THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN APPLIED MATHEMATICS AND MATHEMATICAL STATISTICS

Mathematical aspects of apolipoprotein kinetics, with focus on metabolic diseases

Valentina Fermanelli

CHALMERS

Department of Mathematical Sciences
Chalmers University of Technology
and the University of Gothenburg
Göteborg, Sweden 2020

Mathematical aspects of apolipoprotein kinetics, with focus on metabolic diseases
Valentina Fermanelli
Göteborg 2020

ISBN 978-91-7905-327-7
Doktorsavhandlingar vid Chalmers tekniska högskola, Ny serie nr 4794
ISSN 0346-718X

Department of Mathematical Sciences
Chalmers University of Technology
and the University of Gothenburg
SE-412 96 Göteborg
Sweden
Telephone: +46 (0)31 772 10 00
E-mail: valfer@chalmers.se

Printed by Chalmers Reproservice,
Gothenburg, Sweden, 2020.

Abstract

Biomathematics is a branch of science that aims at describing biological processes in mathematical terms to frame and solve otherwise unsolvable research questions in the biological and medical field.
Cardiovascular diseases are the number one cause of death in the world.
Metabolic diseases as metabolic syndrome and diabetes increase the risk of cardiovascular diseases, due to disrupted lipid metabolism. Apolipoproteins (apo) are particles attached to the lipid-carrying proteins (lipoproteins) and give them properties. ApoC-III inhibits lipoprotein lipase, therefore lowering down the release of the triglyceride from the lipoproteins to the tissues, apoE is a ligand involved in the uptake of lipoproteins from the liver and apoA4 influences insulin secretion. Studying how fast the apolipoproteins are formed and released in the blood (secretion rate (SR)) and how fast they are removed from the blood (fractional catabolic rate (FCR)) enriches our knowledge on lipid metabolism.

Apolipoprotein kinetics studies are only possible thanks to mathematical modelling. So far the three apolipoproteins had not being measured in the same study. We analyse the time series data generated from three experiments with the nonlinear mixed effects modelling framework. The novelty consists in having validated a structural model for all the apolipoproteins across three studies and having chosen statistical error model. A covariance-variance matrix for the random effects has been designed and it has been validated in 16 out of the 18 total occasions (three apolipoproteins with 2 occasions for 3 experiments).

Applications of the mathematical framework to the kinetic studies has led to advancements in the realm of lipid metabolism.
In Paper I and II apoC-III kinetics is studied before and after hypercaloric fructose treatment in abdominally-obese individuals. In Paper I the modelling framework developed in this thesis is presented and further applied to apoA4 and apoE kinetics; a strong bond has been uncovered between triglyceride levels and the SR and PS of apoC-III and an association has been found between apoC-III and apoE kinetic parameters. In Paper II the results for apoC-III are combined with lipoprotein kinetics analysis. Hypercaloric fructose intake leads to an increase in triglyceride levels, but the mechanism behind this phenomenon was so far unknown; in this work, results support the hypothesis that the increase in apoC-III SR causes a rise in the apoC-III concentration with subsequent increase in triglyceride level. In Paper III apoC-III and apoE kinetics are analysed before and after a PSK9-inhibitor-based drug treatment in type-II-diabetic individuals. ApoE FCR increases consistently and apoE diminishes as a result of the treatment. Different elements suggest that the increase in apoE FCR might be related to an increase in VLDL2 FCR.

Keywords:
mathematical modelling, NLMEM, apolipoproteins, apoA4, apoC-III, apoE, type II diabetes, abdominal obesity

List of appended papers

This thesis is based on the work represented by the following manuscripts:

Paper I: **Fermanelli, V.**, Björnson, E., Andersson, L., Sihlbom, C., Thorsell, A., Matikainen, N., Taskinen, M-R., Borén, J., Adiels, M. (2020). Combined analysis of apolipoprotein C-III, A4 and E in obese individuals using proteomics and population kinetics: Effect of fructose intervention. *Manuscript*.

Paper II: Björnson, E., **Fermanelli, V.**, Andersson, L., Adiels, M., Sihlbom, C., Thorsell, A., Matikainen, N., Taskinen, M-R, Borén, J. (2020). Worsening of lipid profile after fructose treatment is driven by increased apoC-III secretion rate. *Manuscript*.

Paper III: **Fermanelli, V.**, Adiels M., Taskinen M.-R. and Borén J.(2020). Inhibition of PCSK9 by evolocumab increases the clearance of apolipoprotein E. *Manuscript*.

Author contributions

Paper I: The model choice, data analysis, implementation of calculations and major part of the writing are mainly attributable to the work of the author.

Paper II: The model choice, data analysis, implementation of calculations for the kinetic analysis of apolipoprotein C-III are mainly attributable to the work of the author. The author contributed to the revision of the manuscript.

Paper III: The model choice, data analysis, implementation of calculations and major part of the writing are mainly attributable to the work of the author.

Paper not included in the thesis

Kawamura T, Kasai H, **Fermanelli V**, Takahashi T, Sakata Y, Matsuoka T, Ishii M, Tanigawara Y. (September 2018) Pharmacodynamic analysis of eribulin safety in breast cancer patients using real-world postmarketing surveillance data, Cancer Science.

Acknowledgements

First and foremost I would like to express my deepest gratitude to my Academic father, **Martin Adiels**. Thank you for your patience, your enthusiasm and all the support you have given me.
Gratitude goes also to Bernt Wennberg and Torbjörn Lundh for their support throughout my Academic development.
I feel indebted to you Marija Cvijovic for our stimulating conversations that have encouraged me throughout my last months of my studies.
Thank you Joachim Almquist, Jonas Hagmar and Helga Kristin Olafsdottir for the scientific discussions and thank you Mats Jirstrand, Philip Gerlee and Adam Malik for the nice company in the lunches of our research group.
I am really grateful to SSF for funding the research project of which I have been part of and to JSPS, STINT and the European Union for funding my two research stays in Japan. Heartfelt gratitude goes to Tanigawara sensei, Nakaoka sensei and Aihara sensei.
I deeply appreciate the opportunity given to us of developing not only as a researcher, but also as a person, therefore I am grateful to Chalmers for offering the GTS courses. Thank you Lyudmyla Turowska and Hjalmar Rosengren for your human skills during the medarbetarsamtal. Thank you Aila Särkkä for being so inspirational with all your hard work for your PhD students and your devotion to helping all the PhD students.
Thank you to all the administrative staff, especially Marie Kühn, Liselotte Fernström, Elisabeth Eriksson and Loredana Colque. I would have not enjoyed my years at Chalmers as much if I wouldn't have had the privilege of getting to know so many wonderful people: Peter Helgesson, Dawan Dara, Hossein Raufi, Roza Maghsood, Malin Palö Forsström, Ivar Simonsson, Anders Hildeman, Claes Andersson, Matteo Molteni, Henrik Imberg, Juan Salvador Inda Diaz, Sunney Fotedar, Anton Johansson, Carl Lundholm, Olof Giselsson, Caroline Granfeldt, Oskar Allerbo, Niek Welkenhuysen, Jonatan Kallus, Anna Johnning, Tobias Magnusson, Edvin Wedin, Elizabeth Wulcan, Anders Södergren, Dennis Eriksson, Martin Raum, Laura Fainsilber, Olof Elias, Carl-Joar Karlsson, Farnaz Ghorbanpour, Joao Pedro Paulos,Philip Townsend, Frida Zachrisson, Lashi Bandara, Jeanette Montell, Damiano Ognissanti, Jovan Pankovski and many others.
Special thanks goes to Mariana Pereira, Marco Longfils, Henrike Häbel, Manh Hung Tran, Medet Nursultanov, Nastaran Dashti, Hanna Oppelmayer, Milo Viviani, Sandra Eriksson Barman, Maximilian Thaller, Gabrijela Obradovic, Linnea Hietala, Johannes Borgqvist, Barbara Maria Schnitzer, Deise Fernanda Barbosa de Mattos, Reyhaneh Pourshahami, Jules Lamers, Masashi Kajita, Muyuan Xu, FengYun Zu, Nina Sviridova, Kayoko Tajima, Kashturi Prakash and Sebastián Herrero.
My gratitude goes also to my old-time friends. I would like to express deep gratitude to my family, especially my grandparents, my parents **Enrico** and **Paola**, my siblings **Daniela** and **Marco**, my brother in law **Stefano** and my niece **Zoe**.
Finally I'd like to express my gratitude to **Atanas** for your support and love. Anything feels possible close to you. I love you.

Contents

1 Background

Already in the 1623 Galileo was suggesting that the universe is written in mathematical language [1]. Mathematics can be used as a descriptive language of aspects of reality and more specifically of some biological processes. Here we use it to describe and analyse apolipoprotein kinetics. In Section 1.1 we introduce differential equations and nonlinear mixed effect models, the main mathematical tools used throughout the thesis. Before diving into the details of the analysis, let us briefly introduce the motivation behind this applications.
Cardiovascular diseases (CVD) are the biggest cause of death in the world [2]. Abdominal obesity and Type II Diabetes are two risk factors for CVD [3, 4]. Dyslipidemia (abnormal composition of blood lipid profile) is one of the links between type 2 diabetes and increased risk of CVD [5], while it represents an aggravating factor for CVD in obesity [6]. Therefore lipid metabolism is a target of research in metabolic diseases, and our focus within this domain will be on apolipoproteins.
Apolipoproteins are proteins that are attached to the lipid carrying particles and that give them different properties affecting their metabolism. Apolipoprotein A4, C-III and E, the proteins that will be studied throughout the thesis, are important for lipid metabolism, however much of our results may be applied to other proteins as well. In order to highlight their role, let us first introduce in Section 1.2 lipid metabolism and lipoproteins and let our main players, apolipoproteins, enter the field.
The scope of the experimental and mathematical framework depicted in this thesis is much broader than the answer to the very specific questions answered in the thesis. In Section 1.3 the horizon of vitally important questions that can be answered in this framework will be laid out, after presenting more in detail the extent of the type II diabetes and obesity pandemic. A few of the past discoveries obtained through this framework will be also presented, with the goal of highlighting the complexity of the study of the metabolic disease and the role played by apolipoproteins.
Mathematical biology makes fruit of the advances in mathematics and statistics to further the knowledge on biological systems. Here it is used to investigate the kinetics of apolipoproteins. We will illustrate the interdependency between the experimental setup and the mathematical description in Section 1.4.

1.1 Mathematical foundation

The awe-inspiring sight of nature has been stimulating the creation of mathematical tools with a descriptive purpose since a long time ago. Pythagoras, Fibonacci and Leonardo da Vinci are all examples thereof. The objects created in this meditative state have shown a wider applicability than the description of the specific aspect of reality, from which they sprang. The generality of the mathematical language represents its power.
Mathematics does not only have a descriptive purpose, but it can help access otherwise unknown aspects of reality and in some cases predict the future.
This thesis is one among plentiful examples of how mathematics can be used for solving an otherwise unsolvable problem and opening the doors to a deeper understanding of reality. Let us therefore introduce the main mathematical object in our thesis, a parametric system of differential equations.

1.1.1 Differential equations

Let x represent a variable dependent on time t. A differential equation provides a description of the interrelationship among the change in the dependent variable x in time (its derivative), x and the independent variable, time. The differential equation is usually accompanied by an initial condition.

$$g\left(t, x, \frac{dx}{dt}, ..., \frac{d^{(n-1)}x}{dt^{(n-1)}}, \frac{d^n x}{dt^n}\right) = 0 \tag{1.1}$$

$$x\left(t_0\right) = x_0 \tag{1.2}$$

where $\frac{d^n x}{dt^n}$ represents the n-th order derivative. Solving a differential equation means finding $x(t)$ that satisfies equations 1.1 and 1.2. The equation 1.1 is implicit, since the dependency of the derivatives on the variables of the system is not clear. If the equation only contemplates the first derivative of x and no other derivatives of higher order, the equation is of first order ($d = 1$). An explicit differential equation shows the dependency of the highest order derivative on the independent variable, on the dependent variable and on its derivatives of lower order. Here we will focus on explicit differential equations of first order.

$$\frac{dx\left(t\right)}{dt} = f\left(x(t), t\right) \tag{1.3}$$

$$x\left(t_0\right) = x_0 \tag{1.4}$$

Such a mathematical description as in equations 1.3 and 1.4 can be used when the dynamics of the system is already clear.
A parametric equation is a family of equations which can be described with a

unified mathematical formula. Different values for the parameters will specify a different equation. The parametrization of an equation is suitable when we know the behaviour of x on a qualitative level, but we ignore the quantitative aspects of the system.
One dependent variable x is not enough to describe the kinetics of apolipoproteins, therefore we will use instead a vector of variables, denoted by $\boldsymbol{x}$ and we will use a system of differential equations instead of a single equation. A parametric system of first-order explicit differential equations is the tool that will be used throughout the thesis to describe the kinetics of the apolipoproteins.

$$\frac{d\boldsymbol{x}\,(t)}{dt} = \boldsymbol{f}\,(\boldsymbol{x}(t), t, \boldsymbol{k}) \tag{1.5}$$

$$\boldsymbol{x}\,(t_0) = \boldsymbol{x}_0\,(\boldsymbol{k}) \tag{1.6}$$

1.1.2 Nonlinear Mixed Effects models

Fixed and mixed effects Many studies are interested in analysing the impact of different factors on an outcome variable. For example in a quality control analysis, the weight of a box of syringes could be measured by different scales. In this case the factors are the boxes and the scales. The effect of each single factor on the outcome could be of one of the two types:

- **fixed**, if we assume that the levels of the factors within the study exhaust the whole extent of the total level for that factor
- **random**, if we assume that the levels of the factors observed in the study are only a sample from a distribution and do not exhaust all the possible levels of the study .

Most likely the choice of an effect as random gives rise also to a fixed effect. For example if we choose the individual to be a random effect, then the total population, of which the individuals in the study are just a sample, will represent the fixed effect. Whether a factor is fixed or random depends on the purpose of the study. For example, if the same individual is measured in different days, each day could be seen as contributing in a specific and unique way to that specific measurement (fixed effect). On the other hand, we expect that the variation within each day is due to a measurement error and we might want to assume that this has a normal distribution centred in 0, since we expect the measurement error to be positive at times and negative at times. In this case the specific days would be a random effect.
A model has mixed effects when it presents a combination of fixed and mixed effects. The fixed effects will appear in the model equations as fixed parameters, while the random effects will translate into random parameters for the model. In our model throughout the thesis the factors and their effects are:

- individual –> random

- population –> fixed

Linear and nonlinear models A model is linear when the partial derivatives with respect to each parameter of the model is independent on the other parameters of the model. It is nonlinear if this derivative is dependent on at least one of the other parameters of the model or if it does not exist or it is discontinuous [7].

Birth and development of NLMEM The expression nonlinear mixed effects models denominate nonlinear models that have mixed effects (both fixed and random).
While models with merely fixed effects can benefit from the least squares estimation, introducing random effects adds a level of complexity, since parameter distributions shall be introduced. Therefore more sophisticated mathematical instruments are needed.
The first application of ANOVA, method for the analysis of variance, to quantify the contribution of different factors in linear mixed effects models, appeared in 1921 in Fisher [8]. The standard two stage method (STS) used to be applied to draw information on the population behaviour from the individual estimates, obtained through the least square minimization. This was a first attempt to draw conclusions on the population kinetics based on data from a sample of individuals. Only in 1977 a parameter estimation method was introduced for NLMEM [9] and applied to pharmacokinetics that would encompass the description of the parameter as random and information on the distributions. The study of the absorption, diffusion and elimination of a drug is still one of the main application fields of NLMEM.
In the last decade of the 20th century, Lindstrom and Bates ([10]), Wolfinger ([11]) and Pinheiro and Bates ([12]) contributed in developing and clarifying the different methods available for the implementation of NLMEM. The NLMEM parameter estimation makes use of a maximum likelihood approach. At the very end of the 20th century, a stochastic version of the expectation-maximization (EM) algorithm was proven to converge [13]; this algorithm (stochastic approximation of EM, denoted as SAEM) is now the basis of the parameter estimation in Monolix, the software used in this work. One recent advance in the NLMEM parameter estimation was the introduction of sensitivity equations to have a more exact gradient estimation [14]. This method was also generalized to NLMEM with stochastic dynamics [15].

1.2 Lipid Metabolism

Lipids represent one of the sources of energy for our body, together with carbohydrates and proteins. Lipids are also important for cell membranes and

hormones.
Lipids are categorized by their structure and function. Fats or triglyceride are a category of lipids. A triglyceride molecule is derived from glycerol and fatty acids. Lipids can be used as an energy source by almost all the tissues. The brain is an exception, in that it is fuelled by carbohydrates and ketone bodies, produced by the liver.
The lipid fat storage is mainly found in adipose tissues and accounts in total for more than 80 per cent of our energy reserves.
Lipid metabolism starts with the intake of fat-containing food, continues with absorption and hydrolysis of fats in the intestine, secretion and circulation of lipid particles in the blood stream. By releasing fatty acids to peripheral tissues, these particles increase their density and eventually get cleared by liver cells.
Other lipid-carrying-particles are synthesized by the liver, they travel through the blood stream and follow the same journey as the dietary lipids.

Since lipids are not water-soluble, they need a means to be transported in the blood stream. Lipoproteins fulfil this role.
On the surface of lipoproteins there are various proteins, called apolipoproteins.
Lipid metabolism is a very wide topic. We choose to focus our attention on one important piece of the puzzle, apolipoproteins kinetics, focusing our attention on apolipoproteins A4, E and C-III.

1.2.1 Lipoproteins

Lipoproteins are spherical particles with a core of triglyceride and cholesterol ester, and a surface monolayer of phospholipids. Lipoproteins do not only transport lipids among different parts of the body. They also facilitate the removal of toxins from infected areas.
There are mainly three categories of lipoproteins.
Chylomicrons are produced by the intestine and their core is represented by an apo B-48 molecule. Chylomicron remnants result from a transfer of triglycerides to the peripheral tissues.
The two remaining categories of lipoproteins are produced in the liver.
Apo B-100 carrying lipoproteins have apoB-100 molecule at their core. Their role is to facilitate the transport of lipids from the liver to the tissues. When losing triglycerides to the peripheral tissues through the action of lipoprotein lipase (LPL), these apolipoproteins decrease their density, therefore we can subdivide this category into further subcategories depending on their density. These are: VLDL (very low density lipoproteins), which can be further subdivided into VLDL1 and VLDL2, IDL (intermediate density lipoproteins) and LDL (low density lipoproteins). The composition of the lipoproteins varies among the different groups, with lower density pointing to an earlier stage of the lipoprotein life cycle.
Chylomicrons remnants and apoB-100-carrying particles are removed from the

circulation through LDL receptors or LPR (LDL receptor related protein).
The last category of lipoproteins is HDL (high density lipoproteins), that transports the cholesterol back to the liver. ApoAI is the apolipoprotein at their core.

1.2.2 Apolipoproteins

ApoB-100, apoB-48 and apo-AI are respectively at the core of ApoB-100 carrying particles, chylomicrons and HDL.
Additional apolipoproteins with diverse roles can be found in the lipoproteins. Here we will focus on apoA4, apoC-III and apoE, object of our analysis.

ApoA4 Mainly produced in the intestine, apoA4 can be found on chylomicrons, in HDL and in a free form and embodies different roles. Here we report several results about apoA4.
Men with coronary artery disease had significant lower apolipoprotein A4 level than the control counterpart [16]. ApoC-II binds with LPL to allow triglyceride release to the tissue. ApoA4 stimulates LPL activity, leading to an increased removal of apoC-II containing particles [17]. This stimulating role is lost in absence of apoC-II containing particles.
Apolipoprotein A4 plays a role in the inverse cholesterol transport [18] and inhibits lipoprotein oxidation [19, 20, 21, 22]. Moreover apolipoprotein A4 stimulates insulin secretion, therefore playing an important role in glucose homeostasis [23].

ApoC-III ApoC-III is part of the family of apolipoproteins C. These proteins are mainly synthetized in the liver and can exchange among the lipid particles, therefore they can be found in all the three categories of lipoproteins. ApoC-III inhibits LPL and hepatic lipase. Not only does apoC-III influence the triglyceride blood concentration by its action on LPL, but underlying links between apoC-III and triglyceride are being unravelled also at a cellular level. A high fat diet failed to stimulate VLDL1 production in apoC-III deficient mice, while this strategy was successful in mice with reconstituted apoC-III expression [24]. Cell culture in lipid-rich environment confirmed the link between apoC-III and VLDL1 secretion [25].
In individuals with abdominal obesity, VLDL1-triglyceride concentration correlates with apoC-III plasma concentration, while it does not correlate with LPL activity [26], highlighting once more the deep connection between apoC-III and triglyceride.

ApoE Apolipoprotein E is mainly synthetized in the liver and in the intestine. ApoE facilitates the removal of lipoproteins by binding with the LDL-family receptors and with LRP [27, 28]. Its role in the removal of lipoproteins is not limited to this. In fact apoE is also a ligand for cell-surface heparan sulfate proteoglycans (HSPGs) which, independently of LDL-family receptors, contributes to the lipoprotein clearance [29].

1.3 Research domain

1.3.1 Obesity and diabetes epidemics

In 2016 the World Health Organization (WHO) reported that 650 million of people, 18 years and older, were obese (13% of the global population in this age range) and that all over the world, the incidence of obesity has almost tripled since 1975 [30], while the number of people with diabetes in 2014 (422 millions) has almost become four-fold the one in 1980 [31]. More than 280000 were the average annual deaths attributable to obesity only in the United States in 1999 [32].

Abdominal obesity Obesity, as normally used in global statistics, indicates when the body weight divided by the squared height (BMI) exceeds 30 kg/m^2. On the other side, abdominal obesity, measured mainly as waist-to-hip circumference ratio or waist circumference, is a stronger predictor than BMI for incidence of cardiovascular diseases [33]. Abdominal obesity is one of the characteristics of metabolic syndrome [34]. While subcutaneous abdominal fat behaves similarly to the other adipose tissues, high level of visceral fat impairs metabolic functioning [35], by upregulating the free fatty acid release from the visceral fat and thereby fostering an accumulation of free fatty acids in other tissues (cardiac, pancreatic, liver and more). This in turn perturbs the normal functioning of these tissues.

Type II Diabetes Measures of abdominal obesity are strongly associated with incidence of type II diabetes. Type II Diabetes is not the only form of diabetes. This family of diseases entails insufficient production (Type I) or disfunctional use of insulin (Type II), an hormone responsible for the blood glucose levels. Type II Diabetes increases the risk of vascular diseases [36].

Consequences for lipid metabolism

Though not universally present in type II diabetic patients [37], dyslipidemia (high fasting and postprandial triglyceride levels, high LDL levels and low

HDL levels) represents the link between this disease and an increased risk of cardiovascular event [5]. Another marker of worsened lipid profile is a high ratio (bigger than 5) of total cholesterol over HDL cholesterol. Dyslipidemia, as measured in this way, is associated to abdominal obesity, across different ethnic groups [38]. Viscerally [39] and abdominally obese individuals [40] have higher triglyceride levels than their normal counterpart.

1.3.2 Purposes of kinetic studies

The experimental and mathematical framework that will be introduced in Section 1.4 and explained in depth in Chapter 3 can answer many questions. While the model used in this thesis is specific to apolipoproteins, different models can be used with the same framework to study lipoproteins. Therefore here we will introduce the plethora of questions that can be answered with this framework, with the goal of highlighting its importance in the study of metabolic diseases, the associated dyslipidemia and drug efficacy.
As will be clear in Section 1.4, this framework allows us to study the kinetics of lipoproteins and apolipoproteins, so their **lifecycle**, and not only their concentration. That is, it is possible to study rates of production and clearance.
Three are the main purposes of kinetic studies and this will be discussed below. An additional paragraph will summarize information on some ameliorative treatments, whose effect can be better understood through the use of kinetic studies.

Kinetics for the healthy individual Our end goal is to study disruptions in apolipoprotein and lipoprotein kinetics in a disease state, but important in this direction is also to study how the kinetics looks like in a normal state. This allows us to define normal ranges of production and clearance.

Comparison with disease state Comparing the kinetics in normal individuals with the disease state let us uncover differences in the kinetics and potentially understand which mechanism is behind the disruption in lipid concentration in the disease state. Beside highlighting disruptive pathways, this also may suggest which pathway to target by treatments and interventions.

Treatment and drugs High fasting and postprandial triglyceride are a marker of dyslipidemia, together with high LDL and low HDL. Several drugs and lifestyle interventions aim at improving the lipid profile, and our interest is in quantifying the effect on lipoproteins and apolipoproteins (since these are involved in lipid metabolism). While measuring the concentration of a biologically interesting quantity, be it apolipoproteins and lipoproteins, highlights the net effect of the treatment, kinetic studies can point out to whether the resulting change in concentration is due to a change in the production or in clearance of

the particles.
Alongside therapies to improve the lipid profile, other experiments are performed in order to prove or disprove an hypothesized deleterious effect of a treatment.
Whether the intervention aims at improving the lipid profile or at confirming a potential negative effect of an agent, apolipoprotein and/or lipoprotein kinetics studies combined with mathematical modelling helps us to draw conclusion on its effects.

Ameliorative interventions Since dyslipidemia increases the risk for cardiovascular diseases, interventions to improve the lipoprotein profile are needed. Though lifestyle interventions in terms of more physical exercise [41] and changes in diet [42] have been proven to be successful in improving the lipid profile, they do not always suffice to obtain optimal levels for the markers, therefore a pharmacological treatment is needed in many cases.
Statins reduce the LDL level [43]. PCSK9 is responsible for the degradation of the LDL receptors and as a consequence evolocumab, one of the drugs based on a PCSK9 inhibitor, lowers LDL cholesterol level [44].
In order to target triglyceride levels, other drugs, like fibrates or niacin, are also needed. Controversies arose on the safety of niacine and fibrates [45] and new therapeutic solutions targeting the triglyceride levels are now under development.
In the light of the tight bond between apoC-III and triglyceride, targeting apoC-III could be a viable option to lower the triglyceride concentration in dyslipidemic individuals. Indeed apoC-III-synthesis-inhibition has been proven to reduce triglyceride levels in a phase 2 clinical study [46].

1.3.3 Kinetic studies in metabolic diseases

Here we report some examples of past discoveries obtained through the study of lipoprotein or apolipoprotein kinetics that enrich our knowledge on metabolic diseases.
It is important to remember that apoB is a core part of chylomicrons and of apoB-100-containing lipoproteins, therefore kinetics of lipoproteins can be studied by taking track of the content of apolipoprotein B in lipoprotein group fractions.

Lipoprotein kinetics

Type II diabetes Individuals with noninsulin-dependent type II diabetes were compared with normal individuals [47]. The type-II-diabetic individuals had higher VLDL apoB levels, due to an increased production rate and decreased conversion from IDL to LDL. IDL apoB concentration was also higher due to a decreased transfer to LDL particles. Finally clearance rate for LDL apoB was significantly lower in diabetic patients.

Abdominal obesity In a study cohort with abdominally obese people, elimination rate of VLDL particles correlates more strongly with the triglyceride concentration than the production rate of triglyceride-rich-lipoproteins [26].

Apolipoprotein kinetics

ApoC-III Compared to nonobese individuals, viscerally obese individuals had higher concentration of VLDL-apoC-III, with lower clearance rate and higher production rate. Lower clearance rate was observed also for the HDL-apoC-III, with a coupled increase in the concentration [48].
In a comparative study between normal and type-2 diabetic with matched body weight, diabetic individuals showed an increased secretion rate and clearance rate of apoC-III compared to the non-diabetic group [49].

ApoE VLDL-apoE kinetics and HDL-apoE kinetics was studied in a group of obese individuals ($BMI > 29.3$). The 7 individuals with type II diabetes showed increased levels of VLDL-apoE and HDL apoE, both of which tied to an increased production rate [50].

1.4 Glimpses of the experimental design and mathematical tools

As pointed out in Section 1.3, only studying the concentration of apolipoproteins and lipoproteins might not be enough to unravel the mechanisms involved in the metabolic disease, therefore we study their kinetics.
Here, for simplicity purposes, we will be specifically talking about apolipoproteins. It is worth noting that the experimental set-up and the mathematical framework is the same when lipoprotein kinetics is analysed.

Experimental set-up Noninvasive measurements can only take place in the blood, therefore we are limited to sampling blood to access apolipoprotein concentration and we would like to use these measurements to reconstruct their lifecycle.
In order to do this, we choose to study the lifecycle of leucine, an amino acid, which is a building block of all proteins, including apolipoproteins.
An isotope of the naturally occurring leucine is given with a bolus injection to the patients and we quantify at different time points its occurrence in the free state and its presence in the apolipoproteins. These studies are called tracer-tracee studies, in that an inserted material (tracer) is used to take track of the dynamics of another material (tracee).

There are two different important stages for the isotope leucine: a state when it is free and a state in which it has been incorporated into the apolipoproteins. The transition from one state to another requires a description. This is resolved through a structural model, which can simply be thought as a picture describing the different stages of the molecule.

Mathematical framework Since our goal is quantification of the biologically interesting parameters, we need a mathematical description of the model to assess how the different states (parts of the picture) are quantitatively connected to each other. In our case, the biological process can be described as a linear system of differential equations, where the independent variable is time and the dependent variables are the quantities of leucine in different stages. The parameters of the system quantify the speed of transition from one stage to another.
We shall also decide how to think about the individuals in the study. Are they just separate individuals, each of them exhibiting a behaviour totally untied from the others, or are they part of a bigger family, each of the individual behaviours being a slight variation on the common distribution of a larger population? Therefore we choose a statistical model for the parameters of the system.
Our key question is on the quantification of the parameters for production and elimination of the apolipoproteins. While the combination of the mathematical model for the kinetics and statistical model for the parameters of the system gives birth to a family of distributions, we are interested in the specific set of parameters that maximizes the probability of observing exactly the specific dataset of measurements (maximum likelihood estimation approach).
Therefore the last ingredient, the experimental data shall be incorporated. Optimization algorithms exploit our mathematical and statistical description in order to obtain the sought-after parameter estimates from the data.
As seen in Section 1.1.2 nonlinear mixed effects models (NLMEM) encapsulate the mathematical description of the system into a statistical framework for the population behaviour as and exploits optimization algorithm for parameter estimation.

2 Aims and outline of the thesis

In this chapter, we lay out the aims of the thesis and subsequently we present its outline, focusing on the specific strategies we adopted to fulfil the overarching mathematical aim.

2.1 Aims of the thesis

Some biological questions would remain unanswered if mathematics and statistics would not provide the tools for describing biological mechanisms and make inference on them.
The overall aim of the thesis is to develop mathematical models and tools to describe the kinetics of three apolipoproteins of interest (apoC-III, apoE and apoA4) and to apply them to clinical data. While a model for apoC-III kinetics had been already developed [48] and then later applied to the study of VLDL-apoE kinetics [51], to our knowledge a general model for overall apoE kinetics and for apoA4 kinetics had not yet been developed. Because of the above-mentioned versatily of the model in [48], one aim was to understand whether a suitable version of this model could be generalized to study the kinetics of all the three apolipoproteins. Additionally, we investigated whether the choices on the statistical model (variance-covariance matrix structure and error model) could also be generalized. The developed framework, complete with structural and statistical model, will be presented in Paper I. Let us now dive into the specific biological questions related to the datasets that we are analysing and the needs from which they arise and keep in mind that these are just some of the questions that can be answered with the tools developed in this thesis.

Despite the abundance of the biological questions that stems from the connection between lipid metabolism and metabolic diseases, very few population kinetic studies have been focusing on lipoproteins and the number of kinetic

studies on apolipoproteins is even more scarce.

As seen in Section 1.2, apolipoproteins confer properties to the lipid carrying particles where they are attached.
Apolipoprotein C-III (ApoCIII or apoC-III) inhibits LPL, therefore reducing the clearance of triglyceride in the blood [52].
Apolipoprotein E instead facilitates the removal of triglyceride particles, since its presence is necessary so triglyceride containing lipoproteins can bind to the LDL receptor and abandon the blood stream [52]. Apolipoprotein A4 increases insulin secretion [23], therefore playing an important role in metabolic diseases.
Our focus will revolve around these three apolipoproteins with the overall aim of deepening our understanding of their kinetics and of how they are nestled in the frame of metabolic impairments.

First let us zoom in our apolipoproteins and obtain the most information on this piece of the puzzle. The antagonism between apoC-III and apoE could suggest the existence of a relationship between their kinetics parameters, but so far their kinetics have been studied separately. The novel experimental set-up of our data allows us to study apoA4, apoC-III and apoE in a composite manner, with the goal of uncovering relationships among their kinetics parameters. This aspect will be tackled in Section 5.2, through the study of the correlation structure of these parameters at baseline.

Now let us zoom out and find where our apolipoproteins of interest belong in the overall puzzle of lipid metabolism. This will be analysed by studying the correlation structure of the apolipoproteins kinetics parameters with individual characteristics of the subjects undergoing the experiment at baseline; this content is also discussed in Section 5.2.

The last biological interest is to check whether three different treatments affect the apolipoprotein kinetics.
More than once, hypercaloric intake of fructose has been shown to generate an increase in the triglyceride content in the blood.
This is confirmed by the large study [53], that our dataset is part of, where an additional increase in apoC-III content is resulted from the treatment. We are interested in investigating whether an increased production rate or a decreased clearance rate has produced this effect and whether additional effects can be found in the kinetics of apoA4 and apoE. While the overall effects for the three apolipoproteins will be described in Paper I, the results concerning apoC-III will be integrated with the results concerning lipoprotein kinetics in Paper II, where we elucidate the previously unknown mechanism behind the increase in triglyceride concentration consequent to fructose treatments.
Evolucomab and liraglutide are two drugs to treat aspects of type II diabetes.
Evolucomab (Evo) inhibits PCSK9, therefore increasing clearance of LDL parti-

cles, by reducing the degradation of LDL receptors. Since apoE binds with LDL receptors, we expect an effect of the drug on apoE kinetics and therefore investigate on it. The results about the effects of evolocumab on the apolipoprotein E and C-III kinetics will be shown and discussed in Paper III.
The action of liraglutide increases insulin secretion and therefore lowers glucose blood levels. The cascade of insulin signalling impacts lipid metabolism. It has already been shown in the larger study [54] that a liraglutide treatment positively affects postprandial lipid metabolism. Our question is on how apolipoprotein kinetics is impacted. This content will be show in Section 5.3.3.

