{z}\,-\,\mathbf{y})^T\,\mathbf{R}^{-1}\,(\mathbf{z}\,-\,\mathbf{y})$$

Thus maximization of the logarithmic form of the likelihood function is equivalent to least squares minimization and we are able to use the Fisher information matrix approach for generating the desired covariance matrix.

As we recall the generic element of the Fisher information matrix $\mathbf{F}$ is given by

$$f_{ij} = E\left\{\left(\frac{\partial \ln L}{\partial p_i}\right)\left(\frac{\partial \ln L}{\partial p_j}\right)\right\} \tag{7.4}$$

but from (7.3) we get

$$\begin{aligned}\frac{\partial \ln L}{\partial p_i} &= -\frac{1}{2}\sum_{l=1}^{N}\frac{1}{\sigma^2(t_l)}\frac{\partial}{\partial p_i}[z(t_l) - y(t_l)]^2 \\ &= \sum_{l=1}^{N}\frac{1}{\sigma^2(t_l)}[z(t_l) - y(t_l)]\frac{\partial y(t_l)}{\partial p_i}\end{aligned} \tag{7.5}$$

since $z(t_l)$ are the measurements and it therefore holds that $\partial z(t_l)/\partial p_i = 0$.

Substituting (7.5) into (7.4) we arrive at the following

$$\begin{aligned}f_{ij} &= E\left\{\left(\sum_{l=1}^{N}\frac{1}{\sigma^2(t_l)}[z(t_l) - y(t_l)]\frac{\partial y(t_l)}{\partial p_i}\right)\right. \\ &\quad \left.\times\left(\sum_{l=1}^{N}\frac{1}{\sigma^2(t_l)}[z(t_l) - y(t_l)]\frac{\partial y(t_l)}{\partial p_j}\right)\right\} \\ &= \sum_{l=1}^{N}\frac{1}{\sigma^4(t_l)}, \frac{\partial y(t_l)}{\partial p_i}, \frac{\partial y(t_l)}{\partial p_j} E\{[z(t_l) - y(t_l)]^2\}\end{aligned} \tag{7.6}$$

And as $E\{\mathbf{e}\mathbf{e}^T\} = \mathbf{R}$, (7.6) is reduced to

$$f_{ij} = \sum_{l=1}^{N}\frac{1}{\sigma^2(t_l)}\frac{\partial y(t_l)}{\partial p_i}\frac{\partial y(t_l)}{\partial p_j} \tag{7.7}$$

It should be noted that the derivatives of (7.7) can be calculated for both, linear and nonlinear models.

This last calculations lead to the following approximation for the desired covariance matrix:

$$\mathbf{V}(\hat{\mathbf{p}}) \;=\; [\mathbf{J}^T \mathbf{R}^{-1} \mathbf{J}]^{-1}$$

where the Jacobi matrix $\mathbf{J}$ is evaluated at $\mathbf{p} = \hat{\mathbf{p}}$.

So, what does this mean for our model?
As we only have one measurement series per patient we don't know the required error variances $\sigma^2(t_l)$. But we can estimate the value of σ^2 through the variance of the residuals

$$s^2 = \frac{Q(\hat{\mathbf{p}})}{n-m} \tag{7.8}$$

which is the residual sum of squares $Q(\hat{\mathbf{p}})$ divided by the degrees of freedom (number of measurements minus the number of parameters).

We therefore arrive at the following approximation for the covariance matrix of the parameter estimates:

$$\mathbf{V}(\hat{\mathbf{p}}) \;=\; [\mathbf{J}^T \mathbf{J}] \frac{n-m}{Q(\hat{\mathbf{p}})}$$

Having estimated $\mathbf{V}(\hat{\mathbf{p}})$ the diagonal elements $v_{ii}(\hat{p}_i)$ of $\mathbf{V}(\hat{\mathbf{p}})$ provide the desired variances of the parameter estimates, so that the accuracy with which the parameter p_i can be estimated may be expressed in terms of its standard deviation by

$$\hat{p}_i \pm \sqrt{v_{ii}(\hat{p}_i)}$$

Parameter precision may also be expressed as the **coefficient of variation** (CV), which is also known as the **fractional standard deviation** (FSD):

$$CV(\hat{p}_i) \;=\; FSD(\hat{p}_i) \;=\; \frac{\sqrt{v_{ii}(\hat{p}_i)}}{\hat{p}_i} \times 100$$

When the CVs of estimated parameter values are unreasonably large (i.e., greater than 100 %) the model is typically considered invalid. Large CVs may arise from limitations in the experimental data such as a small number of measurements or large measurement errors. Large CVs may also arise from utilization of a model that is too complex for the available experimental data. [42, 7]

7.2 The Monte Carlo Technique

Monte Carlo analysis is based on performing multiple evaluations with randomly selected model input, and then using the results of these evaluations to determine the parameter uncertainties.

So, with our measurements given one begins by identifying the underlying model. Then, with the parameter estimates thus found a 'perfect' output of the system at exact the same time points at which the measurements were taken is generated. This 'ideal' output is, first, memorized and then superposed by some random numbers that should represent the 'noise' in the experimental data.
This random numbers are taken from a Gaussian distribution with a mean of zero and a standard deviation given by

$$sd = \sqrt{\frac{Q}{n-m}}$$

with

Q	residual sum of squares
n	number of measurements taken
m	number of independent system parameters

About 100 artificial protocols are generated that way, taking the memorized 'perfect' system output and disturbing it a hundred times with random numbers taken from the above distribution.
The identification procedure is performed for all the artificial protocols and the resulting parameter constallations evaluated statistically for their means and standard deviations. [14, 15, 43]

These standard deviations should be equivalent to the standard errors of the parameters derived by means of the Fisher information matrix method (see also next section). However, since this classical technique has as a necessary condition a Gaussian distribution of the residuals superposed to the solutions, the computer-oriented procedure is more universally applicable.

7.2.1 Using the Monte Carlo Technique on our Example

After the identification procedure was performed in chapter 6.5 on the following data

Time	Concentration
5 min.	276 mg/l
10 min.	227 mg/l
15 min.	203 mg/l
20 min.	190 mg/l
25 min.	184 mg/l
30 min.	174 mg/l
35 min.	176 mg/l
40 min	171 mg/l
45 min.	167 mg/l
50 min.	163 mg/l
55 min.	151 mg/l
60 min.	155 mg/l
75 min.	150 mg/l
90 min.	142 mg/l
105 min.	141 mg/l
120 min.	135 mg/l
150 min.	128 mg/l
180 min.	120 mg/l
240 min.	111 mg/l

yielding the following results

k_{01}	k_{12}	k_{21}	V_1	V_2
0.0039	0.0616	0.0466	7.8836	5.9637

for the parameter estimates, we now want to know more about the confidence we can have on these values obtained.

This was done by means of the Monte Carlo technique: 100 artificial protocols were created, identified and the results thus obtained were evaluated for their means and standard deviations.
All this resulted in the following standard errors:

k_{01}	k_{12}	k_{21}	V_1	V_2
0.0039	0.0616	0.0466	7.8836	5.9637

7.3 Comparing the Fisher Information Matrix Approach to the Monte Carlo Techique

For simplicity reasons in evaluating the Fisher information matrix I chose the following empirical approach for solving the system (5.1)-(5.5):

$$x = A'e^{-\alpha t} + B'e^{-\beta t}$$

so that for the concentration c which is x/V it holds that

$$c = \frac{A'}{V}e^{-\alpha t} + \frac{B'}{V}e^{-\beta t} = Ae^{-\alpha t} + Be^{-\beta t}$$

So we have four independent system parameters to identify: A, B, α and β.

After the identification procedure the parameter estimates

A	α	B	β
167.2249	0.1102	179.2932	0.0002

thus gained are used in two ways:
First, they are used for evaluating the Fisher information matrix by inserting them in the Sensitivity Matrix **J**, and, second, they are used for generating the 'perfect' system output which is then disturbed a 100 times, the artificial protocols being used for identification. Finally the resulting parameter quadruples are statistically evaluated for their means and standard deviations.

The final results show a very good conformity in the found standard errors indicating that in the case of our model both methods are equally practicable (the same argument should also hold for the nonlinear model discussed in chapter 9).

	Monte Carlo	Covariance matrix
A	12.485	12.2214
α	0.0110	0.0115
B	3.2349	3.0284
β	0.0002	0.0002

Table 7.1: Standard deviations of the parameters derived by means of two different methods

Chapter 8

The Matlab Code

8.1 Matlab

The language chosen for the implementation of this problem is MATLAB. MATLAB is a widely used software tool based on numerical vector and matrix manipulation. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation.
MATLAB is an interactive system whose basic data element is an array that does not require dimensioning. This allows you to solve many technical computing problems, especially those with matrix and vector formulations, in a fraction of the time it would have taken to write a program in a scalar noninteractive language such as C or Java.

The name MATLAB stands for MATrix LABoratory. MATLAB was originally written to provide easy access to matrix software developed by the LINPACK and EISPACK projects, which together represent the state-of-the-art in software for matrix computation. Over the years, MATLAB has evolved,also thanks to the input of many users.

The MATLAB language:

This is a high-level matrix/array language with control flow statements, functions, data structures, input/output, and object-oriented programming features. It allows both "programming in the small" as well as "programming in the large" to create complete large and complex application programs. The language features are organized into six directories in the MAT-

LAB Toolbox:

ops	Operator and special characters
lang	Programming language construct
strfun	Character strings
iofun	File input/output
timefun	Time and dates
datatypes	Data types and structures

8.1.1 Toolboxes

MATLAB is both, an environment and a programming language, and one of its great strengths is the fact that the MATLAB language allows you to build your own reusable tools. You can easily create your own special functions and programs (known as M-files) in MATLAB code. As you write more and more MATLAB functions to deal with certain problems, you might be tempted to group related functions together into special directories for convenience. This leads directly to the concept of a *Toolbox*: a specialized collection of M-files for working on particular classes of problems.

The Optimization Toolbox

The Optimization Toolbox contains many commands for the optimization of general linear and nonlinear functions. Optimization is a very broad topic; its purpose is to find the best possible solution to a given problem (which may also include a number of limiting constraints).
Graphically, an optimization problem can be visualized as trying to find the lowest (or highest) point in a complex, highly contoured landscape. An optimization algorithm can thus be likened to an explorer wandering through valleys and across plains in search of the topological extremes. [18]

8.2 The code for the Linear Model

The following contains all the m-files used for

- generating the plot on page 48
- performing the identification procedure

and

- evaluating the accuracy of the found parameter values

by calling the m-files "solveandplot", "identification" and "montecarlo", respectively. All the other m-files are used automatically by the files just mentioned.

opti

```
% time points at which the measurements were made and the corresponding
% measurements
time =[5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 75, 90, 105, 120, 150, 180,
240];
data =[276, 227, 203, 190, 184, 174, 176, 171, 167, 163, 151, 155, 150,...
142, 141, 135, 128, 120, 111];  % measurements taken at the timepoints
contained in 'time'
```

init

```
% Parametervalues used for creating the plot in chapter 3.4.
k01=0.0041;
k12=0.0585;
k21=0.0498;
V1=7.3; V2=6.18;
tau=0.5;
```

solveandplot

```
% solveandplot solves the differential equation system for given values (in
% init) of the parameters. After solving the systems ODEs it plots the
% concentration-curve in the central and peripheral compartment

tspan= [0:0.5:240];     %timespan over which will be integrated
x0= [0;0];
% initial values of the marker amount in the two compartments
init     % calling init gives the parameter values
[t, x] =ode23(@nierenfunktion, tspan, x0,[],k01, k12, k21);

% solving the system with a runge-kutta solver, MATLABs ODE solver
% 23, the ODEs can be found in the function m-file 'kidneyfunction'
hold on
% generating the plot
xlabel('time t (min)')
ylabel('Marker concentration [mg/ml]')
title('Marker concentrations in the central and peripheral compartment')
plot(t,x(:,1)/V1,t,x(:,2)/V2,'-')
```

kidneyfunction

```
% the function m-file nierenfunktion simply gives the two differential
% equations. in the first ODE the function infusion is called. infusion
% returns the input of marker to the system depending on the time

function neu=kidneyfunction(t,x, k01,k12, k21)
neu= [infusion(t)+k12*x(2) - (k01+k21)*x(1);k21*x(1) - k12*x(2)];
```

infusion

```
% the function infusion returns the input of marker (due to injection) to
% the system depending on the time
function inf=infusion(t)
D=2500;
n=length(t);
init
for i=1:n
   if t < tau
      inf(i)=D/tau;
   else
      inf(i)=0;
   end
end
```

