UNIVERSITA' DI PADOVA FACOLTA' DI INGEGNERIA

Dipartimento di Ingegneria dell'Informazione

Scuola di Dottorato di Ricerca in Ingegneria dell'Informazione
Indirizzo: Bioingegneria

CICLO XXI

NONLINEAR MIXED-EFFECTS MODELING OF GLUCOSE-INSULIN METABOLISM

Direttore della Scuola: Ch.mo Prof. Matteo Bertocco

Supervisore: Ch.mo Prof. Claudio Cobelli

Dottorando: Paolo Denti

A Fabio

Frère, ne t'en va plus si loin.
D'un peu d'aide j'ai grand besoin,
Quoi qu'il m'advienne.
Je ne sais où va mon chemin,
Mais je marche mieux quand ta main
Serre la mienne.
(Alfred De Musset)

Riassunto della Tesi

Il diabete mellito è non solo una patologia molto seria, che causa disagi e sofferenze a milioni di persone nel mondo, ma, anche a causa dell'affermarsi di uno stile di vita sedentario e dell'invecchiamento della popolazione, negli ultimi decenni ha raggiunto proporzioni epidemiche, diventando una vera e propria emergenza sanitaria e sociale. Per fronteggiare questo problema, molte risorse sono state dedicate all'attività di ricerca scientifica, che ha permesso una più profonda conoscenza dell'eziologia del diabete. Tuttavia, il diabete è a tutt'oggi ancora inguaribile e molte questioni rimangono aperte, fra cui la completa comprensione dei fattori che causano e fanno progredire la malattia. Anni di ricerca hanno permesso di sviluppare molti sofisticati strumenti per studiare il sistema metabolico glucosio-insulina in vivo e poter così fronteggiare il problema dell'inaccessibilità diretta di alcuni dei fenomeni chiave che controllano la glicemia. Tali strumenti, fra cui protocolli di studio e approcci basati su modello usati per interpretare i dati sperimentali, si sono rivelati armi molto potenti nelle mani dei ricercatori, ma le proporzioni epidemiche della malattia e il parziale cambiamento delle strategie e obiettivi della ricerca hanno sollevato l'esigenza di poter disporre di metodologie meno invasive, più economiche, e quindi più adatte ad essere applicate ad estesi studi clinici. Alcuni strumenti matematici e statistici che sono collettivamente conosciuti con il nome di "approcci di popolazione" sono già stati sviluppati e vengono largamente impiegati in studi di farmacocinetica e farmacodinamica, per lo sviluppo di farmaci. Tali approcci si prefiggono come obiettivo primario di stimare la distribuzione dei parametri di un modello all'interno di una popolazione e pertanto si avvalgono, per la stima individuale, delle informazioni disponibili sull'intero gruppo di soggetti. Sono particolarmente adatti a situazioni in cui il campionamento intensivo in un singolo soggetto non è possibile, e quando l'interesse del ricercatore è focalizzato sulla variabilità inter-individuale. Tuttavia, nonostante le loro interessanti potenzialità, gli approcci di popolazione non sono ancora apprezzati all'interno dell'ambiente di ricerca sulle malattie metaboliche, e la loro applicazione in tali studi è stata molto limitata. Pertanto è necessaria dell'attività di ricerca per saggiare l'effettiva fattibilità e rilevanza dell'utilizzo di tali approcci nello studio del diabete. La ricerca qui presentata risponde a queste esigenze, proponendosi come obiettivo l'applicazione di queste sofisticate tecniche ai modelli di metabolismo del glucosio, prima testandone la fattibilità e adattandole al problema in esame, e poi impiegandole nell'analisi di dati raccolti in studi di popolazione.

Poiché in letteratura sono stati proposti molti diversi algoritmi, come primo passo, un dataset simulato è stato utilizzato per effettuare un confronto delle metodologie quando applicate al modello minimo del glucosio per il Test di Tolleranza IntraVenosa al Glucosio (IVGTT).