2.2 Outline of the thesis

We can describe the apolipoprotein kinetics with a compartmental model and in so doing, moving our biological interest into the mathematical world. The manuscripts are addressed to a public interested in biology, therefore the mathematical details are left out of the manuscripts and can be found in Chapters 3 and 4.
Section 3.1 will depict the mathematical theory on the compartmental models and present the alternative approaches for the estimation of the parameters (standard two stage method (STS) and nonlinear mixed effects models (NLMEM)). In Section 3.2 we present the datasets used in this work together with the model structure and the software used. Chapter 4 is dedicated to the choice of the model. First in Section 4.1 we introduce the proposed model. In Section 4.2 the above-mentioned two alternative methods for parameter estimation (STS and NLMEM) will be explored and we justify our choice of NLMEM as method of estimation.
In Section 4.3 we argue some choices on the model.
From a biological perspective two quantities (the amount of free leucine in the blood and the amount of leucine in each apolipoprotein) are equally eligible to be parameters of the model. We investigate whether choosing one over the other has effect on the goodness of fit of the model.
We then compare two possible descriptions of the system: one model encompassing all the apolipoproteins (forcing some of the parameters to have the same distribution)(combined model) or a model where each apolipoprotein kinetics is analysed separately.
A common variance-covariance structure is needed in order to analyse the data across different visits or studies together. While it might be reasonable to decide for the most general model, a parsimony principle would lead us to have a more specific variance-covariance structure. Therefore we investigate the datasets and design the minimal variance-covariance matrix that is able to encompass the relevant correlations between random effects found in the different studies.
In Section 4.4 we walk through the process of validation of the model.
Chapter 5 describes thoroughly the biological results. The abundance of our data, that contains information for three apolipoproteins, has allowed us to analyse the kinetics of apoA4, apoC-III and apoE for the same group of individuals.

Nevertheless not all the results are so novel and interesting for the scientific world to be extrapolated as manuscripts. Indeed, the analysis of the results is a necessary step for the selection of the new findings. Section 5.1 describes the steps of this analysis and illustrates the visual tools used to perform them. The remaining sections of this chapter describe the results with the aid of these tools.
Finally Chapter 6 summarizes the findings in the manuscripts, while Chapter 7 offers perspectives on the outlined work and make suggestions about further work in the research domain.

3 Methodology

In the human body, at any given moment, we assume that quantities are in a steady state.
This means that, if we were to measure the amount of apolipoproteins in the blood at two time points, we would only detect a negligible difference.
This is the reason behind the use of isotopes of naturally occurring molecules for the study of the kinetics. We assume that the labelled material behaves exactly as the natural occurring one and that its insertion does not influence the behaviour of the system.
The inference on the kinetics is the bridge between the datasets and a deeper understanding of reality. A model is necessary for this purpose; it can be thought as a simplification of a phenomenon, with the goal of shading light on it. After depicting the general framework for the inference on kinetics in Section 3.1, elements specific to this thesis will be presented in Section 3.2.1: the available datasets for the analysis and the software used.

3.1 Inference on the kinetics

In this section we will first introduce the needed ingredients for the inference on the kinetics: compartmental models and datasets. We will then explain the model development process and familiarize more with two alternative ways for performing the parameter estimation, STS and NLMEM, already introduced in Section 1.1. The second one will be explained more in detail and alternative algorithms for its implementation will be laid out. We will then look into the concepts of identifiability and observability and finally present an overview of the tools to diagnose the goodness of fit of the model.

3.1.1 Compartmental models

The following definition of compartmental models is rearranged from [55].
A compartmental model represents the overall description of a system with

multiple compartments and highlights the movement of material among them. Compartments are units with homogeneous chemical and physical characteristics and kinetic properties, not necessarily correspondent to a unified space.
Let us introduce the compartmental model for leucine kinetics, which will be then used in Section 4.2. This represents part of the bigger model used to describe apolipoprotein kinetics, which will be presented in Section 4.1. Each of

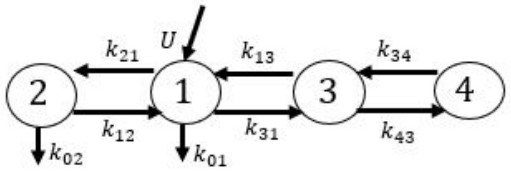

Figure 3.1: Model for natural leucine kinetics

the node in the graph represents a compartment. k_{ij} indicates the transfer rate from compartment j to compartment i and k_{0j} indicates the transfer rate from compartment j to a compartment external to the system. In order to reduce the number of parameters in the model, some relations among the parameters of the system are enforced, as seen in equations 4.9 and 4.10 for the bigger compartmental model and the removal from compartment 2 k_{02} is fixed to 0.011, as done in [56]. The compartmental model can be described with the mathematical language as a system of differential equations:

$$\begin{aligned}
\frac{dQ_1}{dt} &= U - (k_{12} + k_{01} + k_{31}) * Q_1(t) + k_{13} * Q_3(t) + k_{12} * Q_2(t) \\
\frac{dQ_2}{dt} &= -(k_{12} + 0.011) * Q_2(t) + k_{12} * Q_1(t) \\
\frac{dQ_3}{dt} &= -(k_{13} + k_{43}) * Q_3(t) + k_{31} * Q_1(t) + 0.1 * k_{43} * Q_4(t) \\
\frac{dQ_4}{dt} &= -0.1 * k_{43} * Q_4(t) + k_{43} * Q_3(t)
\end{aligned}$$

Leucine kinetics is only one example of a biological mechanism that can be thought of as a deterministic process, where the future state can be predicted if we know the current state and the evolution law. This concept is crystallized in the mathematical language as a dynamical process [57]. A system of differential equations describes the law and the initial condition signifies the current state, as in equations 1.5 and 1.6.
In this work we do not aim at merely describing the biological process, but we are interested in the behaviour of the specific individuals in the experiments and in generalized the obtained results to the population of individuals where they belong. Therefore while the structural model and equations remain the same for all the individuals, the parameters vary for each individual i and we

indicate this with a subscript:

$$\frac{d\boldsymbol{x}_i(t)}{dt} = \boldsymbol{f}(\boldsymbol{x}_i(t), t, \boldsymbol{k}_i) \tag{3.1}$$

$$\boldsymbol{x}_i(t_0) = \boldsymbol{x}_0(\boldsymbol{k}_i) \tag{3.2}$$

$\boldsymbol{f}$ is the function that explains how the quantities in the system change over time and $\boldsymbol{k}_i$ is the parameter vector for the i-th individual.
We call $\boldsymbol{x}$, the vector of dependent variables, state vector and $\boldsymbol{x}_i$ is the state vector for the i-th individual.
The kinetics of the state vector is highly dependent on the parameters of the system $\boldsymbol{k}_i$.
For the endogenous leucine model:

$$\boldsymbol{x} = [Q_1, Q_2, Q_3, Q_4] \tag{3.3}$$

$$\boldsymbol{k} = [U, k_{01}, k_{12}, k_{13}, k_{31}, k_{43}] \tag{3.4}$$

In case the system of equations 3.1 describes the change in the quantity in one compartment in terms of the inflow from other compartments and the outflow from the compartment, the equations are called mass balance equations. If each component of $\boldsymbol{f}$ is a function linear in the components of $\boldsymbol{x}$ the differential system of equations is called linear.

Fixed and random effects The individual i could be considered as a fixed effect if the overall group of individuals exhaust our research interest [7]. In this case each individual i could be considered separately and the individual level would represent a fixed effect. Alternatively, all the individuals in the study could be thought of as a sample of a larger population, that we are interested in analysing and the population would be a fixed effect. In this case, the individual level represents a random effect and the model has mixed effects, in that it comprehends parameters which are fixed, relative to the fixed effects (the population) and random, relative to the random effects (the deviation of each individual from the population average).

Dataset Compartmental models are used in conjunction with experimental data. The trials usually involve the insertion of a material in the body, since a perturbation shall occur in order to detect modifications in the body. The material could be a drug or an isotope of a naturally occurring quantity. Depending on the methods of administration, the information on this quantity can be included in the initial condition, as in our case, or incorporated in the function $\boldsymbol{f}$. The system of differential equations represents the structural model. The goal is to make inference on the system, therefore we need a dataset. Since our interest is not only on the kinetics for one individual, but on the general behaviour of the population, the experiment to generate data will comprehend several individuals. The study of a dynamical system requires that each in-

dividual is sampled at different time points. Let i denote the index for the i-th individual, with $i = 1, ..., N$. Let t_{i_j} denote the time point for the j-th measurement for the i-th individual with $i_j = 0, ..., m_i$.
The dataset $\boldsymbol{d}$ will consist of $\boldsymbol{d}_i$, denoting the data for the i-th individual. $\boldsymbol{d}$ is written in bold because several quantities could be measured.

$$\boldsymbol{d}_i = \{\boldsymbol{d}_{ij}\}_{j=1,...,m_i} \tag{3.5}$$

We can approximate the data with a function of the state vector, model approximation, therefore we can denote it as a measurement function and write:

$$\boldsymbol{y}_{ij} = \boldsymbol{h}\left(\boldsymbol{x}_i\left(t_{i_j}\right)\right) \tag{3.6}$$

The data $\boldsymbol{d}_ij$ will differ from the model approximation $\boldsymbol{y}_{ij}$ due to the measurement error, that we will later encounter.

3.1.2 Choice of the model

The overall goal is making quantitative inference on the kinetics. In order to do this a model is required. The choice of the model is done in two phases:

1. Model development: develop a suitable structural model, a set of equations that can describe the biological process and a statistical model for the parameters, if the estimation method requires it,

2. Model validation: test whether the goodness of fit of the model is consistent through application to other datasets.

In few cases, this could be a linear process, where phase 2 follows phase 1, but in general this can be thought as a circular process. If the model developed in the first step cannot describe the datasets used in the second phase, information gathered in this step can be used to update the structural and statistical model, in phase 1.
Knowledge about biology is used in the structural model development process. The selected model is chosen based on the fit of the model with one or more datasets, therefore parameter estimation is involved in the process. One can adjust the structural model or the statistical model based on the information delivered by the diagnostics tools.
A general aim is that the model is not specific to the dataset under analysis but that its validity continue to hold for different datasets.
Therefore model validation consists in applying the model to one or more datasets different to the one used for model development and check whether the dataset can be described with the same model.

3.1.3 Estimation methods

We will assume that the structural model is fixed, so our goal is to explore the statistical model, if needed, and estimate the average behaviour of the population and its inter-individual variability.

Standard two stage (STS) method

A naive way to get this information is called standard two stage (STS) method. Here we do not choose any statistical model and we use the following procedure:

1. The parameters $\boldsymbol{k}_i$ are estimated separately for each individual, by minimizing the sum of the squared difference between the data and the approximation given by the model.
2. Statistical analysis are performed to obtain the mean of the parameter across the study individuals and its standard deviation.

While the simplicity of this approach makes it highly suitable for implementation, there are some drawbacks that must be taken into account:

1. Obtaining a good estimate of the population average can be difficult when the model comprehends more than one parameter. The standard deviation of the parameter across the different individuals might be very high since the parameter estimates are independent for each individual
2. Time series must contain a big number of time points for each individual [58], since the parameter estimate for the i-th individual can only make use of data for that individual.

Nonlinear mixed effects models

We can imagine that the study individual do not exhaust the extent of our research questions and that they are a sample from a larger population, therefore the individual parameters and the error would contain random components, due to the fact that they must behave according to the distribution of the larger population. To integrate this in the modelling framework, we introduce a statistical model, which encompasses choices for the parameter distribution and for the error model.

Distribution of the parameters We can think of $\boldsymbol{k}_i$, the parameter for the i-th individual, as a realization of a random variable, containing information on the

population average and on the inter-individual variability. This approach is called nonlinear mixed effects models (NLMEM), since it encompasses fixed effects (population averages) and random effects (random parameters for the specific individual).
Almquist et al. introduced a modification to the NLMEM algorithm [14]. Wang neatly explains how the formulas used in NONMEM, one of the most used software for NLMEM, are derived [59]. Most of the formulas in this section are from these two main sources.
A common choice for the parameter distribution is the log-normal.

$$\boldsymbol{k}_i = \boldsymbol{k} * e^{\boldsymbol{\eta}_i} \tag{3.7}$$

where $\boldsymbol{\eta}_i$ is the vector of the individual random effects. In case of the log-normal distribution $\boldsymbol{\eta}_i$ is a realization of a normal variable with mean $\mathbf{0}$ and variance-covariance matrix given by $\boldsymbol{\Omega}$.

Error model choices The randomness in the individual parameter is not the only source of volatility.
In fact the error also contains a random part; the j-th observation of the i-th individual at time t_{ij} is the sum of the measurement function based on the model evaluated at time t_{ij} plus an error component.

$$\boldsymbol{d}_{ij} = \boldsymbol{h}\left(\boldsymbol{x}_i\left(t_{ij}\right)\right) + \boldsymbol{e}_{ij} \tag{3.8}$$

In Monolix, there are several choices of error model. We describe here the choices that we have contemplated in this thesis. The error model for each of the components of the measurement function $\boldsymbol{h}$ might differ from each other, therefore here we report the formulas for the error model for one component of $\boldsymbol{h}$, and for the sake of simplicity we will denoted it as h.

1. constant: the error does not depend on h, but it is constant.

$$d_{ij} = h\left(\boldsymbol{x}_i\left(t_j\right)\right) + ae_{ij} \tag{3.9}$$

2. proportional: the error is proportional to h. The bigger the value of the function, the bigger the error that is expected.

$$d_{ij} = h\left(\boldsymbol{x}_i\left(t_j\right)\right) + bh\left(\boldsymbol{x}_i\left(t_j\right)\right) e_{ij} \tag{3.10}$$

3. combined1: this error model combines the two previous ones, entailing that the total measurement error can be written as a combination of a component which is independent on h and another that is proportional to h.

$$d_{ij} = h\left(\boldsymbol{x}_i\left(t_j\right)\right) + \left(a + bh\left(\boldsymbol{x}_i\left(t_j\right)\right)\right) e_{ij} \tag{3.11}$$

In each of the preceding error models the random component is represented

by e_{ij}, which is a realization of a random variable. In Monolix e_{ij} for each of the components of the measurement function is a standardized Gaussian random variable; in a more general setting $\boldsymbol{e}_{ij}$ can encompass the error component for the different measurement functions as in 3.9. A common choice for the distribution of this random vector is the normal distribution, with mean $\mathbf{0}$ and variance-covariance matrix given by $\boldsymbol{R}$. We assume in this case that the variance-covariance matrix is fixed and does not depend on the individuals nor on the time point.
The fixed parameters of the model are not limited to the vector of the population fixed effect parameters $\boldsymbol{k}$. In fact the parameters for the inter-individual variability in the variance-covariance matrix $\boldsymbol{\Omega}$ and the ones for the error model are also fixed effects. We will denote with $\boldsymbol{\theta}$ the vector containing all the fixed parameters of the system. For simplicity purposes, we will explain more in depth mathematics with the constant error model across all the measurement functions.

Towards parameter estimation, the likelihood function In order to estimate the parameters of the model from the experimental data, a maximum likelihood approach is chosen. The chosen set of parameters will be the one that maximizes the likelihood of obtaining exactly our dataset, if the system described with that set of parameters was to be sampled.
The likelihood function can be written in the following way:

$$L\left(\boldsymbol{\theta}\right) = \prod_{i=1}^{N} \int p_1\left(\boldsymbol{d}_i|\boldsymbol{\theta},\boldsymbol{\eta}_i\right) p_2\left(\boldsymbol{\eta}_i|\boldsymbol{\theta}\right) d\boldsymbol{\eta}_i \tag{3.12}$$

where

$$p_1\left(\boldsymbol{d}_i|\boldsymbol{\theta},\boldsymbol{\eta}_i\right) = \prod_{j=1}^{n_i} \frac{e^{-\frac{1}{2}\boldsymbol{\epsilon}_{ij}^{\intercal}\boldsymbol{R}^{-1}\boldsymbol{\epsilon}_{ij}}}{\sqrt{\det\left(2\pi\boldsymbol{R}\right)}} \tag{3.13}$$

and

$$p_2\left(\boldsymbol{\eta}_i|\boldsymbol{\theta}\right) = \frac{e^{-\frac{1}{2}\boldsymbol{\eta}_i^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{\eta}_i}}{\sqrt{\det\left(2\pi\boldsymbol{\Omega}\right)}} \tag{3.14}$$

We can rewrite the likelihood function as:

$$L\left(\boldsymbol{\theta}\right) = \prod_{i=1}^{N} \int exp\left(l_i\right) d\boldsymbol{\eta}_i \tag{3.15}$$

with:

$$l_i = -\frac{1}{2}\sum_{j=1}^{n_i}\left(\boldsymbol{\epsilon}_{ij}^{\intercal}\boldsymbol{R}^{-1}\boldsymbol{\epsilon}_{ij} + \operatorname{logdet}\left(2\pi\boldsymbol{R}\right)\right) - \frac{1}{2}\boldsymbol{\eta}_i^{\intercal}\boldsymbol{\Omega}^{-1}\boldsymbol{\eta}_i - \frac{1}{2}\operatorname{logdet}\left(2\pi\boldsymbol{\Omega}\right) \tag{3.16}$$

being the individual likelihood function for the i-th individual. There are different ways to proceed in the maximum likelihood estimation.

Laplace Approximation We can use the Laplace approximation to approximate the likelihood function in the following way:

$$L_L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \left(exp\left(l_i\left(\boldsymbol{\eta}_i^*\right)\right) \left(\det[\frac{-\Delta l_i\left(\boldsymbol{\eta}_i^*\right)}{2\pi}] \right)^{-\frac{1}{2}} \right) \tag{3.17}$$

where $\boldsymbol{\eta}_i^*$ is the point which maximizes the individual joint-likelihood l_i. The formula contains the Hessian (matrix of the second derivatives) and the gradient (vector of first derivatives) therefore we need to calculate the first and second derivatives.
We show the derivation of the individual likelihood function with respect to the random parameters:

$$\frac{dl_i}{d\eta_{ik}} = -\frac{1}{2}\sum_{j=1}^{n_i} \left(2\boldsymbol{\epsilon}_{ij}^{\intercal}\boldsymbol{R}^{-1}\frac{d\boldsymbol{\epsilon}_{ij}}{d\eta_{ik}} + \boldsymbol{\epsilon}_{ij}^{\intercal}\boldsymbol{R}^{-1}\frac{d\boldsymbol{R}}{d\eta_{ik}}\boldsymbol{R}^{-1}\boldsymbol{\epsilon}_{ij} + tr[\boldsymbol{R}^{-1}\frac{d\boldsymbol{R}}{d\eta_{ik}}] \right) - \boldsymbol{\eta}_i^{\intercal}\boldsymbol{\Omega}^{-1}\frac{d\boldsymbol{\eta}_i}{d\eta_{ik}} \tag{3.18}$$

$\frac{d\boldsymbol{\epsilon}_{ij}}{d\eta_{ik}}$ are the derivatives of the residuals with respect to the random effects. These can be written as:

$$\frac{d\boldsymbol{\epsilon}_{ij}}{d\eta_{ik}} = \frac{d(\boldsymbol{d}_{ij} - \boldsymbol{y}_{ij})}{d\eta_{ik}} = -\left(\frac{\partial \boldsymbol{h}}{\partial \eta_{ik}} + \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}_{ij}}\frac{d\boldsymbol{x}_{ij}}{d\eta_{ik}} \right) \tag{3.19}$$

therefore derivatives of the state variables with respect to the random effects are needed. We also need the second derivatives of the likelihood function, therefore we need also second derivatives of the state variables with respect to the random effects. Additionally mixed derivatives of the state variables with respect to a component of the fixed parameter vector and a random effect are also needed. These equations that express the change in the state variable with respect to the model parameters are called sensitivity equations. The approach of estimating the gradient with the analytic solution, rather than with numerical approximation, introduced in Almquist et al. [14], allows for an improved gradient precision and for reduced computational times.
Not all the parameter estimation algorithms use the same algorithm. In order to proceed with the calculations in a faster way, some algorithms apply the first order approximations, that neglects the terms containing the second derivatives. The FO (First order approximation) calculates the likelihoods by using $\boldsymbol{\eta}_i = \mathbf{0}$ for all the individuals, since this is the population mean.
The FOCE (First order conditional estimation) instead calculates the likelihood at a specific value of $\boldsymbol{\eta}_i$ that maximizes the individual likelihood, but neglect the dependency of the $\boldsymbol{R}$ matrix on the random effects. The FOCE-I (FOCE with interaction) takes into account this dependency.

The following are the steps of the optimization algorithm. The second step shall be continued until an established convergence criterion is met.

1. Preliminary step:
   (a) initialize the values of the fixed parameters
   (b) set the individual random parameters to be 0
2. Fixed parameter search: repeat this step until a chosen convergence criteria is verified
   (a) Inner Optimization for each individual:
      i. find the random parameter maximizing the individual log-likelihood function
   (b) find the population parameters maximizing the population likelihood, based on the Laplace approximation

SAEM algorithm

We have used the software Monolix to perform our analysis.
Monolix uses a different algorithm to obtain the parameter set that maximizes the likelihood function. The algorithm, called SAEM (Stochastic Approximation of Expectation Maximization) is a modified version of the EM (Expectation Maximization) algorithm.
The goal of the EM algorithm, usually used for incomplete data, is to obtain the maximum likelihood parameter estimate. The EM algorithm consists of two steps, iterated until convergence is reached:

1. E-step: we calculate a conditional expectation of the likelihood with respect to the current conditional distribution of the latent variables given the state vector and the current value for the fixed parameter vector .
2. M-step: we maximize the function obtained in the E-step in order to set the new value for the next iteration of the algorithm.

The random effects, which cannot be observed, can be considered as latent variables, therefore the EM algorithm is a suitable solution for parameter estimation in NLMEM. The SAEM algorithm presents a modification of the EM algorithm, with a stochastic twist: it adds a simulation step in the expectation phase. The full algorithm and proof of its convergence can be found here [60].
The parameter search implements a Monte Carlo Markov Chain (MCMC) method to draw samples from the conditional distributions. There are two phases of the parameter search: the exploratory phase, which aims at exploring the parameter space to find a good first approximation of the parameter set and the smoothing phase, that ensures a smoother convergence to the parameter

vector that maximizes the population likelihood. Here we report the algorithm for the fixed effects parameters for a simple model, where the parameters for inter-individual variability and error model have been fixed therefore the vector $\boldsymbol{\theta}$ containing the parameters that shall be estimated only encompasses the fixed population parameters $\boldsymbol{k}$.
The subscripts for $\boldsymbol{\theta}$ in the algorithm equations will signify the two phases of the algorithm (E stands for exploration and S for smoothing). Here is how the complete algorithm works:

1. Preliminary step:
   (a) initialize the values of the parameters $\boldsymbol{\theta}_E^0$
2. Exploratory phase: for l in $\{1, \ldots, M_E\}$
   (a) draw $\boldsymbol{\psi}_i$ from $p\left(\boldsymbol{\psi}|\boldsymbol{d}_i, \boldsymbol{\theta}_E^l\right)$ for i in $1, ..., N$ with MCMC
   (b) estimate the new population parameters through

$$\boldsymbol{\theta}_E^{l+1} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\psi}_i \tag{3.20}$$

   (c) update l: $l = l + 1$
3. Smoothing phase: set $l = 0$ and continue until convergence is reached:
   (a) $\boldsymbol{\theta}_S^0 = \boldsymbol{\theta}_E^{M_E}$ where M is the number of iterations in the exploratory phase
   (b) draw $\boldsymbol{\psi}_i$ from $p\left(\boldsymbol{\psi}|\boldsymbol{d}_i, \boldsymbol{\theta}_S^l\right)$ for i in $1, ..., N$ with MCMC
   (c) estimate the new population parameters through:

$$\boldsymbol{\theta}_S^{l+1} = \boldsymbol{\theta}_S^l + \frac{1}{l}\left[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\psi}_i^{(l)} - \boldsymbol{\theta}_S^l\right] \tag{3.21}$$

   (d) update l: $l = l + 1$

3.1.4 Identifiability and observability

The measurement functions are not always enough to uniquely determine the parameters or the states of a model.
That's where the study of identifiability and observability analysis comes into play. Hermann et al. in [61] and Villaverde in [62] clearly outline these concepts, that we report here.
A dynamical system can be described as a system of differential equations with an initial condition and a measurement function, as in equations 3.1, 3.2 and 3.6. A model is globally observable when the initial condition uniquely determines

the measurement function vector.
In mathematical terms we can state it as follows:

$$\begin{aligned} \forall \boldsymbol{x}_0, \boldsymbol{k} \text{ with } \frac{d\boldsymbol{x}^0}{dt} &= \boldsymbol{f}(\boldsymbol{x}^0(t), t, \boldsymbol{k}) \text{ and } \boldsymbol{x}^0(0) = \boldsymbol{x}_0(\boldsymbol{k}) \\ \nexists \boldsymbol{x}_1, \boldsymbol{k}_1 \text{ with } \frac{d\boldsymbol{x}^1}{dt} &= \boldsymbol{f}(\boldsymbol{x}^1(t), t, \boldsymbol{k}_1) \text{ and } \boldsymbol{x}^1(0) = \boldsymbol{x}_1(\boldsymbol{k}_1) \\ \text{and } \boldsymbol{x}_1(\boldsymbol{k}_1) \neq \boldsymbol{x}_0(\boldsymbol{k}) &\text{ and } \boldsymbol{k}_1 \neq \boldsymbol{k} \text{ and } \boldsymbol{h}(\boldsymbol{x}^0(t)) = \boldsymbol{h}(\boldsymbol{x}^1(t)) \; \forall t \geq t_0 \end{aligned} \tag{3.22}$$

The parameter k^e, e-th component of $\boldsymbol{k}$, of a model is locally structurally identifiable if for all the parameter sets, there is no parameter vector with a different value for k^e in a neighbourhood of this parameter, so that the value of the measurement vector is the same for all the time points.
In mathematical terms we can write it as:

$$\begin{aligned} \forall\, \boldsymbol{x}_0, \boldsymbol{k} \,\exists\, U(\boldsymbol{k}) \text{ such that } \nexists\, \boldsymbol{k}_1 \text{ in } U(\boldsymbol{k}) \text{ with } k^e \neq k_1^e \text{ such that} \\ \boldsymbol{h}(\boldsymbol{x}^0(t)) = \boldsymbol{h}(\boldsymbol{x}^1(t)) \;\forall\, t \geq t_0 \text{ with} \\ \frac{d\boldsymbol{x}^0}{dt} = \boldsymbol{f}(\boldsymbol{x}^0(t), t, \boldsymbol{k}) \text{ and } \boldsymbol{x}^0(0) = \boldsymbol{x}_0(\boldsymbol{k}) \text{ and} \\ \frac{d\boldsymbol{x}^1}{dt} = \boldsymbol{f}(\boldsymbol{x}^1(t), t, \boldsymbol{k}_1) \text{ and } \boldsymbol{x}^1(0) = \boldsymbol{x}_1(\boldsymbol{k}_1) \end{aligned} \tag{3.23}$$

A model is locally structural identifiable if all the parameters are locally structural identifiable.
Local structural identifiability can be seen as a special case of observability, if we consider for each component of the parameter vector $\boldsymbol{k}$ additional states k_i^e with $\frac{dk_i^e}{dt} = 0$.
The identifiability and observability analysis can be performed by a software. Our choice was the software presented in [63].

3.1.5 Model diagnostics and model selection criteria

Diagnostics through comparison of model approximation and observed data
There are several tools for model evaluation based on the comparison of the observed data with the model-based approximation.
Individual fit The easiest way to do such a comparison is plotting for each individual the model-generated curve together with the observed data. This is a good way of checking whether the model could fit well the data for the specific individual. If we look at this graph for each of the individuals we can check whether the model can fit all of them.
When using this tool for model selection, care must be applied. Usually focusing just on one individual and its potential worse fit might not be a good measure to choose one model over another. The choice based on this graph can be justified only if this worse fit is consistent across many individuals. Very similar

models, on the other side, do not provide fits which are consistently worse for many individuals. Therefore the individual fit is usually not a sufficient tool to discriminate between two similar models.
When the number of individuals in the sample is high, checking the individual graphs becomes expensive. Two additional drawbacks of this diagnostic tool are that it is difficult to grasp from these graphs whether there exists a tendency of the model to overestimate or underestimate the observed data and whether the model can describe the average behaviour of the population. Two additional graphs, observed vs predicted data and tracer curves, respectively address the first and second newly cited drawbacks. Since they both consider all the individuals simultaneously, they are also less expensive to look at.
Observed vs predicted data This scatter plot shows on the x axis the observed data and on the y axis the model prediction. A $y = x$ line enables to compare whether the observed and predicted values are similar. This graph can highlight whether the observed data points are consistently underestimated or overestimated.
Tracer curves For each of the time points the average observed data and the standard deviation is calculated across the individuals. The individual model-predicted curves are averaged. The graph shows the mean ($\pm$ standard deviation) for each time point and the averaged individual curves , therefore it indicates whether the model can predict the average behaviour of the population.
This graph does not provide specific information about the individual fit.

Visual Predictive Check While the individual fit graph can highlight the goodness of fit for a single individual and the tracer curves show if the model can explain the average behaviour of the individuals, the Visual Predictive Check (VPC) does not take into account the individual parameters but it simulates data to check whether the observed median, the 90% percentile and the 10% percentile are inside of the confidence intervals of the respective percentiles originated with simulated data. The simulated data takes into account the statistical model, the inter-individual variability and the error model.
The time points of the observed data are divided into different bins. New individuals are simulated from the distributions and 10%, 50% and 90% percentile are calculated for the chosen bins. A 90% confidence interval is then built for these percentiles for the bins. The same percentiles are calculated for the observed data for the same bins and a graph is built to check whether the observed percentiles for different time points are inside of the confidence interval of the simulated percentiles.
This graph is informative, because it shows whether the observed percentiles are in line with the ones from the simulated data. This tells whether the observed data can be statistically expected from the model, so the information delivered in this graph is much deeper than the one obtained from the tracer curve, that only explains whether the average population behaviour can be described by averaging the individual curves.
In case the individuals are subject to different dosages, the VPC might not

be informative. Therefore a prediction-corrected VPC has been introduced to normalize the value in the bins [64] and keep the predictive value of the VPC.

Akaike information criterion (AIC) The AIC takes into account m, the number of parameters of the model and $ln(L^*)$, the logarithm of the maximum likelihood estimate:

$$AIC = 2m - 2ln(L^*) \tag{3.24}$$

The connected strategy is to choose the model that minimizes the AIC. Models with a low number of parameters and high value of the likelihood function are advantaged. A principle of parsimony is perfectly married to the maximum likelihood principle: if two models have the same maximum likelihood value, the one with a lower number of parameters will be chosen and if two models have the same number of parameters, the one with the highest maximum likelihood will be chosen.
The AIC springs from information theory and can be seen as a modification to the Kullback-Leibler divergence, that measures the discrepancy of a model with respect to the true function that generated the data. More details on the theoretical foundations of the principle can be found in [65] and in [66], where the principle is defined.

Checking the range of the parameters NLMEM describes the individual behaviour as a slight variation on the average population behaviour. After obtaining the individual estimates, it is good practice to check their standard deviation. In fact, in some cases, the abundance of parameters with random effects could have the variability gravitate in a cluster of parameters and reduce the variability of other parameters. One shall therefore check whether the obtained range of parameters is in line with the expected range.

Additional model selection guiding measures The choice of the model is not only concerning the structural model. More specific choices shall be made on the statistical model: the error model and the statistical model for the population parameters, which entails the choice of the variance-covariance matrix and of the distribution for the parameters. Therefore the three following measures are also important:
Relative standard error of the error model parameters The confidence interval range for a parameter is dependent on the standard error, therefore a big relative standard error (r.s.e) indicates a big uncertainty in the parameter estimates. The model might be too complex to support for a more certain estimate of the parameter, therefore parameters with a small value and big r.s.e. are good candidates to be removed from the error model to simplify it.
Correlation between the random effects The standard choice in Monolix is to have a diagonal variance-covariance matrix. If a specific test, described later, shows that two random effects might be correlated, one can modify the variance-covariance matrix structure and add the hypothesized correlation couple.

Normality of the random effects The model rests on the assumption that the random parameter have a normal distribution therefore one has to check whether this assumption truly holds.
These features will be explained more in detail during the model selection process.

Motivation behind the choice of the figures and tables shown

The different statistical models analysed in Section 4.3 do not show significant differences in the individual fit of the model to the data. Therefore the only individual curve shown is one comparing the STS estimation with the NLMEM estimation.
The lack of discrepancy between the different models is shown through the overlapping of the tracer curves. Other curves (observable vs predicted or VPC) are shown when they provide us further elements to lead to a decision on the model.

3.2 Data and tools

In this section, we present the datasets and tools that have made the findings in this thesis possible. The specific experiments analysed will be presented shortly. Mathematical modelling is the tool to make inference on the biologically interesting questions: a software is needed to perform the analysis. Space will be given to it in this section.

3.2.1 Datasets

The study design and experimental procedure is not in the scope of this thesis. The experimental setup of the tracer-tracee study is explained in Section 1.4 and mathematically described in Section 4.1, therefore the description of the datasets in this section will be short, briefly describe the treatment and focus on the quality of the datasets.