identification

```
% identification calculates the parameter values so that the resulting
% solution best fits the measured data and compairs them by plotting
% both, the optimized curve as well as the measurements

kopt0=[0.005 0.05 0.05 7.27];
% kopt0 contains the initial guesses of the parameter values
tspan=[0:1:240];
x0=[0;0];
x1=0;
x2=0;
opti;     % opti contains the measured data

% Calling the Marquardtlevenbergalgorithm
options=optimset('LargeScale','off','Display','iter','LevenbergMarquardt','on',
'TolX',0.0001,'TolFun',0.0001);
[kopt, resnorm,residual] = lsqnonlin(@error, kopt0,[], [],options,x1,x2,data,time);
```

```
% the returned array 'kopt' contains the optimized parameter values
k01=kopt(1)
k12=kopt(2)
k21=kopt(3)
V1=kopt(4)
V2=V1*k21/k12
resnorm
residual;
clearance=k01*V1*1000

% plotting of the optimized curve compaired to the measured data
[t x] =ode23(@kidneyfunction, tspan, x0,[],k01,k12,k21);
plot(t,x(:,1)/V1,time,data,'o')
title('Data points and identified function')
xlabel('time t(min)')
ylabel('concentration in the central compartment')
```

error

```
% calculation of the nonsquared difference between the measured data and
% the modeled solution at the very same time points
function f=error(kopt,x1,x2,data,time)
k01=kopt(1);     % kopt gives the initial guesses of the parameter values
k12=kopt(2);
k21=kopt(3);
V1=kopt(4);
tspan= [0:1:240];
x0= [0;0];
tau=0.5;
[t,x] =ode45(@kidneyfunction, tspan, x0,[],k01, k12, k21);
% solving the system

for i=1:length(time)
   wert(i)=x(time(i),1)/V1;
   % the array wert contains the modeled values at the same time
   % points the measurements were made
end
f=wert-data;
% the difference between the measured and the calculated data is
% returned
```

montecarlo

```
% Montecarlo performs the monte carlo method on the found optimal
% solution to see how sensitive the system reacts to small changes in
% the parameter values. 100 artificial protocols are generated and the
% marquardtlevenberg-algorithm is called for each of them.
clear;
init;
opti;
dataneu=data;
awert=data;
% data are disturbed a 100 times
for k=1:100
   marquardtlevenberg;
   % for each of the 100 times, the optimized parameter values need
   % to be remembered in order to calculate the mean values and the
   % standart deviations
   mck01(k)=k01;
   mck12(k)=k12;
   mck21(k)=k21;
   mcV1(k)=V1;
   mcV2(k)=V1*k21/k12;

   resid=sum(error(kopt,x1,x2,data,time).*error(kopt,x1,x2,data,time));
   % the sum of the squared differences that is needed
   % for the generation of the artificial protocols

   % solve the differential equation with the new set of parameters
   tspan=[0:1:240];
   x0= [0; 0];
   [t, x] =ode45(@kidneyfunction, tspan, x0,[], mck01(k), mck12(k), mck21(k));

   for i=1:length(time)
      xwert(i)=x(time(i),1)/V1;
      % only the values at the time points where
      % the measurements were taken are needed
   end

   if k==1
      % the 'original' function value ('awert') needs to be
      % remembered for it is that value that is disturbed
      awert=xwert;
      r = normrnd(0,sqrt(resid/(length(data)-4)),[1,length(data)]);
      % the error r that is given to the optimal function values is calculated
```

```
    end

    dataneu = awert + r;
    % the artificial data is generated by adding the calculated error
    % to the 'original' value
end

% calculation of the mean values and standard deviations for each
% parameter:
meank01=mean(mck01)
meank12=mean(mck12)
meank21=mean(mck21)
meanV1=mean(mcV1)
meanV2=mean(mcV2)
stdk01=std(mck01)
stdk12=std(mck12)
stdk21=std(mck21)
stdV1=std(mcV1)
stdV2=std(mcV2)
```

marquardlevenberg

```
kopt0=[0.005 0.05 0.05 7.27]; % initial guesses of the parameter values k01,
% k12, k21 and V1
D=2500;
x1=0;
x2=0;
% calling the Marquardtlevenbergalgorithm
options=optimset('LargeScale','off','Display','off','LevenbergMarquardt','on',
'TolX',0.0001,'TolFun',0.0001);
[kopt, resnorm,residual] = lsqnonlin(@error, kopt0,[], [],options,x1,x2,dataneu,time);
k01=kopt(1); % kopt contains the optimized parameter values
k12=kopt(2);
k21=kopt(3);
V1=kopt(4);
```

8.3 Comparison: Fisher Information Matrix - Monte Carlo

The following m-files have been used for calculating the values of the standard deviations of the various parameters found in the table on page 72by means of two different methods: the Fisher information matrix approach and the Monte Carlo technique. For simplicity reasons an empirical solution approach has been chosen, so that although named similar to the files in the last section, most files are **not** the same.
Evaluation of the desired standard deviations is started by calling the m-file "comparison". "comparison" calculates the covariance matrix as the inverse of the Fisher information matrix, takes as the standard deviations the square roots of the diagonal entries of this covariance matrix and writes them in the array "parametervariances". Then, "comparison" calls the m-file "montecarlo" for evaluation of the standard errors by the Monte Carlo approach to be able to finally compare the results thus gained.

comparison

```
identification;
% Covariance matrix method: the columns of the matrix S contain
% the partial derivatives with respect to the parameters, evaluated
% at the same time points the measurements were taken

for u=1:length(time)
   S(u,1)=exp(-alpha*time(u));
   S(u,2)=(-time(u))*A*exp(-alpha*time(u));
   S(u,3)=exp(-beta*time(u));
   S(u,4)=-time(u)*B*exp(-beta*time(u));
end
V=inv(S.'*S)*resnorm/15;
for i=1:4
   a(i)=sqrt(V(i,i));
end
parametervariances=a
montecarlo
```

identification

```
% identification calculates the parameter values so that the resulting
% solution best fits the measured data and compairs them by plotting both,
% the optimized curve as well as the measurements
clear
```

```
popt0= [3000.061000.01];
% popt0 contains the initial guesses of the parameter values
opti;     % opti contains the measured data
% Calling the Marquardtlevenbergalgorithm:
options=optimset('LargeScale','off','Display','iter','LevenbergMarquardt','on','
TolX',0.001,'TolFun',0.001);
[popt, resnorm,residual] = lsqnonlin(@error, popt0,[], [],options,data,time);
% the returned array popt contains the optimized parameter values
A=popt(1);
alpha=popt(2);
B=popt(3);
beta=popt(4);
par= [A, alpha, B, beta]
resnorm;
residual;
t=1:240;
c=A*exp(-alpha*t)+B*exp(-beta*t);
plot(t,c,time,data,'o')
```

error

```
% calculation of the nonsquared difference between the measured data and
% the modelled solution at the very same time points
function f=fehler(popt,data,time)
A=popt(1);     % popt gives the initial guesses of the paramter values
alpha=popt(2);
B=popt(3);
beta=popt(4);
for i=1:length(time)
  c(i)=A*exp(-alpha*time(i))+B*exp(-beta*time(i));
end
f=c-data;     % the difference between the measured and the calculated data
% is returned
```

montecarlo

```
% montecarlo performs the monte carlo method on the found optimal
% solution to see how sensitive the system reacts to small changes in
% the parametervalues. 100 artificial protocols are generated and the
% marquardtlevenberg-algorithm is called for each of them.
opti;
dataneu=data;
```

```
awert=data;
% data are disturbed a 100 times
for k=1:100
   marquardtlevenberg;
   % for each of the 100 times, the optimized parametervalues
   % need to be remembered in order to calculate the mean
   % value and the standart deviation
   mcA(k)=A;
   mcalpha(k)=alpha;
   mcB(k)=B;
   mcbeta(k)=beta;

   if k==1
      % the 'original' function value needs to be remembered
      % for it is that value that is disturbed
      awert=A*exp(-alpha*time)+B*exp(-beta*time);
      resid=sum(fehler(popt,data,time).*fehler(popt,data,time));
      % the sum of the squared differences that is needed for the generation
      % of the artificial protocols
   end

   % solve the system with the new set of parameters
   r = normrnd(0,sqrt(resid/(length(data)-4)),[1,length(data)]);
   % the error that is given to the optimal function values is calculated
   dataneu = awert + r;
   % the artificial data is generated by adding the calculated
   % error to the 'original' value
end

resid;
% calculation of the mean values and standard deviations for each parameter

meanA=mean(mcA);
meanalpha=mean(mcalpha);
meanB=mean(mcB);
meanbeta=mean(mcbeta);
mittelwerte= [meanA, meanalpha, meanB, meanbeta]
stdA=std(mcA);
stdalpha=std(mcalpha);
stdB=std(mcB);
stdbeta=std(mcbeta);
stdabweichungen= [stdA, stdalpha, stdB, stdbeta]
t=1:240;
c=meanA*exp(-meanalpha*t)+meanB*exp(-meanbeta*t);
```

```
plot(t,c,time,data,'o')
```

marquardtlevenberg

```
% marquardtlevenberg calculates the parameter values so that the resulting
% solution best fits the measured data and compairs them by plotting both,
% the optimized curve as well as the measurements
popt0=[200 0.05 100 0.002];
% popt0 contains the initial guesses of the parametervalues
opti;     % opti contains the measured data

% Calling the Marquardtlevenbergalgorithm
options=optimset('LargeScale','off','Display','off','LevenbergMarquardt','on',
'TolX',0.001,'TolFun',0.001);
[popt, resnorm,residual] = lsqnonlin(@error, popt0,[], [],options,dataneu,time);
% the returned array contains the optimized parameter values
A=popt(1);
alpha=popt(2);
B=popt(3);
beta=popt(4);
res=resnorm;
```

Chapter 9

Non-linear Pharmacokinetics

9.1 Reaction Kinetics

In pharmacokinetics there are usually only three kinds of reactions: first order reactions, zero order reactions and reactions obeying a Michaelis Menten kinetics.

9.1.1 First Order Reactions:

Most processes of reabsorption, distribution and elimination of drugs follow a reaction of first order, a so-called linear kinetics, meaning that the speed at which the change of drug-concentration occurs is only dependent on the concentration itself:

$$\frac{dC}{dt} = -K * C$$

dC/dt	change of concentration during time
K	velocity constant of first order $[h^{-1}]$
C	concentration of the drug $[mcg/ml]$

Plotting the concentration against time yields a concave curve, while plotting the same relationship in a semilogarithmic way results in a descendent straight.[1]
The velocity of the concentration-change per unit time equals the product of velocity constant and concentration.

The pharmacokinetical processes of reabsorption, distribution and elimination of most useable drugs (over 90%) obey such a first order kinetics. [4]

[1] see fig. 9.1

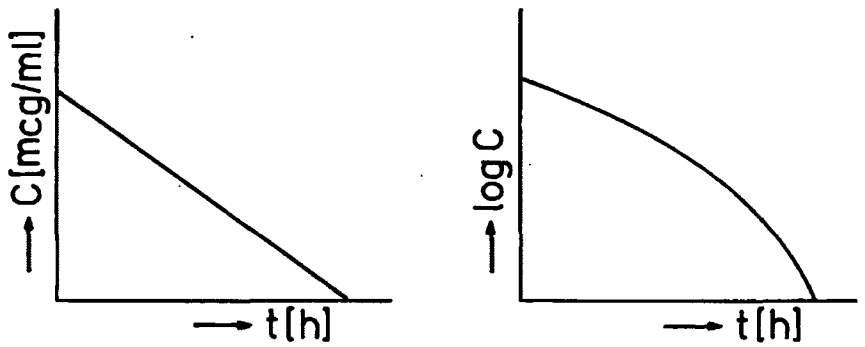


Figure 9.1: Concentration modeled by a first order kinetics

9.1.2 Zero Order Reactions:

If the rate at which absorption or elimination occurs is not dependent on the concentration, meaning that per unit time a constant amount of drug is absorbed/eliminated (and not a constant *proportion* as in the first order reaction), then we are talking about a zero order kinetics. In this case we have the following relation:

$$\frac{dC}{dt} = -K^0$$

K^0 velocity constant of zero order $[mcd/h]$

The accordingly numerical and semilogarithmic plots of concentration against time can be found in fig. 9.2. Examples for such a zero order kinetics are the elimination of Ethanol or the rise of the blood level curve during an intravascular drop infusion.

9.1.3 Michaelis Menten Kinetics:

Metabolic processes of biotransformation require specific enzyme systems, which are of limited capacity. Also, transport enzymes used for active transport of drugs (often against a concentration gradient) can only transport a certain amount before saturation occurs. This can best be described by a Michaelis-Menten-kinetics:

$$\frac{dC}{dt} = -\frac{V_m * C}{K_m + C}$$

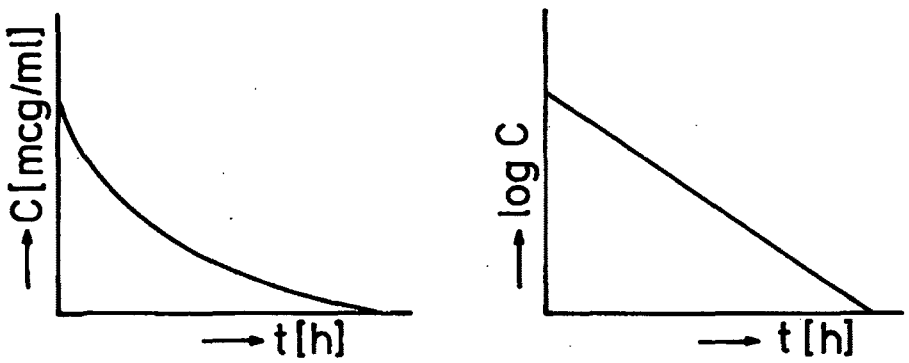


Figure 9.2: Concentration modeled by a zero order kinetics

V_m theoretical maximum rate of the process
K_m Michaelis-Menten-constant

Plotting concentration versus time results in a curve, that is mildly concave in the upper part, but is strongly concave in the lower part (see fig. 9.3). In the semilogarithmic plot one observes convex behaviour at first but then the curve merges into a straight line as time proceeds.