First-Order Conditional Estimation (FOCE) si è rivelato come l'algoritmo più soddisfacente, in quanto ha fornito i risultati più accurati e robusti in caso di scarsità o rumorosità dei campioni. Successivamente, per validare i risultati trovati su dati reali, l'analisi è stata ripetuta su un dataset più esteso, relativo a 204 soggetti sani testati con IVGTT. Per poter saggiare la bontà delle soluzioni fornite dai vari algoritmi, è stato impiegato un sistema di stima della likelihood function basato su campionamento Monte Carlo. Questa analisi, non solo ha permesso di confermare la scelta di FOCE come metodo preferenziale, ma si è anche rivelata come un potente strumento per valutare la precisione delle stime dei parametri di popolazione. Successivamente, è stato messo a punto e ottimizzato un modello di popolazione, conservando nella matrice di covarianza solo i termini di correlazione fra i parametri SI-P2 e SG-VOL. Questo modello è servito come base per la successiva integrazione di covariate nel modello. Al momento dell'esecuzione degli esperimenti, infatti, sono stati raccolti alcuni dati sui pazienti, fra cui altezza, peso, sesso, età, glicemia e insulinemia basali, informazioni sul grasso corporeo. È stata effettuata una analisi per determinare quali fra queste variabili potessero essere usate per spiegare parte della variabilità nei valori dei parametri del modello minimo fra i diversi soggetti. Il risultato è un modello che integra queste informazioni direttamente nelle sue equazioni, mentre i coefficienti di regressione per ognuno dei predittori diventano veri e propri parametri del modello e il loro valore viene ottimizzato insieme agli altri parametri di popolazione. L'analisi effettuata ha trovato come buoni predittori per SI e P2 l' età, l'insulinemia basale e il grasso addominale, che in ambo i parametri riescono a spiegare una buona fetta della variabilità inter-individuale. Sia l'impiego di metodologie di popolazione, sia l'introduzione delle covariate nel modello, permettono di aumentarne il potere predittivo, e sono in grado di usare informazioni indipendenti dai soli dati sperimentali. Questo permette di mettere a punto dei protocolli di studio meno invasivi, meno costosi, e pertanto più adatti ad un impiego su larga scala: ulteriore ricerca potrebbe avere come obiettivo l'ottimizzazione di una sampling schedule ridotta, che si avvantaggi dell'utilizzo degli approcci di popolazione. Ad ogni modo, il dataset utilizzato in questa analisi comprende solo soggetti sani, ed è quindi caratterizzato da una quantità limitata di variabilità di popolazione. Pertanto, sarebbe necessario ripetere l'analisi su altri dataset, per poter confermare questi risultati, in particolare sulle covariate.

Inoltre, in una sezione successiva, un metodo di popolazione è stato applicato anche ad un altro problema diverso, la stima del Disposition Index (DI) del glucosio. Questo è un indice calcolato combinando sensitività e responsività all'insulina, che serve per testare l'effettiva efficacia del sistema di controllo della glicemia. Ci sono due versioni proposte per la formula, una semplificata, che consiste semplicemente nel prodotto (da cui il nome di Legge Iperbolica), e una con un parametro aggiuntivo α ad esponente della sensitività all'insulina. Per poter calcolare il DI medio in una popolazione, e per poter saggiare quale delle due

formule sia effettivamente più adatta, in letteratura si trovano alcuni approcci basati su un fit geometrico. Tuttavia, alcune approssimazioni sono utilizzate per semplificare il fit, e sono molte le questioni metodologiche spesso sottovalutate. Pertanto viene presentato qui un nuovo metodo Total Least Squares (TLS) che affronta il problema senza l'impiego di approssimazioni. Grazie ad alcune simulazioni, si è effettuato un paragone fra i vari metodi disponibili, e il nuovo algoritmo è risultato migliore rispetto ai predecessori. Tuttavia, tutti gli algoritmi basati su fit si fondano sull'ipotesi che i soggetti appartenenti alla popolazione abbiano lo stesso valore di DI, e l'unica fonte di incertezza nei dati sia dovuta alla stima degli indici di secrezione e sensitività. Questa ipotesi sembra una forte semplificazione e, in effetti, l'analisi di un dataset reale sembra confermare la presenza di variabilità di popolazione nei valori del DI. Ulteriori simulazioni hanno confermato che tutti metodi basati su fit, TLS compreso, falliscono quando la variabilità di popolazione è presente. Pertanto, è stato ideato un altro metodo basato su approcci di popolazione e, in particolare, su NonLinear Mixed-Effects Models (NLMEM), che è in grado di separare la variabilità nei dati, poiché fondato su ipotesi meno restrittive. Tale algoritmo stima i parametri della distribuzione di probabilità congiunta degli indici di secrezione e sensitività, e poi estrae le informazioni sul DI dalla matrice di covarianza. NLMEM si è rivelato equivalente a TLS quando non c'è variabilità di popolazione, ma di gran lunga più affidabile quando le ipotesi per il fit geometrico non sono rispettate, pertanto si è deciso di utilizzarlo sul dataset reale per testare la validità della legge iperbolica. Anche se una validazione su altri dataset è auspicabile per validare i risultati qui presentati, il modello con il parametro aggiuntivo α sembra spiegare i dati in maniera più soddisfacente, e il valore del parametro α sembra dipendere dalla coppia di parametri usata per la definizione del DI, più che dalla popolazione in esame (anziani piuttosto che giovani). Inoltre, nello studio qui proposto, il punto di partenza sono stati i valori degli indici di secrezione già calcolati, insieme con la loro precisione, grazie ad un metodo tradizionale; un approccio ancora più potente consisterebbe nell'utilizzare un modello di popolazione per stimare contemporaneamente sia gli indici di secrezione che sensitività, sia i parametri della loro distribuzione di popolazione, da cui ricavare le informazioni sul DI.