Fructose This study is part of a bigger study [53].
Before and after the 12 weeks of intervention, when 75 g of fructose were consumed daily, the subjects underwent the tracer-tracee experiment. They received a fat-rich meal, 2 hours after being injected with a bolus of leucine (time 0) and blood samples were taken at different time points. The chemical analysis of the samples, explained in detail in the Appendix, resulted in the time-series data used for the parameter estimation.
Table 3.1 shows the composition of the data for the different apolipoproteins. While apoC-III and apoE kinetic parameters analysis can rely on 20 individuals,

	apoA4 visit 1	apoA4 visit 2	apoC-III visit 1	apoC-III visit 2	apoE visit 1	apoE visit 2
n Subjects	17.0	17.0	20.0	20.0	20.0	20.0
tot ape(leucine)	304.0	304.0	358.0	358.0	358.0	358.0
av ape(leucine)	17.9	17.9	17.9	17.9	17.9	17.9
tot ape(apo)	94.0	99.0	163.0	168.0	171.0	180.0
av ape(apo)	5.5	5.8	8.2	8.4	8.6	9.0
tot apo pool	169.0	168.0	197.0	196.0	193.0	199.0
av apo pool	9.9	9.9	9.8	9.8	9.6	10.0
tot points	567.0	571.0	718.0	722.0	722.0	737.0
av points	33.4	33.6	35.9	36.1	36.1	36.8

Table 3.1: For each apolipoprotein and for each visit, number of individuals, time points, average number of time points for the different measurements and total number of data points for fructose treatment data; ape stands for atom percent excess: while ape(leucine) indicates m_1 in equation 4.40, ape(apo) indicates m_2 as described in equation 4.41

only 17 individuals are suitable for the kinetic study of apoA4. As will be clear in Section 4.1, where the model and the measurement functions will be described, the measurement denoted with ape(apo) is the key measurement for the estimation of the kinetic parameters. The average number of points for this measurement is low in apoA4 for both visits, therefore the quality of data for apoA4 is low.

Evolocumab The dataset is part of a bigger study [67].
Before and after 12 weeks of evolocumab treatment, the subjects received a fat-rich meal and a bolus injection of leucine. The chemical analysis of the blood samples extracted after the injection generated time series data for the different individuals. 14 subjects fulfilled the study criteria and of these, 13 individuals had a suitable data configuration to perform the kinetic analysis.
Conjunctively with the evolocumab treatment, the subjects introduced statin and metformin. Table 3.2 describes the data for the different apolipoproteins in terms of number of datapoints. Similarly to the fructose study, the low number of average points for ape(apo) implicates a low quality of data for apoA4.

Liraglutide The dataset is part of a previous study [54].
Before and after 16 weeks of liraglutide treatment, on background of metformin and statins, the subjects underwent the tracer-tracee experiment. Analysis of the chemical composition of the data generated the times-series data used for the parameter estimation.
The composition of the data in terms of abundance of data points is described in Table 3.3 for the different apolipoproteins. Consistently with the fructose and evolocumab experiments, the average number of data points for apoA4 is lower than the ones for apoC-III and apo-E, therefore the quality of data for apoA4 is lower than the one for apoC-III and apoE.

	apoA4 visit 1	apoA4 visit 2	apoC-III visit 1	apoC-III visit 2	apoE visit 1	apoE visit 2
n Subjects	13.0	13.0	13.0	13.0	13.0	13.0
tot ape(leucine)	214.0	209.0	214.0	209.0	214.0	209.0
av ape(leucine)	16.5	16.1	16.5	16.1	16.5	16.1
tot ape(apo)	80.0	77.0	100.0	97.0	109.0	109.0
av ape(apo)	6.2	5.9	7.7	7.5	8.4	8.4
tot apo pool	128.0	129.0	126.0	126.0	128.0	129.0
av apo pool	9.8	9.9	9.7	9.7	9.8	9.9
tot points	422.0	415.0	440.0	432.0	451.0	447.0
av points	32.5	31.9	33.8	33.2	34.7	34.4

Table 3.2: For each apolipoprotein and for each visit, number of individuals, time points, average number of time points for the different measurements and total number of data points for evolocumab treatment data; ape stands for atom percent excess: while ape(leucine) indicates m_1 in equation 4.40, ape(apo) indicates m_2 as described in equation 4.41

3.2.2 Monolix

The software used for the estimation of the parameters is Monolix2019R2. More on the algorithm that is implemented in the software has been described in the Section 3.1.3.

	apoA4 visit 1	apoA4 visit 2	apoC-III visit 1	apoC-III visit 2	apoE visit 1	apoE visit 2
n Subjects	12.0	12.0	13.0	13.0	13.0	13.0
tot ape(leucine)	214.0	214.0	232.0	232.0	232.0	232.0
av ape(leucine)	17.8	17.8	17.8	17.8	17.8	17.8
tot ape(apo)	86.0	74.0	95.0	100.0	108.0	112.0
av ape(apo)	7.2	6.2	7.3	7.7	8.3	8.6
tot apo pool	120.0	120.0	130.0	130.0	130.0	129.0
av apo pool	10.0	10.0	10.0	10.0	10.0	9.9
tot points	420.0	408.0	457.0	462.0	470.0	473.0
av points	35.0	34.0	35.2	35.5	36.2	36.4

Table 3.3: For each apolipoprotein and for each visit, number of individuals, time points, average number of time points for the different measurements and total number of data points for liraglutide treatment data; ape stands for atom percent excess: while ape(leucine) indicates m_1 in equation 4.40, ape(apo) indicates m_2 as described in equation 4.41

4 Choice of the model

In this chapter, the proposed model will be introduced in Section 4.1 and in Section 4.2 it will be used to analyse the kinetics of the three apolipoproteins (A4, C-III and E); here the two alternative methods of estimation (STS method and NLMEM) will be compared. In Section 4.3, the analysis will focus on specific choices on the structural and statistical model and in Section 4.4 the model will be validated. Some of the choices can be generalized for the three apolipoproteins, while for the remaining ones, a procedure will be offered to navigate among the different options, so that the choice will be perfectly suitable for the dataset under analysis.

4.1 Proposed model

Motivation

The model developed in [48] to describe apolipoprotein C-III kinetics, has been previously used to describe apoE-VLDL kinetics as well [51]. We suggest that with some adjustments, this model can be used for apoC-III, total apoE plasma kinetics and apoA4 kinetics.

Approach

- introduce some modifications to the model in [48]
- segment the datasets so that one part can be used for model development and the remaining part can be used for model validation
- decide between the two approaches for parameter estimation (STS and NLMEM)
- investigate choices for the structural model and statistical model in Section 4.3.
- in Section 4.4 validate the model with the choices established in Section 4.3

Since leucine is a building block of all the synthetized proteins, this amino acid is a common choice to take track of the synthesis and elimination of proteins. In fact we use an isotope of this molecule (labelled molecule) to study

the kinetics of the apolipoproteins. By tracking the quantity of free labelled leucine and of labelled leucine in the apolipoproteins and comparing these to the natural occurring quantities we can extract information on the whole kinetics of apolipoproteins. Production rate, time delay until the release in the blood stream and clearance from the blood can be inferred by analysing the time series measurements.
Important to notice is that the inference on the lifecycle of apolipoproteins is only possible through the combination of tracer/tracee experiments, mathematical modelling and statistical inference, since the journey of the leucine molecule can be described in a mathematical language and statistics bridges this description with the experimental data in order to obtain a quantitative understanding of the system.

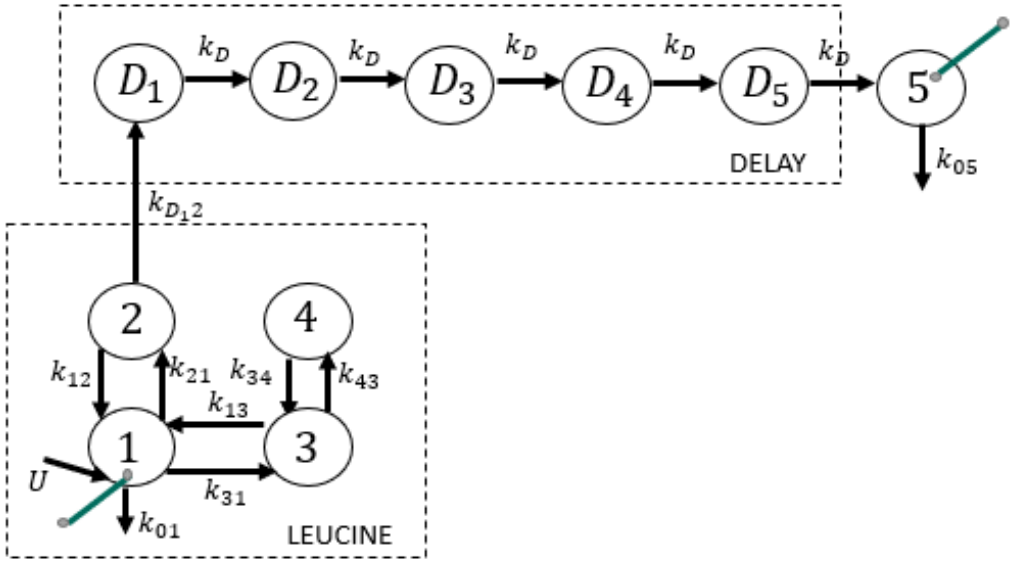

Figure 4.1: Model for endogenous leucine, from its free state in the blood (compartment 1) to the release into the blood as part of the apolipoproteins (compartment 5)

Compartmental models for tracer-tracee experiments We describe the kinetics of the naturally occurring leucine for the i-th individual with a system of differential equations:

$$\frac{d\boldsymbol{Q}_i(t)}{dt} = \boldsymbol{f}(\boldsymbol{Q}(t), t, \boldsymbol{k}_i) + \boldsymbol{U}_i \tag{4.1}$$

$$\boldsymbol{Q}_i(t_0) = \boldsymbol{Q}_{i,0}(\boldsymbol{k}_i) \tag{4.2}$$

We assume that the insertion of the labelled leucine q does not perturb the behaviour of the system, therefore the total quantity $q + Q$ will follow the same dynamics of Q alone. The only difference between Q and q consists in the input. While the blood compartment of natural occurring leucine Q_1 is always replenished with a constant input U (in compact vectorial form denoted by $\boldsymbol{U}$,

see Figure 4.1), the labelled leucine input only occurs at the initial time point. The stable-isotope of leucine is administered with an injection and this implies a quick distribution of the material, therefore all the quantity administered can be thought of as concentrated at the initial time. In light of this, we can incorporate the information about the dosage in the initial condition.

$$\frac{d\left(\boldsymbol{Q}+\boldsymbol{q}\right)_i(t)}{dt} = \boldsymbol{f}\left(\left(\boldsymbol{Q}+\boldsymbol{q}\right)_i(t), t, \boldsymbol{k}_i\right) + \boldsymbol{U}_i \tag{4.3}$$

$$\left(\boldsymbol{Q}+\boldsymbol{q}\right)_i(t_0) = \left(\boldsymbol{Q}+\boldsymbol{q}\right)_{i,0}(\boldsymbol{k}_i) \tag{4.4}$$

When describing biochemical mechanisms, a linear differential equation is often enough to describe the dependency of the change in the substrate concentration with respect to its concentration. Biochemical reactions are subject to saturation if too much material overloads the enzymatic action; in this case one would need to switch from linearity to a more complex mathematical description (for example Michaelis-Menten kinetics). The tracer-tracee study is designed to avoid saturation, therefore the inserted quantity q will be much smaller than the naturally occurring quantity Q ($q << Q$).
This ensures the linearity of $\boldsymbol{f}$ with respect to the state variables and the parameters $\boldsymbol{k}_i$.
Because of linearity, the state variables q and Q will appear separately and not be mixed, therefore we can split the system of differential equations in two twin systems, one where only Q appears (tracee equations) and the other where only q appears (tracer equations), that will be solved together with the respective initial conditions. The assumption that the naturally occurring quantities are at steady state implies that the change over time is zero, therefore the system of differential equation for the natural leucine Q will turn into a system of equations 4.6.

$$\frac{d\boldsymbol{q}_i(t)}{dt} = \boldsymbol{f}\left(\boldsymbol{q}_i(t), t, \boldsymbol{k}_i\right) \tag{4.5}$$

$$0 = \boldsymbol{f}\left(\boldsymbol{Q}_i(t), t, \boldsymbol{k}_i\right) + \boldsymbol{U}_i \tag{4.6}$$

$$\boldsymbol{q}_i(t_0) = \boldsymbol{q}_{i,0}(\boldsymbol{k}_i) \tag{4.7}$$

$$\boldsymbol{Q}_i(t_0) = \boldsymbol{Q}_{i,0}(\boldsymbol{k}_i) \tag{4.8}$$

The model used for the description of the kinetics of leucine from the free state to its being part of the apolipoproteins, shown in Figure 4.1, is based on the model used in [48] to describe the kinetics of apolipoprotein C-III. Because of the lack of mathematical details in the paper where the model is presented, we hypothesize that the model is congruent to the one in Figure 1(B) in the article, therefore we conclude that two of our choices, pointed out below, are modifications of the original model. Before elucidating them, let us explain the meaning of the different compartments.

Leucine kinetics Free leucine kinetics can be described with 4 compartments. Compartment 1 represents free leucine in the blood; the different phases of the leucine decay curve suggest that there is an exchange with a free leucine pool, represented by compartments 3 and 4. Compartment 2 has a fast material exchange with compartment 1 and is where free leucine will flow before entering the liver. In order to reduce the number of parameters in the model, some relations among the parameters of the system are fixed.

Population kinetics Apolipoprotein E, C-III and A4 are released in the blood after being produced in the liver (apoE and apoC-III) and in the intestine (apoA4). Protein synthesis and secretion from free leucine is represented by a series of 5 delay compartments. The delay is connected to a compartment representing the apolipoproteins in the blood.
Our datasets comprehend only measurements of apolipoproteins, but do not specify whether the apolipoproteins are free or located in any specific lipoprotein group (LDL, VLDL, IDL or HDL). One could hypothesize a more complex model where the compartments representing the apolipoproteins release into the blood are two or more and these could indicate the different lipoproteins group (for example VLDL1 and VLDL2). The resolution of our data is not enough to support a more complex model, therefore we choose to settle for the one here described, where only one compartment represents the apolipoproteins in the blood.

Model equations Two relationships are enforced to limit the number of possible parameters, as done in [68]:

$$k_{21} = k_{12} \tag{4.9}$$

$$k_{34} = 0.1k_{43} \tag{4.10}$$

The tracer kinetics is described by the following system of differential equations:

$$\frac{dq_1}{dt} = k_{13} * q_3(t) + k_{12} * q_2(t) - (k_{12} + k_{01} + k_{31}) * q_1(t) \tag{4.11}$$

$$\frac{dq_2}{dt} = k_{12} * q_1(t) - (k_{11} + k_{D_12}) * q_2(t) \tag{4.12}$$

$$\frac{dq_3}{dt} = k_{13} * q_1(t) + 0.1 * k_{43} * q_4(t) - (k_{13} + k_{43}) * q_3(t) \tag{4.13}$$

$$\frac{dq_4}{dt} = k_{43} * q_3(t) - 0.1 * k_{43} * q_4(t) \tag{4.14}$$

$$\frac{dq_{D_1}}{dt} = k_{D_12} * q_2(t) - k_D * q_{D_1}(t) \tag{4.15}$$

$$\frac{dq_{D_2}}{dt} = k_D * q_{D_1}(t) - k_D * q_{D_2}(t) \tag{4.16}$$

$$\frac{dq_{D_3}}{dt} = k_D * q_{D_2}(t) - k_D * q_{D_3}(t) \tag{4.17}$$

$$\frac{dq_{D_4}}{dt} = k_D * q_{D_3}(t) - k_D * q_{D_4}(t) \tag{4.18}$$

$$\frac{dq_{D_5}}{dt} = k_D * q_{D_4}(t) - k_D * q_{D_5}(t) \tag{4.19}$$

$$\frac{dq_5}{dt} = k_D * q_{D_5}(t) - k_{05} * q_5(t) \tag{4.20}$$

$$\tag{4.21}$$

The following system describes the tracee kinetics. In compartment 1 there is a constant input of material.

$$\frac{dQ_1}{dt} = U + k_{13} * Q_3(t) + k_{12} * Q_2(t) - (k_{12} + k_{01} + k_{31}) * Q_1(t) \tag{4.22}$$
$$\frac{dQ_2}{dt} = k_{12} * Q_1^{toliver}(t) - (k_{12} + k_{D_12}) * Q_2(t) \tag{4.23}$$
$$\frac{dQ_3}{dt} = k_{13} * Q_1(t) + 0.1 * k_{43} * Q_4(t) - (k_{13} + k_{43}) * Q_3(t) \tag{4.24}$$
$$\frac{dQ_4}{dt} = k_{43} * Q_3(t) - 0.1 * k_{43} * Q_4(t) \tag{4.25}$$
$$\frac{dQ_{D_1}}{dt} = k_{D_12} * Q_2(t) - k_D * Q_{D_1}(t) \tag{4.26}$$
$$\frac{dQ_{D_2}}{dt} = k_D * Q_{D_1}(t) - k_D * Q_{D_2}(t) \tag{4.27}$$
$$\frac{dQ_{D_3}}{dt} = k_D * Q_{D_2}(t) - k_D * Q_{D_3}(t) \tag{4.28}$$
$$\frac{dQ_{D_4}}{dt} = k_D * Q_{D_3}(t) - k_D * Q_{D_4}(t) \tag{4.29}$$
$$\frac{dQ_{D_5}}{dt} = k_D * Q_{D_4}(t) - k_D * Q_{D_5}(t) \tag{4.30}$$
$$\frac{dQ_5}{dt} = k_D * Q_{D_5}(t) - k_{05} * Q_5(t) \tag{4.31}$$

where $Q_1^{toliver}$ indicates all the leucine that flow to the liver, since this is not limited to the leucine in the blood. Since the tracee system is in steady state, the system of differential equations becomes a system of linear equations and solving for the state variables (obtaining the analytical solution for the linear system of equations) provides the quantities of the tracee at steady state as a function of the model parameters.

$$Q_1^{toliver} = \frac{Q_1}{p_1} \tag{4.32}$$
$$Q_2 = \frac{k_{12} * Q_1^{toliver}}{k_{D_12} + k_{12}} \tag{4.33}$$
$$Q_{D_1} = \frac{k_{D_12} * Q_2}{k_D} \tag{4.34}$$
$$Q_{D_2} = Q_{D_1} \tag{4.35}$$
$$Q_{D_3} = Q_{D_2} \tag{4.36}$$
$$Q_{D_4} = Q_{D_3} \tag{4.37}$$
$$Q_{D_5} = Q_{D_4} \tag{4.38}$$
$$Q_5 = \frac{k_D * Q_{D_5}}{k_{05}} \tag{4.39}$$

Since the free leucine pool in the blood is not the only source of leucine for

the assembly of apolipoproteins, we insert p_1 in the equation, which indicates the proportion of leucine in the apolipoproteins coming from the above cited source. We assume that this proportion is tracer-free.

The measurement functions for the tracer-tracee study can be described as follow:

$$m_1(t) = 100\frac{q_1(t)}{q_1(t) + Q_1} \tag{4.40}$$

$$m_2(t) = 100\frac{q_5(t)}{q_5(t) + Q_5} \tag{4.41}$$

$$m_3(t) = Q_5 \tag{4.42}$$

Differences between the model and the original model The proposed model is based on the one suggested in [48]; two are the main differences:

1. **Parameter** k_{01}: in the original model, no exit was contemplated from the compartment representing the free leucine in the blood, supposedly due to a different approach to data analysis: supposedly they could have used use the first two measurement functions (in equations 4.40 and 4.41) for the parameter estimation and subsequently adjusted the results with the pool size. We introduce the parameter k_{01}, which can also be found in [49] in a model describing the apoC-III kinetics.

2. **Analytic solution for** Q_5: In the measurement functions Q_1 and Q_5 appear. Apparently, one could have chosen Q_1 and Q_5 as model parameters and it would have not been necessary to introduce the steady state approximation to write Q_5 as a function of Q_1. Instead in our model Q_1 is a parameter and an additional parameter p_1 is introduced. While this could look like an unnecessary move, the choice is justified by the fact that Q_5 is measured, therefore we choose not to enforce a distribution on it. The validity of this choice will be proven when in the Section 4.3 we will compare the Q_1 and the Q_5 model. Having Q_5 as a parameter worsens the predictability of the model.

Biological parameters We are interested in knowing how fast the apolipoproteins are produced and how fast they are removed from the blood stream. This is what determines their concentration, which is also the object of our interest. Therefore while continuing to look at the other parameters from a mathematical modelling perspective, the true biological concern will be regarding the fractional catabolic rate, the pool size and the secretion rate. In the study time is measured in hours (h). The dosage, initial condition of q_1 in the system of differential equation, is provided in mg leucine/kg; all the quantities of the system are therefore normalized with the weight. The state variables in the

system have therefore unity of measure mg leucine/kg. In order to obtain the biologically relevant parameters with the right unity of measure, we need to rescale the parameters of the system. Let $R_{\frac{leu}{apo}}$ be the amount of mg of leucine in each mg of apolipoprotein. Let 0.45 be the conversion rate between body weight in kg and dL of blood in the body.

$$FCR = k_{05} * 24 \tag{4.43}$$

$$PS = Q_5/R_{\frac{leu}{apo}}/0.45 \tag{4.44}$$

$$SR = k_{05} * Q_5 * 24/R_{\frac{leu}{apo}} \tag{4.45}$$

Now the unities of measure for FCR, PS and SR are respectively pools/day, mg apolipoprotein/dL, mg apolipoprotein/(day kg).

Identifiability and Observability analysis of the model The model is locally structurally identifiable, which means that all the parameters are locally structurally identifiable. The parameters attaining the leucine pool are nonobservable and so are the state variables for these compartments.

Segmenting the data for model selection and model validation The total available data consists of three datasets. Therefore we segment it in two parts: the dataset regarding the fructose intervention experiment will be used for the choice of the method (Section 4.2) and model development (Section 4.3) and the experiments concerning the evolocumab and liraglutide treatments will be used for model validation purposes (Section 4.4). Since the first dataset is more abundant, we aim at discovering more possible features through this dataset and then at transferring this knowledge to the more restricted second and third dataset, from which characteristic features might not be noticed.

4.2 Choice of the method

Motivation
Scrutinize different methods for the parameter estimation

Approach

- investigate NLMEM vs STS for leucine model
- investigate NLMEM vs STS vs mixed approach for the complete apolipoprotein data

Result

- results for rich datasets are comparable for STS and NLMEM, even though STS shows a bigger variability
- in case of poor datasets, NLMEM exploits information about the population behaviour to help fitting the data, therefore rendering an improved fit.

For the leucine compartments

We have performed the STS estimation and the NLMEM estimation first by taking into account only the data corresponding to the measurements of the enrichment (the ratio of tracer over the whole amount (tracer + tracee)) of free leucine (equation 4.40).
The compartmental model for the free endogenous leucine alone consists of the compartments 1-4 of the apolipoprotein model (see Figure 3.1 at page 18).
For identifiability purposes [56], the removal rate from compartment 2 k_{02} had been fixed to 0.011. We can assume that the system describing endogenous leucine is in steady state, while the one for the tracer, labelled leucine, will be described with a similar system of equations without natural leucine replenishment U, and therefore we will choose to consider Q_1 (instead of U) as a parameter of the model.
By analysing the results in the tables and figures, we can make the observations:

- the standard deviation of the parameter estimate for the STS method is much larger (see Table 4.1). This is confirmed in the scatter plot of the individual parameter estimates (Figure 4.2).

- NLMEM have slightly more dispersed values in the observed vs predicted data (Figure 4.3). The lack of structure for the parameters leads STS to overfitting, while this cannot be true for NLMEM, since the parameters must belong to the same distribution.

parameter	STS mean	NLMEM mean	STS std	NLMEM std
k_{01}	3.54	3.23	1.08	0.31
k_{12}	3.15	2.45	1.69	0.48
k_{13}	1.73	2.35	0.81	0.24
k_{31}	2.79	2.33	1.43	0.13
k_{43}	1.04	0.91	1.16	0.01
Q_1	1.97	2.1	0.67	0.32

Table 4.1: Mean and standard deviation of the kinetic parameters for the leucine model estimated with STS and with NLMEM. Unity of measure for Q_1 is mg leucine/kg, while for the remaining parameters is 1/h

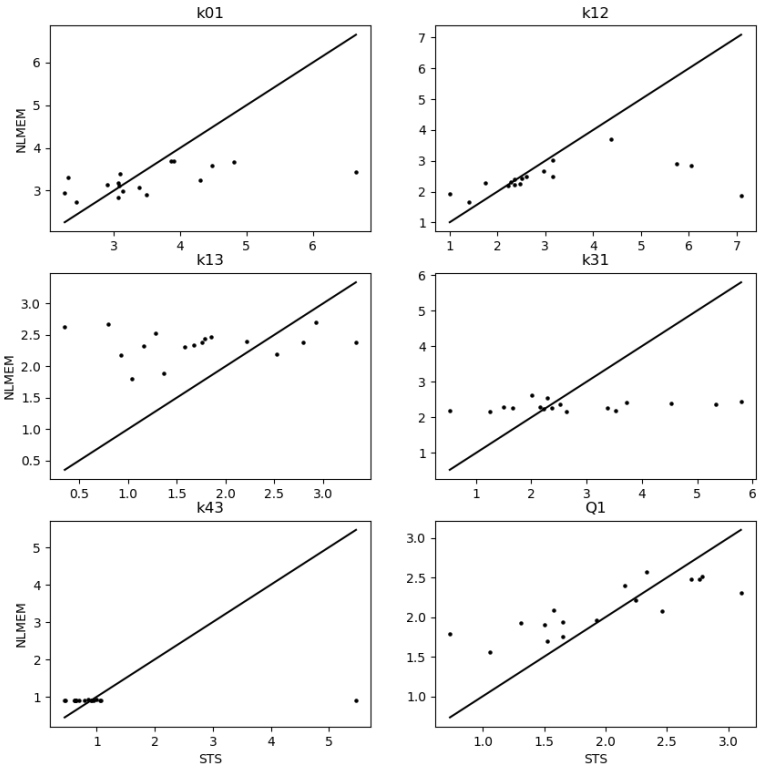

Figure 4.2: Scatter plot of the individual parameter estimates for the free leucine compartment

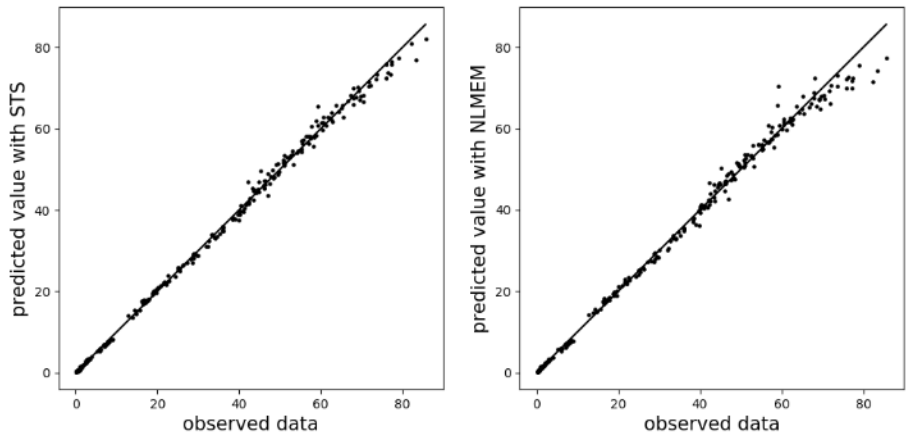

Figure 4.3: Plot of observed vs predicted data with the two estimation methods for the free leucine compartment measurement (equation 4.40) at visit one

parameters	STS mean	NLMEM mean	mixed mean	STS std	NLMEM std	mixed std
apoA4 FCR	3.62	3.81	3.85	1.49	0.51	0.02
apoA4 PS	5.39	5.4	5.4	1.46	1.34	1.43
apoA4 SR	8.6	9.06	9.34	4.26	1.8	2.44
apoC-III FCR	1.51	1.5	1.46	0.54	0.4	0.29
apoC-III PS	6.65	6.62	6.65	2.83	2.83	2.95
apoC-III SR	4.1	4.12	4.21	1.24	1.1	1.67
apoE FCR	4.65	4.61	4.6	1.18	0.9	1.04
apoE PS	2.81	2.81	2.83	1.19	1.25	1.3
apoE SR	5.56	5.59	5.66	1.73	1.93	2.17

Table 4.2: Mean and standard deviation of the kinetic parameters for the different apolipoproteins estimated with STS, NLMEM and mixed approach. FCR, PS and SR have respectively unity of measure pools/day, mg apolipoprotein and mg apolipoprotein/day

For the full model

Here we compare three alternative methods. Additionally to the STS method and NLMEM framework, we will also use a mixed approach, where all the parameters, apart from the parameters directly related to the compartment 5, will be estimated with STS. Only the elimination rate of apolipoproteins (k_{05}) and the rate of release of the apolipoproteins into the blood (k_D) will be estimated with a NLMEM approach. This choice stems from the fact that these two parameters are the ones who mostly affect the biological parameters, as it can be seen from equations 4.43 - 4.45 at page 42

From this comparison, the following observations stem:

- As seen in the leucine case, we can notice from Table 4.2 a bigger standard deviation for the parameter estimates with STS method.

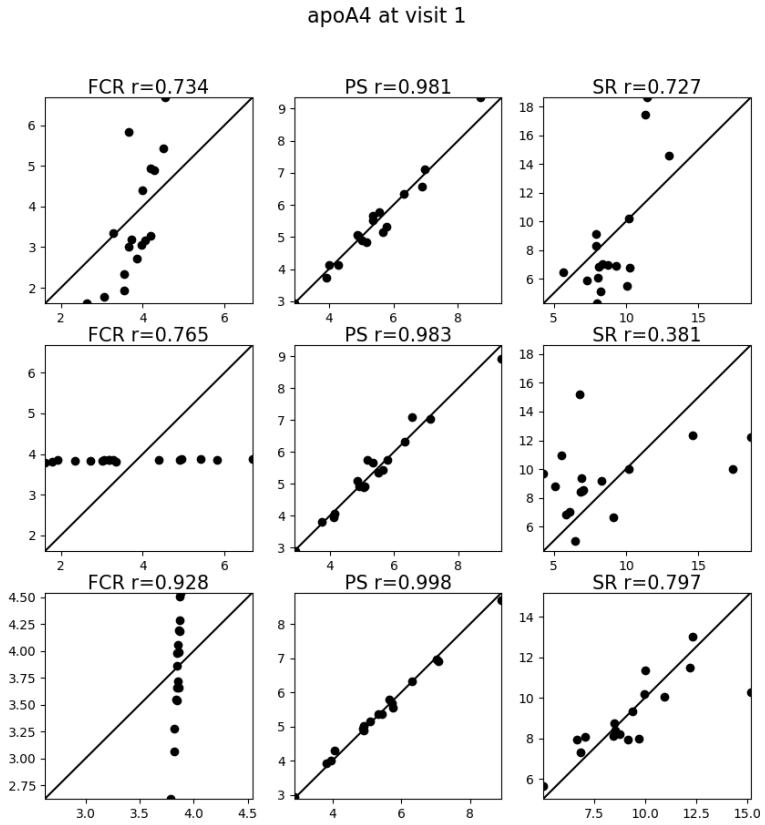

Figure 4.4: Scatter plot of the individual parameter estimate for the biological parameters for apoA4 at visit 1 with x and y axis indicating the used method for estimation (first row: x axis NLMEM, y axis STS; second row: x axis STS y axis mixed; third row: x axis mixed, y axis NLMEM). The digits in the title express the Pearson's correlation coefficient of the two estimates. FCR, PS and SR have respectively unity of measure pools/day, mg apolipoprotein and mg apolipoprotein/day

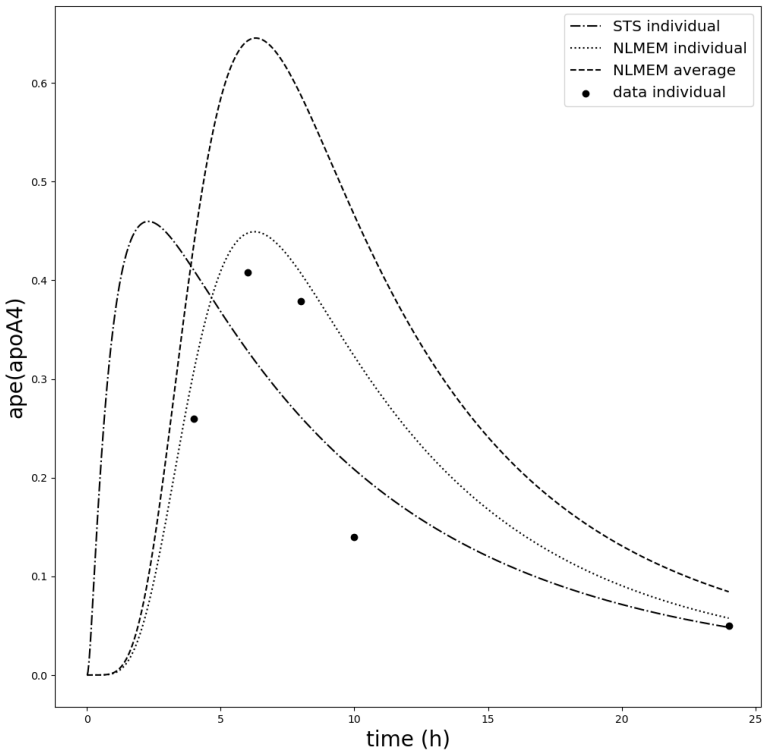

Figure 4.5: Curve estimated with NLMEM (and average curve from the other individuals in the population) and curve estimated with STS approach. NLMEM can make fruit of the population information to overcome scarcity of data for one individual.