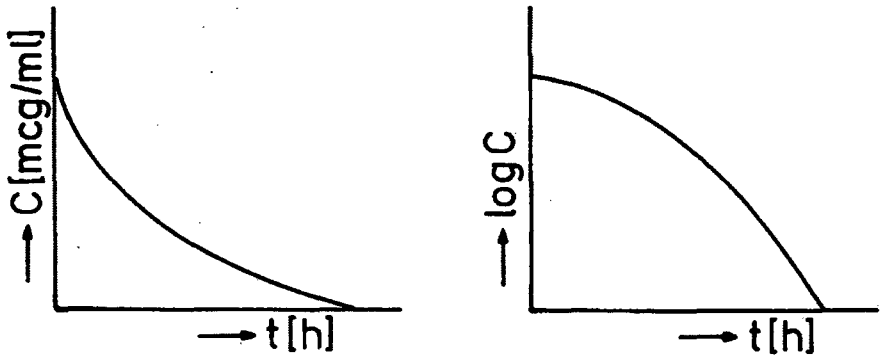


Figure 9.3: Concentration modeled by a Michaelis-Menten kinetics

The Michaelis-Menten constant equals that drug concentration at which the rate of the process is half the theoretical maximum ($V_m/2$). If the concentration is much less than K_m, C can be neglected in the denominator. With V_m and K_m both being constants one can unite them to a new constant K, resulting in a first order kinetics:

$$\frac{dC}{dt} = -K * C$$

Is, on the other hand, the concentration of a much higher value than the Michaelis-Menten constant, then K_m can be neglected and one arrives at the following approximation equation:

$$\frac{dC}{dt} = -V_m$$

which is like a zero order kinetics. [4, 8]

9.2 Saturation in Tubular Secretion

Glomerular filtration is an essential part in the formation of urine. However, the final composition of urine actually excreted depends largely on the transport of solutes and water across the renal cells of the tubuli. This transfer goes from the lumina of the tubuli to the efferent capillaries and vice versa and may be by passive diffusion or by active transport.

Clearly and as already said a couple of times before, a sufficiently high GFR is an essential condition for a normally operating kidney. This, on the other hand, doesn't mean that with a inconspicuous GFR there is nothing to worry about. With its autoregulation ability each nephron is capable of maintaining a relatively steady filtration rate rather independent from the current blood pressure.[2]
In many cases it might be interesting to know how much the actual renal plasma and blood flow really is. But how can this be measured?

For the determination of the glomerular filtration rate a marker substance is administered that is freely filtered and neither reabsorbed or secreted nor used for any metabolic processes. Such a substance has a fractional excretion (C_X/GFR) of 1 as its clearance equals the GFR.
For substances which are net-reabsorbed in the tubules it holds that $\frac{C_X}{GFR} < 1$ while substances which are net-secreted have a fractional excretion (FE) that is greater than 1. An $FE > 1$ is typical for substances that need to be removed from the body quickly like toxins, waste products or extrinsic

[2] see chapter 1.2.2 for more information

substances.
One such substance that is quickly removed from the organism is p-amino hippuric acid (PAH). With an FE that can be as high as 5, removal so rapidly takes place that hardly any PAH can be found in the renal vein. About 90% of the arterial incoming p-amino hippuric acid is excreted, so the renal plasma flow nearly equals the PAH-clearance:

$$ERPF = \frac{V_u * U_{PAH}}{C_{PAH}}$$

$ERPF$	effective renal plasma flow $[ml/min]$
U_{PAH}	PAH concentration in urine $[mg/ml]$
C_{PAH}	PAH concentration in plasma $[mg/ml]$
V_u	urine volume excreted $[ml/min]$

If we now take into consideration that only 90% of the arterial incoming PAH appears in the urin, we only need to divide the PAH-clearance by 0.9 to get an accurate estimate of the renal plasma flow:

$$RPF = \frac{ERPF}{0.9} \quad [ml/min]$$

Once we know the renal plasma flow, renal blood flow can be calculated using the so-called hematocrit ($\backsimeq 45\,\%$)[3]:

$$RBF = \frac{RPF}{1 - Hct}$$

A necessary condition for the determination of the renal blood flow by means of the PAH-clearance is that really 90% of the p-amino hippuric acid are excreted by the kidneys. This, however, is only the case at relatively low PAH-plasma concentrations ($K_M \approx 10\,\mu mol/l$), because at higher plasma concentration saturation occurs. [2, 1]
This leaves us with two possibilities: either, the dose administered is small enough that the PAH-plasma concentration doesn't rise above the critical value where saturation occurs, or, the model needs to be expanded into a nonlinear version.

[3]see appendix: glossary

9.3 The Non-linear Two-Compartment Model

The differential equations are nearly the same as in the linear case except that now the elimination rate k_{01} depends upon the concentration c:

$$\frac{dx_1}{dt} = f(t) - (k_{01}(c) + k_{21})\,x_1 + k_{12}\,x_2 \tag{9.1}$$

$$\frac{dx_2}{dt} = k_{21}\,x_1 + k_{12}\,x_2 \tag{9.2}$$

with k_{01} being:

$$k_{01}(c) = \begin{cases} k_2 & \text{if} \quad 0 \leq c < c_{krit} \\ k_1 + (k_2 - k_1)\frac{c_{krit}}{c} & \text{if} \quad c_{krit} \leq c < \infty \end{cases}$$

Equations (4.1) and (4.2) describe the marker kinetics in the organism in the following way:
Firstly, the rate of change of the marker amount in the central compartment $\dot{x_1}$ is determined by the chosen input strategy $f(t)$, the loss of marker from the well perfused central to the less perfused peripheral compartment by diffusion, the gain of marker from the peripheral to the central compartment by backdiffusion, and, finally, the elimination from the central compartment through the renal excretion mechanism. The markers may differ in the nature of their excretion mechanisms, but are chosen such that no extrarenal elimination routes exist.
Secondly, the rate of change of the markers amount $\dot{x_2}$ is due to diffusive gain from the central compartment and diffusive loss to the peripheral compartment. The diffusion processes are supposed to be proportional to the amounts in the two volumes. [44]

The input function $f(t)$ is described by the following equation:

$$f(t) = D/\tau, \quad if \;\; 0 \leq t < \tau \tag{9.3}$$

The initial marker amounts are given by:

$$x_1(0) = c_1(0)\,V_1 = x_{10} \tag{9.4}$$

$$x_2(0) = c_2(0)\,V_2 = c_2(0)\,V_1\,(k_{21}/k_{12}) = x_{20} \tag{9.5}$$

with $c_1(t) = x_1(t)/V_1$ and $c_2(t) = x_2(t)/V_2$.

The symbols in the expressions have the following meaning:

$f(t)$	the input strategy as a function of time t
x_1	the amount of the marker in the central compartment
x_2	the amount of the marker in the peripheral comparmtent
k_{21}	the relative rate of transport from compartment 1 to 2
k_{12}	the relative transport rate from compartment 2 to 1 and tubular secretion
D	the priming dose
τ	the infection duration
c_{krit}	the critical concentration in the piecewise linearized Michaelis-Menten law
k_1	the relative transport rate due to glomerular filtration
k_2	the relative transport rate due to the sum of glomerular filtration
V_1	the volume of the central compartment
V_2	the volume of the peripheral compartment
c	the concentration of the marker in the central compartment
c_2	the concentration of the marker in the peripheral compartment

Normally, determining the parameter values of a nonlinear model is a rather time consuming enterprise with a higher computational cost than identifying the parameters of a linear model. One parameter more or less can make a significant difference in the time needed to run the Marquardt-Levenberg algorithm.
As for our nonlinear model it is possible to spare the identification of the parameter k_1 , so we only have five instead of six parameters to compute: Since k_1 is the parameter associated with the relative transport rate due to glomerular filtration solely, this parameter equals k_{01} which can be gained by identification of the linear model.

Having two measurement series, one for sinistrin and one for PAH, one uses the measurements made for sinistrin to calculate the renal clearance thus also estimating the parameter k_{01} which then can be used for the nonlinear case. For the nonlinear model using the PAH measurements, k_1 is set to the estimated value of k_{01} and the identification procedure, therefore, only needs to be performed on the remaining 5 independent parameters $k_{12}, k_{21}, k_2, ckrit$ and V_1.

9.4 Examples

As heard before, if we want to fasten the identification procedure, we are in need of measurements, not only of PAH, but of sinistrin as well. The

following data was gained after a dose of 2500 mg Sinistrin and a mean infusion time of 7.5 min and after administration of 2200 mg PAH given over an infusion time of 7.5 min.:

MARKER:	SINISTRIN
Time	**Concentration**
5 min.	222 mg/l
10 min.	301 mg/l
15 min.	236 mg/l
20 min.	200 mg/l
25 min.	169 mg/l
30 min.	150 mg/l
35 min.	140 mg/l
40 min	135 mg/l
50 min.	110 mg/l
60 min.	92 mg/l
75 min.	80 mg/l
90 min.	65 mg/l
105 min.	59 mg/l
120 min.	50 mg/l
140 min.	45 mg/l
160 min.	31 mg/l
180 min.	30 mg/l
200 min.	22 mg/l
220 min.	20 mg/l

MARKER:	PAH
Time	**Concentration**
5 min.	70 mg/l
10 min.	92 mg/l
15 min.	58 mg/l
20 min.	45 mg/l
25 min.	31 mg/l
30 min.	23 mg/l
35 min.	21 mg/l
40 min	19 mg/l
50 min.	16 mg/l
60 min.	15 mg/l
75 min.	8 mg/l
90 min.	7 mg/l
105 min.	6 mg/l
120 min.	5 mg/l
140 min.	4 mg/l
160 min.	3 mg/l
180 min.	2 mg/l
200 min.	1 mg/l
220 min.	0.5 mg/l

Identification using the sinistrin-measurements yielded the following parameter estimates:

k_{01}	k_{12}	k_{21}	V_1	V_2
0.0212	0.0437	0.0324	6.1111	4.5365

The parameter estimate of k_{01} is taken and the parameter k_1 (of the nonlinear model) is set to that value. Then identification is performed using the PAH-data resulting in:

k_{12}	k_{21}	k_2	$ckrit$	V_1	V_2
0.0432	0.0585	0.0564	74.9914	13.5327	18.3266

These parameter estimates lead to the consequent value for the GFR and the renal plasma flow (RPF):

$GFR \approx 129.34$

$RPF \approx 763.375$

The curves fitted to the experimental data can be seen in fig.9.4 and 9.5, respectively.

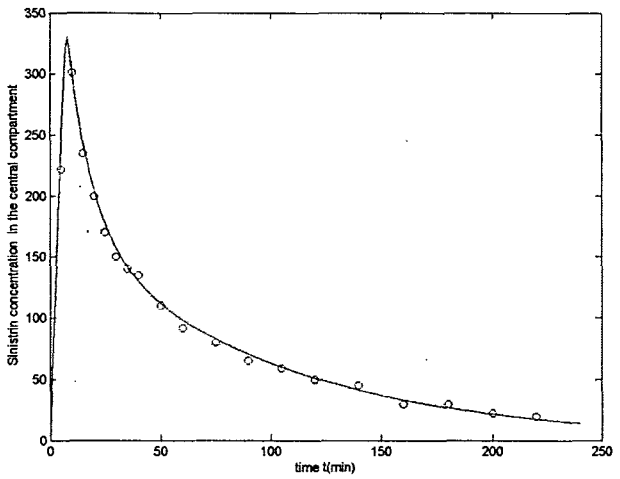


Figure 9.4: Plotting the fitted curve to the experimental data derived from an injection of sinistrin

For the determination of the standard errors, Monte Carlo analysis was performed resulting in the following errors:

	k_{12}	k_{21}	k_2	$ckrit$	V_1	V_2	RPF
Mean values:	0.0490	0.0656	0.0577	76.4251	13.7628	18.4134	791.6219
St.deviations:	0.0206	0.0266	0.0168	3.2623	1.7828	3.1327	83.7755

The same procedure is performed for another set of data derived from a different patient than the data before but with the same dosage sizes and injection times.