Riassumendo, in questo lavoro si sono messi in luce i vantaggi dell'applicazione di approcci di popolazione nello studio nel diabete. Le potenzialità sono molte, dal miglioramento delle stime dei parametri individuali grazie all'uso dei prior di popolazione o di covariate e la relativa possibilità di mettere a punto protocolli di studio più leggeri, fino all'analisi di situazioni in cui la struttura gerarchica della variabilità è un aspetto cruciale.

Abstract

Diabetes mellitus is not only a very serious disease, causing discomfort and pain to millions across the world, but with the aging of the population and the prevalence of a sedentary lifestyle, it is assuming the proportion of a real epidemic, becoming a public health and social emergency. In answer to this call, research on diabetes has been intensely carried out in the past decades and the knowledge and understanding of its etiology have been significantly improved. However, investigation is still ongoing, many important questions are still unanswered, and the causes eliciting the pathogenesis and progression of the diseases are not yet fully understood. During all these years of research, sophisticated tools have been developed to study the glucose-insulin metabolic system in vivo, and cope with the inaccessibility to direct measurement of some of the key phenomena underlying the glycemic control. Such tools, as complex test protocols and model-based approaches used to interpret the experimental data, have proven powerful weapons of investigation, but, considering the epidemic proportion of the disease, there is great demand for approaches that are less invasive, less expensive and therefore more suitable for large clinical studies. Mathematical and statistical techniques that collectively go under the name of "population approaches" have already been developed and are largely employed for pharmacokinetics and pharmacodynamics studies in drug development. However, in spite of their interesting potential, they have not found significant application yet in the context of metabolism research. Thus, investigation is required to probe the feasibility and relevance of such approaches in the study of diabetes. The research presented here addresses these issues, and is aimed at applying these sophisticated techniques to the modeling of glucose metabolism, first assessing the applicability of these approaches and tailoring them to the problem under investigation, then employing them to the analysis of data from population studies.

First, a simulated but physiologically plausible dataset is created based on previous real data and employed as a benchmark to assess the applicability of population approaches to the Intra Venous Glucose Tolerance Test (IVGTT) minimal model of glucose disappearance.

Various population algorithms have been proposed in the literature, therefore a thorough comparison of the available methodologies is performed, and a sparse data situation is replicated to test the robustness of these methods in such cases. The results select the First-Order Conditional Estimation as method of choice and show its robustness to poor sampling. Then, a larger real dataset is employed and analyzed with the same techniques, this time assessing the quality of the results with a Monte Carlo sampling approach to profile the likelihood function. Then the population model is optimized, to provide a base model for the following covariate analysis. In fact, at the time of the experiments, demographic data about

the subjects has been collected, and the purpose of the covariate analysis is to determine whether some of these variables are significantly correlated with the model parameters and can be successfully used to explain part of the differences among the subjects. After a first exploratory regression analysis, different models are tested, integrating the most significant covariates directly as predictors into the model. In agreement with previous findings in literature, basal insulinemia, age and visceral abdominal fat are shown to be good predictors of insulin sensitivity and their introduction in the model is able to account for about a third of the between-subject variability of the values of this parameter. The use of covariates enhances the explanatory power of the model and opens the way for devising new lighter experimental protocols. One of the main benefits of the population approaches, in fact, consists in their ability to borrow information across the population and use it to improve the individual parameter estimates. As a result, the experimental protocols can be less demanding, both in invasiveness and economic cost, allowing in this way a broader use in large clinical studies. The results presented here, in fact, show that population approaches are very robust and able to cope with sparse data situations. In addition, the use of covariates in the model enhances even further the power of such techniques and makes them very appealing approaches to the study of glucose-insulin metabolism.