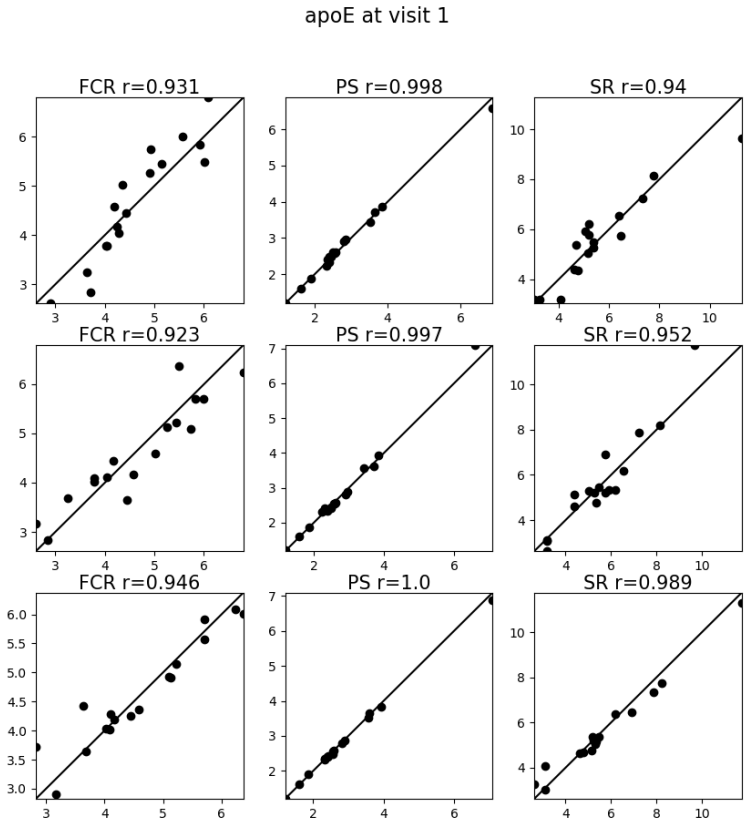

Figure 4.6: Scatter plot of the individual parameter estimate for the biological parameters for apoE at visit 1 with x and y axis indicating the used method for estimation (first row: x axis NLMEM, y axis STS; second row: x axis STS y axis mixed; third row: x axis mixed, y axis NLMEM). The digits in the title express the Pearson's correlation coefficient of the two estimates. FCR, PS and SR have respectively unity of measure pools/day, mg apolipoprotein and mg apolipoprotein/day

- While for apoE the parameter estimates with the different methods correlate strongly with each other and they mostly fall along the line $y = x$ (as shown in Figure 4.6), the same does not hold for apoA4 (as seen in Figure 4.4). In fact for apoA4 different methods yield quite diverse estimates, with lower correlation coefficient (Pearson's r as shown in the figure) and that do not fall onto the line $y = x$. This is due to the low quality of the apoA4 data. In fact, as shown in Table 3.1 on page 31, apoA4 analysis can rely on a smaller number of points for the enrichment curve. ApoC-III has similar quality of data as apoE and the different estimates correlate significantly with each other and fall along the line $y = x$, therefore, since the corresponding graph for apoC-III would only confirm what already stated for apoE, we omit to show it.

- The example individual in Figure 4.5 crystallizes one of the strengths of NLMEM. Here we can see that while STS cannot go beyond the data that it is analysing, NLMEM can transcend the limitations of the specific individual and make fruit of the overall information from the population to obtain the estimates and this leads to a more reasonable dynamics for the data.

Conclusions Based on the observations above, we conclude that the best estimation method is NLMEM. In fact it can exploit the overall signal coming from the population in order to fit the individual data, which is especially useful when data are scarce.

4.3 Model development

In this section we will explore the model selection process. Each subsection will inspect one specific choice in this procedure. A box will introduce the main points of the exploration: the motivation, the approach used and the results. Details will be given in the remaining part of the subsection.

4.3.1 Choice between Q1 and Q5 model

Motivation
From the biological point of view two different quantities are eligible to be used as parameters. Even though, we would like to avoid enforcing a probability distribution on the measured quantity (Q_5) as pointed out in Section 4.1, analysis shows better observability for the model that includes Q_5 as a parameter.

Approach
Perform the kinetic analysis with the two alternative models

Result
Results are very similar for the estimated parameters, but we choose the model where Q_1 is a parameter due to better predictability

More on the motivation In the model equations at page 40 Q_1 (the steady state amount of free leucine in the blood) appears as a parameter and we describe Q_5 (the steady state quantity of leucine in apolipoproteins) as a function of Q_1, scaled by the parameters of the system. We compare the identifiability and observability for this model, already discussed at the end of Section 4.1 with the one of an alternative model where instead Q_5 is a parameter in the system and Q_1 is a function of Q_5.
While grandly the results about identifiability and observability coincide between the two models (both are identifiable and the leucine compartments and leucine parameters are non-observable), one detail differs, in that Q_5 is observable in the model containing it as a parameter, while Q_1 is not in the model where it appears as a parameter. This stems from the fact that Q_5 is actually measured. While we are aware that having Q_5 as a parameter would mean enforcing a distribution for this parameter, we explore whether this model, with Q_5 as a parameter, Q_5 model, can yield an improvement over the model with Q_1 as a parameter, the Q_1 model.

Analysis of the results The parameter estimates are very similar to each other as shown in Figure 4.7. Therefore we choose to look at the VPC (visual predictive checkplot).

While the comparison of the parameter estimates doesn't highlight clear differences between the two models, discrimination between the fit of the two models comes from the comparison of the VPCs, in Figure 4.8. The plot for the Q_5 model indicates in fact that the simulated data using the estimated parameters would yield too much variability, so that the 90th percentile of the observed data (blue upper segmented curve) would not fall inside of the 90 percent confidence interval of the 90th percentile (blue curve); in fact due to the blue curve falling inside of the red area (for the 90th percentile of the median), the red area delimited by the blue curve is highlighted in bright red.

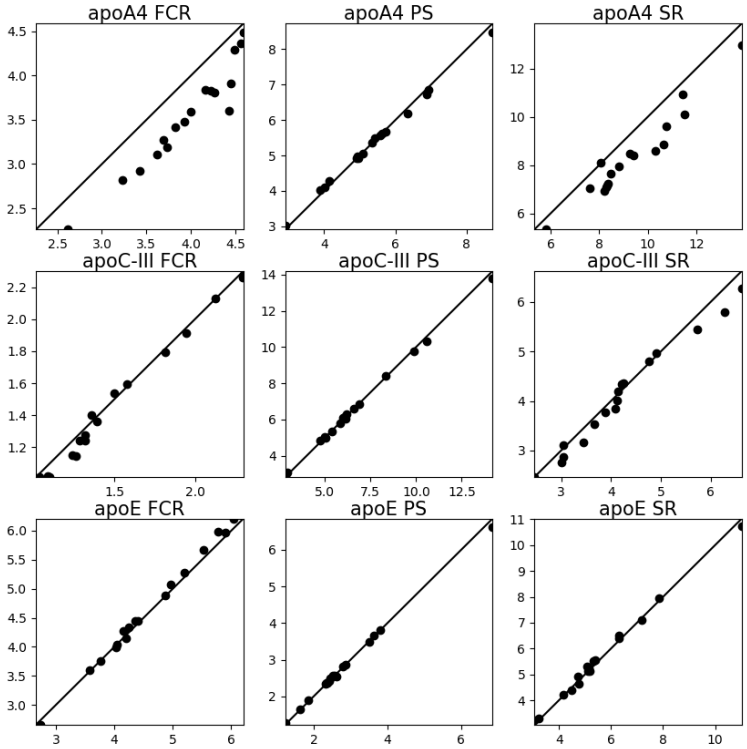

Figure 4.7: Scatter plot of the individual parameter estimates with the two models for the biological parameters at visit 1. On the x axis the Q_1 model and on the y axis the Q_5 model. FCR, PS and SR have respectively unity of measure pools/day, mg apolipoprotein and mg apolipoprotein/day

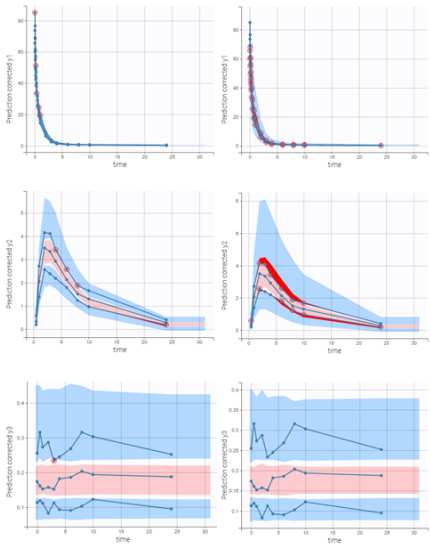

Figure 4.8: Prediction-corrected visual predictive checkplots for apoE at visit 1. On the left the VPC for the Q_1 model and on the right the one for the Q_5 model. The VPCs for m_1(equation 4.40), m_2(equation 4.41) and m_3(equation 4.42) respectively in rows 1,2 and 3.

Conclusion The Q_5 model shows a reduced predictability, therefore our final choice is the Q_1 model.

4.3.2 Choice between the combined and the separate model

Motivation

Since the unique study design allows to consider the kinetics of all the apolipoproteins simultaneously, we contemplate using a combined model, where we analyse the complete data set.

Approach

Perform the kinetic analysis with the two alternative models, a combined model, where all the measurements comprehending the three apolipoproteins are analysed simultaneously, and separate model, where each apolipoprotein is studied separately

Result

- no substantial difference can be noticed from the results
- we opt for the separate model in order to maximize the information obtained from the dataset.

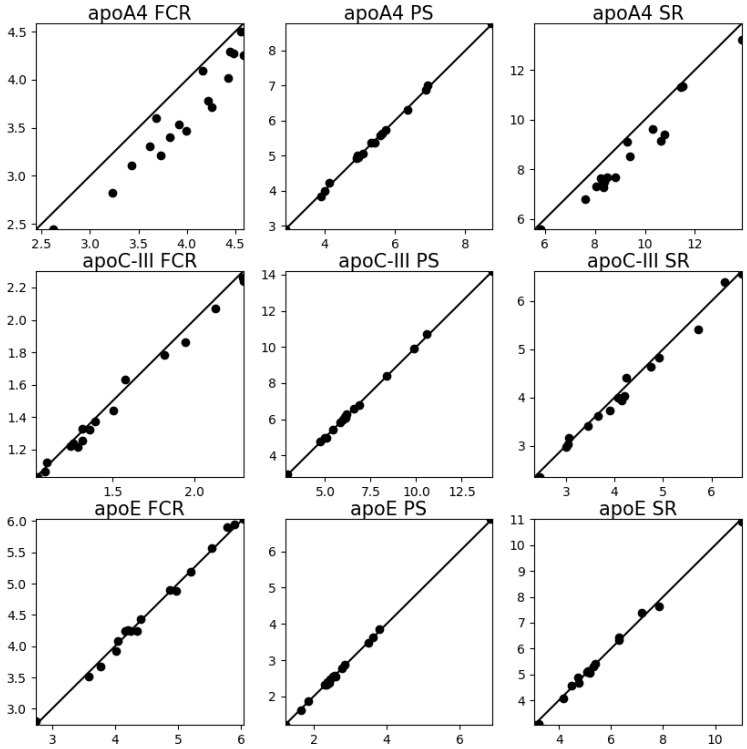

Figure 4.9: Scatter plot for the individual estimates of the biological parameters at visit one. On the x axis the separate model and on the y axis the combined one. FCR, PS and SR have respectively unity of measure pools/day, mg apolipoprotein and mg apolipoprotein/day

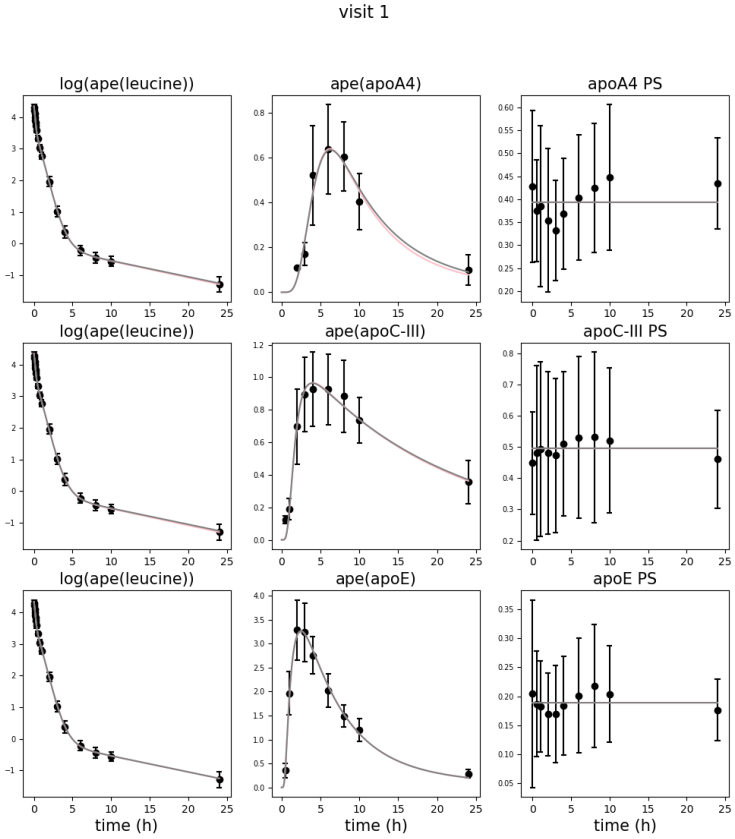

Figure 4.10: Tracer curves for the three apolipoproteins (the rows 1,2,3 show apoE, apoC3 and apoA4 respectively, with columns 1,2,3 showing the logarithm of the enrichment of leucine (logarithm of m_1 in equation 4.40), the enrichment of the apolipoprotein (m_2 in equation 4.41) and the apolipoprotein pool size (m_3 in equation 4.42)), with the two different methods for visit one. The curves (pink one representing the separate model and grey one the combined model) almost overlap with each other. The observed data are presented as mean and the error bars represent the standard deviation.

Analysis of the results As seen in the scatter plot for the individual estimates of the biological parameters (Figure 4.9) the estimates are comparable between the two different models.
The tracer curves in Figure 4.10 are almost overlapping, showing that both of the models can describe well the average behaviour of the individuals in our data sets.

Conclusion No considerable differences are noticeable from the use of the combined model instead of the separate one.
The choice between the two models, therefore, is dependent solely on the quality of the data. When using the combined model, we need to use the data for the three apolipoproteins together. As seen in Table 3.1 on page 31 three individuals with complete set of data for apoE and apoC3 lack suitable data for apoA4, therefore we choose the separate model, in that it permits for a maximization of the information extracted from our datasets.

4.3.3 Choice of the error model

Motivation
choice of the error model can influence the range of the parameters, most importantly the parameter k_{05}, which is biologically interesting , since it is tightly related to FCR (cfr equation 4.43)

Approach
compare different choices for the error model

Result

- simplifying the error model in most of the cases brings a slight improvement of the Akaike criterion and even when it does not, simplification of the model is advised
- we follow the criteria suggested in Monolix guidelines, while checking that the potential worsening of the Akaike criterion is negligible.

Possible choices for the error model and suggested approach The predicted data from the model do not usually fall exactly onto the observed data. The discrepancy between the prediction from the model and the observed data is explained in the model as measurement error (cfr equation 3.8 at page 22).
Let us briefly recapitulate here three of the four different choices for the error model contemplated in Monolix, already cited in Section 3.1.3 at page 22, where the mathematical equations can be found:

- constant: the error does not depend on the magnitude of the function describing the kinetics, but it is constant.

	GEM	DDEM
apoA4 visit 1	412.858	412.436
apoA4 visit 2	412.124	420.11
apoC-III visit 1	376.352	372.9
apoC-III visit 2	305.004	305.607
apoE visit 1	279.26	278.119
apoE visit 2	229.673	228.627

Table 4.3: Akaike criterion for the two approaches for the error model for each of the apolipoproteins. GEM stands for generic error model, while DDEM is the data-driven error model.

- proportional: the error is proportional to the magnitude of the function. The bigger the value of the function, the bigger the error that is expected.

- combined1: this error model combines the two previous ones, entailing that the total measurement error can be written as a combination of a component which is independent on the magnitude of the function and another that is proportional to the function.

In order to choose a suitable error model for the data the guidelines of Monolix suggest to start with the combined1 approach, the more general one, and to try to remove parameters that have a high relative standard error and compare the results obtained with different error models. We choose the threshold for the relative standard error as 100, so we will remove the parameters for the error model with r.s.e. bigger than 100.

Approach and results We follow exactly the approach suggested. The model with generic error model (combined1 for all the three measurement functions) will be called GEM and the model with error model chosen by removing the parameters with RSE bigger than 100 is called DDEM (data driven error model).

The scatter plot do not show a significant difference between the two error models. The data driven error model leads to a slight improvement of the Akaike criterion in most cases.
For apoC-III visit 2 the data driven error model has a slightly worse Akaike criterion, but the difference is negligible.
The difference between the data driven error model and the generic one for apoA4 at visit 2 is the largest. Here, though, the relative standard error of the parameter (more than 8e15) and its value (about 2e-16) calls for a simplification of the model. The following table shows the chosen error model for the different apolipoproteins and the different visits.

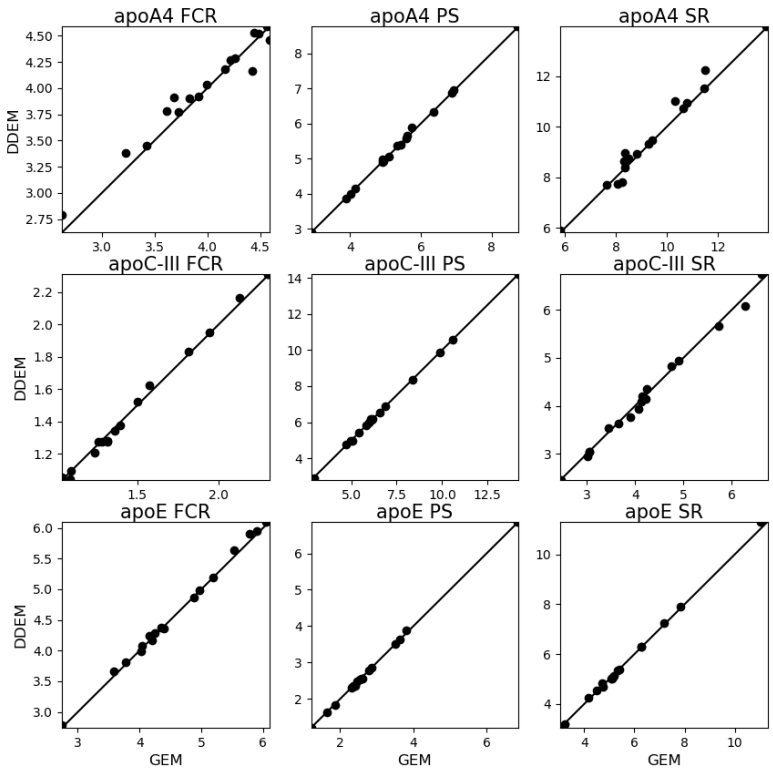

Figure 4.11: Scatter plot for the individual estimates of the biological parameters for visit 1. On the x axis is the estimate with the DDEM, while the y axis provides the estimate with the GEM model. FCR, PS and SR have respectively unity of measure pools/day, mg apolipoprotein and mg apolipoprotein/day

apo	EM leucine	EM apo	EM PS
apoA4 visit 1	combined1	combined1	proportional
apoA4 visit 2	combined1	combined1	proportional
apoC-III visit 1	combined1	combined1	proportional
apoC-III visit 2	combined1	constant	proportional
apoE visit 1	combined1	combined1	proportional
apoE visit 2	combined1	combined1	proportional

Table 4.4: Error model for the three measurement functions for the data driven error model. EM leucine stands for the error model of the leucine enrichment. The next two columns show the error model for the leucine enrichment in the apolipoprotein and for the pool size.

4.3.4 Choice among different variance-covariance matrix structures

Motivation
the choice of the variance-covariance matrix structure can have an impact on the range of the parameters

Approach
run different calculations with different variance-covariance matrix structures:

- diagonal matrix
- full matrix
- selected matrix according to the significance of the correlations between the estimated random effects
- a minimal variance-covariance matrix dictated by the correlations in common for most of the apolipoproteins

Result

- even though an intuitive approach would be to choose as variance-covariance matrix the one suggested by analysing the p-values of the suggested correlations, this might not be the best choice
- poor data can neglect highlighting a possible correlation structure

Variance-covariance matrix in Monolix In the NLMEM approach a big role is played by the variance-covariance matrix. This encompasses possible correlations among the parameters for the system.
A significant correlation between two variables indicate that the magnitude of the variables are related to one another; a positive (negative) correlation between variable A and B highlights that when variable A has large values, B will have large (small) values and when A has small values, B will have small (large) values. Correlations between two variables do not imply causality, but the relationship could possibly be investigated with further experiments to uncover this possible connections.
The correlations between the random parameters are encoded in the variance-covariance matrix. As a standard choice in Monolix the variance-covariance matrix is diagonal (does not contain any correlation between the random parameters) and correlations can be added manually. In Monolix, correlations in the variance-covariance matrix are transitive; if the correlations between variable A and variable B are inserted in the variance-covariance structure as well as correlation between variable B and variable C, the correlation will be automatically inserted also between variable A and C. This will create a correlation group in the variance-covariance matrix consisting of variables A, B and C. In the tab tests in the result section, Monolix shows the results of the following tests:

- normality of the random effects
- test to check whether the correlation between two random effects is significant

While we will go back to the first one later, let us now focus on the second one. Monolix suggests to use this as a guiding principle on whether correlations shall be added. One shall check the p-values of the test for the correlation. In case the p-value for the correlation between a couple of random effects is smaller than a chosen significance level, we shall try to add a correlation between the random effects, thereby making the structure of the variance-covariance matrix more complex. After adding it and performing the calculation one shall check whether the p-value of the correlation is still smaller than the significance level. We will continue adding correlation couples until all of the remaining couples will have p-values for the correlation bigger than the significance value. This principle with a chosen significance level $\alpha = 0.05$ will be used to select a possible variance-covariance matrix for our model (selected). This procedure will be operated separately for each apolipoprotein and for each visit.
While the diagonal matrix is the simplest variance-covariance matrix possible, the full matrix is the most complex one. In this case, since the number of parameters of our compartmental model is 10, each of them having a random effect, and in consideration of the symmetry of the variance-covariance matrix we would need to insert and estimate 45 additional parameters.
After running the estimations with the selected approach described above, we have noticed that despite the diversity of the data for the different apolipoproteins, many of the obtained variance-covariance matrices would contain similar correlation structures. The correlation group k_{01}, p_1 and Q_1 would be in common for many apolipoproteins for one or more visits. Therefore, on the side of the three described approaches (diagonal matrix, full matrix, selected matrix), we added a fourth one, consisting of this minimal variance-covariance matrix.

Normality test Due to our choice of distribution for the parameters, the estimation algorithm rests on the assumption that the random effects are normally distributed, while the parameters of our compartmental model are log-normally distributed.
It is important to check whether this assumption holds.
In case the test delivers non-normality for one or more random effects, we will test whether removing an outlier will restore the normality of the random effect. In fact non-normality can in some cases be only due to the extreme value of one individual [69]. In this case, we do not proceed further with scrutinizing the model, since we accept that the non-normality is due to only one value far from the mean. Monolix repeats the test for normality for many different samples for the individuals, drawn from their conditional distribution. An individual whose parameter estimate has an extreme value is likely to have an individual conditional distribution shifted away from the population mean, therefore this individual could mislead the algorithm to think that the overall distribution

of the random effect is non-normal, while this conclusion is mainly due to its extreme distribution.
The removal of the outlier followed by a persisting non-normality for the parameter could be a symptom of poor variance-covariance structure or error model, therefore in this case we will take a closer look and analyse alternative choices for the error model or for the variance-covariance structure for the model.

apo	full	diagonal	selected	minimal
apoA4 visit 1	475.788	411.712	384.764	396.884
apoA4 visit 2	425.951	419.653	404.618	411.388
apoC-III visit 1	510.909	493.955	516.035	478.386
apoC-III visit 2	425.592	408.668	401.507	400.593
apoE visit 1	478.766	403.965	383.994	383.822
apoE visit 2	309.159	314.675	291.059	291.059

Table 4.5: Akaike criterion for the calculations with different variance-covariance matrices. The selected model has variance-covariance matrix where all the possible correlation couples are analysed and one by one the ones with p-values <0.05 are included, if the correlations are still valid after inserting the correlation couple. The minimal approach has a variance-covariance matrix including the correlation group k_{01}, p_1 and Q_1

Results For many of the occasions, the minimal variance-covariance model yields the best Akaike criterion (see Table 4.5). Exceptions are the estimations for apoA4.
For apoA4 in the first visit, the correlation structure (k_{01}, p_1 and Q_1) would be suggested by adding the couples with significant correlation, but they would not be the ones chosen at the beginning, because smaller p-values for other correlations would have higher priority.
The interesting result here is that for apoA4 in visit 2 the correlation structure is not suggested by the p-values (p-values for the test of the three couples in the minimal correlation group are 0.08, 0.15 and 0.14), but when the correlation structure is added, the correlations become significant. Additionally, we can see that, despite the rise in the Akaike criterion in the minimal model compared to the selected model, in the first one all the parameters follow a log-normal distribution, while in the selected model two parameters fail in the test for log-normality. This implies that the selection criterion based on the p-values for the correlation test (denominated by selected) might not bring the most suitable variance covariance matrix.

For apoE at visit 2 and apoC-III at visit 1, one parameter fails in the test for log-normality. By removing an outlier, though, none of the random effects fail to show normality, therefore the choice of the variance-covariance structure is plausible.

apo	full	diagonal	selected	minimal
apoA4 visit 1	1	0	0	0
apoA4 visit 2	0	1	2	0
apoC-III visit 1	0	1	2	1
apoC-III visit 2	1	2	2	0
apoE visit 1	1	2	0	0
apoE visit 2	2	1	1	1

Table 4.6: For each of the visits and each of the choices for the variance-covariance matrix, number of random effects that fail in the test for normality

Conclusions Due to the improvement of the Akaike criterion for apoE and apoC-III and due to the fact that despite the increase in the Akaike criterion, the correlations in the chosen minimal matrix (k_{01}, p_1 and Q_1) are significant for apoA4, we are safe to choose this as the standard choice for the variance-covariance model.

4.4 Model validation

Motivation

- need to test whether the strategy for choice of error model yield a good fit also for the remaining datasets
- need to test whether the choice of the variance covariance matrix allows a good fit for the datasets from the two experiments not used in the model selection process

Approach

run different calculations with different variance-covariance matrices and error models:

- general error model (GEM)
- data driven error model (as suggested in Section 4.3.3) (DDEM)
- general error model with the minimal variance-covariance matrix as suggested in Section 4.3.4 (GEM + COV)
- data driven error model with minimal variance-covariance matrix as suggested in Section 4.3.4 (DDEM + COV)

Result

- in most of the cases the best performing model is the DDEM + COV
- this model is not a one-size-fit-all measure, therefore we need to keep a flexible mindset and do specific adjustments as the specific dataset requires.

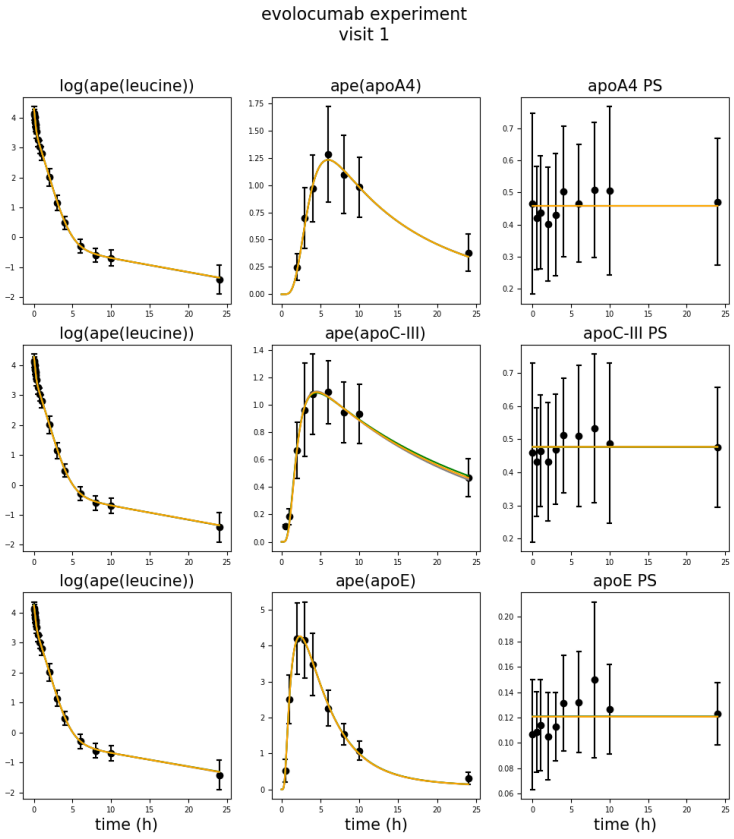

Figure 4.12: Tracer curves for the three apolipoproteins (the rows 1,2,3 show apoE, apoC3 and apoA4 respectively, with columns 1,2,3 showing the log enrichment of leucine, the enrichment of the apolipoprotein and the apolipoprotein size), with the two different methods at visit 1 for the evolocumab experiment. The curves (pink, grey, green and orange representing respectively GEM, DDEM, GEM + COV and DDEM + COV) almost overlap with each other. The observed data are presented as mean (points), with error bars representing the standard deviation.

4.4.1 Process and specific results

Evolocumab The Akaike criterion for the DDEM + COV is lower compared to the one for the other models; only exception is apoA4 at visit one, where the difference between the Akaike for DDEM and DDEM + COV is not major.
We notice that there are two occurrences where one or more parameters fail to be proven to have a log-normal distribution. We will consider these separately and describe our inspection strategy.
The two occurrences are:
ApoE at visit 2: We first remove an individual to check whether the lack of nonnormality was caused by the presence of an outlier.
Despite removing the outlier, the nonnormality still persists, therefore we challenge one of our assumptions of our choice of the variance-covariance model. We remove one of the correlations. Still one of the parameters lacks lognormality. The nonnormality is though, due to the presence of an outlier. Therefore we

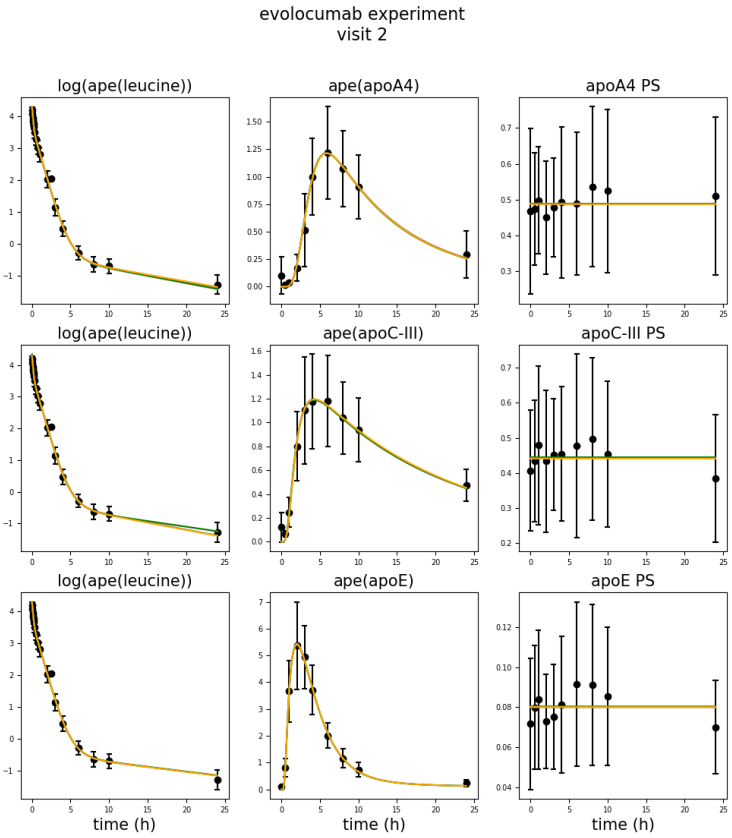

Figure 4.13: Tracer curves for the three apolipoproteins (the rows 1,2,3 show apoE, apoC3 and apoA4 respectively, with columns 1,2,3 showing the log enrichment of leucine, the enrichment of the apolipoprotein and the apolipoprotein size), with the two different methods at visit 2 for the evolocumab experiment. The curves (pink, grey, green and orange representing respectively GEM, DDEM, GEM + COV and DDEM + COV) almost overlap with each other. The observed data are presented as mean (points), with error bars representing the standard deviation.

apo	GEM	DDEM	GEM + COV	DDEM + COV
apoA4 visit 1	361.0	361.0	345.0	345.0
apoA4 visit 2	352.0	340.0	345.0	345.0
apoC-III visit 1	286.0	285.0	254.0	254.0
apoC-III visit 2	279.0	221.0	265.0	200.0
apoE visit 1	174.0	173.0	150.0	148.0
apoE visit 2	36.0	35.0	21.0	19.0

Table 4.7: Akaike criterion for each of the visits with different models for the evolocumab experiment (GEM and DDEM are respectively the generic error model and the data-driven one, +COV indicates integrating the chosen minimal variance-covariance matrix in the calculations)

apo	GEM	DDEM	GEM + COV	DDEM + COV
apoA4 visit 1	0	0	0	0
apoA4 visit 2	0	1	1	1
apoC-III visit 1	0	0	0	0
apoC-III visit 2	1	0	2	0
apoE visit 1	1	0	0	0
apoE visit 2	0	1	2	2

Table 4.8: For each of the visits and each of the methods for the evolocumab experiment, number of random effects that fail in the test for normality

apo	GEM	DDEM	GEM + COV	DDEM + COV
apoA4 visit 1	284.0	296.0	281.0	357.0
apoA4 visit 2	179.0	225.0	161.0	233.0
apoC-III visit 1	301.0	296.0	300.0	293.0
apoC-III visit 2	220.0	236.0	231.0	230.0
apoE visit 1	262.0	271.0	260.0	276.0
apoE visit 2	141.0	91.0	189.0	71.0

Table 4.9: Akaike criterion for each of the visits with different models for the liraglutide experiment (GEM and DDEM are respectively the generic error model and the data-driven one, +COV indicates integrating the chosen minimal variance-covariance matrix in the calculations)

accept the model with simplified variance-covariance structure.
ApoA4 at visit 2: Removing the individual with more extreme value for the parameter does not result in log-normality for all the parameters, therefore we challenge our model. For the choice of the error model, we had established, by following the guidelines in Monolix, that an error model shall be simplified if the relative standard error(rse) of one of its parameters is bigger than 100, as explained in Section 4.3.3. For apoA4 at the first visit, we had simplified the error model by eliminating a parameter with rse slightly above the threshold (108). Now, in order to obtain log-normality for all the parameters, instead of simplifying the variance-covariance structure, as done with the case of apoE visit 2, we reinsert the parameter for the error model that we had eliminated due to the selection criteria. Doing so ensures that all the parameters are log-normally distributed.