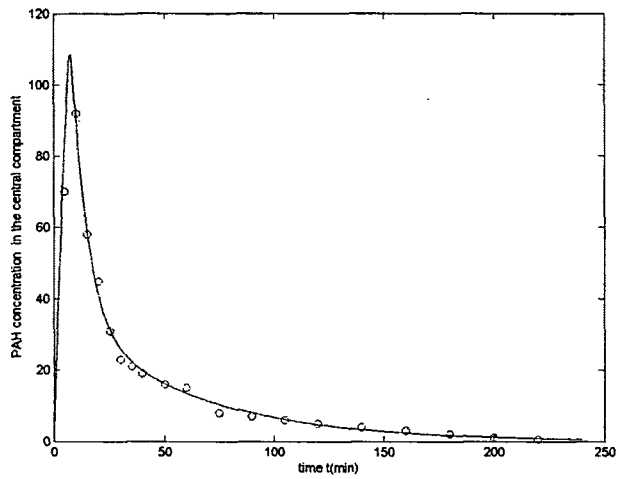


Figure 9.5: Plotting the fitted curve to the experimental data derived from an injection of PAH

MARKER:	SINISTRIN
Time	**Concentration**
5 min.	335 mg/l
10 min.	278 mg/l
20 min.	258 mg/l
30 min.	255 mg/l
35 min.	249 mg/l
40 min.	230 mg/l
45 min.	225 mg/l
50 min	211 mg/l
60 min.	190 mg/l
70 min.	178 mg/l
90 min.	215 mg/l
110 min.	152 mg/l
130 min.	142 mg/l
140 min.	132 mg/l
160 min.	120 mg/l
180 min.	110 mg/l
200 min.	103 mg/l
220 min.	95 mg/l
240 min.	90 mg/l

MARKER:	PAH
Time	**Concentration**
5 min.	135 mg/l
10 min.	88 mg/l
20 min.	80 mg/l
30 min.	85 mg/l
35 min.	59 mg/l
40 min.	49 mg/l
45 min.	52 mg/l
50 min	42 mg/l
60 min.	29 mg/l
70 min.	33 mg/l
90 min.	25 mg/l
110 min.	19 mg/l
130 min.	17 mg/l
140 min.	16 mg/l
160 min.	12 mg/l
180 min.	8 mg/l
200 min.	7 mg/l
220 min.	6 mg/l
240 min.	5.5 mg/l

Identified parameter values:

Linear model:

k_{01}	k_{12}	k_{21}	V_1	V_2
0.0064	0.1412	0.0359	6.8746	1.75

Non-linear model:

k_{12}	k_{21}	k_2	$ckrit$	V_1	V_2
0.0868	0.1475	0.0513	54.9962	7.2334	12.1554

Monte Carlo analysis yielded the subsequent standard errors for the particular parameters:

	k_{12}	k_{21}	k_2	$ckrit$	V_1	V_2	RPF
Mean values:	0.0891	0.1518	0.0585	54.8903	7.2682	12.7791	424.7715
St.deviations:	0.0902	0.0928	0.0178	2.3358	0.8291	4.3489	73.7437

With $GFR = k_{01} * V_1 * 1000$ and $RPF = k_2 * V_1 * 1000$ the following values are obtained:

$GFR \approx 44.25$ $\quad$ $RPF \approx 370.877$

As we can see from the results, this patient suffers from severe renal problems, his kidney not being able to function properly. The value calculated for the gfr as well as the found rpf is clearly below average.

The curve fitted to the experimental data can be seen in fig.9.6.

9.5 The Code

identification

identilinear
k1=k01;

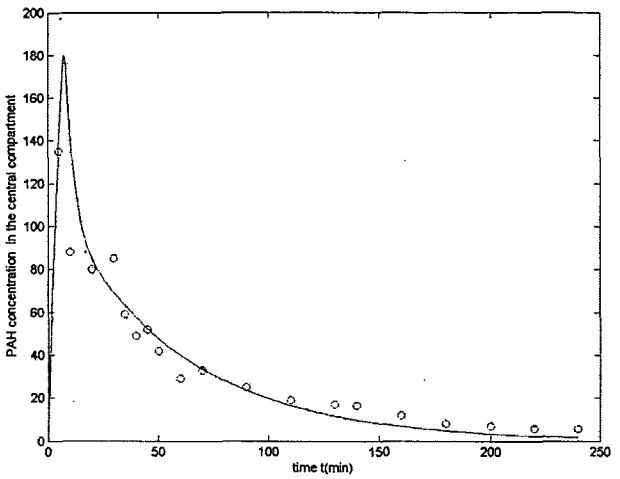


Figure 9.6: Plotting the fitted curve to the experimental data derived from an injection oPAH

```
% kopt0 contains the initial parameter values
kopt0= [0.05 0.05 7.27 0.05 75];
tspan= [0:1:240];
x0= [0; 0];
x1=0;
x2=0;
opti2;     % opti2 contains the measured data

% calling the Marquardlevenbergalgorithm
options=optimset('LargeScale','off','Display','iter','LevenbergMarquardt','on',
'TolX',0.0001,'TolFun',0.0001);
[kopt, resnorm,residual] = lsqnonlin(@error, kopt0,[], [],options,x1,x2,data,time,k1);
k12=kopt(1)
k21=kopt(2)
V1=kopt(3)
k2=kopt(4)
ckrit=kopt(5)
V2=V1*k21/k12
clearence=k2*V1*1000
resnorm
residual;
% plotting the 'optimized' curve compared to the measured data
```

```
[t x] =ode23(@kidneyfunction, tspan, x0,[],k12,k21,V1,k1,k2,ckrit);
plot(t,x(:,1)/V1,time,data,'o')
```

identilinear

```
% identilinear calculates the parameter values of the corresponding linear
% model so that the resulting solution best fits the measured data and
% compairs them by plotting both, the optimized curve as well as the
% measurements
% kopt0 contains the initial guesses of the parameter values
kopt0= [0.005 0.05 0.05 7.27];
tspan= [0:1:240];
x0= [0;0];
x1=0;
x2=0;
opti;     % opti contains the measured data of the marker sinistrin

% Calling the Marquardtlevenbergalgorithm
options=optimset('LargeScale','off','Display','iter','LevenbergMarquardt','on',
'TolX',0.0001,'TolFun',0.0001);
[kopt, resnorm,residual] = lsqnonlin(@errorr2, kopt0,[], [],options,x1,x2,data,time);
% the returned array kopt contains the optimized parameter values
k01=kopt(1)
k12=kopt(2)
k21=kopt(3)
V1=kopt(4)
```

error2

```
% calculation of the nonsquared difference between the measured data and
% the modelled solution at the very same time points
function f=error2(kopt,x1,x2,data,time)
k01=kopt(1);     % kopt gives the initial guesses of the paramter values
k12=kopt(2);
k21=kopt(3);
V1=kopt(4);
tspan= [0:1:240];
x0= [0;0];
tau=0.5;
[t,x] =ode45(@kidneyfunction2, tspan, x0,[],k01, k12, k21);
% solving the system
for i=1:length(time)
```

```
    wert(i)=x(time(i),1)/V1;
    % the array wert contains the modelled values at the same
    % time points the measurements were made
end
f=wert-data;     % the difference between the measured and the
% calculated data is returned
```

kidneyfunction2

```
% the function m-file kidneyfunction2 simply gives the two differential
% equations. in the first ODE the function infusion is called. infusion
% returns the input of marker to the system depending on the time
function neu=kidneyfunction2(t,x, k01,k12, k21)
neu= [infusion2(t)+k12*x(2) - (k01+k21)*x(1);k21*x(1) - k12*x(2)];
```

opti

```
% timepoints at which the measurements were made and the corresponding
% measurements
time =[5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 75, 90, 105, 120, 150, 180,
240];
data =[276, 227, 203, 190, 184, 174, 176, 171, 167, 163, 151, 155, 150,...
142, 141, 135, 128, 120, 111];
% measurements taken at the time points contained in time
```

init

```
% init contains the administration information
tau=7.5;
D=2200:
```

kidneyfunction

```
% the function m-file nierenfunktion simply gives the two differential
% equations. in the first ODE the function infusion is called. infusion
% returns the input of marker to the system depending on the time
function neu=nierenfunktion(t,x,k12, k21,V1, k1, k2, ckrit)
neu=[infusion(t)+k12*x(2) - (k01(x(1)/V1,k1,k2,ckrit)+k21)*x(1);k21*x(1)
- k12*x(2)];
```

infusion

```
% the function infusion returns the input of marker (due to injection) to
% the system depending on the time
function inf=infusion2(t)
D=2200;
n=length(t);
init
for i=1:n
   if t<tau
     inf(i)=D/tau;
   else
     inf(i)=0;
   end
end
```

k01

```
function out=k01(c,k1,k2,ckrit)
if c < ckrit
   out=k2;
else
   out=k1+(k2-k1)*ckrit/c;
end
```

error

```
% calculation of the error between model and measurements
function f=fehler(kopt,x1,x2,data,time,k1)
k12=kopt(1);
k21=kopt(2);
V1=kopt(3);
k2=kopt(4);
ckrit=kopt(5)
; tspan= [0:1:240];
x0= [0;0];
tau=0.5;
[t,x] =ode45(@nierenfunktion, tspan, x0,[], k12, k21,V1, k1, k2, ckrit);
for i=1:length(time)
   wert(i)=x(time(i),1)/V1;
end
f=wert-data;
```

montecarlo

```
% Montecarlo performs the monte carlo method on the found optimal
% solution to see how sensitive the system reacts to small changes in
% the parameter values. 100 artificial protocols are generated and the
% marquardtlevenberg-algorithm is called for each of them.
clear xwert;
init;
opti;
identilinear
k1=k01:
dataneu=data;
awert=data;
% data are disturbed a 1000 times
for k=1:10     marquardtlevenberg;
   mck12(k)=k12;
   mck21(k)=k21;
   mcV1(k)=V1;
   mcV2(k)=V1*k21/k12;
   mck2(k)=k2;
   mcckrit(k)=ckrit;

   resid=sum(fehler(kopt,x1,x2,data,time,k1).*fehler(kopt,x1,x2,data,time,k1));
   % solve the differential equation with the new set of parameters
   tspan= [0:1:240];
   x0= [0;0];
   [t,x] =ode45(@kidneyfunction, tspan, x0,[], mck12(k), mck21(k),
   mcV1(k), k1, mck2(k), mcckrit(k));

   for i=1:length(time)
      xwert(i)=x(time(i),1)/V1;
   end
   error=xwert-dataneu;
   r = normrnd(0,sqrt(resid/(length(data)-4)),[1,length(data]);

      if k==1     % remembering the first function value
     awert=xwert;
   end
   dataneu = awert + r;
end
plot(t,x(:,1)/V1,time,dataneu,'o',time,awert,'x')
resid

% calculating the mean and standard deviations
```

```
meank12=mean(mck12)
meank21=mean(mck21)
meanV1=mean(mcV1)
meanV2=mean(mcV2)
meank2=mean(mck2)
meanckrit=mean(mcckrit)
stdk12=std(mck12)
stdk21=std(mck21)
stdV1=std(mcV1)
stdV2=std(mcV2)
stdk2=std(mck2)
stdckrit=std(mcckrit)
```

marquardtlevenberg

```
kopt0= [0.05 0.05 7.27 0.05 75];
D=2500;
x1=0;
x2=0;
% calling the Marquardtlevenbergalgorithm
options=optimset('LargeScale','off','Display','off','LevenbergMarquardt','on',
'TolX',0.0001,'TolFun',0.0001);
[kopt, resnorm,residual] = lsqnonlin(@error, kopt0,[], [],options,x1,x2,dataneu,time);
k12=kopt(1);
k21=kopt(2);
V1=kopt(3);
k2=kopt(4);
ckrit=kopt(5);
```

Chapter 10

Protocol Discussion

As one can imagine taking measurement series isn't very amusing, neither for the patient nor for the doctor in charge. Measurements are taken over several hours, each measurement involved with taking a blood sample that should be collected at exact the required time points.
Nonetheless, it is essential that measurements are taken over a sufficiently long period of time. If the procedure is broken off too early, this could result in an overestimation of the clearance value that could lead to believe that the renal filtration performance is much better than it really is.

In clinical practice guide lines for an optimal protocol have been created. At first, the measurement density should be high while later on, the time between the taken measurements may increase. So, blood samples should be taken after $5, 10, 15, 30, 60, 90, 120, 180$ minutes after the injection of the marker.
For all the examples shown earlier, sufficiently long protocols have been chosen. In the following we will see what would have happened if the measurements hadn't lasted the required amount of time but had been broken off after $60, 75, \ldots$ minutes.

For this we take the measurements of our very first example gained from an injection with overall dose size $D = 2500$ and injection duration of $\tau = 0.5$ min:

Time	Concentration
5 min.	276 mg/l
10 min.	227 mg/l
15 min.	203 mg/l
20 min.	190 mg/l
25 min.	184 mg/l
30 min.	174 mg/l
35 min.	176 mg/l
40 min	171 mg/l
45 min.	167 mg/l
50 min.	163 mg/l
55 min.	151 mg/l
60 min.	155 mg/l
75 min.	150 mg/l
90 min.	142 mg/l
105 min.	141 mg/l
120 min.	135 mg/l
150 min.	128 mg/l
180 min.	120 mg/l
240 min.	111 mg/l

We now start with only taking into account the measurements made in the first hour, but then we subsequently use more and more of the experimental data until at the last identification run the whole measurement series is used.

The different clearance values due to the different protocol lengths and the according errors can be seen in fig.10.1 and 10.2, respectively.