In addition, a population approach is proposed to solve the problem of the estimation of the Disposition Index (DI) of glucose tolerance in a population. Since both insulin sensitivity and beta-cell response must be taken into account to assess the actual efficiency of the glucose disposal system, the DI was proposed to condense the information conveyed by both these parameters in a single value. Traditionally, approaches based on a geometrical fit are used to determine the value of DI in a population of subjects characterized by the same degree of glucose tolerance. However, all these methods rely on the assumption that all the subjects in the population share exactly the same value of DI and are therefore not able to account for the population variability, which is inevitably inherent to biological data. In this work, a NonLinear Mixed-Effects Approach is proposed to analyze the distribution of the insulin sensitivity and beta-cell response indices across a population, and then obtaining the information on the DI from the population features thus estimates. Comparisons on simulated datasets between the newly proposed method and its competitors prove that a proper model of the variability structure is essential to avoid severe bias in the estimates.

Table of Contents

Chapter 1
Introduction and Outline of the Work

Population approaches are gathering growing interest as a valid alternative to traditional individual estimation methods, such as Weighted Least Squares (WLS), to biomedical model parameter estimation, especially in cases where a sparse sampling schedule or data with large variability make traditional paradigms inadequate to provide plausible estimates. In such cases, and if data are available in multiple subjects, population approaches are able to borrow information from the entire population and include it in the estimation procedure to improve individual parameter estimates. In addition, they are a very powerful research instrument to properly investigate and possibly identify the sources of variability in a population, making them a much more appealing approach than traditional statistical analysis of individual results.

The population approach has become very common in pharmacokinetics and pharmacodynamics studies in drug development, but it is not as of now so widely used in the context of metabolic models, such as those representing the glucose-insulin system, which have become a very precious tool for research about diabetes. In fact, the delicate metabolic system that ensures a stable glycemia is not only extremely complex, but the chemical pathways and tissues involved are hard to probe, and imply invasive experiments. In order to circumvent this lack of in vivo accessibility of significant variables, very sophisticated mathematical models for data analysis have been developed in the last decades, which have proved to the researchers very powerful instruments of investigation. Initially the experimental protocols, such as the euglycemic/hyperinsulinemic clamp, were very invasive, expensive and therefore unsuitable for epidemiological studies. Since the introduction and acceptance from the clinical community of protocols such as the intravenous and oral glucose tolerance test (IVGTT and OGTT), interpreted with the so-called minimal models, a much larger amount of data has become available, preparing the ground for the application of population approaches. The advantages are not only a deeper insight in the causes underlying population variability, but also the possibility of devising even lighter protocols.

In this investigation, we assess the actual advantages of the application of population approaches to study of the glucose-insulin system in many ways. In Chapter 3, we use

simulated data to test the performance of different population methods when applied to the minimal model of glucose disappearance, and the behavior in both serial and scarce sampling situations is explored. In Chapter 4 we apply the methodology to a real dataset, and propose a population framework for the glucose minimal model of disappearance. This model is used as a base for the covariate selection process that is described in Chapter 5. At the time of the experiment, in fact, some demographic information was collected about the subjects such as age, gender, demographic measures and body fat distribution. In our work, the variables resulting more significantly correlated with the parameters are integrated in the model and are used to explain part of the between-subject variability, improving in this way the predictive power of the model. Finally, in Chapter 6, we propose a new population-based approach to tackle the problem of the determination of the disposition index. This is an important tool which is used to assess the efficiency of the overall glucose-insulin metabolic system by considering both insulin secretion and insulin sensitivity.

Chapter 2
Background

Physiological Background

Glucose Homeostasis

Glucose is the primary source of metabolic energy for the majority of cells, and the ability to access supply sugar is essential for the survival of cells and therefore, the entire body. In particular, some cells, such as neurons and erythrocytes, are almost totally dependent on it. The brain requires a reasonably stable glycemia for its normal functioning, to the point that blood concentrations of glucose of less than 30 mg/dl or greater than 300 mg/dl can cause confusion, unconsciousness and convulsions, possibly having extreme consequences such as coma and death.