Liraglutide The Akaike criterion shows a big improvement for the DDEM + COV for apoE in visit 2, an improvement for apoC-III at visit 1 and for apoA4 at visit 2. The other cases will be further analysed together with apoC-III at visit 1, since it presents a parameter for which the distribution is not log-normal.
ApoE at visit 1 The Akaike criterion does not improve with our choice of the model and we have one parameter with non log-normal distribution. Since the simplification for the error model was necessary (the relative standard error of the removed parameter was above e16), we can now consider DDEM as term of

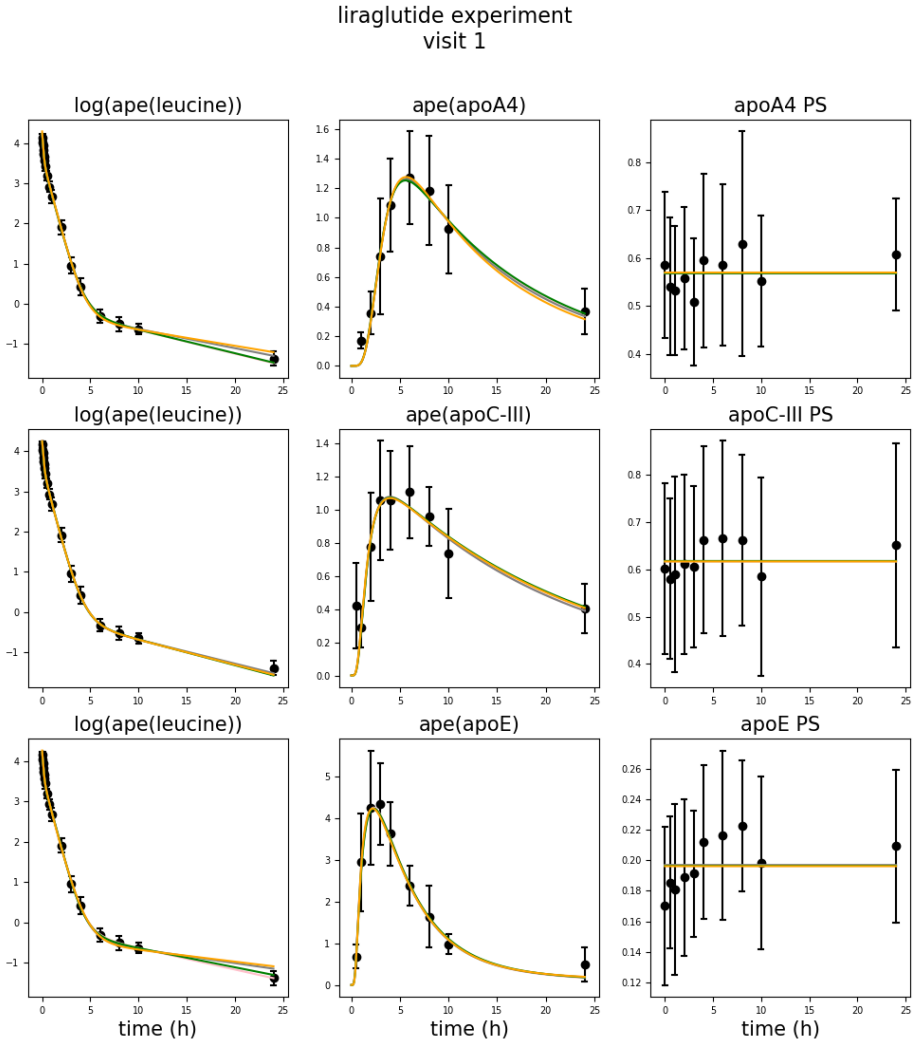

Figure 4.14: Tracer curves for the three apolipoproteins (the rows 1,2,3 show apoE, apoC3 and apoA4 respectively, with columns 1,2,3 showing the log enrichment of leucine, the enrichment of the apolipoprotein and the apolipoprotein size),with the two different methods at visit 1 for the liraglutide experiment. The curves (pink, grey, green and orange representing respectively GEM, DDEM, GEM + COV and DDEM + COV) almost overlap with each other. The observed data are presented as mean (points), with error bars representing the standard deviation

apo	GEM	DDEM	GEM + COV	DDEM + COV
apoA4 visit 1	0	0	0	0
apoA4 visit 2	0	1	0	1
apoC-III visit 1	0	0	2	1
apoC-III visit 2	1	2	1	2
apoE visit 1	0	0	0	1
apoE visit 2	1	0	1	0

Table 4.10: For each of the visits and each of the methods for the liraglutide experiment, number of random effects that fail in the test for normality

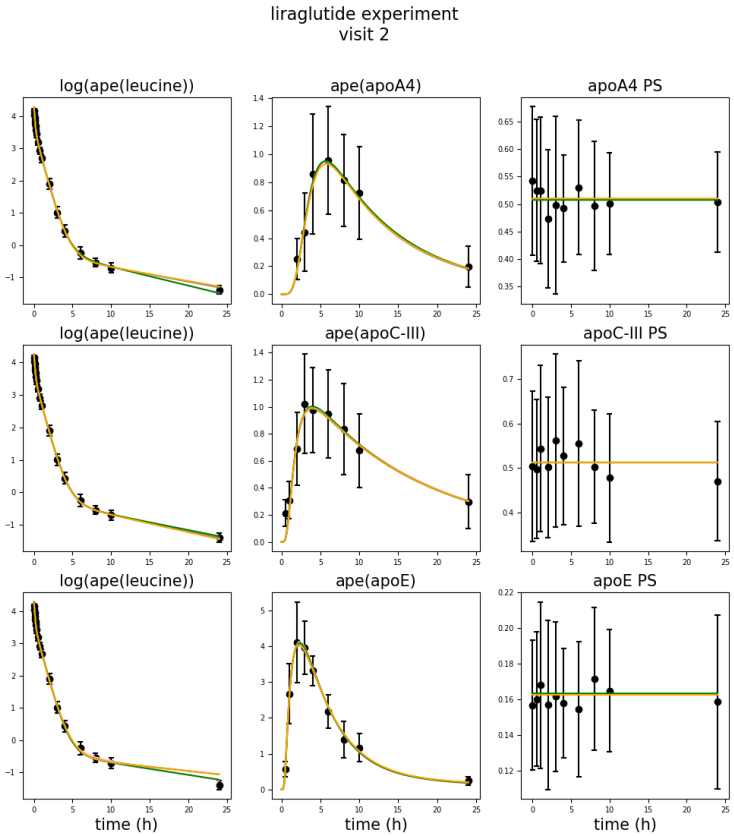

Figure 4.15: Tracer curves for the three apolipoproteins (the rows 1,2,3 show apoE, apoC3 and apoA4 respectively, with columns 1,2,3 showing the log enrichment of leucine, the enrichment of the apolipoprotein and the apolipoprotein size),with the two different methods at visit 2 for the liraglutide experiment. The curves (pink, grey, green and orange representing respectively GEM, DDEM, GEM + COV and DDEM + COV) almost overlap with each other. The observed data are presented as mean (points), with error bars representing the standard deviation

comparison. The difference in AIC of DDEM + COV when compared to DDEM is not so major, so we aim to reach log-normality for all the parameters byremoving the outlier. With this step, we obtain lognormality for the distribution of all the parameters.
ApoC-III at visit 1 Despite the Akaike criterion improves with the DDEM + COV model, a parameter is not positive in the lognormality test. By removing an outlier, we obtain lognormality for all the parameters.
ApoC-III at visit 2 The Akaike criterion for DDEM + COV is just slightly worse than for the more general model and the parameter estimate are similar, therefore we do not look further into the model. By removing an outlier, we obtain normality for all the parameters, therefore we keep this model.
ApoA4 at visit 1 The Akaike criterion for the DDEM + COV is much worse than the one for GEM and DDEM. We then try a model with simplified variance-covariance structure. This, while bringing an improvement of the Akaike criterion, narrows too much down the range for the FCR. Therefore, in this case, we choose the simplified error model without any correlation structure in the variance-covariance matrix, DDEM.
ApoA4 at visit 2: Our standard choice for the model (DDEM + COV) has a much bigger Akaike criterion than GEM + COV. Since the relative standard error of one of the removed parameters for the error model was not so far from the threshold (180), we choose to reinsert this parameter again. Since this brings an improvement in the Akaike criterion (165) and all the parameters are log-normally distributed, we choose this as the final model.

4.4.2 Conclusions

For most of the apolipoproteins and visits, the model selection choices depicted in the Section 4.3.3 and Section 4.3.4 (that when combined generate the DDEM + COV model) bring to an improvement of Akaike criterion.
In case that this criterion is much worse for the DDEM + COV model compared to the alternative models, we suggest first to look at the error model, by checking if we have simplified too much (we reinsert one parameter that we had eliminated, with rse not far from 100). Secondly, we can look at the correlation structure in the variance-covariance matrix. Can this be simplified?
Another measure of goodness of fit that we always would like to check is the log-normality of the parameters. First, we shall check whether by removing one outlier for the specific parameter, the distribution becomes log-normal. In case this doesn't happen, we suggest to act on the error model or on the correlation structure in the variance-covariance matrix.
All in all, only 2 of the 12 occasions (apoA4 at visit 2 for both evolocumab and liraglutide studies) have required reinserting one parameter for the error model, for which the relative standard error was bigger than 100 (108 and 180), only one occasion have required simplifying the correlation structure (evolocumab for apoE visit 2) and only one occasion have required totally removing the correlation structure (liraglutide for apoA4 at visit 1).
We conclude therefore that our model choice strategy (DDEM + COV) can

describe largely the extent of possible behaviours for the apolipoproteins, but that this is not a one-fix-all measure, therefore the results of the models shall always be checked and adjustments shall be made to fit to the specificity of the data.

5 Results

Our conclusions on the physiology behind the changes due to the treatments can stem only from a thorough analysis of the results of the experiment. Therefore several facets shall be taken into account:

- correlations at baseline
- overall effects of the treatments
- linearity between individual characteristics (and their changes) to the kinetic parameters

In this chapter we will first introduce in detail the tools used for each analysis and then we will use them to facilitate the description of the results of the experiments.

5.1 Analysis of the results

The analysis of the results entails different parts. First we are interested merely in analysing the correlation structure among the kinetic parameters and among the kinetic parameters and the individual characteristics at baseline. Secondly, we compare whether changes in the kinetic parameters or in the individual characteristics between before and after the treatment. Lastly, we look at the correlations between the kinetics parameters and relevant characteristics for lipid metabolism at baseline and after the treatment and additionally check whether the changes in the variables are related.
Here we describe which figures and which tables we will use for the analysis of the different pieces of the puzzle.

5.1.1 Correlations at baseline

Motivation

Baseline represents the natural state of the individuals, before the treatment perturbs their system. Therefore monitoring the correlations among the kinetic parameters and among these parameters and the individual characteristics and enriches our knowledge on the kinetics of the apolipoproteins. Moreover they represent a standard against which the correlations at visit 2 shall be compared to highlight whether the treatment brought disruptions to the system.

Tools

- Table 1

Table 1 In this table, the names of the variables in the significant correlations couples are shown. Additionally the Pearson's correlation coefficient (later referred to as r) and the p-value (later referred to as p) for the significance of the correlation are shown together with the normalized slope for the linear regression of the variable as a function of the covariate. One table shows the correlations among the kinetic parameters and one those among kinetic parameters and individual characteristics.
Our chosen significance level is $\alpha = 0.05$, therefore p-values which are smaller than this threshold will indicate that the Pearson's correlation coefficient is significant. For ease of read purposes, we will abuse the mathematical denotation, by stating that two variables correlate or that a correlation exists between two variables instead of fully acknowledging that the two variables correlate significantly or that their Pearson's r is significant. The same denotation will be used in Section 5.4.

5.1.2 Effects of the treatment

Motivation

In order to check whether the treatment had an effect on the individual characteristics or on the kinetic parameters, we need to compare them between before and after and establish which changes are significant

Tools

- Tables for the individual characteristics and kinetic parameters
- Scatter plots of kinetic parameters and selected individual characteristics with values before and after the treatment

Tables to compare between before and after the treatment This table displays the following elements: the mean and standard deviation of the quantities for each of the visits (before and after the treatment), the increase in mean in percentage and the p-value for the paired t.test with null hypothesis being that there is no change between the two visits.
We provide one table for the kinetic parameters and one for relevant individual characteristics.

Scatter plots of kinetic parameters and selected individual characteristics with values before and after the treatment In this plot, x and y axis indicate the value of the variable for the first respectively the second visit. The title of the picture shows the mean percentage change together with the p-value for the significance of the change. This represents a visual aid for the table and helps us to check whether the significance or non significance of the difference might be due to some outlier. Additionally it raises our awareness on strong changes in the range of the values for the parameters.

5.1.3 Correlation of kinetic parameters with selected covariates

Motivation

Correlations between covariates and kinetic parameters could be a symptom of causality, especially if the relation is stable throughout several conditions (before and after treatment). With the same purpose, monitoring the correlation between changes for covariates and kinetic parameters is highly important. High triglyceride levels, high LDL levels and low HDL levels are the markers of dyslipidemia, an hallmark of type-II-diabetes, that increases the risk for cardiovascular diseases [5]. Therefore we study the correlation of these individual characteristics with the kinetic parameters.

Tools

- Picture 1
- Picture 2

Picture 1 In Picture 1 the x axis represents the covariate while the y axis represents the kinetic parameter. For each individual, two points occur, the green one describing the first visit and the red the second visit. A line of the corresponding color indicates the linear regression of the kinetic parameter as a function of the covariate with title to the sub-figure showing the p-values for the two sided hypothesis test with null hypothesis being that the slope is equal for both the visits.
In case the p-value is smaller than 0.05, a correlation exists between the two

variables and it is reasonable to assume a linear relationship between the kinetic parameter and the covariate.

Picture 2 We look at this plot to check whether the changes of the covariate and the kinetic parameters correlate or not.
The x axis represents the covariate, while the y axis indicates the kinetic parameter. In the title the name of the kinetic parameter plotted in the sub-figure appears, together with the p-value for the test with null hypothesis that the slope of the regression line is 0. Therefore a p-value smaller than 0.05 indicates that the correlation coefficient between the two variables is significant.

5.2 Correlations at baseline

Now we shall look at the correlations among kinetics parameters and between kinetic parameters and the individual characteristics.

5.2.1 Fructose

covariate	variable	Pearson's r	p-value	slope
apoA4 PS	apoA4 FCR	-0.705	0.002	-0.208
apoA4 SR	apoA4 FCR	-0.509	0.037	-0.182
apoC-III PS	apoC-III FCR	-0.668	0.001	-0.261
apoC-III SR	apoC-III PS	0.924	<0.001	1.148
apoE SR	apoE PS	0.804	<0.001	0.972
apoE PS	apoA4 FCR	-0.544	0.024	-0.09
apoA4 SR	apoA4 PS	0.961	<0.001	1.169
apoE FCR	apoC-III FCR	0.541	0.014	0.473
apoE PS	apoC-III FCR	-0.688	<0.001	-0.311
apoE SR	apoC-III FCR	-0.482	0.031	-0.263
apoE FCR	apoC-III PS	-0.535	0.015	-1.197
apoE PS	apoC-III PS	0.699	<0.001	0.809
apoE FCR	apoC-III SR	-0.453	0.045	-0.815
apoE PS	apoC-III SR	0.465	0.039	0.432

Table 5.1: Correlations among the kinetic parameters at baseline of the fructose treatment

First let us take a close look at the correlation among the kinetic parameter for the same apolipoprotein and then enlarging the view to encompass the relationships among apolipoproteins. ApoA4 and apoC-III pool sizes both correlate to their respective SR and FCR, but the positive correlation with the SR is much stronger (for apoA4 r=0.961 and p<0.001 and for apoC-III r=0.924 and p<0.001) than the negative correlation with the FCR (for apoA4 r=-0.705

and p=0.002 and for apoC-III r=-0.668 and p=0.001). One can notice that the p-values do not reflect the difference in absolute value of the Pearson's correlation coefficient. This is due to the discrepancy of the number of individuals for whom the two measured apolipoprotein concentrations would represent a suitable dataset for the analysis (20 for apoC-III and 17 for apoA4) (cfr. Table 3.1 at page 31). We now describe the relationships between kinetic parameters for different apolipoproteins. ApoA4 appears to be solitary, exhibiting only one non strong correlation between the FCR and the PS of apoE. Several associations exist between the parameters for apoE and the ones for apoC-III. Two are the strongest ones: between apoE PS and apoC-III FCR (r=-0.688 and p<0.001) and between apoE PS and apoC-III PS (r = 0.699 and p<0.001).

covariate	variable	Pearson's r	p-value	slope
Cholesterol	apoC-III FCR	-0.71	<0.001	-0.715
Cholesterol	apoC-III PS	0.738	<0.001	1.9
Cholesterol	apoC-III SR	0.601	0.005	1.245
Cholesterol	apoE FCR	-0.533	0.016	-0.613
Cholesterol	apoE PS	0.638	0.002	1.42
LDL cholesterol	apoC-III FCR	-0.584	0.007	-0.48
LDL cholesterol	apoC-III PS	0.623	0.003	1.309
LDL cholesterol	apoC-III SR	0.511	0.021	0.865
LDL cholesterol	apoE PS	0.489	0.029	0.89
HDL cholesterol	apoA4 PS	0.684	0.002	0.497
HDL cholesterol	apoA4 SR	0.646	0.005	0.386
Triglyceride	apoC-III FCR	-0.542	0.014	-0.219
Triglyceride	apoC-III PS	0.865	<0.001	0.895
Triglyceride	apoC-III SR	0.792	<0.001	0.659
Triglyceride	apoE PS	0.711	<0.001	0.636
Triglyceride	apoE SR	0.523	0.018	0.387
ApoA1	apoA4 PS	0.707	0.002	0.999
ApoA1	apoA4 SR	0.688	0.003	0.799
ApoA1	apoE FCR	-0.507	0.027	-0.553

Table 5.2: Correlations among the individual characteristics and the kinetic parameters at baseline of the fructose treatment

Many correlations can be found between the kinetic parameters and the individual characteristics.
We can divide the individual characteristics in two groups, according to the correlation pattern: cholesterol, LDL cholesterol and triglyceride correlate with many of the kinetic parameters of apoE and apoC-III, while HDL cholesterol and ApoA1 levels correlate mostly with apoA4 kinetic parameters. We first describe the first correlation group.
Cholesterol correlates with all the parameters of apoC-III (with FCR r=-0.71 and p<0.001, with PS r=0.738 and p<0.001 and with SR r=0.601 and p=0.005) and with the pool size of apoE (p=0.638 with p=0.002). LDL cholesterol correlates with apoC-III FCR(r=-0.584 with p=0.007) and with apoC-III PS (r=0.623 and p=0.003). Triglyceride correlates with the pool sizes of apoC-III (r=0.865 and

p<0.001) and apoE (r=0.711 and p<0.001) and with the SR of apoC-III (r=0.792 and p<0.001). ApoA1 correlates with two kinetic parameters for apoA4: PS (r = 0.707 and p=0.002) and SR (r=0.688 and p=0.003).

In summary, the pool sizes of all the apolipoproteins is strongly determined by the SR. ApoC-III PS and apoE PS correlate and both correlate with the apoC-III FCR.
Triglyceride correlate with apoC-III SR and PS, while LDL cholesterol correlates with apoC-III PS and FCR.

5.2.2 Evolocumab

covariate	variable	Pearson's r	p-value	slope
apoA4 SR	apoA4 PS	0.735	0.004	0.89
apoC-III SR	apoC-III PS	0.83	<0.001	0.999
apoE PS	apoE FCR	-0.579	0.038	-0.265
apoE SR	apoE PS	0.924	<0.001	1.126
apoC-III SR	apoA4 SR	0.581	0.037	0.582
apoE SR	apoA4 SR	0.617	0.025	1.111
apoE FCR	apoC-III FCR	0.7	0.008	1.709

Table 5.3: Correlations among the kinetic parameters at baseline of the evolocumab treatment

For all the apolipoproteins, the pool size correlates positively with the secretion rate (for apoA4 r=0.735 and p=0.004, for apoC-III r=0.83 and p<0.001 and for apoE r=0.924 and p<0.001).
Additionally the fractional catabolic rates of apoE and of apoC-III correlate with each other (r=0.7 and p=0.008).

covariate	variable	Pearson's r	p-value	slope
Weight	apoA4 SR	-0.61	0.027	-0.971
Cholesterol	apoA4 FCR	0.643	0.018	1.468
Cholesterol	apoC-III PS	0.63	0.021	1.901
HDL cholesterol	apoC-III SR	0.572	0.041	0.68
Triglyceride	apoA4 SR	0.692	0.009	0.708
Triglyceride	apoC-III PS	0.858	<0.001	1.055
Triglyceride	apoC-III SR	0.76	0.003	0.777
ApoA1	apoC-III PS	0.645	0.017	1.44
ApoA1	apoC-III SR	0.769	0.002	1.425

Table 5.4: Correlations among the individual characteristics and the kinetic parameters at baseline of the evolocumab treatment

The triglyceride levels correlate strongly with apoC-III PS (r=0.858 with p<0.001) and SR (r=0.76 with p=0.003). They correlate less strongly also with apoA4 SR

(r=0.692 and p=0.009). An association exists also between apoA1 and apoC-III SR (r=0.769 with p=0.002).

In summary, the major determinant of the pool size for each apolipoprotein is the secretion rate. ApoE and apoC-III fractional catabolic rates correlate. Triglyceride levels correlate with apoC-III PS and secretion rate, as noticed in the fructose experiment.

5.2.3 Liraglutide

covariate	variable	Pearson's r	p-value	slope
apoA4 SR	apoA4 PS	0.993	<0.001	0.981
apoC-III PS	apoC-III FCR	-0.707	0.007	-0.187
apoC-III SR	apoC-III PS	0.965	<0.001	1.109
apoE PS	apoE FCR	-0.774	0.002	-0.278
apoE SR	apoE PS	0.95	<0.001	1.327
apoC-III FCR	apoA4 FCR	0.645	0.023	0.182
apoE FCR	apoC-III SR	-0.579	0.038	-2.666

Table 5.5: Correlations among the kinetic parameters at baseline of the liraglutide treatment

We can notice strong correlations between the pool sizes of the apolipoproteins and their secretion rates, as seen in the other experiments (for apoA4 r=0.993 and p<0.001, for apoC-III r=0.965 and p<0.001 and for apoE r=0.95 with p<0.001). Additionally apoE PS has a strong correlation with its FCR (r=-0.774 and p=0.002).

covariate	variable	Pearson's r	p-value	slope
HDL cholesterol	apoE PS	-0.592	0.033	-0.436
HDL cholesterol	apoE SR	-0.667	0.013	-0.352
Triglyceride	apoC-III PS	0.795	0.001	0.778
Triglyceride	apoC-III SR	0.775	0.002	0.66
Triglyceride	apoE PS	0.6	0.03	0.309
Triglyceride	apoE SR	0.651	0.016	0.24
Glucose	apoA4 PS	0.705	0.01	0.637
Glucose	apoA4 SR	0.713	0.009	0.652
Liver fat	apoC-III FCR	0.659	0.014	0.139

Table 5.6: Correlations among the individual characteristics and the kinetic parameters at baseline of the liraglutide treatment

Once again, we can notice a strong correlation of the triglyceride levels with apoC-III pool size (r=0.795 with p=0.001) and secretion rate (r=0.775 with p=0.002). Additionally apoA4 SR correlates with the glucose level (r=0.713 and p=0.009).

In summary, the stronger determinant of the pool size for all the apolipoproteins is the secretion rate. ApoE PS does additionally exhibit a strong correlation with the FCR. Triglyceride levels are strongly correlated with the apoC-III PS and SR.

5.2.4 General conclusions

Across all the studies some correlation patterns remain unvaried. The pool size of all the apolipoproteins is mostly determined by the secretion rate. In the fructose and evolocumab studies, associations exist between kinetic parameters of apoE and of apoC-III.
Triglyceride levels correlate in all the three studies with apoC-III PS and SR. Weaker correlations exist also in two of the three studies between triglyceride levels and apoE PS and SR.
In the evolocumab and liraglutide experiment, the individuals are on statin treatment; this might induce changes in LDL kinetics so the possible link to apoE and apoC-III kinetics, detectable in the fructose experiment, could be overshadowed.
The associations between LDL cholesterol and apoC-III in the fructose study might be secondary to effects on triglyceride. The negative correlation of LDL cholesterol with apoC-III FCR might be due to the removal of LDL particles together with apoC-III. This would also explain the positive correlation between the pool size of apoC-III and the one of LDL.

5.3 Effects of the treatments

A general description of the results of the experiments will be laid out, with summary focusing only on the kinetic parameters and individual characteristics whose change between before and after the treatment is significant.

5.3.1 Fructose

General description of results

Individual characteristics The individuals taking part to the experiment had mean body weight at visit one of around 100 kg. After the treatment they experience a 0.8 percent body weight increase which is statistically significant, but not truly biologically relevant. The mean total cholesterol level was 4.63 mmol/L at the first visit and this undergoes a statistically significant increase in mean of 4 percent, which lacks biological significance. Average LDL cholesterol level at the beginning of the study was 3.42 mmol/L and becomes 3.51 mmol/L

	visit 1 mean	visit 1 std	visit 2 mean	visit 2 std	mean change	p-value
Weight	99.74	12.45	100.56	12.86	1%	0.034
Cholesterol	4.63	0.87	4.81	0.96	4%	0.04
LDL cholesterol	3.42	0.79	3.51	0.82	3%	0.266
HDL cholesterol	1.02	0.36	1.06	0.34	3%	0.192
Triglyceride	1.47	0.69	1.77	0.82	20%	0.001
Glucose	5.48	0.4	5.47	0.32	0%	0.95
Liver fat	6.24	4.83	7.99	5.17	28%	<0.001
ApoA1	1.31	0.26	1.4	0.24	6%	0.002

Table 5.7: Comparison of relevant individual characteristics before and after the fructose experiment. Weight is expressed in kg, cholesterol, LDL cholesterol, HDL cholesterol, triglyceride and glucose are in mmol/L, while liver fat is in percentage and apoA1 is in mg/L

	visit 1 mean	visit 1 std	visit 2 mean	visit 2 std	mean change	p-value
apoA4 FCR	4.14	0.31	4.2	1.28	1%	0.851
apoA4 PS	10.74	2.7	9.33	2.94	-13%	0.036
apoA4 SR	2.34	0.48	1.94	0.31	-17%	0.01
apoC-III FCR	1.43	0.27	1.35	0.33	-6%	0.119
apoC-III PS	8.8	4.26	11.5	4.95	31%	<0.001
apoC-III SR	6.34	2.47	7.89	3.02	24%	0.003
apoE FCR	4.83	1.04	4.97	0.96	3%	0.43
apoE Pool	5.71	2.39	5.84	2.08	2%	0.771
apoE SR	1.42	0.49	1.49	0.43	5%	0.581

Table 5.8: Comparison of kinetic parameters before and after the fructose treatment. FCRs are expressed in pools/day, PS is expressed in mg/dL and SR in mg/(kg day)

after the treatment, this change being not significant. HDL cholesterol level also undergoes a nonsignificant increase (increase in mean amounts to 3 percent). The triglyceride level increases considerably (the increase in mean amounts to 20 percent) and significantly. Glucose level stays almost unvaried going from an average of 5.48 mmol/L to an average of 5.47 mmol/L. Mean liver fat goes from an average of 6.24% to an average of 8%, thereby experiencing a 28 percent increase in average which is significant with p-value less than 0.001.
Another significant increase is that of ApoA1 level, going from an average of 1.31 mg/L to an average of 1.4 mg/L, corresponding to 6 percent in average.

Kinetic parameters The kinetic parameters have been calculated as a result of the tracer-tracee experiment before and after the treatment period. Few parameters have experienced a significant change. ApoA4 FCR has slightly increased (with 1 percent increase in the mean), while the pool size and the secretion rate of apoA4 have experienced a statistically significant decrease (respectively of

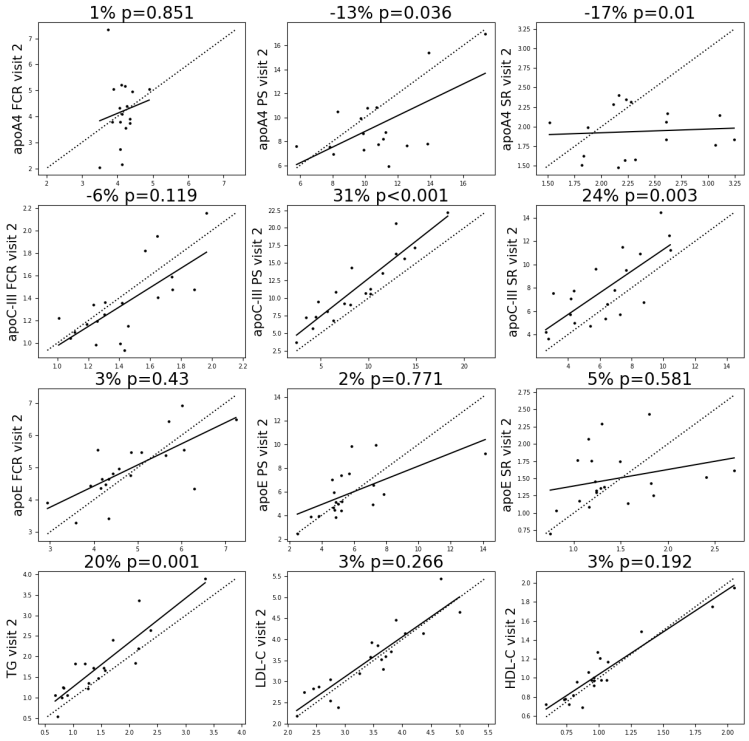

Figure 5.1: Kinetic parameters and relevant covariates before (x axis) and after (y axis) fructose treatment

13 and 17 percent), which, due the proximity of the p-value to the statistical significance level and due to the scarcity of the apoA4 data, as shown in Table 3.1 at page 31, might not correspond to truly relevant biological results.
ApoC-III FCR has slightly decreased (with a 6 percent decrease in average), while the pool size and secretion rate of apoC-III have both increased significantly (respectively with 31 and 24 percentage increase in average). All of the kinetic parameters for apoE have experienced an increase, but none of them were significant. (FCR, PS and SR respectively having a 3, 2 and 5 percent increase in average).

Figure 5.1 largely confirms the results of the tables. It is interesting to notice that for apoE pool size there is an outlier, with very high value at the first visit, that shifts the balance from the common trend of an increase in pool size at the second visit.

Summary of results for fructose Some individual characteristics have had considerable changes between the two visits. The most significant one is the increase in liver fat with 28 percent increase in average ($p<0.001$). Additionally we notice an increase in triglyceride with increase in average of 20 percent

(p=0.0015) and an increase in ApoA1 (increase in average of 6 percent with p=0.002).

Several kinetic parameters have also been affected by the fructose treatment. While the results for the apoA4 kinetic parameters might not be reliable due to the low quality of data, changes in apoC-III kinetic parameters are relevant and reliable. The most significant one is the increase in apoC-III pool size (increase in average of 31 percent and p<0.001). An increase in apoC-III secretion rate (increase in average of 24 percent) is significant with p-value 0.003.

5.3.2 Evolocumab

	visit 1 mean	visit 1 SD	visit 2 mean	visit 2 SD	mean change	p-value
Weight	84.3	17.5	84.4	16.8	1%	0.95
Cholesterol	4.15	0.546	2.46	0.561	-41%	p<0.001
LDL cholesterol	2.3	0.434	0.699	0.362	-70%	p<0.001
HDL cholesterol	1.4	0.389	1.64	0.431	17%	p<0.001
Triglyceride	1.71	0.553	1.48	0.606	-13%	0.27
Glucose	6.67	0.649	6.96	1.13	4%	0.18
Liverfat	7.59	7.09	7.87	8.31	4%	0.72
ApoA1	142	25.2	152	24.9	7%	0.005
ApoB	79.3	12.8	37.6	9.85	-53%	p<0.001

Table 5.9: Comparison of relevant individual characteristics before and after the evolocumab treatment. Weight is expressed in Kg, cholesterol, LDL cholesterol, HDL cholesterol, TG and glucose are in mmol/L, while liver fat is in percentage and apoA1 and apoB are in mg/L.