As the results show, the shorter the protocols lasted the higher the estimated clearance value was. This can also be seen in the plots where much higher end-concentrations in the central compartment were achieved when the protocols have been broken off early. For these plots see fig.10.3 to fig.10.10.

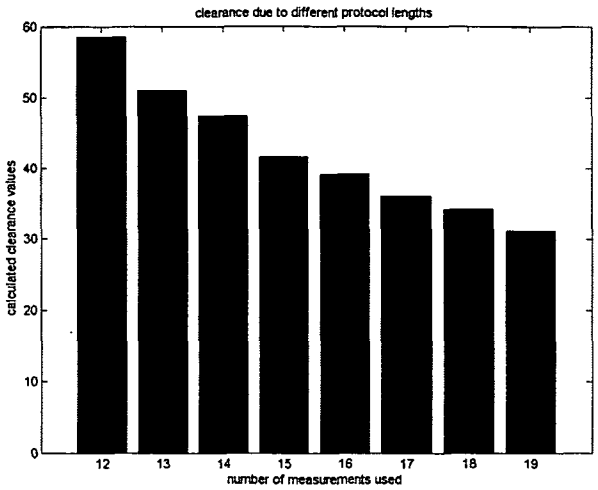


Figure 10.1: The different clearance values due to different protocol lengths

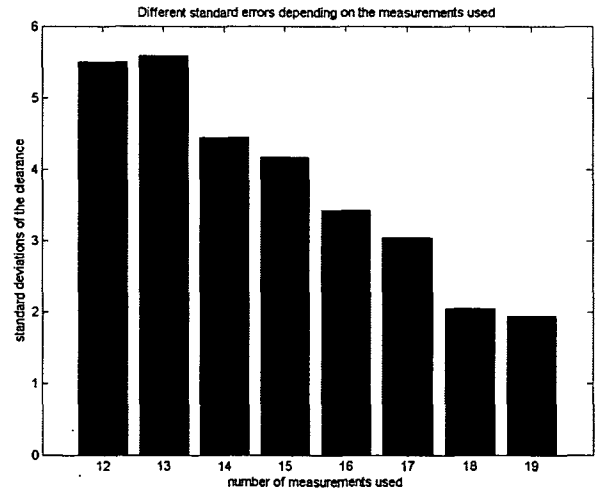


Figure 10.2: The according standard deviations to the clearance values of fig. 10.1

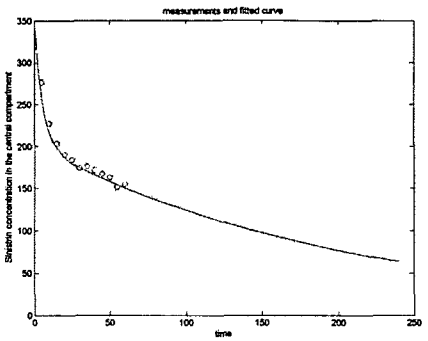


Figure 10.3: Plot of the fitted curve of a shortened protocol of 1 hour

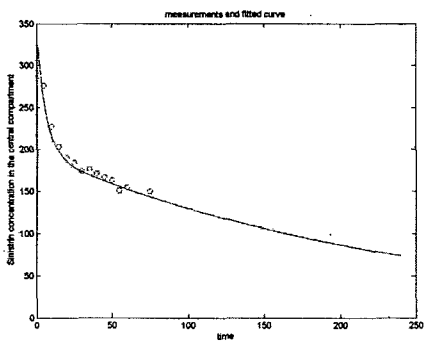


Figure 10.4: Plot of the fitted curve of a shortened protocol of 75 minutes

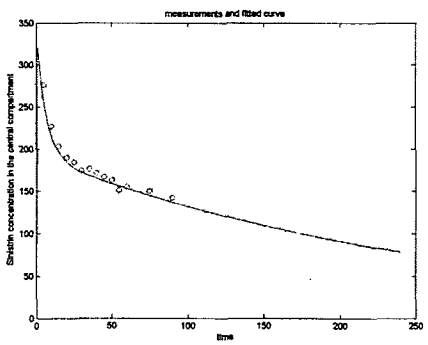


Figure 10.5: Plot of the fitted curve of a shortened protocol of 90 minutes

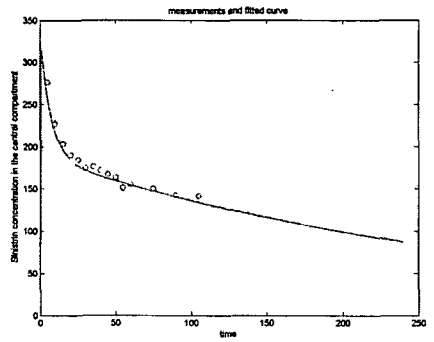


Figure 10.6: Plot of the fitted curve of a shortened protocol of 105minutes

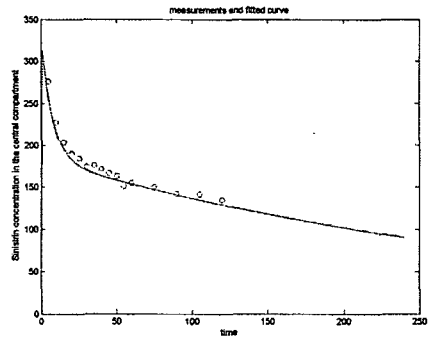


Figure 10.7: Plot of the fitted curve of a shortened protocol of 120 minutes

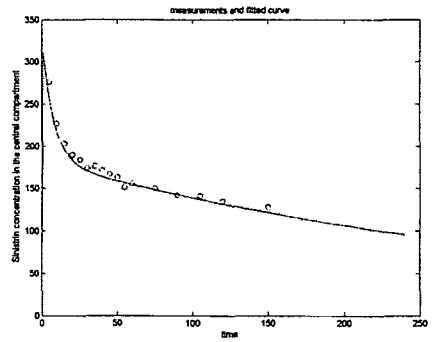


Figure 10.8: Plot of the fitted curve of a shortened protocol of 150 minutes

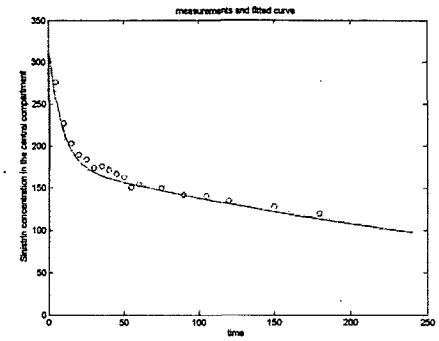


Figure 10.9: Plot of the fitted curve of a shortened protocol of 180 minutes

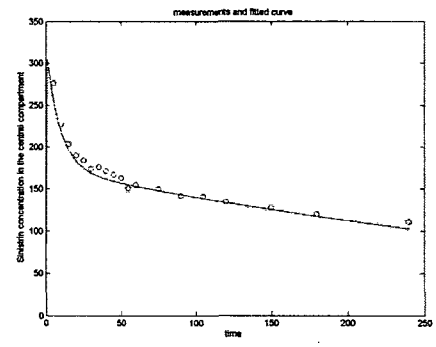


Figure 10.10: Plot of the fitted curve of a shortened protocol of 240 minutes

An even more drastic example is the following one:

Time	Concentration
5 min.	228 mg/l
10 min.	173 mg/l
15 min.	149 mg/l
20 min.	132 mg/l
25 min.	124 mg/l
30 min.	120 mg/l
35 min.	117 mg/l
40 min	112 mg/l
45 min.	110 mg/l
50 min.	106 mg/l
55 min.	104 mg/l
60 min.	102 mg/l
75 min.	99 mg/l
90 min.	92 mg/l
105 min.	94 mg/l
120 min.	99 mg/l
150 min.	100 mg/l
180 min.	98 mg/l
210 min	99 mg/l
240 min.	99 mg/l
290 min.	98 mg/l
350 min.	98 mg/l
410 min	96 mg/l
470 min.	100 mg/l

The different clearance values due to the different protocol lengths and the according errors can be seen in fig.10.11 and 10.12, respectively.

As can be seen the difference between a shortened and a sufficiently long protocol in this case is more than 1000 percent!

As the results show, the shorter the protocols lasted the higher the estimated clearance value was. This can also be seen in the plots where much higher end-concentrations in the central compartment were achieved when the protocols have been broken off early. For these plots see fig.10.13 to fig.10.20.

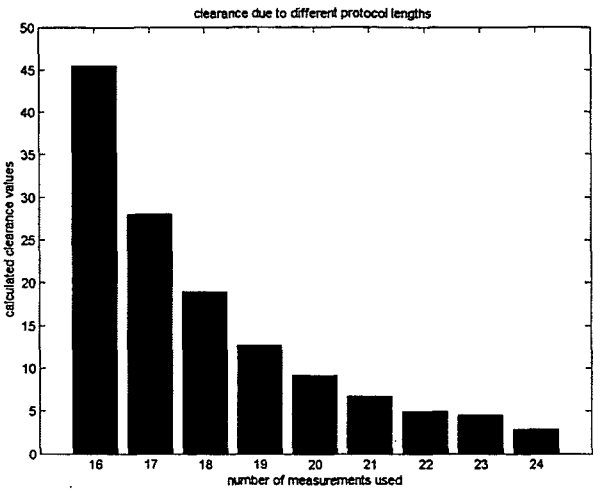


Figure 10.11: The different clearance values due to different protocol lengths

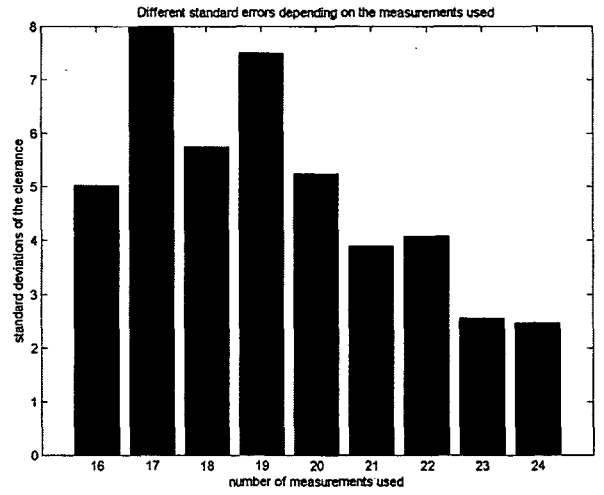


Figure 10.12: The according standard deviations to the clearance values of fig. 10.11