General description of results

Individual characteristics The individuals have an average body weight of about 84 kg at the beginning of the study and that remains almost unvaried at the end of the study. Significant differences concern all the cholesterol levels, with cholesterol, LDL cholesterol, HDL cholesterol undergoing a change in average respectively amounting -41, -70 and 17 percent. Triglyceride levels vary also significantly, going from 1.61 mmol/L to 1.32 mmol/L, thereby experiencing a 14 percent decrease in average. Glucose levels undergo a slight increase in average amounting 4 percent. Liver fat undergoes a slight decrease in average (3 percent). ApoB undergoes a significant decrease (with decrease in average of 53 percent). ApoA1 undergoes a significant increase (the average increases by 7 percent).

	visit 1 mean	visit 1 std	visit 2 mean	visit 2 std	mean change	p-value
apoA4 FCR	2.28	0.69	2.76	0.62	21%	0.018
apoA4 PS	12.5	5.01	13.3	3.92	6%	0.35
apoA4 SR	1.46	0.48	1.94	0.62	33%	0.005
apoC-III FCR	1.33	0.33	1.47	0.31	10%	0.03
apoC-III PS	11.52	4.58	10.03	5.55	-13%	0.252
apoC-III SR	7.82	2.58	7.35	3.06	-6%	0.564
apoE FCR	6.22	0.64	9.01	1.8	45%	<0.001
apoE PS	3.65	0.82	2.43	0.78	-33%	0.001
apoE SR	1.19	0.22	1.12	0.3	-6%	0.494

Table 5.10: Comparison of kinetic parameters before and after the evolocumab treatment. FCRs are expressed in pools/day, PS is expressed in mg/dL and SR in mg/(Kg day)

Kinetic parameters ApoA4 experiences significant increases in FCR and SR, the increase in average respectively amounting to 21 and 33 percent (p-values respectively 0.018 and 0.005). Due to the low quality of the apoA4 data, these changes might not be reliable. ApoA4 pool size increases in average by 6 percent, this change being non significant.
The fractional catabolic rate of apoC-III increases significantly by 10 percent in average. The pool size and secretion rate of apoC-III decrease non-significantly, respectively going from 11.52 to 10.03 mg/dL and from 7.82 to 7.35 mg/(kg day).
For apoE the FCR increases significantly, from 0.26 to 0.38 pools/day, the increase in average amounting to 45 percent. The pool size of apoE decreases significantly with increase in average of 33 percent. The secretion rate for apoE decreases only slightly (going from 0.74 to 0.69 mg/(Kg day)).

The information delivered from Figure 5.2 is very valuable. For apoC-III PS and SR we can notice how the parameters tend mostly to decrease from visit one to visit two, but the presence of an outlier, the same individual in both cases, shifts this balance and erase the possibility of a significant decrease for these two parameters from visit one to visit two. For LDL cholesterol, the massive decrease is highlighted by reduction of the scatter plot to a very narrow area of the graph.

Summary of results for evolocumab Few changes between before and after the evolocumab treatment are strongly significant. First is the decrease in ApoB (with decrease in average of 53 percent and $p<0.001$). Secondly, the decrease(in average) 70 percent of LDL cholesterol ($p<0.001$). Significant is also a decrease in cholesterol (decrease in average 41 percent with $p<0.001$). Other significant changes are an increase in HDL cholesterol (increase in average 17 percent with $p<0.001$). Moreover we can notice a decrease in ApoA1 (in average 7 percent and $p=0.005$).

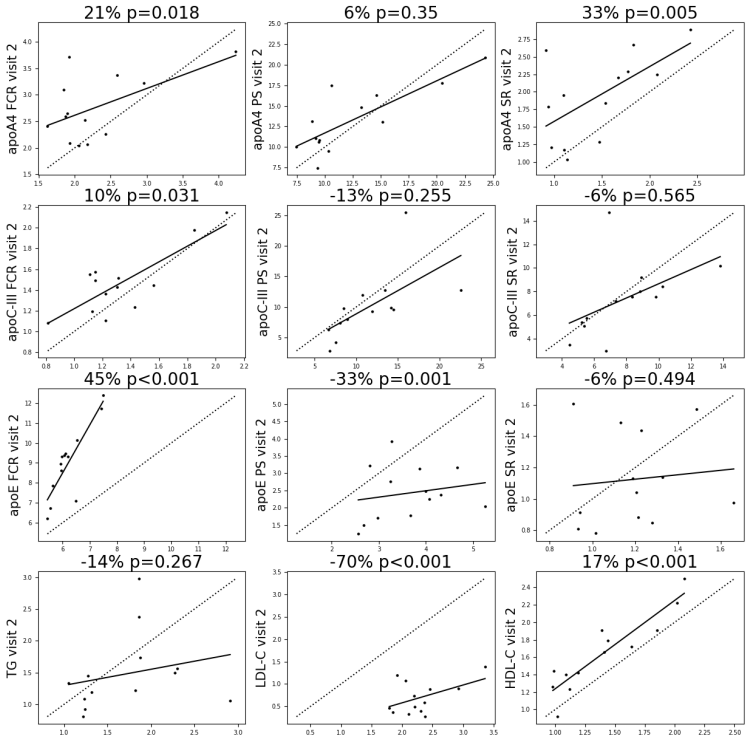

Figure 5.2: Kinetic parameters and relevant covariates before (x axis) and after (y axis) evolocumab treatment

Many of the biological parameters are affected by the treatment. ApoE reports an increase (average 45 percent) of its fractional catabolic rate (p<0.001), while its pool size diminishes significantly (p=0.001 with decrease in average amounting 33 percent). ApoC-III is affected in its FCR that increases with increase in average of 10 percent (p=0.03). ApoA4 experiences an increase in FCR and SR, but these might not be truly reliable due to the scarcity of the apoA4 data, as noticable in Table 3.2 at page 32.

5.3.3 Liraglutide

General description of results

Individual characteristics After the liraglutide treatment weight decreases significantly (p<0.001) going from a mean of 100 kg to a mean of 98 kg. The cholesterol level decreases in a non-significant way, going from 3.86 mmol/L in average to 3.96 mmol/L. LDL cholesterol level decreases significantly with a decrease in mean of 12 percent. HDL cholesterol level remains almost unvaried in mean (1.74 mmol/L at visit one and 1.68 mmol/L at visit two). The

	visit 1 mean	visit 1 std	visit 2 mean	visit 2 std	mean change(%)	p-value
Weight	100.29	10.65	98.17	10.6	-2	<0.001
Cholesterol	3.76	0.73	3.56	0.7	-5	0.061
LDL cholesterol	2.01	0.74	1.78	0.64	-12	0.01
HDL cholesterol	1.24	0.34	1.24	0.37	0	0.958
Triglyceride	1.73	0.69	1.65	0.67	-5	0.454
Glucose	9.51	2.33	7.81	1.53	-18	<0.001
Liver fat	15.6	7.57	11.6	6.07	-26	<0.001

Table 5.11: Comparison of relevant individual characteristics before and after the liraglutide treatment. Weight is expressed in Kg, cholesterol, LDL cholesterol, HDL cholesterol, TG and glucose are in mmol/L, while liver fat is in percentage

triglyceride levels remain stable, their average going from 1.74 mmol/L to 1.53 mmol/L. Glucose level and liver fat level decrease significantly, the first going from 9.51 mmol/L to 7.81 mmol/L) and the second going from 15.6% to 11.6% (respectively 18 and 26 percent decrease in mean).

	visit 1 mean	visit 1 std	visit 2 mean	visit 2 std	change (%)	p-value
apoA4 FCR	2.31	0.07	2.99	0.49	29	<0.001
apoA4 PS	15.49	3.71	13.86	2.8	-11	0.166
apoA4 SR	1.91	0.46	2.19	0.42	14	0.1
apoC-III FCR	1.5	0.15	1.91	0.43	27	0.002
apoC-III PS	11.55	4.48	9.84	3.88	-15	0.066
apoC-III SR	8.98	3.03	9.4	2.96	5	0.624
apoE FCR	5.73	0.42	5.72	0.88	0	0.963
apoE PS	5.93	1.21	4.91	1.03	-17	0.01
apoE SR	1.79	0.26	1.47	0.2	-18	<0.001

Table 5.12: Comparison of kinetic parameters before and after the liraglutide treatment. FCRs are expressed in pools/day, PS is expressed in mg/dL and SR in mg/(Kg day)

Kinetic parameters Two of the kinetic parameters for apoA4 (FCR and SR) have undergone a seemingly relevant change between before and after the treatment (respectively increasing in mean by 29 and 14 percent), only the first one being significant, while the second being above the significance level (0.1). The pool size have decreased by 11 percent in mean, this decrease not being significant. As already mentioned for the fructose and evolocumab treatment, apoA4 parameter estimates might not be so reliable as the estimates for apoC-III or apoE.
The FCR of apoC-III has increased considerably (with increase in average of 27 percent) and significantly. The pool size has undergone a decrease in mean of 15 percent, this being slightly above the significance level. The SR of apoC-III have increased in average by 5 percent, but this increase is not significant.

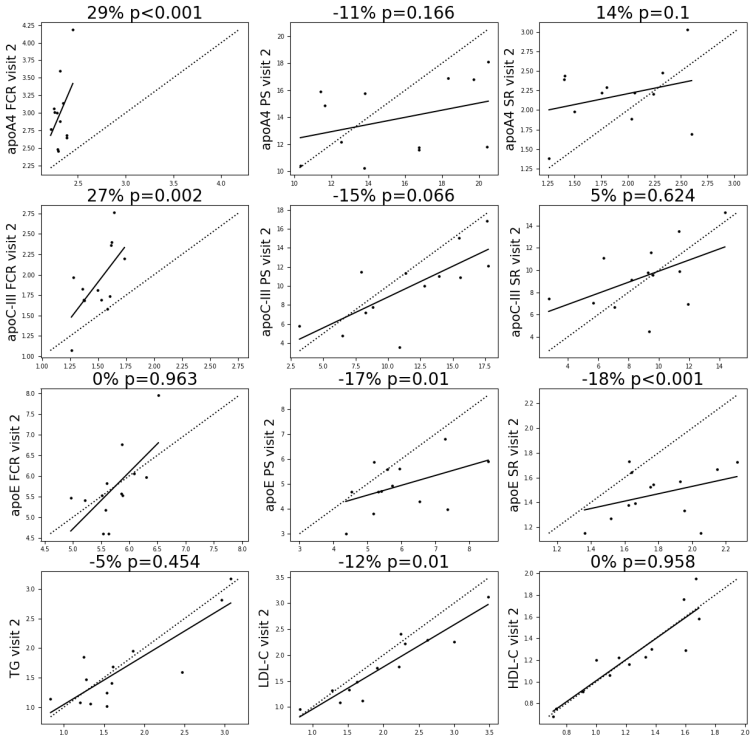

Figure 5.3: Kinetic parameters and relevant covariates before (x axis) and after (y axis) liraglutide treatment

For apoE, while the FCR remains almost unvaried, the pool size and the SR decrease significantly (respectively with a decrease in average of 17 and 18 percent).

Other than confirming the results for the significant and non-significant changes between the two visits, an incongruence of range is to be noticed for the FCR of apoA4 in Figure 5.3.
The range for the FCR of apoA4 is very narrow for the first visit and the values are very low, while they become much higher in the second visit, with the range being larger.
This incongruity could be ascribed to the poor quality of the data, as seen in Table 3.3 at page 33.

Summary of results for liraglutide Few of the individual characteristics have decreased significantly between before and after the liraglutide treatment. LDL cholesterol, glucose and liver fat levels have undergone significant and substantial decreases (respectively 12, 18 and 26 percent decrease). The slight decrease in body weight is significant, with average decrease of 2 percent.

The liraglutide treatment has also brought significant changes in some of the kinetic parameters. We notice significant changes in all the three apolipoproteins; we focus on the changes in apoC-III and apoE due to the low quality of the apoA4 data. The FCR of apoC-III undergoes an increase in average of 27 percent, while the decrease in pool size (15 percent in average) is slightly above the significance level (p=0.066). Pool size and SR for apoE decrease respectively by 17 and 18 percent, both of the changes being significant.

5.4 Correlation of kinetic parameters with selected covariates

As a result of the treatments, significant changes occur in some of the kinetic parameters of the three apolipoproteins under analysis and the individual characteristics.
We would like to investigate whether existing baseline correlations between an individual characteristic and the kinetic parameter are maintained with the intervention or whether new ones arise (Picture 1) and whether the changes in these two variables between the first and second visit correlate (Picture 2).

5.4.1 Fructose

HDL cholesterol

Figure 5.4 (Picture 1):
HDL cholesterol correlates at visit one with apoA4 PS, but this relationship is not kept at the second visit. Slightly above the significance level is the p-value for the linear relationship between FCR of apoE and the HDL cholesterol at visit one.

Figure 5.5 (Picture 2):
There are no significant correlations between the changes in HDL-cholesterol and the changes in the kinetic parameters.

Conclusions
No strong correlations can be found between HDL cholesterol levels and the kinetic parameters, nor between their changes.

LDL cholesterol

Figure 5.6 (Picture 1):
Several of the regression lines describing the kinetic parameters as a linear

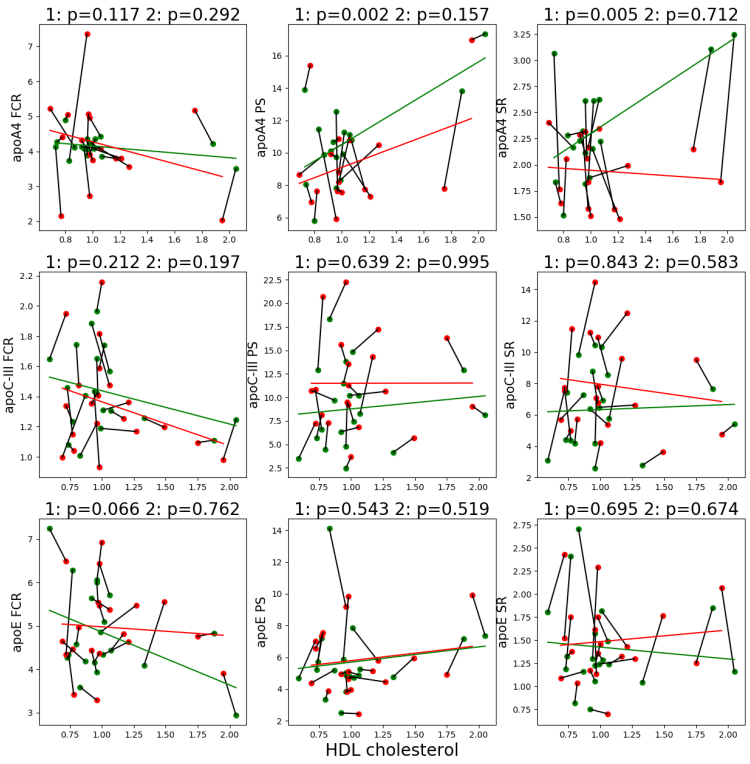

Figure 5.4: Individual HDL cholesterol levels (x axis) and kinetic parameters (y axis) before (green) and after (red) fructose treatment

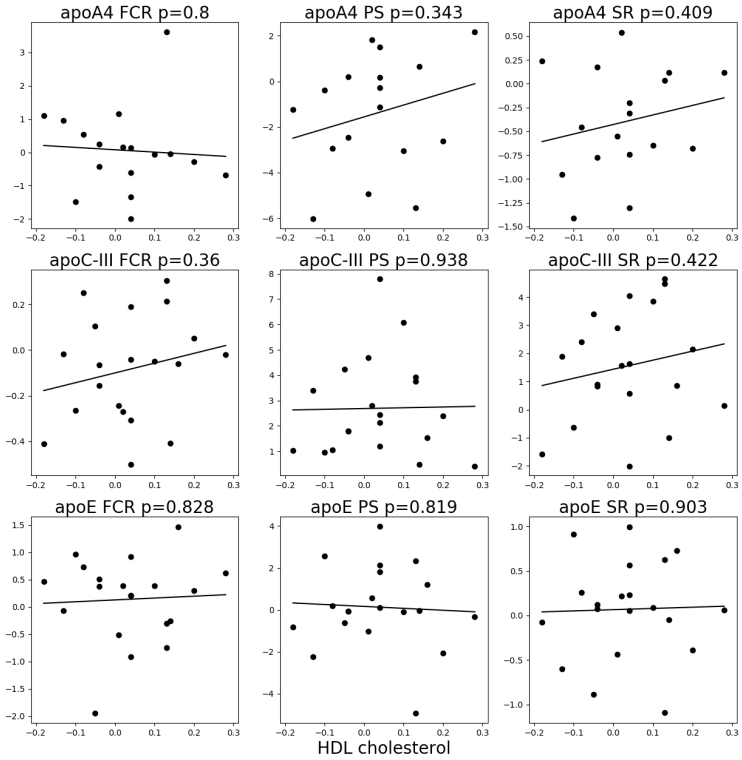

Figure 5.5: Individual changes in HDL cholesterol levels (x axis) and kinetic parameters (y axis) between before and after fructose treatment

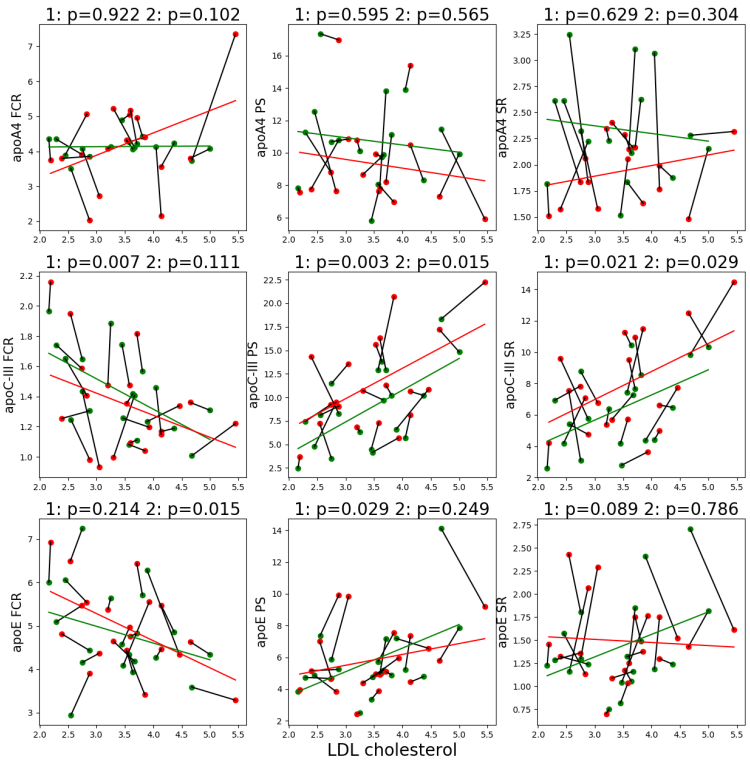

Figure 5.6: Individual LDL cholesterol levels (x axis) and kinetic parameters (y axis) before (green) and after (red) fructose treatment

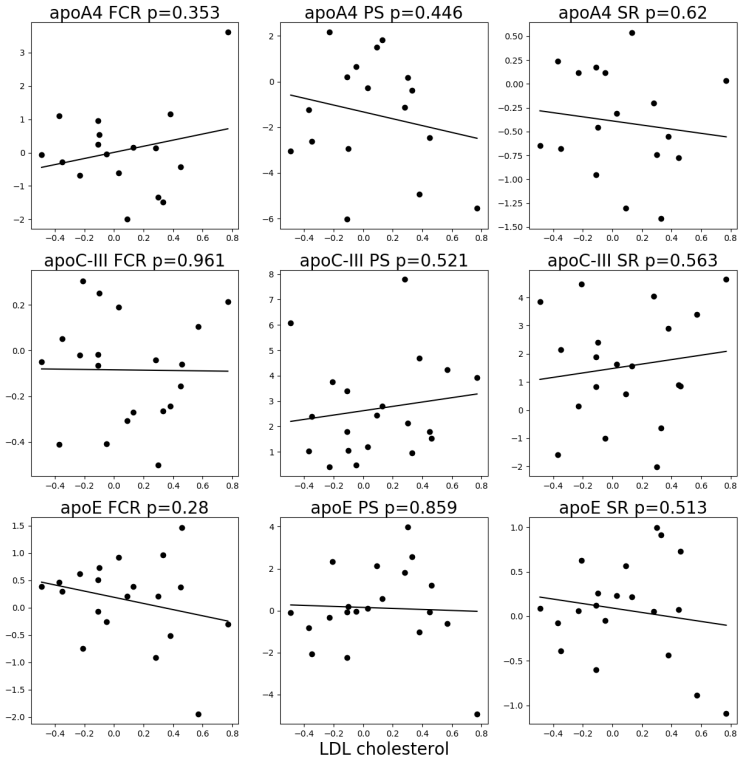

Figure 5.7: Individual changes in LDL cholesterol levels (x axis) and kinetic parameters (y axis) and kinetic parameters between before and after fructose treatment

function of the LDL cholesterol have statistical significance. At visit one the correlation between LDL cholesterol and apoC-III FCR is significant. For both of the visits, the correlation between LDL cholesterol and the pool size of apoC-III and the SR of apoC-III is significant.
Regarding apoE, a significant correlation with apoE FCR is present at visit two, while a correlation with the pool size of apoE is present at visit one.

Figure 5.7 (Picture 2):
The picture fails to highlight any significant correlations between the changes in LDL-cholesterol and the changes in the kinetic parameters.

Conclusions
LDL cholesterol strongly correlates with apoC-III FCR and apoC-III PS at visit one. The relationships respectively loses its significance and become weaker at visit two.
There is no significant correlation between changes in LDL cholesterol and the ones in the kinetic parameters.

Triglyceride levels

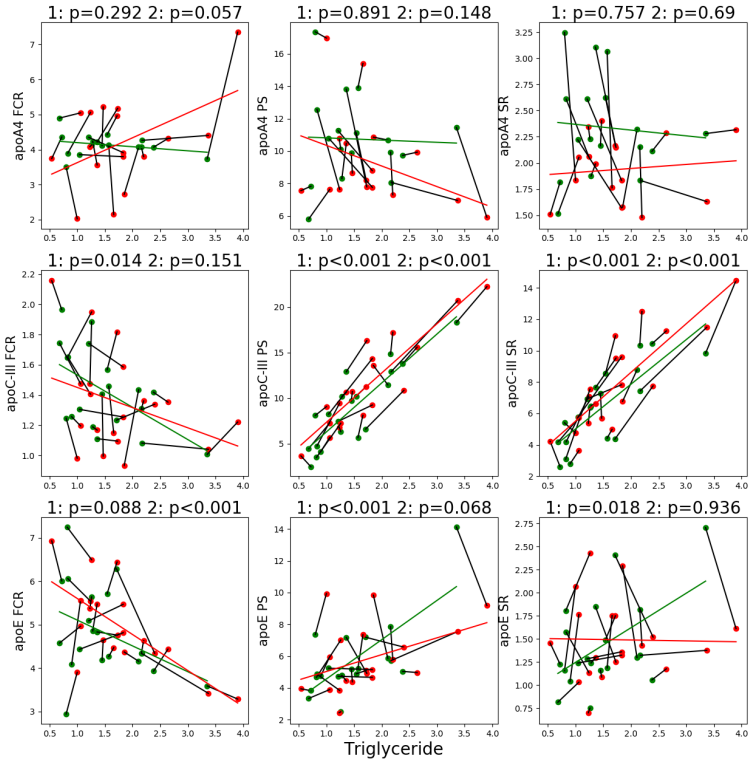

Figure 5.8: Individual triglyceride levels (x axis) and kinetic parameters (y axis) before (green) and after (red) fructose treatment

Figure 5.8 (Picture 1):
The p-value for the hypothesis test on the slope in the linear relationship between the triglyceride levels and the FCR of apoA4 is slightly above the significant level at visit one. The correlation between apoC-III FCR and the triglyceride levels at visit one is significant. The relationship between the pool size and the SR of apoC-III and the triglyceride levels deserves special attention. The correlation between them is significant and the relationship does not change significantly between the two visits. In fact we see that the green point move to the red point parallel to the regression line, implying that the relationship remains unvaried between the two visits.
For apoE each of the parameters have a significant correlation with the triglyceride levels, for only one of the two visits, the FCR at visit two (visit one only being slightly above the significance level), PS at visit one (visit two having p-value of 0.07) and SR at visit one.

Figure 5.9 (Picture 2):
For apoC-III pool size and SR, their changes correlate significantly with the

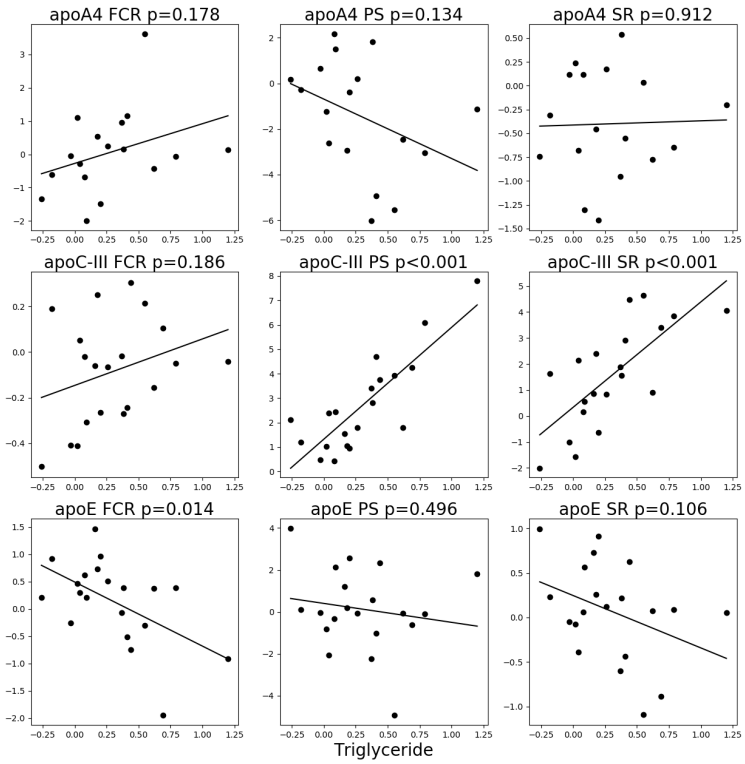

Figure 5.9: Individual changes in triglyceride levels (x axis) and kinetic parameters (y axis) between before and after fructose treatment

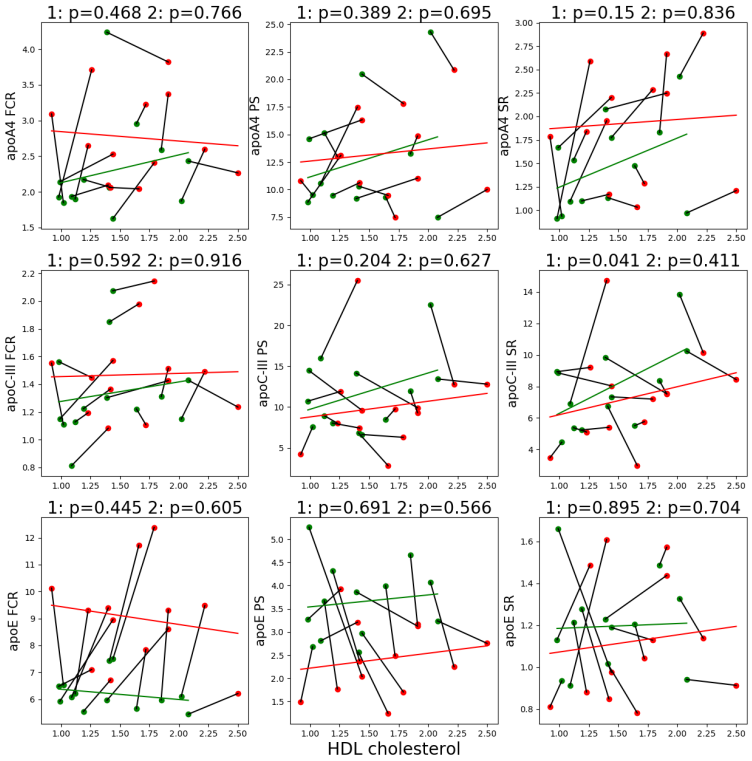

Figure 5.10: Individual HDL cholesterol levels (x axis) and kinetic parameters (y axis) before (green) and after (red) evolocumab treatment

change in triglyceride levels. For the SR of apoC-III, the p-value of the hypothesis test on the intercept being 0 is 0.44, therefore we can imply a merely proportional relationship between the changes in apoC-III pool size and the triglyceride levels.
For apoE the change in FCR can be described linearly as a function of the change in triglyceride levels.

Conclusions
ApoC-III PS and SR correlate strongly with the triglyceride levels at visit one and at visit two, keeping an almost unvaried linear relationship. The changes of the kinetic parameters and of triglyceride levels correlate too.
Additionally triglyceride levels correlate positively with apoE PS at visit one. At visit two they correlate negatively with apoE FCR.

5.4.2 Evolocumab

HDL cholesterol

Figure 5.10 (Picture 1):

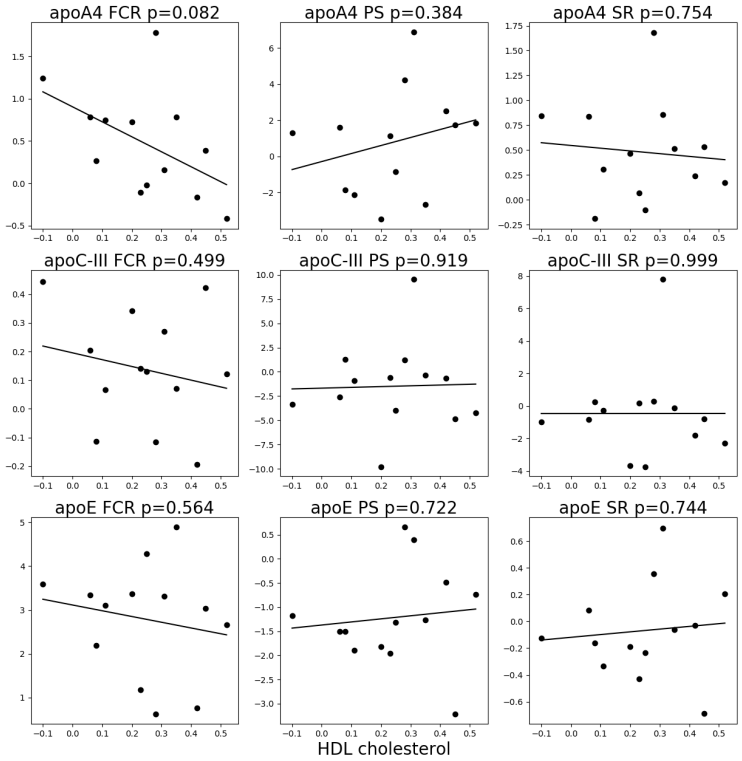

Figure 5.11: Individual changes in HDL cholesterol levels (x axis) and kinetic parameters (y axis) between before and after evolocumab treatment

Only the correlation between apoC-III secretion rate and HDL cholesterol is significant at visit one, while the other kinetic parameters show non significant relationship to HDL cholesterol.

Figure 5.11(Picture 2):
From the picture, we can see that there are no significant correlations between the changes in HDL cholesterol and the changes in the kinetic parameters.

Conclusions
HDL cholesterol levels at visits one or two do not show any strong correlation with the kinetic parameters, nor such associations exist among the changes of these quantities.

LDL cholesterol

Figure 5.12 (Picture 1):
Slightly above the significance value is the p-value for the correlation between the LDL cholesterol and the FCRs of apoA4, apoC-III and apoE, the first ones at visit one and the latter at visit two.

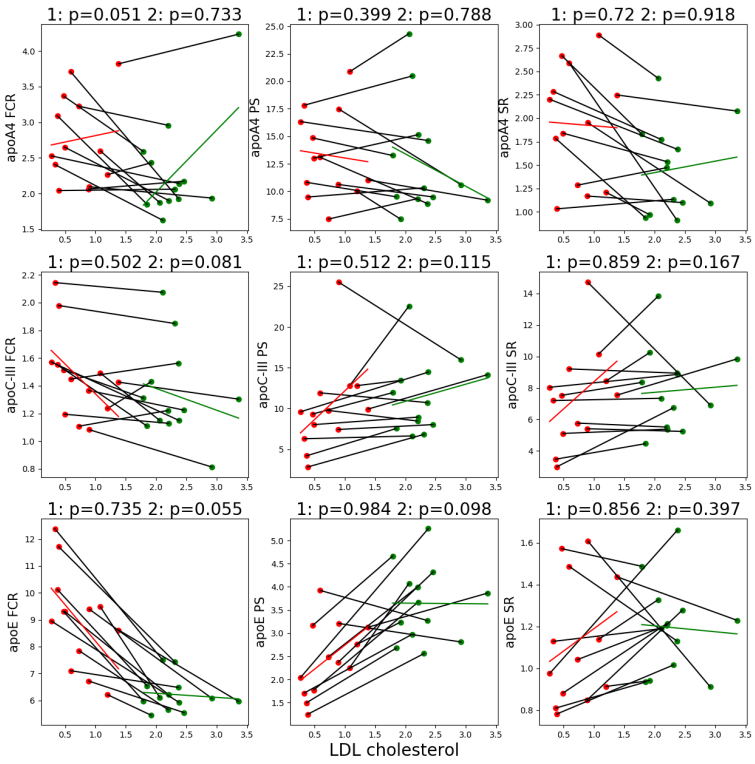

Figure 5.12: Individual LDL cholesterol levels (x axis) and kinetic parameters (y axis) before (green) and after (red) evolocumab treatment

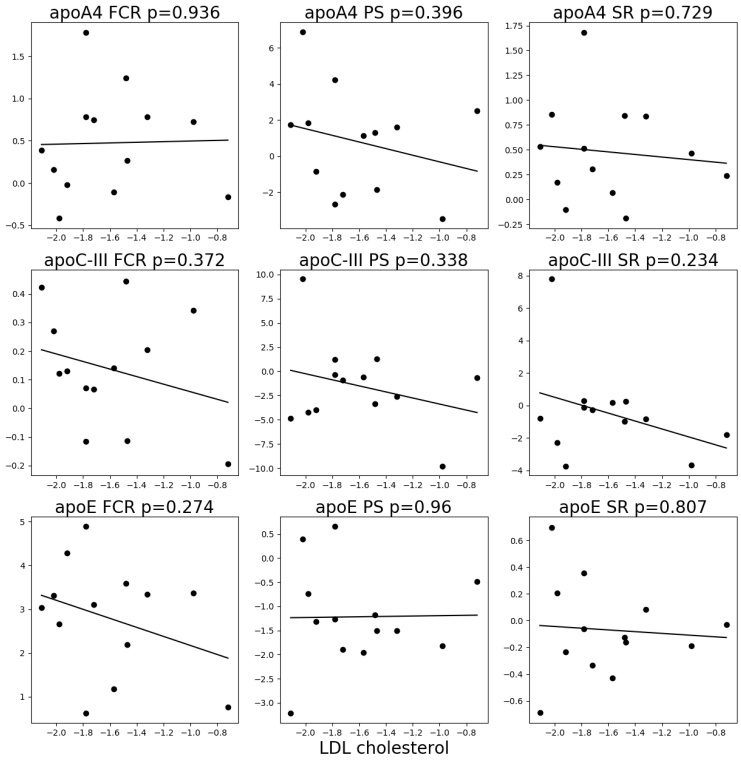

Figure 5.13: Individual changes in LDL cholesterol levels (x axis) and kinetic parameters (y axis) between before and after evolocumab treatment

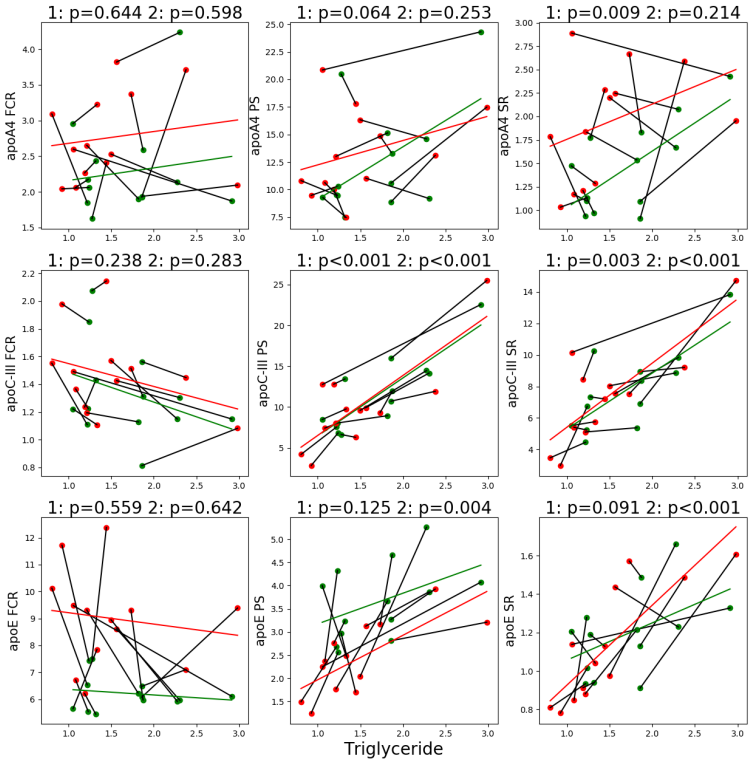

Figure 5.14: Individual triglyceride levels (x axis) and kinetic parameters (y axis) before (green) and after (red) evolocumab treatment

Figure 5.13 (Picture 2):
The picture fails to show any significant correlations between the changes in LDL-cholesterol and the changes in the kinetic parameters.

Conclusions
LDL cholesterol does not show significant correlations with the kinetic parameters for the apolipoproteins and such relation is non-existent also between the changes in these variables.

Triglyceride levels

Figure 5.14 (Picture 1):
The p-value for the correlation of apoA4 PS with triglyceride is slightly above the significance value at visit one. Significant is the correlation between the SR for apoA4 and the triglyceride level.
The significant linear relationship between the pool size and the secretion rate of apoC-III and the triglyceride levels are worth noticing. Both of them are significant at both of the visits. The points move parallel to the green line, resulting in the red lineoverlapping with the green line, for the apoC-III PS and

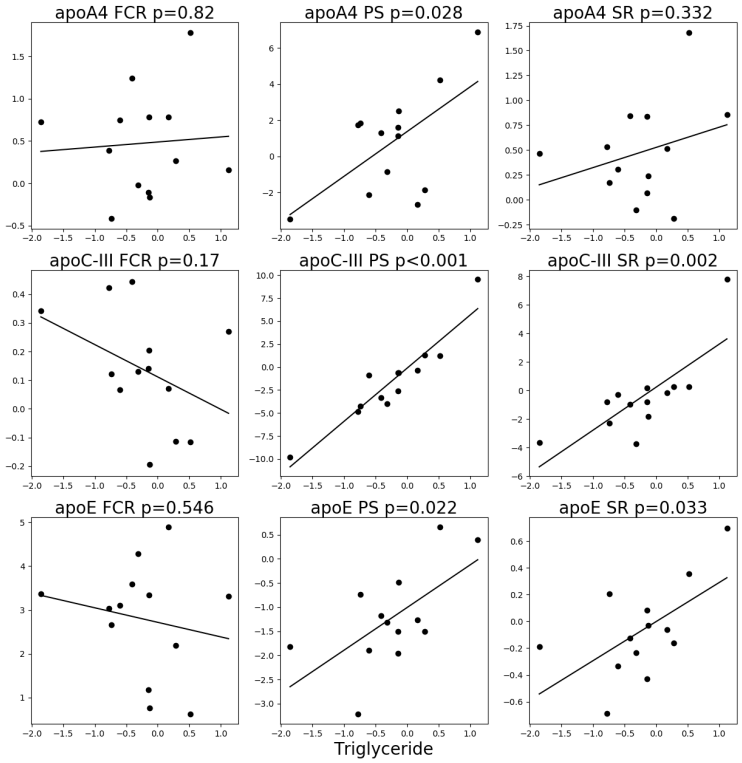

Figure 5.15: Individual changes in triglyceride levels (x axis) and kinetic parameters (y axis) between before and after evolocumab treatment

almost overlapping for the apoC-III SR.
For apoE, the PS and secretion rate can be described linearly as a function of the triglyceride only for the second visit.

Figure 5.15 (Picture 2):
The change in pool size for apoA4 can be described linearly as a function of the change in triglyceride levels.
Significant is also the linear relationship between the changes in PS and SR for apoE and apoC-III and the triglyceride levels.

Conclusions
Triglyceride levels correlate strongly with apoC-III PS and SR at both visits, keeping an almost unvaried linear relationship. The changes in triglyceride levels and in these two kinetic parameters also correlate. At visit two, triglyceride levels correlate with apoE PS and SR.

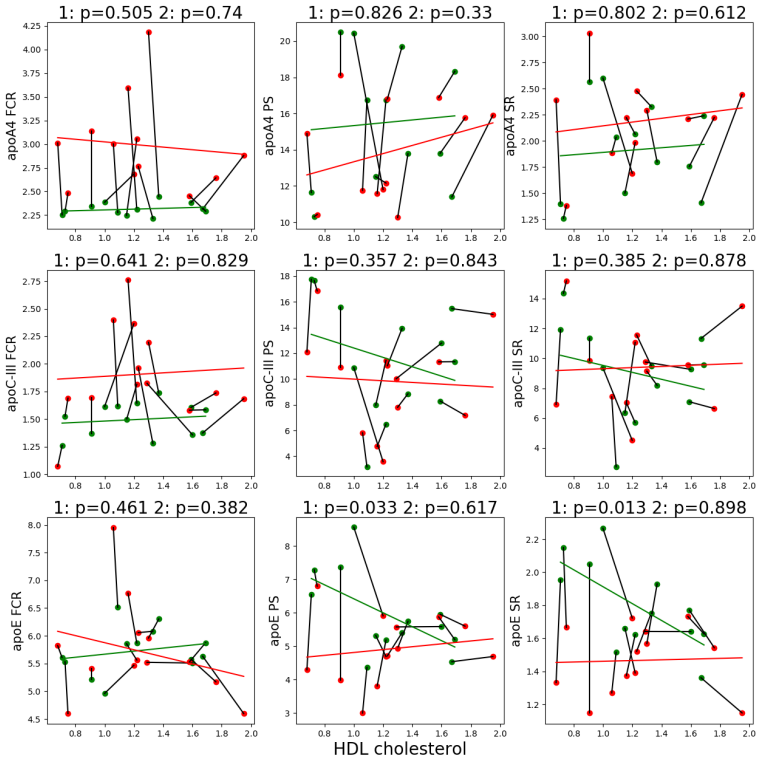

Figure 5.16: Individual HDL cholesterol levels (x axis) and kinetic parameters (y axis) before (green) and after (red) liraglutide treatment

5.4.3 Liraglutide

HDL cholesterol

Figure 5.16 (Picture 1):
The PS and SR of apoE are the only ones who have a significant linear relationship with the HDL cholesterol level and this holds only for the first visit.

Figure 5.17 (Picture 2):
None of the correlations between changes in HDL cholesterol and the kinetic parameters are significant.

Conclusions

No strong correlation can be noticed between HDL cholesterol and the kinetic parameters, nor between the changes in HDL cholesterol and the changes in the apolipoprotein kinetic parameters.

LDL cholesterol

Figure 5.18 (Picture 1):

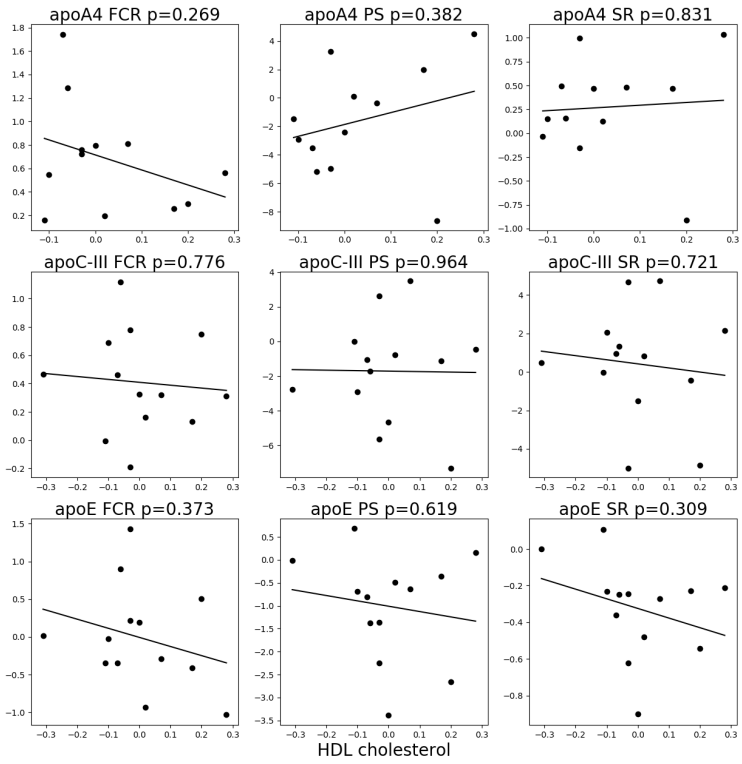

Figure 5.17: Individual changes in HDL cholesterol levels (x axis) and kinetic parameters (y axis) between before and after liraglutide treatment

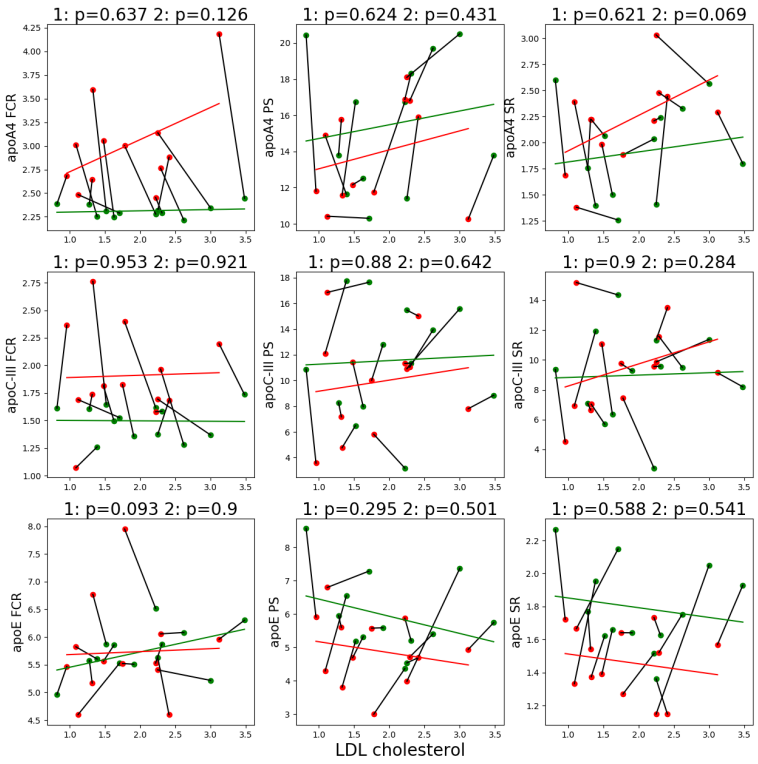

Figure 5.18: Individual LDL cholesterol levels (x axis) and kinetic parameters (y axis) before (green) and after (red) liraglutide treatment

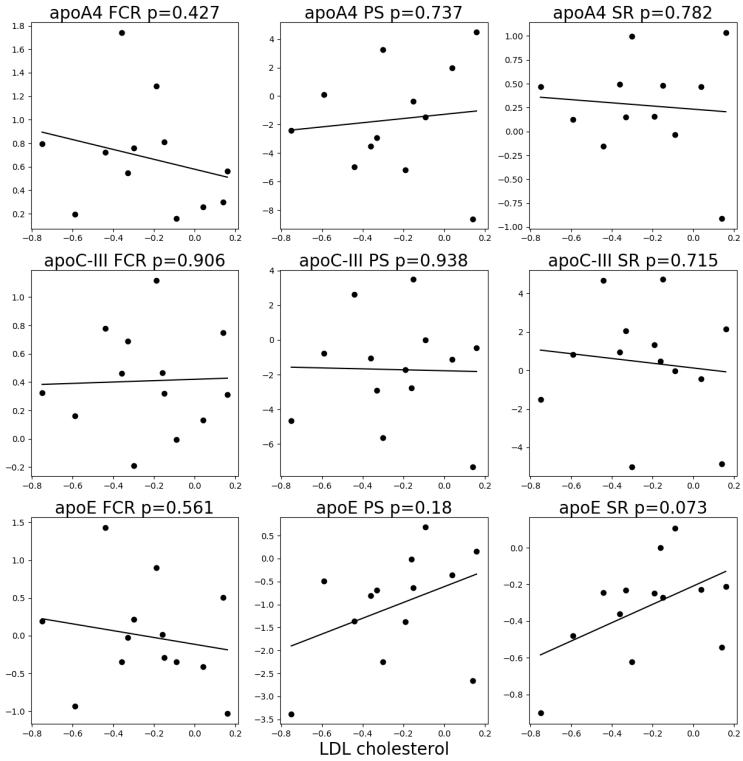

Figure 5.19: Individual changes in LDL cholesterol levels (x axis) and kinetic parameters (y axis) between before and after liraglutide treatment

Slightly above the significance level is the p-value for the correlation between apoA4 SR and LDL cholesterol level.

Figure 5.19 (Picture 2):
There is no significant correlation of the changes in kinetic parameters with changes in the LDL cholesterol level. It can be noticed, though, that the p-value for the linear relationship between SR of apoE and LDL cholesterol level is very close to the significance level (0.073).

Conclusions
There are no significant correlations between LDL cholesterol and the kinetic parameters at visit one nor at visit two, nor between the changes among visit one and two.

Triglyceride levels

Figure 5.20(Picture 1):
The FCR for apoC-III at visit two correlate with the triglyceride levels (p=0.049). At visit one the p-value is not so far from the significance value (p.value 0.085). The pool size of apoC-III correlates with the triglyceride levels in both of the

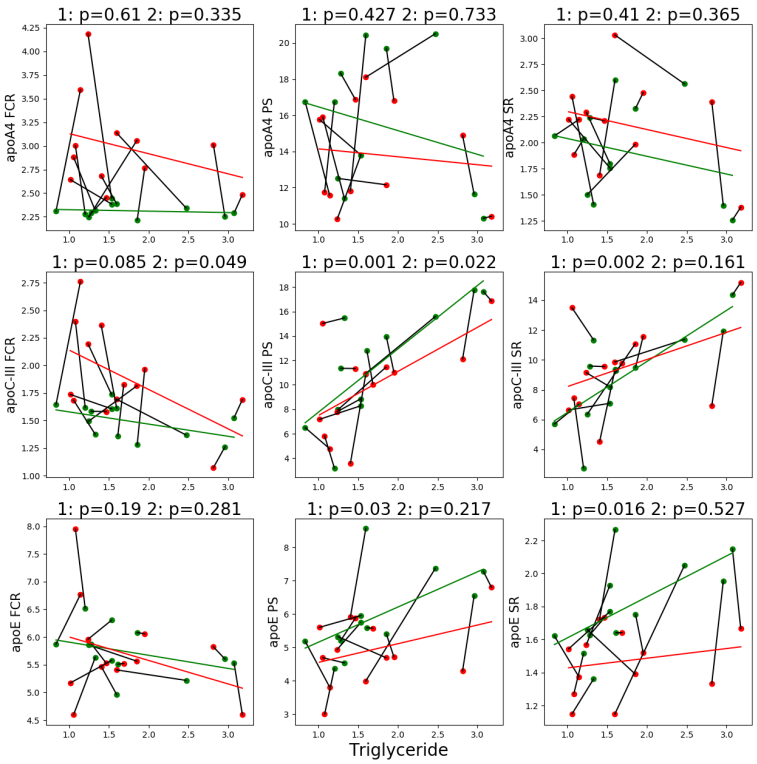

Figure 5.20: Individual triglyceride levels (x axis) and kinetic parameters (y axis) before (green) and after (red) liraglutide treatment

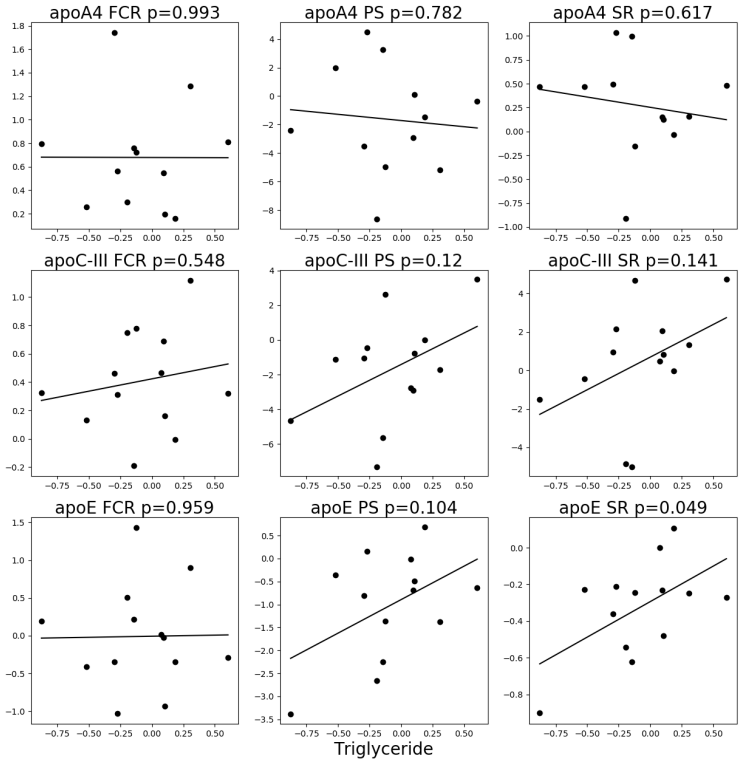

Figure 5.21: Individual changes in triglyceride levels (x axis) and kinetic parameters (y axis) between before and after liraglutide treatment

visits, while its SR correlates with this covariate only at visit one.
Significant are the correlations of apoE PS and apoE SR with the triglyceride levels, but only at visit 1.

Figure 5.21(Picture 2):
The only significant correlation is the one between the changes in SR of apoE and in the triglyceride levels.

Conclusions
The triglyceride levels correlate with the PS and SR of apoC-III but this relationship respectively becomes weaker and loses its significance at visit two.

5.4.4 General conclusions

In the fructose experiment, LDL cholesterol correlates strongly with apoC-III FCR and PS only at visit one. Triglyceride levels correlate at visit one with apoC-III PS and SR, consistently for all the apolipoproteins. In the fructose

and evolocumab experiments these correlations are also kept at the second visit, with an almost unvaried linear relationship among these quantities and the changes in these variables between the visits correlate with the changes in triglyceride levels. This highlights a tight bond between triglyceride, apoC-III PS and SR.
In the fructose and evolocumab experiments, associations could be found at some of the visits between triglyceride levels and kinetic parameters of the apolipoproteins, but these are not consistent across the studies.

6 Summary of papers

6.1 Paper I and Paper II

Increased fasting and postprandial triglyceride concentration after hypercaloric fructose treatment had been observed in systematic reviews and meta-analysis [70, 71, 72], but so far it was not elucidated how.
Though apoC-III is known to be an inhibitor of TRL kinetics, a kinetic study of apoC-III in a hypercaloric fructose intake setting had not been performed yet. This is to our knowledge the first example of such a study.
Datasets for apolipoproteins A-4, C-III and E were created from blood samples through a spectrometry-based proteonomics technology. In Paper I a newly-developed modelling framework is applied to the kinetic study of the three apolipoproteins and the results are reported. Due to the combined analysis, correlation structure of the different apolipoproteins with each other can be analysed. ApoC-III and apoE fractional catabolic rates correlate at baseline as well as the pool sizes of the two apolipoproteins. ApoC-III and apoE are related at baseline with triglyceride levels and for apoC-III the relation with triglyceride holds also after the treatment. ApoC-III concentration increases due to an increased secretion rate.

In Paper II the results for apolipoprotein C-III for 11 individuals are combined with the results of non-steady state kinetics of apoB48- and of apoB100-containing lipoproteins.
As a result of the Fructose hypercaloric treatment many indicators change as shown in Figure 6.1.
ApoC-III increases in concentration (+31.3%, p=0.002) largely due to an increased secretion rate (+30.4%, p=0.01).
The changes in apoC-III concentration correlate with changes in VLDL1 triglyceride (r = 0.75) and in VLDL-apoB100 (r=0.68) and plasma triglyceride (r = 0.79) as it can be seen in Figure 6.2.
In fact, the increased plasma triglycerides was caused by delayed lipolysis as measured by the decreased VLDL1 to VLDL2 fractional transfer rate. In contrast, VLDL1 triglyceride secretion was not increased by fructose intervention. We can conclude that fructose intervention induced an increased secretion of

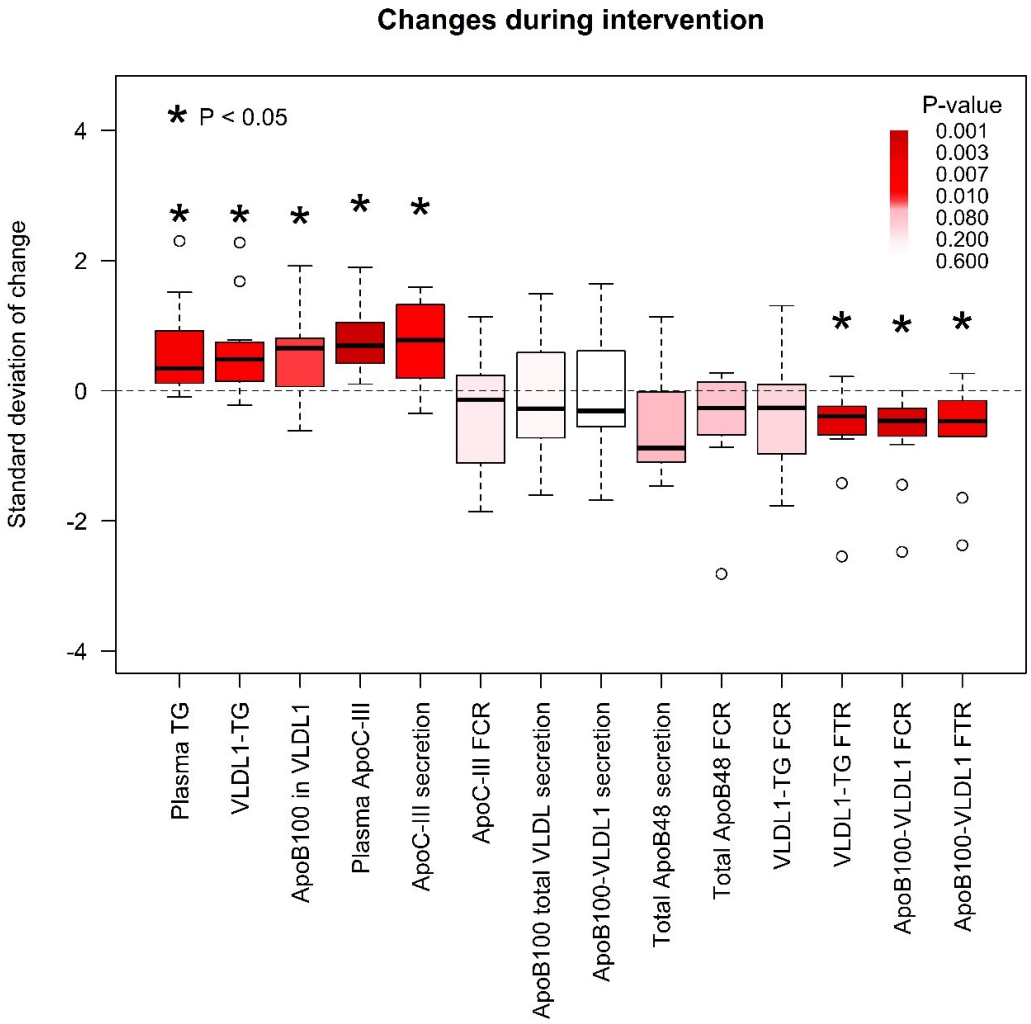

Figure 6.1: Boxplots for the changes in relevant individual characteristics and kinetics parameters

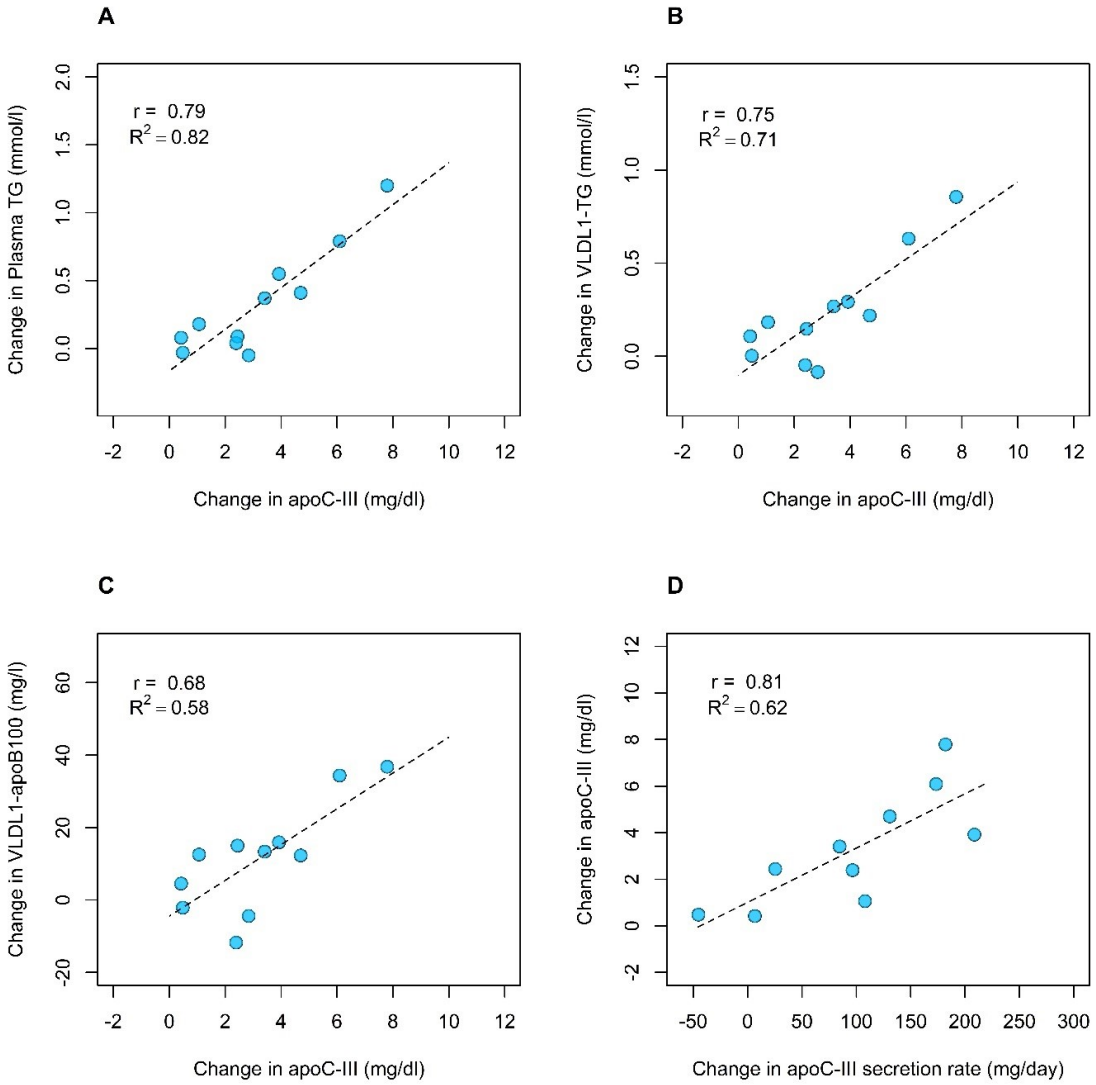

Figure 6.2: Linear regression among changes in individual characteristics and changes in apoC-III PS (A-C) and SR (D)

apoC-III resulting in increased apoC-III concentration and delayed lipolysis of plasma triglycerides.

6.2 Paper III

PCSK-9 fosters the degradation of LDL receptors. PSCK-9-inhibitor based drugs therefore increase LDL-receptor availability. Dyslipidemia is a common characteristic among type-II-diabetic patients. One of its hallmark is the increased LDL concentration.
Apolipoprotein E (apoE) is a binder for LDL receptor and apolipoprotein C-III (apoC-III) is an inhibitor of lipoprotein lipase, especially active in the VLDL1

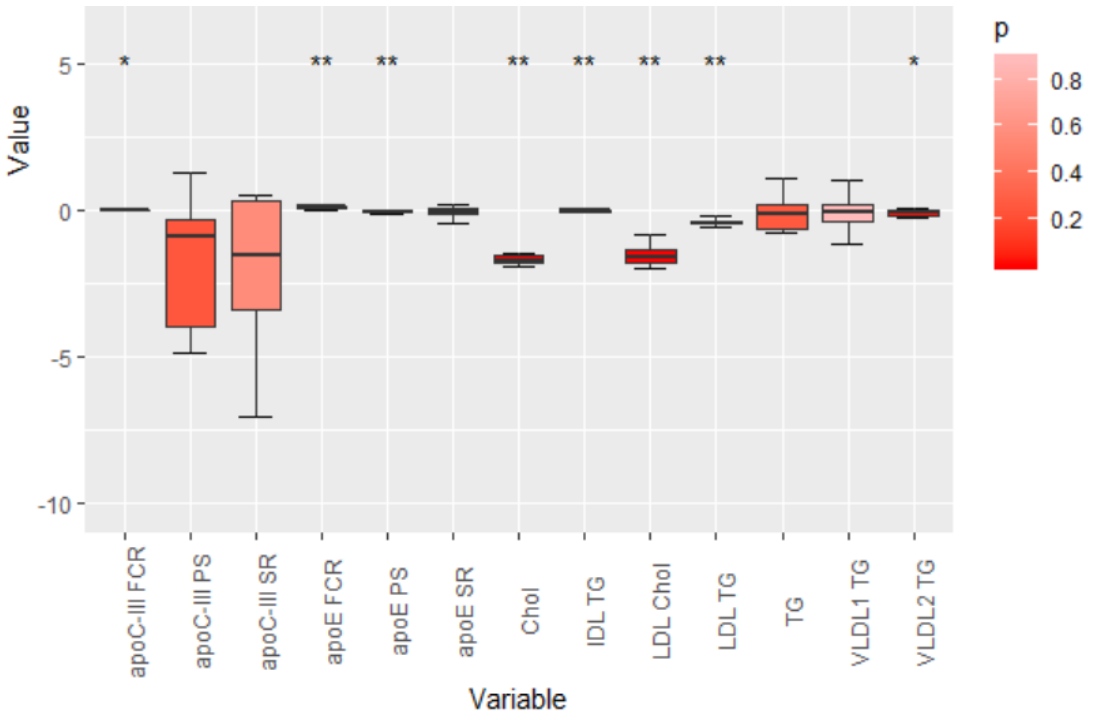

Figure 6.3: Boxplots for the changes in the individual characteristics and relevant kinetic parameters. One star indicates a pvalue smaller than 0.05 while two stars indicate a pvalue smaller than 0.01

fraction [73].
This study analyses apoC-III and apoE postprandial kinetics before and after evolocumab treatment on the background of statin therapy.
The data generated with proteomic technology are analysed with nonlinear mixed effects models approach. The model developed in Paper I was used to analyse apoE and apoC-III kinetics separately.
The treatment has led to an improvement of the lipid profile and to changes in many kinetic parameters as can be seen in Figure 6.3. ApoE concentration is significantly reduced (-33% with p=0.001), largely due to an increased fractional catabolic rate for apoE (45% increase with p<0.001). The correlation structure of the individual characteristics and the apolipoprotein kinetics parameters together with considerations about the contribution of the different lipoprotein fractions to the apolipoprotein E FCR hint to the fact that the observable effects on the fasting lipoprotein concentrations and on apoE might be largely due to an increased VLDL2 FCR. Only the combination of our results with lipoprotein kinetics study can confirm or disprove this hypothesis.