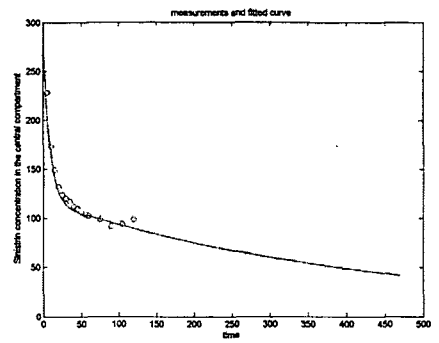


Figure 10.13: Plot of the fitted curve of a shortened protocol of 2 hours

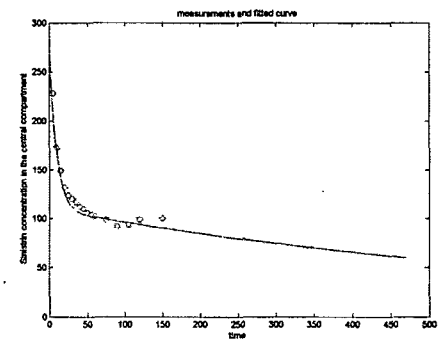


Figure 10.14: Plot of the fitted curve of a shortened protocol of 150 minutes

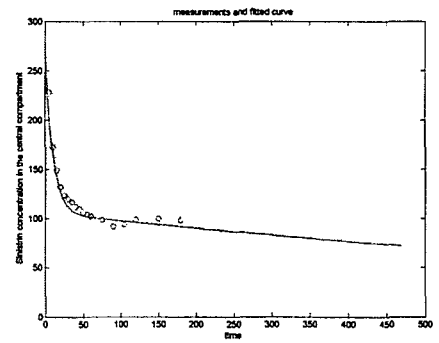


Figure 10.15: Plot of the fitted curve of a shortened protocol of 180 minutes

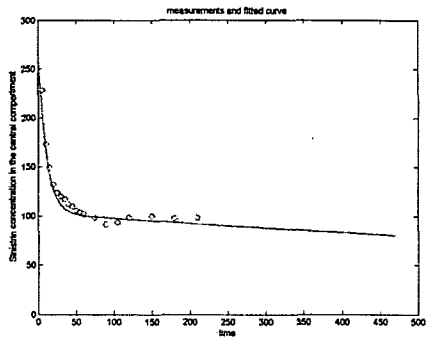


Figure 10.16: Plot of the fitted curve of a shortened protocol of 210 minutes

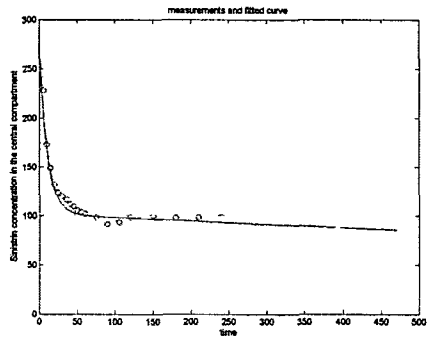


Figure 10.17: Plot of the fitted curve of a shortened protocol of 240 minutes

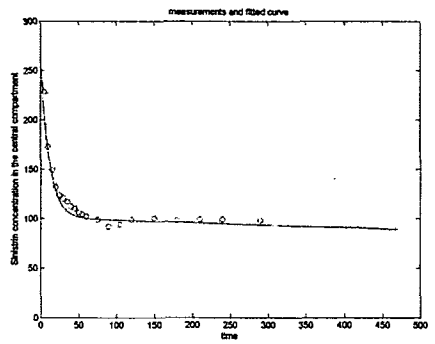


Figure 10.18: Plot of the fitted curve of a shortened protocol of 290 minutes

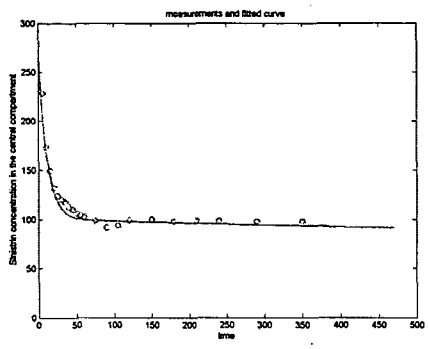


Figure 10.19: Plot of the fitted curve of a shortened protocol of 350 minutes

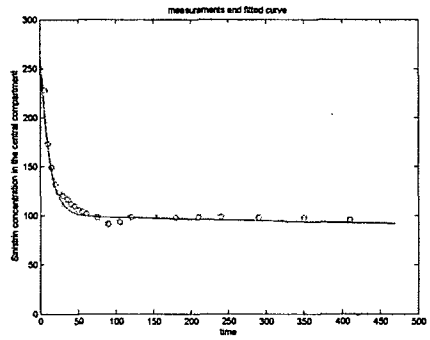


Figure 10.20: Plot of the fitted curve of a shortened protocol of 410 minutes

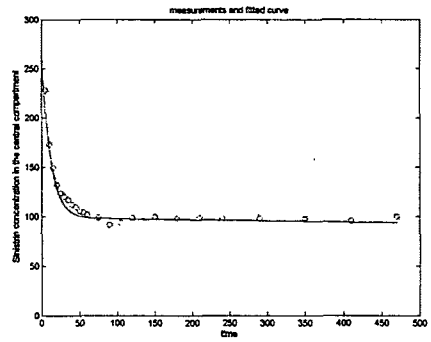


Figure 10.21: Plot of the fitted curve of a shortened protocol of 470 minutes

But where does this obvious need for long protocols come from?
One reason is that the markers injected do not consist of homogenous substances, but are a mixture of different particles with different masses and, therefore, different kinetics. Some particles take part in a more vivid exchange between blood and tissues and, therefore, reaching the state of equilibrium between the compartments more quickly than others. If there was only one substance injected, a shortened protocol would be sufficient as the time course of a single substance is predictable. But dealing with a mixture of different kinetics with one substance being eliminated much faster than other ones, the elimination process needs to be observed more thoroughly, which also includes a longer protocol.

But lets see for ourselves by means of the following example:
Using the measurements from our last example an identification procedure was performed that yielded our paramenter estimates. These estimates in turn were used to create artificial protocols in the following way:

Firstly, all parameters but k_{12} are left unchanged while the value of k_{12} is at first lowered and then increased by 0.017, therefore creating two sets of parameters:

k_{01}	k_{12}	k_{21}	V_1	V_2
$2.98 * 10^{-4}$	0.0157	0.0534	9.5176	32.3720

and

k_{01}	k_{12}	k_{21}	V_1	V_2
$2.98 * 10^{-4}$	0.0497	0.0534	9.5176	10.2262

These two parameters sets are used to create two artificial measurement series of which a third one can be derived by taking their arithmetic mean. Fig 10.22 shows the three curves along with the according artificial measurements, the red one being the mean of the two green ones.
The upper curve, hereby, represents a substance (part of the marker) which is slowly distributed while the lower curve stands for a rather rapid distribution. The middle curve then represents a mixture of different kinetics.

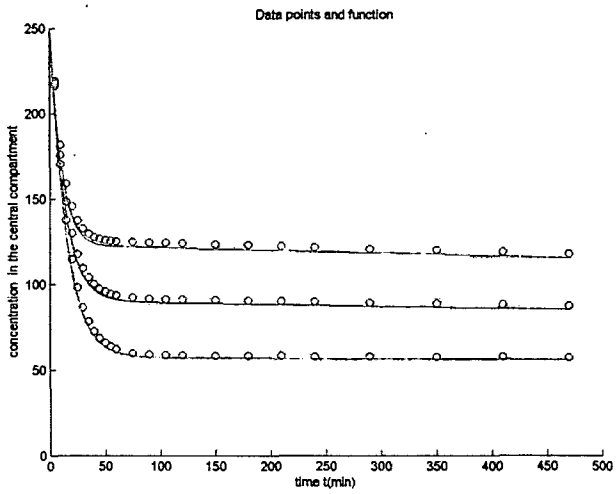


Figure 10.22: The artificially created data sets and the curves they lie on

Secondly, identification was performed twice on all three sets of data, first for a shortened protocol of 12 measurements (1 hour) and then for the full measurement series. As the results show, a significant dependence between the short time and full time protocol is achieved in case of the "mixed curve" while hardly any difference exists dealing with the other two.

	clearance for shortened protocol	clearance for full protocol
upper curve	1.7362	2.5745
lower curve	2.2614	2.8028
mixed curve	16.8199	3.4943

This experiment was repeated for a shortened protocol of 16 measurements (240 minutes) the results of which - although not as clear and drastic as for 12 measurements - show a similar outcome:

	clearance for shortened protocol	clearance for full protocol
upper curve	2.5833	2.5745
lower curve	2.1850	2.8038
mixed curve	6.4799	3.4943

As before, the results of the two generated curves lie within the same range, while the results of the mixed curve show a greater difference between the values of the shortened and the full protocol, the value of the short protocol being nearly as high as double the value of the long one.

The very same procedure was repeated for the parameter k_{21} with very similar results:

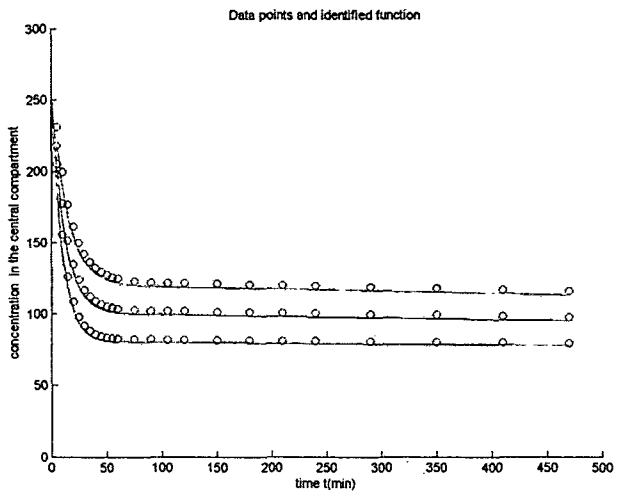


Figure 10.23: The artificially created data sets and the curves they lie on

	clearance for shortened protocol	clearance for full protocol
upper curve	2.4277	2.7419
lower curve	4.1432	2.7927
mixed curve	14.5190	3.1965

and the following values for the 2 hour to full length protocol comparison:

	clearance for shortened protocol	clearance for full protocol
upper curve	1.6056	2.7419
lower curve	2.4285	2.7927
mixed curve	5.4284	3.1965

This underlines the need for sufficiently long protocols as a marker mostly doesn't exist of a homogenous substance but of different ones with different behavior.

But its not only the different transfer rates of certain marker substances that make a relatively long duration of protocol necessary.
Also, many drugs are administered perorally rather than injected into a vein. While with an injection the drug directly enters the blood stream, this isn't the case dealing with peroral administration. The drug moves from the mouth trough the esophagus into the gastrointestinal tract from where it enters the blood stream moving to the liver (First Pass Effect [1]). In the liver the compounds of the drug may be altered by metabolism creating substances with different kinetics than the "original" drug.
What makes it even more difficult is that we cannot exactly say how and in how many metabolites a drug may be altered. Also, all metabolites behave differently not only than the "mother-substance" but also than all the other metabolites which makes predictions about the fate of the pharmacon very difficult.

The next huge problem is that metabolism not only is nearly unpredictable but also differs from person to person. While a substance in one person may not be altered that much and only a few metabolites are created, the very same drug can be handled very differently in a person with a higher rate of metabolism.
It's also in these cases that much better and accurate results should be achieved by choosing sufficiently long protocols.
Of course, this should be subject of further investigation and - hopefully - more research is done on this topic in the future.

10.1 The Implementation

10.1.1 Different protocol lenghts

This sections contains the code that was used to calculate the various clearance values and their according standard deviations due to different protocol lengths.

[1] See chapter 2.4.1

clearance

```
clear
opti;
%opti contains the measurements and the time points at which
% these measurements were taken
anzahl=8;     %number of the last measurements that are to be ignored

% At the beginning only a few measurements are used for calculating
% the optimized parameter values. Each time another measurement point
% is added, so that finally all of the measured data is used for evaluating
% the parameters.
close all     % closes all windows that are open at this time
tspan= [0 240];
x0= [0;0];

for i=anzahl:-1:0
   datashort=data(1:length(data)-i);
   timeshort=time(1:length(time)-i);
   identification
   figure
   [t x] =ode23(@kidneyfunction, tspan, x0,[],k01(anzahl+1-i),
   k12(anzahl+1-i),k21(anzahl+1-i));
   plot(t,x(:,1)/V1(anzahl+1-i),timeshort,datashort,'o')
   title('measurements and generated curve')
   xlabel('Zeit')
   ylabel('Sinistrinconcentration in the central compartment')
   montecarlo
end
figure
u=(length(data)-anzahl:1:length(data));
bar(u,clearance,'r')
title('Clearenceberechnungen mit unterschiedlich vielen Messdaten')
xlabel('Anzahl der verwendeten Messdaten')
ylabel('Berechnete Clearancewerte')
clearance
```

opti

```
% timepoints at which the measurements were made and the corresponding
% measurements
time = [5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 75, 90, 105, 120, 150,
180, 240];
```

```
data = [276, 227, 203, 190, 184, 174, 176, 171, 167, 163, 151, 155, 150,...
142, 141, 135, 128, 120, 111];
% data contains the measurements taken at the time points contained in time
```

identification

```
% kopt0 contains the initial parameter guesses
kopt0= [0.01 0.05 0.05 7.27];
x1=0;
x2=0;
% calling the Marquardtlevenbergalgorithm
options=optimset('LargeScale','off','Display','iter','LevenbergMarquardt','on',
'TolX',0.0001,'TolFun',0.0001);
[kopt, resnorm,residual] = lsqnonlin(@error, kopt0,[], [],options,x1,x2,datashort,timeshort);
k01(anzahl+1-i)=kopt(1);
k12(anzahl+1-i)=kopt(2);
k21(anzahl+1-i)=kopt(3);
V1(anzahl+1-i)=kopt(4);
V2(anzahl+1-i)=V1(anzahl+1-i)*k21(anzahl+1-i)/k12(anzahl+1-i);
residuennorm(anzahl+1-i)=resnorm;
residuen(anzahl+1-i,1:length(residual))=residual;
[t,x]=ode45(@kidneyfunction, tspan, x0,[],k01(anzahl+1-i), k12(anzahl+1-
i), k21(anzahl+1-i));
n=length(x(:,1));
u=x(:,1);
clearence(anzahl+1-i)=k01(anzahl+1-i)*V1(anzahl+1-i)*1000;
```

error

```
% Calculation of the error between the measured and calculated data
function f=error(kopt,x1,x2,data,time)
k01=kopt(1);
k12=kopt(2);
k21=kopt(3);
V1=kopt(4);
tspan= [0:1:240];
x0= [0;0];
[t,x] =ode45(@kidneyfunction, tspan, x0,[],k01, k12, k21);
for i=1:length(time)
   wert(i)=x(time(i),1)/V1;
end
f=wert-data;
```

kidneyfunction

```
% kidneyfunction contains the models differential equations
function neu=kidneyfunction(t,x, k01, k12, k21)
neu=[infusion(t)+k12*x(2) - (k01+k21)*x(1);k21*x(1) - k12*x(2)];
```

infusion

```
function inf=infusion(t)
D=2500;
   % overall dose administered n=length(t);
tau=0.5;
   % duration of the injection for i=1:n
   if t¡tau
     inf(i)=D/tau;
   else
     inf(i)=0;
   end
end
```

init

```
% parameter values
k01=0.0041;
k12=0.0585;
k21=0.0498;
V1=7.3; V2=6.18;
tau=0.5;
```