7 Conclusions and future work

7.1 Contributions to the field

Apolipoproteins are key-role players in metabolic diseases.
Despite advances in computational tools in general, implementation of these has been slow in many fields. Data in this thesis is generated with a novel, semi-automatic, low-cost proteomics based method. When combined, these instruments allow us to rapidly determine plasma kinetics of apolipoproteins. The key ingredient with which the software shall be fed is the structural model and several small choices shall be done on the statistical model. Here resides the contribution of the author to the research field. For the first time a full framework for apolipoprotein modelling (comprehensive of structural model and statistical model) has been developed and validated for the study of the three apolipoproteins: apoA4, apoC-III and apoE. This systematic approach, illustrated in the box, could be applied in the future to other studies of apolipoprotein kinetics. The kinetic data for apolipoproteins A4, C-III and E have been analysed for the first time before and after a fructose treatment and a evolocumab treatment.
Kinetic studies combined with mathematical modelling help us make inference on biological processes. In this work, the field of application was the study of dyslipidemia in type II diabetic individuals and abdominally obese individuals. Due to the analysis through mathematical modelling, advances in the knowledge in the biological domain have been unlocked. Even though all the studies reported significant changes in some parameters for apoA4 kinetics, the quality of data was worse for apoA4, therefore we have chosen to mostly focus on the results regarding the apoE and apoC-III kinetics in the papers.
In Paper I we present the developed modelling framework and we apply it to the kinetic study of apolipoproteins A4, C-III and E before and after a fructose treatment. Paper II highlights that increased secretion of apoC-III causes the increase in triglyceride levels, by reducing the fractional transfer rate of VLDL1 particles to VLDL2. Thus, further knowledge can be gained when we combine our results with other results from the same individuals.
In Paper III evolocumab, PCSK9-inhibitor based drug have been shown to decrease the level of apoE concentration in the blood, by increasing apoE FCR. ApoC-III FCR is also increased to a smaller extent.

Model equations

Transversal through the different apolipoproteins, as shown in pages 39- 40 with measurement function as in page 41

Standard statistical model

- distribution of the random effects: log-normal
- error model: remove parameters with relative standard error (rse) > 100
- variance covariance matrix : to the diagonal matrix add a correlation group for the random effects of the parameters p_1, Q_1 and k_{01}

Procedure

- Do the analysis with different statistical models (GEM, DDEM, GEM + COV, DDEM + COV) as explained in Section 4.4
- Do all the models provide good individual fits? In this case proceed with the following step.
- Do all the models provide plausible ranges for the values of the biological parameters? Proceed with the following step.
- Compare the Akaike criterion for the different models. If DDEM + COV model delivers the best Akaike criterion (or if it is only slightly worse than others) and all the parameters show log-normality, eventually after removing an outlier, choose this model. If it is much worse for the DDEM + COV model or if the test for normality does not hold even after removing an outlier, go through the following procedure. At every step, before checking the Akaike criterion, make sure that the new estimation provides good fit to the individual curves and the ranges for the parameters are not too narrow.
  - check whether one parameter for the error model with RSE not far from 100 was eliminated. Reinsert it while keeping the covariance matrix COV (this will become the DDEM2+COV) and check the Akaike criterion again. If no superfluous removal had been performed in the error model choice, the DDEM is the new terms of comparison, therefore the Akaike criterion for the new tested models shall have the AIC for the DDEM as a comparison term. If the AIC was not improved inthis step and/or a parameter show log-normality even after an eventual removal of an outlier, proceed with the following step in parallel with the DDEM+COV and DDEM2+COV
  - Simplify the correlation structure by removing the included correlation couple in the variance covariance matrix with highest p-value. If Akaike criterion is better than for the DDEM (or DDEM2) model or if it just slightly worse, choose this model (DDEM or DDEM2) with simplified variance-covariance matrix, otherwise perform another calculation with DDEM2 model (if necessary) (without the minimal variance-covariance matrix) and compare the obtained Akaike criterion with the AIC for the different models. Choose the model that has substantially better AIC and the normality holds for all the parameters (eventually after removing an outlier).

7.2 Perspectives on the research field

While the application of mathematics to problems in biology has a huge potential in advancing understanding of the human body, the combination of the two fields sometimes penalizes mathematical modelling. In fact the target audience of the message is the biological community, so even though mathematics is essential to obtain the results on the biology, most of the time it is not sensational enough to deserve a detailed description of the model in the methods section of the paper. Therefore choices like the distribution of the random effects or the choices on the error model are most of the time not cited and the equations are not presented. This makes it hard to evaluate what the novelty of the work of the modeller, since similar choices might have been made before from other researchers but not been reported explicitly in papers.
Another aspect that needs improvement is the communication in the collaborative setting between the medical doctors or biologists with the mathematicians. A deep discussion shall be opened at the onset of the research questions so that the experimental design to investigate them will generate an informative dataset. The risk with not complying with this practice is that huge amounts of resources might be wasted if conclusions cannot be drawn due to a poor quality of data.
Another important point is that data shall be shared, so that it can be further analysed for other research purposes. This would be very important also for an educational purpose. The wider availability of real-world data could make the discipline of biomathematics more appealing for younger students.

7.3 Future work

Peptides are sequences of amino acids which are part of a protein. In the spectrometry analysis, more than one peptide specific to each apolipoprotein have been singled out and measured; each of them representing a signal for the concentration of the protein. In this work, we have selected for each apolipoprotein the peptide with the most abundant number of time points. One possible extension for future work is to consider the whole set of peptides for each apolipoprotein. The combination of different signals could potentially lead to more robust estimates.
In this work, we have analysed separately the data for the two visits and then compared them. One could alternatively think about the discrepancy between the parameters at the two visits as a fixed effect and integrate it into the statistical model. In this case, one could analyse the data for the same apolipoprotein for the two visits together. In order for the response from this comprehensive model to be informative (showing which parameters have truly been affected) and not represent only a computational burden, one shall accurately choose which parameters shall contain the fixed effect.
In this work, three datasets were analysed, one regarding abdominally-obese

individuals and the remaining two comprehending individuals with type-II-diabetes. One possible continuation of this work would be to compare the kinetic parameters for the apolipoproteins across the different groups. One could check whether the parameters are consistent across the two diabetic groups and whether differences exist between these and the abdominally obese group. This requires refinement of the data, for example through weight and gender matching, so that the results of the comparison can truly be informative.
Major players in lipid metabolism, apolipoprotein action can only be fully understood when connected to lipoprotein analysis, the particles where they are attached. In Paper II, the results of the apolipoprotein kinetic data analysis, from the author, are combined with the results of the lipoprotein analysis. The dots are connected in this way and a comprehensive understanding of the treatment effect can be drawn.
The integration of the results in Paper III with the ones about lipoprotein kinetics is required to access to the full potential of the study.
In the integrative analysis in Paper II the results are combined after the parameters have been estimated separately with two different models (one for the apolipoprotein and one for the lipoproteins), but one can envision having a unique model that combines apolipoprotein kinetics and lipoprotein kinetics. If this combined model would be practicable, some quantities could potentially be used as covariate and with the same cost for the experiment more individuals could be analysed.

Bibliography

[1] Galileo Galilei. *Il saggiatore*. G. Barbèra, 1864.

[2] WHO. *Cardiovascular Diseases (CVDs)*. 2017. URL: https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (visited on 05/17/2017).

[3] Salim Yusuf et al. "Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study". In: *The lancet* 364.9438 (2004), pp. 937–952.

[4] Ken Gu, Catherine C Cowie, and Maureen I Harris. "Diabetes and decline in heart disease mortality in US adults". In: *Jama* 281.14 (1999), pp. 1291–1297.

[5] Marja-Riitta Taskinen and Jan BorÚn. "New insights into the pathophysiology of dyslipidemia in type 2 diabetes". In: *Atherosclerosis* 239.2 (2015), pp. 483–495.

[6] Boudewijn Klop, Jan Willem F Elte, and Manuel Castro Cabezas. "Dyslipidemia in obesity: mechanisms and potential targets". In: *Nutrients* 5.4 (2013), pp. 1218–1240.

[7] Peter L Bonate et al. *Pharmacokinetic-pharmacodynamic modeling and simulation*. Springer, 2011.

[8] Ronald A Fisher. "On the'probable error'of a coefficient of correlation deduced from a small sample". In: *Metron* 1 (1921), pp. 1–32.

[9] Lewis B Sheiner, Barr Rosenberg, and Vinay V Marathe. "Estimation of population characteristics of pharmacokinetic parameters from routine clinical data". In: *Journal of pharmacokinetics and biopharmaceutics* 5.5 (1977), pp. 445–479.

[10] Mary J Lindstrom and Douglas M Bates. "Nonlinear mixed effects models for repeated measures data". In: *Biometrics* (1990), pp. 673–687.

[11] Russ Wolfinger. "Laplace's approximation for nonlinear mixed models". In: *Biometrika* 80.4 (1993), pp. 791–795.

[12] José C Pinheiro and Douglas M Bates. "Approximations to the log-likelihood function in the nonlinear mixed-effects model". In: *Journal of computational and Graphical Statistics* 4.1 (1995), pp. 12–35.

[13] Bernard Delyon, Marc Lavielle, and Eric Moulines. “Convergence of a stochastic approximation version of the EM algorithm”. In: *Annals of statistics* (1999), pp. 94–128.

[14] Joachim Almquist, Jacob Leander, and Mats Jirstrand. “Using sensitivity equations for computing gradients of the FOCE and FOCEI approximations to the population likelihood”. In: *Journal of pharmacokinetics and pharmacodynamics* 42.3 (2015), pp. 191–209.

[15] Helga Kristin Olafsdottir et al. “Exact Gradients Improve Parameter Estimation in Nonlinear Mixed Effects Models with Stochastic Dynamics”. In: *The AAPS journal* 20.5 (2018), p. 88.

[16] Florian Kronenberg et al. “Low apolipoprotein A-IV plasma concentrations in men with coronary artery disease”. In: *Journal of the American College of Cardiology* 36.3 (2000), pp. 751–757.

[17] IJ Goldberg et al. “Lipoprotein ApoC-II activation of lipoprotein lipase. Modulation by apolipoprotein A-IV.” In: *Journal of Biological Chemistry* 265.8 (1990), pp. 4266–4272.

[18] Jie Qu et al. “Apolipoprotein A-IV: A Multifunctional Protein Involved in Protection against Atherosclerosis and Diabetes”. In: *Cells* 8.4 (2019), p. 319.

[19] Robert D Cohen et al. “Reduced aortic lesions and elevated high density lipoprotein levels in transgenic mice overexpressing mouse apolipoprotein A-IV.” In: *The Journal of clinical investigation* 99.8 (1997), pp. 1906–1916.

[20] Xiaofa Qin et al. “Apolipoprotein AIV: a potent endogenous inhibitor of lipid oxidation”. In: *American Journal of Physiology-Heart and Circulatory Physiology* 274.5 (1998), H1836–H1840.

[21] Maria A Ostos et al. “Antioxidative and Antiatherosclerotic Effects of Human Apolipoprotein A-IV in Apolipoprotein E–Deficient Mice”. In: *Arteriosclerosis, thrombosis, and vascular biology* 21.6 (2001), pp. 1023–1028.

[22] G Ferretti et al. “Effect of human Apo AIV against lipid peroxidation of very low density lipoproteins”. In: *Chemistry and physics of lipids* 114.1 (2002), pp. 45–54.

[23] Fei Wang et al. “Apolipoprotein A-IV improves glucose homeostasis by enhancing insulin secretion”. In: *Proceedings of the National Academy of Sciences* 109.24 (2012), pp. 9641–9646.

[24] Zemin Yao. “Human apolipoprotein C-III-a new intrahepatic protein factor promoting assembly and secretion of very low density lipoproteins”. In: *Cardiovascular & Haematological Disorders-Drug Targets (Formerly Current Drug Targets-Cardiovascular & Hematological Disorders)* 12.2 (2012), pp. 133–140.

[25] Meenakshi Sundaram et al. “Expression of apolipoprotein C-III in McA-RH7777 cells enhances VLDL assembly and secretion under lipid-rich conditions”. In: *Journal of lipid research* 51.1 (2010), pp. 150–161.

[26] Jan Borén et al. “Kinetic and related determinants of plasma triglyceride concentration in abdominal obesity: multicenter tracer kinetic study”. In: *Arteriosclerosis, thrombosis, and vascular biology* 35.10 (2015), pp. 2218–2224.

[27] DA Chappell et al. “Low density lipoprotein receptors bind and mediate cellular catabolism of normal very low density lipoproteins in vitro.” In: *Journal of Biological Chemistry* 268.34 (1993), pp. 25487–25493.

[28] Jheem D Medh et al. “Lipoprotein lipase-and hepatic triglyceride lipase-promoted very low density lipoprotein degradation proceeds via an apolipoprotein E-dependent mechanism”. In: *Journal of lipid research* 41.11 (2000), pp. 1858–1871.

[29] Robert W Mahley and Yadong Huang. “Atherogenic remnant lipoproteins: role for proteoglycans in trapping, transferring, and internalizing”. In: *The Journal of clinical investigation* 117.1 (2007), pp. 94–98.

[30] WHO. *Obesity and overweight*. 2020. URL: `https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight` (visited on 03/03/2020).

[31] WHO. *Diabetes*. 2020. URL: `https://www.who.int/news-room/fact-sheets/detail/diabetes` (visited on 10/30/2018).

[32] David B Allison et al. “Annual deaths attributable to obesity in the United States”. In: *Jama* 282.16 (1999), pp. 1530–1538.

[33] Kai-Ting Chang et al. “Which obesity index is the best predictor for high cardiovascular disease risk in middle-aged and elderly population?” In: *Archives of gerontology and geriatrics* 78 (2018), pp. 165–170.

[34] Robert H Eckel, Scott M Grundy, and Paul Z Zimmet. “The metabolic syndrome”. In: *The lancet* 365.9468 (2005), pp. 1415–1428.

[35] Jean-Pierre Després and Isabelle Lemieux. “Abdominal obesity and metabolic syndrome”. In: *Nature* 444.7121 (2006), pp. 881–887.

[36] Emerging Risk Factors Collaboration et al. “Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies”. In: *The Lancet* 375.9733 (2010), pp. 2215–2222.

[37] Joe M Chehade, Margaret Gladysz, and Arshag D Mooradian. “Dyslipidemia in type 2 diabetes: prevalence, pathophysiology, and management”. In: *Drugs* 73.4 (2013), pp. 327–339.

[38] Fred Paccaud et al. “Dyslipidemia and abdominal obesity: an assessment in three general populations”. In: *Journal of clinical epidemiology* 53.4 (2000), pp. 393–400.

[39] Charles Couillard et al. “Postprandial triglyceride response in visceral obesity in men.” In: *Diabetes* 47.6 (1998), pp. 953–960.

[40] Surapon Tangvarasittichai et al. “Abdominal obesity associated with elevated serum Butyrylcholinesterase activity, insulin resistance and reduced high density lipoprotein-cholesterol levels”. In: *Indian Journal of Clinical Biochemistry* 30.3 (2015), pp. 275–280.

[41] William E Kraus et al. “Effects of the amount and intensity of exercise on plasma lipoproteins”. In: *New England Journal of Medicine* 347.19 (2002), pp. 1483–1492.

[42] Paul T Williams et al. “Changes in lipoprotein subfractions during diet-induced and exercise-induced weight loss in moderately overweight men.” In: *Circulation* 81.4 (1990), pp. 1293–1304.

[43] Steven E Nissen et al. “Statin therapy, LDL cholesterol, C-reactive protein, and coronary artery disease”. In: *New England Journal of Medicine* 352.1 (2005), pp. 29–38.

[44] Marc S Sabatine et al. “Evolocumab and clinical outcomes in patients with cardiovascular disease”. In: *New England Journal of Medicine* 376.18 (2017), pp. 1713–1722.

[45] Ronald M Krauss. “Lipids and lipoproteins in patients with type 2 diabetes”. In: *Diabetes care* 27.6 (2004), pp. 1496–1504.

[46] Daniel Gaudet et al. “Antisense inhibition of apolipoprotein C-III in patients with hypertriglyceridemia”. In: *New England Journal of Medicine* 373.5 (2015), pp. 438–447.

[47] L Duvillard et al. “Metabolic abnormalities of apolipoprotein B-containing lipoproteins in non-insulin-dependent diabetes: a stable isotope kinetic study.” In: *European journal of clinical investigation* 30.8 (2000), pp. 685–694.

[48] Dick C Chan et al. “Plasma apolipoprotein C-III transport in centrally obese men: associations with very low-density lipoprotein apolipoprotein B and high-density lipoprotein apolipoprotein AI metabolism”. In: *The Journal of Clinical Endocrinology & Metabolism* 93.2 (2008), pp. 557–564.

[49] Martin Adiels et al. “Role of apolipoprotein C-III overproduction in diabetic dyslipidaemia”. In: *Diabetes, obesity and metabolism* 21.8 (2019), pp. 1861–1870.

[50] K Bach-Ngohou et al. “Apolipoprotein E kinetics: influence of insulin resistance and type 2 diabetes”. In: *International journal of obesity* 26.11 (2002), pp. 1451–1458.

[51] Esther MM Ooi et al. “Effect of fenofibrate and atorvastatin on VLDL apoE metabolism in men with the metabolic syndrome”. In: *Journal of lipid research* 53.11 (2012), pp. 2443–2449.

[52] Kenneth R Feingold and Carl Grunfeld. “Introduction to lipids and lipoproteins”. In: *Endotext [Internet]*. MDText. com, Inc., 2018.

[53] M-R Taskinen et al. “Adverse effects of fructose on cardiometabolic risk factors and hepatic lipid metabolism in subjects with abdominal obesity”. In: *Journal of internal medicine* 282.2 (2017), pp. 187–201.

[54] Niina Matikainen et al. “Liraglutide treatment improves postprandial lipid metabolism and cardiometabolic risk factors in humans with adequately controlled type 2 diabetes: A single-centre randomized controlled study”. In: *Diabetes, Obesity and Metabolism* 21.1 (2019), pp. 84–94.

[55] David H Anderson. *Compartmental modeling and tracer kinetics*. Vol. 50. Springer Science & Business Media, 2013.

[56] Martin Berglund et al. "Investigations of a compartmental model for leucine kinetics using non-linear mixed effects models with ordinary and stochastic differential equations". In: *Mathematical medicine and biology: a journal of the IMA* 29.4 (2012), pp. 361–384.

[57] Yuri A Kuznetsov. *Elements of applied bifurcation theory*. Vol. 112. Springer Science & Business Media, 2013.

[58] Martin Berglund et al. "Improved estimation of human lipoprotein kinetics with mixed effects models". In: *PloS one* 10.9 (2015).

[59] Yaning Wang. "Derivation of various NONMEM estimation methods". In: *Journal of Pharmacokinetics and pharmacodynamics* 34.5 (2007), pp. 575–593.

[60] Estelle Kuhn and Marc Lavielle. "Maximum likelihood estimation in nonlinear mixed effects models". In: *Computational statistics & data analysis* 49.4 (2005), pp. 1020–1038.

[61] Robert Hermann and Arthur Krener. "Nonlinear controllability and observability". In: *IEEE Transactions on automatic control* 22.5 (1977), pp. 728–740.

[62] Alejandro F Villaverde. "Observability and structural identifiability of nonlinear biological systems". In: *Complexity* 2019 (2019).

[63] Johan Karlsson, Milena Anguelova, and Mats Jirstrand. "An efficient method for structural identifiability analysis of large dynamic systems". In: *IFAC proceedings volumes* 45.16 (2012), pp. 941–946.

[64] Martin Bergstrand et al. "Prediction-corrected visual predictive checks for diagnosing nonlinear mixed-effects models". In: *The AAPS journal* 13.2 (2011), pp. 143–151.

[65] Hirotogu Akaike. "Information theory and an extension of the maximum likelihood principle". In: *Selected papers of hirotugu akaike*. Springer, 1998, pp. 199–213.

[66] Hirotugu Akaike. "A new look at the statistical model identification". In: *IEEE transactions on automatic control* 19.6 (1974), pp. 716–723.

[67] Marja-Riitta Taskinen et al. "Impact of proprotein convertase subtilisin/kexin type 9 inhibition with evolocumab on the postprandial responses of triglyceride-rich lipoproteins in type II diabetic subjects". In: *Journal of Clinical Lipidology* (2019).

[68] CJ Packard et al. "Apolipoprotein B metabolism and the distribution of VLDL and LDL subfractions". In: *Journal of lipid research* 41.2 (2000), pp. 305–317.

[69] Daniele Coin. "Testing normality in the presence of outliers". In: *Statistical Methods and Applications* 17.1 (2008), pp. 3–12.

[70] Laura Chiavaroli et al. "Effect of fructose on established lipid targets: a systematic review and meta-analysis of controlled feeding trials". In: *Journal of the American Heart Association* 4.9 (2015), e001700.

[71] D David Wang et al. “Effect of fructose on postprandial triglycerides: a systematic review and meta-analysis of controlled feeding trials”. In: *Atherosclerosis* 232.1 (2014), pp. 125–133.

[72] Rodrigo CO Macedo et al. “Effects of fructose consumption on postprandial TAG: an update on systematic reviews with meta-analysis”. In: *British Journal of Nutrition* 120.4 (2018), pp. 364–372.

[73] Alejandro Gugliucci et al. “Short-term isocaloric fructose restriction lowers apoC-III levels and yields less atherogenic lipoprotein profiles in children with obesity and metabolic syndrome”. In: *Atherosclerosis* 253 (2016), pp. 171–177.

[74] Marit Granér et al. “Epicardial fat, cardiac dimensions, and low-grade inflammation in young adult monozygotic twins discordant for obesity”. In: *The American journal of cardiology* 109.9 (2012), pp. 1295–1302.

[75] Marja-Riitta Taskinen et al. “Dual metabolic defects are required to produce hypertriglyceridemia in obese subjects”. In: *Arteriosclerosis, thrombosis, and vascular biology* 31.9 (2011), pp. 2144–2150.

[76] Jesper Lundbom et al. “Long-TE 1H MRS suggests that liver fat is more saturated than subcutaneous and visceral fat”. In: *NMR in Biomedicine* 24.3 (2011), pp. 238–245.

[77] Kristin D Neff and Roos Vonk. “Self-compassion versus global self-esteem: Two different ways of relating to oneself”. In: *Journal of personality* 77.1 (2009), pp. 23–50.

[78] Dawson Church, Midanelle A De Asis, and Audrey J Brooks. “Brief group intervention using Emotional Freedom Techniques for depression in college students: A randomized controlled trial”. In: *Depression research and treatment* 2012 (2012).

[79] Mahvash Shahidi et al. “Laughter yoga versus group exercise program in elderly depressed women: a randomized controlled trial”. In: *International journal of geriatric psychiatry* 26.3 (2011), pp. 322–327.

[80] Leonore De Wit et al. “Depression and obesity: a meta-analysis of community-based studies”. In: *Psychiatry research* 178.2 (2010), pp. 230–235.

[81] Peta Stapleton et al. “Online Delivery of Emotional Freedom Techniques for Food Cravings and Weight Management: 2-Year Follow-Up”. In: *The Journal of Alternative and Complementary Medicine* 26.2 (2020), pp. 98–106.

[82] Adam Biener, John Cawley, and Chad Meyerhoefer. *The high and rising costs of obesity to the US health care system*. 2017.

[83] Dawson Church, Garret Yount, and Audrey J Brooks. “The effect of emotional freedom techniques on stress biochemistry: a randomized controlled trial”. In: *The Journal of nervous and mental disease* 200.10 (2012), pp. 891–896.

[84] Akiko Fujisawa et al. “Effect of laughter yoga on salivary cortisol and dehydroepiandrosterone among healthy university students: A randomized controlled trial”. In: *Complementary therapies in clinical practice* 32 (2018), pp. 6–11.

[85] Per Björntorp. "Do stress reactions cause abdominal obesity and comorbidities?" In: *Obesity reviews* 2.2 (2001), pp. 73–86.

[86] Wendy C King et al. "Alcohol and other substance use after bariatric surgery: prospective evidence from a US multicenter cohort study". In: *Surgery for Obesity and Related Diseases* 13.8 (2017), pp. 1392–1402.

[87] Joseph E Cavanaugh and Andrew A Neath. "The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 11.3 (2019), e1460.

[88] Yongcheng Ren et al. "Prevalence of hypertriglyceridemic waist and association with risk of type 2 diabetes mellitus: a meta-analysis". In: *Diabetes/metabolism research and reviews* 32.4 (2016), pp. 405–412.

[89] SA Ritchie and JMC Connell. "The link between abdominal obesity, metabolic syndrome and cardiovascular disease". In: *Nutrition, Metabolism and Cardiovascular Diseases* 17.4 (2007), pp. 319–326.

Appendix

Chemical analysis for the fructose intervention

Study design

Postprandial kinetic study After an overnight fast, subjects were admitted and a baseline blood sample were drawn at approximately 7:30 am. A bolus injection of [5,5,5-2H3]-leucine (7 mg/kg) was given at 8:00 am. At 10:00 am, the subjects received a fat-rich meal served with a cocoa-fat rich emulsion containing 40 g of olive oil (Amway, Firenze). The meal consisted of bread, cheese, ham, boiled eggs, fresh red pepper, low-fat (1%) milk, orange juice and tea or coffee. Altogether the meal contained 63 g carbohydrate, 69 g fat and 40 g protein, and was consumed within 10 min. The subjects remained physically inactive during the study but were allowed water ad libitum. Blood was drawn between 8:00 am and 6:00 pm as described in [49].

Intra-abdominal fat depots Liver fat content was measured using proton magnetic resonance spectroscopy [75, 76] and subcutaneous abdominal and intra-abdominal fat were measured using magnetic resonance imaging [74] as previously described.

		Peptide Sequence	Retention time	Precursor peptide	Fragment ions	
apoCIII	M0	DALSSVQESQVAQQAR	28.7	858.929196	300.155397	387.187425
apoCIII	M3	DALSSVQESQVAQQAR		860.438611	303.174227	390.206255
apoCIII	ISTD	DALSSVQESQVAQQAR		863.933331	300.155397	387.187425
apoE	M0	LQAEAFQAR	18.4	517.274898	242.149918	313.187031
apoE	M3	LQAEAFQAR		518.784313	245.168748	316.205861
apoE	ISTD	LQAEAFQAR		522.279032	242.149918	313.187031
apoA4	M0	IDQNVEELK	19.2	544.285128	389.239461	260.196868
apoA4	M3	IDQNVEELK		545.794543	392.258291	263.215698
apoA4	ISTD	IDQNVEELK		548.292228	397.25366	268.211067

Table 1: List of the precursor ions for the corresponding pairs of endogenous, d3-endogenous and labelled peptides and selected fragment ions

Measurement of apoC-III, apoE and apoA4 enrichments

Internal standard peptides Stable-isotope-labeled SpikeTides TQL peptides for apoC-III (DALSSVQESQVAQQAR), apoE (LQAEAFQAR) and apoA4 (IDQNVEELK), were purchased from JPT Peptide Technologies (Berlin, Germany). The stable isotope labels (13C615N4 for Arg, 13C615N2 for Lys) were at the C-terminal Arg and Lys residues and a trypsin-cleavable quantification tag (QTag). Purity for each peptide was >95%. The stable-isotope-labeled peptides were used as internal standards (ISTDs) for absolute quantification of apoC-III, apoE and apoA4 respectively. The peptides (1 nmol) were dissolved and diluted in 1% sodium deoxycholate in 100 mmol/l triethylammonium bicarbonate (TEAB) to an ISTD master mix at concentrations of 100 fmol/μl (apoC-III), 15 fmol/ μ L (apoE) and 24 fmol/μL (apoA4). Initially calibration curves were prepared, in which the linearity of the mass spectrometric (MS) determination of the peptides was verified with the internal standard (ISTD) peptides spiked into representative pooled plasma samples. A total of seven plasma digests were prepared as described below with an equimolar mixture of the apoC-III peptide. The concentration interval for peptides was 7 to 469 fmol. This corresponds to a concentration in the plasma of 15 to 938 fmol/μL. For apoE and apoA4 concentration for peptides ranged from 5 to 234 fmol, corresponding to a plasma concentration of 11 to 469 fmol/μL. The MS response was found to be linear within this concentration range for all peptides.

Tryptic digestion of plasma Plasma samples from blood samples drawn before and during the fat-rich mixed meal test were thawed at room temperature and vortexed for 10 min. Plasma (10 μl) was diluted in 340 μl 100 mmol/l TEAB, and 15 μl of diluted plasma was mixed with 10 μl 5% sodium deoxycholate and 3 μl 50 mmol/l DL-dithiothreitol (DTT, Sigma-Aldrich). The samples were reduced for 20 min at 60°C and then alkylated with 1.5 μl 200 mmol/l S-Methyl methanethiosulfonate (MMTS, Sigma-Aldrich) for 20 min at room temperature. The reaction mixture was then diluted in 10 μl 100 mmol/l TEAB and 15 μl ISTD master mix was added. This mixture was digested overnight with 10 μl (0.06 mg/ml) trypsin (Pierce) at 37°C. A second aliquot of 10 μl trypsin (0.06 mg/ml) was added and digestion continued for 3 h at 37°C. H2O (10 μl) and 20 μl 10% trifluoroacetic acid was added to stop digestion and precipitate sodium deoxycholate. Samples were centrifuged for 15 min at 3700 rpm, and the supernatants were desalted on a SOLAμ HRP 96-well plate (Thermo Fisher, USA) using a positive pressure manifold (Biotage) and lyophilized to dryness. For Liquid Chromatography Mass Spectrometry (LC-MS) analysis, the samples were resuspended in 15 μl solution containing 3% acetonitrile and 3% formic acid. The final samples for injection contained the labelled peptides at a concentration of 100 fmol/μl for apoC-III, 15 fmol/μL for apoE and 23.3 fmol/μL for apoA4.

Parallel reaction monitoring (PRM) for absolute quantitation The PRM analyses were performed using a Q-Exactive mass spectrometer interfaced to an

Easy-nLC 1200 (Thermo Fisher, USA). The LC conditions for separation and MS parameters were optimized to enhance the detection of the fragment ions containing d3-leucine. Peptides were trapped on an Acclaim Pepmap 100 C18 trap column (100 μm x 2 cm, particle size 5 μm, Thermo Fisher, USA) and separated on an in-house packed analytical column (150 x 0.075 mm I.D) packed with 3 μm Reprosil-Pur C18-AQ particles (Dr. Maisch, Germany) using a linear gradient of buffers A and B, from 10% to 35% B over 12 min followed by an increase to 100% B for 8 min and a washing step at a flow of 300 nl/min. Solvent A was 0.2% formic acid in water and solvent B was 80% acetonitrile in 0.2% formic acid. Ions were injected into the mass spectrometer under a spray voltage of 1.8 kV in positive ion mode. For the PRM method, an orbitrap resolution of 35 000 (at m/z 200), a target automatic gain control value of 5e6, a maximum fill time of 80 ms, and a quadrupole isolation window of 2 Th were used. The precursor ions of the endogenous, the D3endogenous and labeled peptides were targeted at their 2+ charge state, triggered by a scheduled inclusion list. Fragmentation was performed with normalized collision energy (NCE) of 27. Raw mass spectrometry data (Thermo Fisher, USA) were exported to Skyline (version 3.x.x) for extraction of the fragment ions and peak area integration. The integration boundaries were manually verified for all peaks in the samples. A list of the precursor ions for the corresponding pairs of endogenous, d3-endogenous and labelled peptides and selected fragment ions is given in Table 1. All the ion fragments used for determine turnover and the ratio of endogenous and d3-endogenous contained d3-leucine. To determine the linearity of the peptides in the standard curve samples, a regression analysis was performed on the individual fragments from the peptides. The ISTD peptides were spiked in the individual plasma samples at concentrations close to 1:1 to those determined for their endogenous counterparts in the pool of samples analyzed in the standard curve. The concentration of the endogenous peptides in the individual study samples were calculated from the concentration of the labelled peptide (100, 15 and 23.3 fmol/μl for apoC-III, apoE and apoA4 respectively, quantification based on single point measurement concept) the Ratio-To-Standard (endogenous/labelled peptide) x Concentration labelled peptide was used. The concentrations measured for the endogenous peptides were converted into the actual concentration of the corresponding proteins in the initial blood samples.

Biochemical analysis Triglyceride and cholesterol concentrations in total plasma and lipoprotein fractions as well as in triglyceride-rich lipoprotein (TRL) cholesterol and remnant-lipoprotein (RLP) cholesterol were analyzed using assays (Denka Seiken, Tokyo, Japan) and the Konelab 60i analyzer (Thermo Fisher, USA). RLP-C captures lipoproteins not binding with anti-apoA-I and anti-apoB-100 as remnant lipoproteins [5]. The TRL-C quantifies cholesterol in both chylomicron remnants, and VLDL-IDL. Fasting and postprandial apoB48 levels in total plasma were measured by ELISA (Shibayagi, Shibukawa, Japan). Concentrations of plasma glucose and insulin were measured with hexokinase method (Roche Diagnostic Gluco-quant, Germany) and electrochemiluminescence (Roche sandwich immunoassay on a Cobas autoanalyzer), respectively.

Plasma levels of apoC-III were measured immunoturbidometrically (Kamiya Biomedical Company, Seattle, WA) and β-hydroxybutyrate concentrations were measured by an enzymatic method with β-hydroxybutyrate FS kit (DiaSys Diagnostic Systems, Holzheim, Germany) on a Konelab 60i analyzer.