montecarlo

```
clear xwert;
dataneu=datashort;
% not all of the measured data is used for determining the models
% parameter values, the shorter data-array is called datashort, while
% dataneu will be the disturbed data-array, but is the same in the
% beginning
awert=datashort;
% data are disturbed a 100 times
for k=1:10
```

```
    marquardtlevenberg;
    mck01(k)=k01;
    mck12(k)=k12;
    mck21(k)=k21;
    mcV1(k)=V1;
    mcV2(k)=V1*k21/k12;
    mcClearance(k)=k01*V1*1000;

    resid=sum(fehler(kopt,x1,x2,datashort,timeshort).*fehler(kopt,x1,x2,datashort,timeshort));
    % solve the differential equation with the new set of parameters
    tspan= [0:1:240];
    x0= [0;0];
    [t,x] =ode45(@kidneyfunction, tspan, x0,[],mck01(k), mck12(k), mck21(k));

    for j=1:length(timeshort)
      xwert(j)=x(timeshort(j),1)/V1;
    end
    error=xwert-dataneu;
    r = normrnd(0,sqrt(resid/(length(data)-4)),[1,length(datashort)]);

      if k==1 % the inital function values need to be remembered
      for it is always the array awert that is disturbed a 100 times
      awert=xwert;
    end

      dataneu = awert + r;
end
% calculation of the mean values and standard deviations of the various
% parameters:
meank01(anzahl+1-i)=mean(mck01);
meank12(anzahl+1-i)=mean(mck12);
meank21(anzahl+1-i)=mean(mck21);
meanV1(anzahl+1-i)=mean(mcV1);
meanV2(anzahl+1-i)=mean(mcV2);
meanClearance(anzahl+1-i)=mean(mcClearance);
stdk01(anzahl+1-i)=std(mck01);
stdk12(anzahl+1-i)=std(mck12);
stdk21(anzahl+1-i)=std(mck21);
stdV1(anzahl+1-i)=std(mcV1);
stdV2(anzahl+1-i)=std(mcV2);
stdClearance(anzahl+1-i)=std(mcClearance)
```

marquardtlevenberg

```
% kop0 contains all the initial guesses for the parameter values
kopt0= [0.005 0.05 0.05 7.27];
D=2500;     %overall dose administered by injection
x1=0;
x2=0;
% calling the Marquardtlevenbergalgorithm
options=optimset('LargeScale','off','Display','off','LevenbergMarquardt','on',
'TolX',0.0001,'TolFun',0.0001);
[kopt, resnorm,residual] = lsqnonlin(@error, kopt0,[], [],options,x1,x2,dataneu,timeshort);
k01=kopt(1);
k12=kopt(2);
k21=kopt(3);
V1=kopt(4);
```

10.1.2 Protocol Problem

The following code was used for our last experiments dealing with different marker kinetics, proving the need for long(er) protocols in order to arrive at reliable results.

protocol

```
clear
opti2
x0= [0;0];
taskb
remember=k21;
hold on;
tspan= [0:1:470];
timeneu=time(1:12);
k21=remember+0.017;
[t x]=ode23(@nierenfunktion, tspan, x0,[],k01,k12,k21);
dataneu=(x(time,1)/V1).'
k21=remember-0.017;
[t x]=ode23(@nierenfunktion, tspan, x0,[],k01,k12,k21);
datanew=(x(time,1)/V1).'
datamix=(dataneu+datanew)/2  %generating the mean of the two curves

dataneushort=dataneu(1:12);  %shortened 'measurement' string
figure
hold on
identi
```

```
parcomp(1,:)=[k01, k12, k21, V1];  %identification using all them measure-
ments
[t x]=ode23(@nierenfunktion, tspan, x0,[],k01,k12,k21);
plot(t,x(:,1)/V1,'g',time,dataneu,'o')
title('Data points and function')
xlabel('time t(min)')
ylabel('concentration in the central compartment')
clearance(1,1)=k01*V1*1000;
identi2
parcomp(2,:)=[k01, k12, k21, V1]; %identification using only the first 10
values
clearance(1,2)=k01*V1*1000;
[t x]=ode23(@nierenfunktion, tspan, x0,[],k01,k12,k21);
plot(t,x(:,1)/V1,'g',time,dataneu,'o')
tspan= [0:1:470];

dataneu=datanew;
dataneushort=datanew(1:12);
identi
parcomp(3,:)=[k01, k12, k21, V1]; %identification of the 2nd curve
clearance(1,3)=k01*V1*1000;
[t x]=ode23(@nierenfunktion, tspan, x0,[],k01,k12,k21);
plot(t,x(:,1)/V1,'g',time,dataneu,'o')
identi2
parcomp(4,:)=[k01, k12, k21, V1];
clearance(1,4)=k01*V1*1000;
[t x]=ode23(@nierenfunktion, tspan, x0,[],k01,k12,k21);
plot(t,x(:,1)/V1,'g',time,dataneu,'o')

dataneu=datamix;
dataneushort=dataneu(1:12);
identi
parcomp(5,:)=[k01, k12, k21, V1]; %identification of the mixed data
clearance(1,5)=k01*V1*1000;
[t x]=ode23(@nierenfunktion, tspan, x0,[],k01,k12,k21);
plot(t,x(:,1)/V1,'r',time,dataneu,'o')
identi2
parcomp(6,:)=[k01, k12, k21, V1];
clearance(1,6)=k01*V1*1000;
[t x]=ode23(@nierenfunktion, tspan, x0,[],k01,k12,k21);
plot(t,x(:,1)/V1,'r',time,dataneu,'o')
parcomp
clearance
```

identi

```
%identi calculates the parameter values so that the resulting solution best
%fits the measured data and compairs them by plotting both, the optimized
%curve as well as the measurements
kopt0= [0.005 0.05 0.05 7.27];
%kopt0 contains the initial guesses of the parametervalues
tspan= [0:1:470];
x0= [0;0];
x1=0;
x2=0;
%Calling the Marquardtlevenbergalgorithm
options=optimset('LargeScale','off','Display','iter','LevenbergMarquardt','on',
'TolX',0.0001,'TolFun',0.0001);
[kopt, resnorm,residual] = lsqnonlin(@fehler, kopt0,[], [],options,x1,x2,dataneu,time);
%the returned array contains the optimized parameter values
k01=kopt(1)
k12=kopt(2)
k21=kopt(3)
V1=kopt(4)
V2=V1*k21/k12
resnorm
residual;
clearence=k01*V1*1000
```

identi2

```
%identi2 calculates the parameter values so that the resulting solution best
%fits the measured data and compairs them by plotting both, the optimized
%curve as well as the measurements
kopt0= [0.005 0.05 0.05 7.27];
%kopt0 contains the initial guesses of the parametervalues
tspan= [0:1:timeneu(12)];
x0= [0;0];
x1=0;
x2=0;
%Calling the Marquardtlevenbergalgorithm
options=optimset('LargeScale','off','Display','iter','LevenbergMarquardt','on',
'TolX',0.0001,'TolFun',0.0001);
[kopt, resnorm,residual] = lsqnonlin(@fehler, kopt0,[], [],options,x1,x2,dataneushort,timeneu);
%the returned array contains the optimized parameter values
k01=kopt(1)
k12=kopt(2)
```

```
k21=kopt(3)
V1=kopt(4)
V2=V1*k21/k12 resnorm
residual;
clearence=k01*V1*1000
```

Appendix A

Definitions and Glossary

Absorption of drugs is the process of uptake of the compound from the site of administration into the systemic circulation. A prerequisite for absorption is that the drug be in aqueous solution. The only relatively rare exception is absorption by pinocytosis.

Accumulation is the increase of drug concentration in blood and tissue upon multiple dosing until steady state is reached.

Afferent means conveying toward a center.

ATPase adenosinetriphosphatase; an enzyme that catalyzes the splitting of adenosine triphosphate, with liberation of inorganic phosphate.

Basal means pertaining to or situated near a base.

Blood-, Plasma-, or Serum-Levels demonstrate the concentration in blood, plasma or serum upon administration of a dosage form by various routes of administration. Blood-, plasma- or serum-level curves are plots of drug concentration versus time on numeric or semi-log graph paper.

Central Compartment is the sum of all body regions (organs and tissues) in which the drug concentration is in instantaneous equilibrium with that in blood or plasma. The blood or plasma is always part of the central compartment.

Chloride is a salt of hydrochloric acid; any binary compound of chlorine in which the latter is the negative element.

Chlorine a yellowish green, gaseous element of suffocating odor; symbol Cl; atomic weight 35.453. L: chlorum. It is disinfectant, dexolorant, and an

irritant poison. It is used for disinfecting, fumigating and bleaching.

Clearence is the hypothetical volume of distribution in ml of the unmetabolized drug which is cleared per unit of time (ml/min or ml/h) by any pathway of drug removal (renal, hepatic and other pathways of elimination).

A Compartment in pharmacokinetics is an entity which can be described by a definite volume and a concentration of drug contained in that volume. In pharmacokinetics, experimental data are explained by fitting them to compartment models.

Concentration Gradient is the difference in the concentration in two phases usually separated by a membrane.

Contractile means having the power or tendency to contract in response to a suitable stimulus.

Cortex is the outer layer of an organ or another body structure, as distinguished from the internal substance, or an external layer as the bark of a tree or the rind of a fruit.

Cortical means pertaining to or of the nature of a cortex or bark.

Disposition is the loss of drug from the central compartment due to transfer (distribution) into other compartments and/or elimination and metabolism.

Distal: farther from any point of reference; opposed to proximal.

Diuresis is the increased secretion of urine.

Dose size is the amount of drug in mcg (μg), mg, units or other dimensions to be administered.

A Drug is a chemical compound of synthetic, semisynthetic, natural or biological origin which interacts with human or animal cells. The interactions may be quantified, whereby these resulting actions are intended to prevent, to cure or to reduce ill effects in the human or animal body, or to detect disease-causing manifestations.

Drug Release or **Liberation** is the delivery of the active ingredient from a dosage form into solution. The dissolution medium is either a biological fluid or an artificial test fluid (in vitro). Drug release is characterized by the speed (liberation rate constant) and the amount of drug appearing in solution.

Drug-Receptor Interaction is the combining of a drug molecule with the receptor for which it has affinity, and the initiation of a pharacologic response by its intrinsic activity.

Efferent means conveying away from a center, centrifugal.

Elimination Half-Life of a drug is the time in hours necessary to reduce the drug concentration in the blood, plasma or serum to one-half after equilibrium is reached. The elimination half-life may be influenced by: dose size, variation in urinary excretion (pH), intersubject variation, age, protein binding, other drugs and diseases (especially renal and liver diseases).
Loss of drug from the body, as described by the elimination half-life, means the elimination of the administered parent drug molecule (not its metabolites) by urinary excretion, metabolism or other pathways of elimination (lung, skin, etc.)

Endothelium is the layer of epithelial cells that lines the cavities of the heart and of the blood and lymph vessels, and the serous cavities of the body, originating from the mesoderm. **Epithelium** is the covering of internal and external surfaces of the body, including the lining of vessels and other small cavities. It consists of cells joined by small amounts of cementing substances. Epithelium is classified into types on the basis of the number of layers deep and the shape of the superficial cells.

Excretion of drugs is the final elimination from the bodys systemic circulation via the kidney into urine, via bile and saliva into intestines and into feces, via sweat, via skin and via milk.

Extravascular Administration refers to all routes of administration except those where the drug is directly introduced into the blood stream. An example for an extravascular route would be oral or rectal administration.

Hematocrit is the volume percentage of erythrocytes in the whole blood. The name hematocrit was originally applied to the apparatus or procedure used in its determination, but today it is also (or primarily) used to designate the result of the determination. Abbreviation: htc.

Hepatic Clearence is the hypothetical volume of distribution in ml of the unmetabolized drug which is cleared in one minute via the liver.

Homeostasis is the maintenance of a steady state which characterizes the internal environment of the healthy organism. An important function of homeostasis is the regulation of the fluid medium and volume of the cell.

Hypertonic is a biological term denoting a solution which when bathing body cells causes a net flow of water across the semipermeable cell membrane out of the cell.

Hypothalamus

Hypophyse

Hypotonic is a biological term denoting a solution which when bathing body cells, causes a net flow of water across the semipermeable cell membrane into the cell.

Intravascular Administration refers to all routes of administration where the drug is directly introduced into the blood stream.

Intrinsic Clearence is the theoretical unrestricted maximum clearence of unbound drug by an elimination organ.

Isotonic is a biological term denoting a solution in which body cells can be bathed without a net flow of water across the semipermeable cell membrane.

The LADMER-System deals with the complex dynamic processes of liberation of an active ingredient from the dosage form, its absorption into systemic circulation, its distribution and metabolism in the body, the excretion of the drug from the body and the achievement of response.

Lag Time is the period of time which elapses between the time of administration and the time a measurable drug concentration is found in blood.

Lamina is a thin flat plate or layer; a general term for such a structure or a layer of a composite structure.

Lumen is the cavity or channel within a tube or tubular organ.

Medulla is the inmost part; a general term for the inmost portion of an organ or structure.

Metabolism is the sum of all the physical and chemical processes by which living substance is produced and maintained (anabolism), and also the transformation by which energy is made available for the uses of the organism (catabolism).

Metabolite is any substance produced by metabolism or by a metabolic

process.

Michaelis Menten Kinetics uses equations to characterize certain phenomena such as protein binding, adsorption, and nonlinear or saturation processes often observed with increasing dose sizes.

Mitochondria are small spherical to rod-shaped components found in the cytoplasm of cells, enclosed in a double membrane. They are the principle sites of the generation of energy (in the form of ion gradients and adenosine triphosphate [ATP] synthesis) resulting from the oxidation of foodstuffs. Mitochondria are organelles with genetic continuity and contain an extracellular source of DNA.

Nonlinear Kinetics or **Saturation Kinetics** refers to a change of one or more of the pharmacokinetic parameters during absorption, distribution, metabolism and excretion by saturation or overloading of processes due to increased dose sizes.

Osmolality and **Osmolarity** The osmolal concentration of solution is called osmolality when the concentration is expressed as osmoles per **kilogram** of water; it is called osmolarity when it is expressed as osmoles per **liter** of solution. In dilute solutions such as the body fluids, these two terms can be used almost synonymously because the differences are tiny.

Osmoles The total number of particles in a solution is measured in terms of osmoles. One osmole (osm) is equal to 1 mole (mol) $(6.02 * 10^{23})$ of solute particles. Therefore, a solution containing 1 mole of glucose in each liter has a concentration of 1 osm/L.

Osmosis is the passage of pure solvent from a solution of lesser to one of greater solute concentration when the two solutions are separated by a membrane which selectively prevents the passage of solute molecules, but is permeable to the solvent.

Peripheral Compartment is the sum of all body regions (i.e., organs, tissues or part of it) to which a drug eventually distributes, but is not in instantaneous equilibrium. The peripheral compartment is sometimes further subdivided into a **shallow** and a **deep** compartment.

Pharmacokinetics deals with the changes of drug concentration in the drug product and changes of concentration of a drug and/or its metabolite(s) in the human or animal body following administration, i.e., the changes of drug concentration in the different body fluids and tissues in the dynamic system of liberation, absorption, distribution, body storage, binding,

metabolism and excretion.

Potassium is a metallic element of the alkali group, many of whose salts are used in medicine. It is a soft, silver-white metal, melting at $58°F$; anatomic number: 19, anatomic weigth: 39.102. L: *kalium*

Protein Binding is the phenomenon which occurs when a drug combines with plasma protein (particularly albumin) or tissue protein to form a reversible complex. Protein binding is usually nonspecific and depends on the drug's affinity to the protein molecule, the number of protein binding sites. protein and drug concentration. Drugs can be displaced from protein binding by other compounds having higher affinity for the binding sites.

Proximal: nearest; closer to any point of reference; opposed to distal.

A Receptor is a site in the biophase to which drug molecules can be bound. A receptor (= substrate) is usually a protein or proteinaceous material.

Renal Clearence is the hypothetical plasma volume in ml (volume of distribution) which is cleared of the unmetabolized drug in one minute via the kidney.

Sodium is a soft, silver white, alkaline metallic element; symbol: Na; atomic number: 11; atomic weight: 22.990. With a valence of 1, it has a strong affinity for oxygen and other nonmetallic elements. Sodium provides the chief cation of the extracellular body fluids. It is the most widely used salt in medicines. L: *natrium.*

Steady State is a level of drug accumulation in blood and tissue upon multiple dosing when input and output are at equilibrium. The steady state drug concentrations fluctuate (oscillate) between a maximum and a minimum steady state concentration within each dosing intervall.

Stellate: shaped like a star.

Total Clearence describes the clearence of the hypothetical plasma volume in ml (volume of distribution) of a drug per unit time due to excretion via kidney, liver, lung, skin, etc. and metabolism.

Urea is the diamide of carbonic acid, $NH_2 - CO - NH_2$, a white, crystallizable substance found in the urine, blood, and lymph. It is the chief nitrogenous constituent of the urine, and the chief nitrogenous end-product of the metabolism of proteins. It is formed in the liver from amino acids,CO and from compounds of ammonia.

Urinary Recycling is the phenomenon that occurs when drugs filtered through the glomeruli are reabsorbed from the tubuli into systemic circulation.

Vasoconstriction is the diminuition of the caliber of vessels, especially constriction of arterioles leading to decreased blood flow to a part.

Vasodilation means dilation of a vessel, especially dilation of arterioles leading to in creased blood flow to a part.

Volume of Distribution is not a "real" volume but an artifact - a hypothetical volume of body fluid that would be required to dissolve the total amount of drug at the same concentration as that found in blood. It is a proportionality constant relating the amount of drug in the body to the measured concentration in biological fluid (blood, plasma or serum).

List of Figures

Appendix B

References

[1] A.C. GUYTON, J.E. HALL: *Textbook of medical pysiology - 10th ed..* W.B. Saunders Company, 2000; ISBN 0-7216-8677-X.

[2] R. KLINKE, S. SILBERNAGl: *Lehrbuch der Physiologie - 3.Auflage.* Thieme, 2001; ISBN 3-13-796003-7.

[3] J.P.FRIEL: *Dorland's Illustrated Medical Dictionary - 25th ed..* W.B. Saunders Company, 1974; ISBN 0-7216-3148-7.

[4] H.P. KOCH, W.A. RITSCHEL: *Synopsis der Biopharmazie und Pharmakokinetik.* Ecomed, 1986; ISBN 3-609-64970-4.

[4] H.R. SCHWARZ: *Numerische Mathematik.* B.G. Teubner Stuttgart, 1986; ISBN 3-519-02960-X.

[5] J. STOER: *Numerische Mathematik 1 - 6. Auflage.* Springer-Verlag Berlin Heidelberg, 1993; ISBN 3-540-56213-3.

[6] P. VALKO, S. VAJDA: *Advanced scientific computing in BASIC with applications in chemistry, biology and pharmacology.* Elsevier Science Publishers B.V., 1989; ISBN 0-444-87270-1.

[7] E.R. CARSON, C. COBELLI, L. FINKELSTEIN: *The mathematical modeling of metabolic and endocrine systems.* John Wiley & Sons, Inc, 1983; ISBN 0-471-08660-6.

[8] W. A. RITSCHEL: *Handbook of Basic Pharmacokinetics - 3d ed..* Hamilton Press, Inc. 1986; ISBN 0-914768-43-3.

[9] J.vE. A. MCINTOSH: *Mathematical Modeling and Computers in Endocrinology.* Springer-Verlag Berlin; ISBN 0-387-09693-0.

[10] R. TRAPPL: *Cybernetics and Systems Research - vol.2.* World Scientific 1992; ISBN 981-02-1992-X.

[11] K. N. S. SUBRAMANIAN: *Kinetic Models of Trace Element and Mineral Metabolism during Development.* CRC Press, Inc. 1995; ISBN 0-8493-4793-X.

[12] K, NORWICH: *Molecular Dynamics in Biosystems* Pergamon Press 1977; ISBN 0-08-020420-I.

[13] T. N. BRYANT: *Computers in Microbiology - a practical approach.* Irl Press 1989; ISBN 1-85221-086-9.

[14] A. SALTELLI, K. CHAN, E. M. SCOTT: *Sensitivity Analysis* J. Wiley and Sons, Inc. 2001; ISBN 0-471-99892-3

[15] J. E. GENTLE: *Random Number Generation and Monte Carlo Methods - 2nd ed.* Springer-Verlag 2003; ISBN 0-387-00178-6.

[16] F. KAIJA, OKAYAMA: *Compartmental Analysis - Medical Applications and Theoretical Background* S. Karger AG; ISBN 3-8055-3696-8.

[17] O. RICHTER, D. SOENDGERATH: *Parameter Estimation in Ecology* VCH 1990; ISBN 3-527-27954-7.

[18] A. ANGERMANN, M. BEUTSCHEL: *Matlab - Simulink - Stateflow - 2nd ed.* Oldenbourg Verlag 2003; ISBN 3-486-27377-9.

[19] E. R. CARSON, K. R. GODFREY, J. REEVE: *A Review of Modelling and the Role of Dynamic Tracer Studies in Metabolism* In: Quantitative Approaches to Metabolism, J. Wiley and Sons, 1982.

[20] W. ESTELBERGER, W. PETEK, H. POGGLITSCH: *Model-based Determination of Renal Clearance from Temporal Venous Plasma Profils of Markers* In: Cybern.Syst.Res. Vol 2; 1992: Apr. 21-24; Vienna, Singapore: World Scientific 1992.

[21] W. ESTELBERGER, W. PETEK, H. POGGLITSCH: *Simulation der sttigbaren und hemmbaren Kinetik renal-tubulr eliminierter Pharmaka* In: Biomedizinische Technik; 1992 September 17-19; Graz.Berlin: Schiele und Schn, 1992.

[22] W. ESTELBERGER, W. PETEK, S. ZITTE, A, MAURIC, S. HORN, H. HOLZER, H. POGGLITSCH: *Determination of the Glomerular Filtration Rate by Identification of Sinistrin Kinetics* In: Eur. J. Clin. Chem. Clin. Biochem. 1995; 33:20-209.

[23] W. ESTELBERGER, W. PETEK, S. ZITTE, A, MAURIC, H. POGGLITSCH: *Kinetic Clearance Determination* 10th IFCC European Congress of Clinical Chemistry 1993 April 25-29; Niece. Ann. Biol. Clin. 1993.

[24] W. ESTELBERGER, S. ZITTA, T. LANG, F. MAYER, A. MAURIC, S. HORN, H. HOLZER, W. PETEK, G. REIBNEGGER: *System Identification of the Low-Dose-Kinetic of p-Aminohippuric Acid* In: Eur. J. Clin. Chem. Clin. Biochem. 1995; 33:847-853.

[25] J. T. FINKENSTAEDT, M. P. O'MEARA, J. P. MERRIL: *Observations on the Volume of Distribution of Inulin in Anuric Subjects* J. Clin. Invest. 1953.

[26] W. A. KNORRE: *Pharmakokinetik* Braunschweig/Wiesbaden: Vieweg 1981.

[27] E. GLADTKE VON HATTINGBERG H.M.: *Pharmakokinetik* Berlin-Heidelberg-New York: Springer-Verlag.

[28] E. KREYSZIG: *Advanced Engineering Mathematics* New York: Wiley 1993.

[29] K. THOMASEK: *Robust Sampling Design for Assessing Kinetic Parameters of Clinical Relevance: A Case Study on Kinetic Function Test in Modeling and Control in Biomedical Systems* In: Patterson BW,editor. Proc. IFAC Symp; 1994 March 27-30; Galveston, Texas. Madison: Omnipress 1994.

[30] J. BUYS K. VON GADOW: *A Pascal Program for Fitting Nonlinear Regression Models on a Microcomputer* In: EDV in Medizin und Biologie 1987; 4.

[31] J. P. BOSCH, S. LEW, D. GLABMAN, A. LAUER: *Renal Hemodynamic Changes in Humans* In: American Journal of Medicine 81, 1986.

[32] J. NELDER, R, MEAD: *A Simple Method for Function Minimization* In: Computer Journal 4, 1965.

[[33] C. M. METZLER: *Statistical Properties of Estimates of Kinetic Parameters* In: Pharmacokinetics during Drug Development: Data Analysis and Evaluation Techniques. Fischer, Stuttgart, 1982.

[34] G. MASCHIO, L. OLDRIZZI, C. RUGIU, V. DE BIASE: *Dynamic Evaluation of Renal Function: A Chimera for Nephrologists?* In: Journal of Nephrology 4, 1989.

[35] T. THEORELL: *Kinetics and Distribution of Substances administered to the Body ——. The Intravasculat Modes of Administration* In: Archives of Internal Pharmacodynamics, 57:226.

[36] K. B. BISCHOFF, R. L. DEDRICK: *Generalized Solution to Linear, Two-Compartment, Open Model for Drug Distribution* In: Journal of Theor. Biol. 1970.

[37] C. BIANCHI, C. DONADIO, G. TRAMONTI: *Noninvasive Methods for the Measurements of Total Renal Function* Nephron 1981; 28:53-7.

[38] H. BAUER: *Probability Theory* Berlin, New York: de Gruyter 1996.

[39] R. STORM: *Wahrscheinlichkeitsrechnung, Mathematische Statistik und Statistische Qualittskontrolle* Fachbuchverlag Leipzig-Kln 1995.

[40] DESIMPLOPOLUS

[41] GANONG

[42] G. BAURA: *System Theory and Practical Applications of Biomedical Signals* Wiley-Interscience 2002.

[43] S. ZITTA, K.STOSCHITZKY, R. ZWEIGER, T. LANG, H. HOLZER, F. MAYER, G. REIBNEGGER, W. ESTELBERGER: *Determination of Renal Reserve Capacity by Identification of Kinetic Systems* In: Mathematical and Computer Modelling of Dynamical Systems, Swets and Zeitlinger, 1999.

[44] W. ESTELBERGER, S. ZITTA, T. LANG, F. MAYER, A. MAURIC, S. HORN, H. HOLZER, W. PETEK, G. REIBENEGGER: *System Identification of the Low-Dose Kinetics of p-Aminohippuric Acid* In: Eur J Clin Chem Clin Biochem 1995; 33:847-853. Berlin - New York: Walter de Gruyter and Co, 1995.