This is the reason why glucose, in various forms and compounds, is assumed with digestion, stored in the body, and constantly supplied, in the right quantity, to all the cells via the circulatory system. The glucose metabolic system is so complex and efficient that, in a healthy individual, glycemia is tightly regulated in the relatively narrow range between about 4 and 6 mM (mmol/l), 71 to 108 mg/dl. The normal blood glucose level is about 90mg/dl, so the normal total amount of glucose circulating in the blood of an adult male is, assuming an average blood volume of 5 liters, about 4.5 g. Many are the causes that, in everyday life, perturb this delicate equilibrium, the most common physical exercise and the ingestion of a meal. After a meal, the glucose tends to fluctuate to higher levels, due to the gastric and intestinal absorption of carbohydrates of low molecular weight present in the diet or derived from more complex molecules. When this happens, a very complex chain of events is initiated, that tends to restore the normoglycemia. First, high levels of glucose per se, tend to inhibit the glucose endogenous production, which takes place in the liver and in the kidneys and is caused by the processes of gluconeogenesis (generation of glucose from non-carbohydrate carbon substrates) and glycogenolysis (production of glucose by catabolism of glycogen). This phenomenon is called glucose effectiveness. At the same time, the high level of glucose concentration activates an appropriate insulin response in the endocrine pancreas; the magnitude of the β-cell response to different levels of hyperglycemia characterizes the β-cell responsivity. The secretion of insulin and the consequent rise of this hormone's level in blood enhance the effects already elicited by the hyperglycemia to suppress endogenous glucose production and, in addition, it increases the uptake and utilization of glucose in insulin dependent

tissues (namely liver, muscles and fat tissue cells). The actual promptness and efficiency with which the tissues respond to hyperinsulinemia is called as insulin sensitivity.

Although a much more complex series of events at cellular and biochemical level is involved, a simple pictogram with a conceptualized depiction of the glycemic regulation process is provided in Figure 2.1. If any of the underlying processes is not working correctly, this can lead to disease.

Diabetes Mellitus

Diabetes mellitus is the name given to a number of metabolic disorders having as the most evident effect chronic hyperglycemia. The etymology of the word, in fact, refers directly to one observable result of persistent high levels of blood sugars. The word diabetes comes from Greek and it's related to the concept of passing through (urine), whereas mellitus is Latin for sweet, so the word refers to polyuria (excessive urination) and glycosuria (presence of sugar in the urine), which were in antiquity the first and most easily detectable manifestations of the disease. Besides the abovementioned, other symptoms of diabetes include polydipsia (excessive thirst), unexplainable weight loss and blurred vision. On the long run, the effects of chronic hyperglycemia can be much more destructive: blindness, renal failure, increased risk of cardiovascular disease and peripheral neuropathy that can lead to ulceration and amputations. In western countries, the prevalence of diabetes mellitus has grown significantly in the last decades reaching a level of about 8%, a level that is expected to double in the next decades [2, 81], with severe consequences for public health and enormous costs related to its treatment.

Even though hyperglycemia is the principal factor that all forms of diabetes have in common, there are distinct causes underlying the phenomenon, so diabetes is divided into categories according to the different etiopathogeneses. The most common types of diabetes (1 and 2) are briefly exposed in the following paragraphs. Extensive literature is available on diabetes, so it goes beyond the purpose of this work to give an extensive review of one of the most broadly studied disease in current medical research. However, a schematic description is provided, with a focus on the details pertaining the mathematical modeling, which is the main topic of this dissertation.

Type 1 Diabetes

Type 1 (also insulin-dependent or juvenile) diabetes accounts for about 5 to 10% of all cases [2] and is an autoimmune disease that causes the loss of pancreatic β-cells. The lack of the set of these specialized cells, which are the only secretors of insulin,

causes the absence of this hormone in the blood and therefore the substantial inability of the subject to avoid hyperglycemia. As already explained, in fact, even though glucose shows per se a capability of inhibiting its own production, insulin cannot be substituted as the main regulator of glycemia. This is why the treatment for this kind of diabetes includes insulin injection to compensate for the lack of endogenous insulin, thus the denomination Insulin Dependent Diabetes. The other formerly used name, however, juvenile diabetes, is somehow misleading in the sense that, even though the destruction of the β-cells happens in very young age, this is not always the case and the illness is nowadays recognized to be caused by both genetic and environmental factors.

Even though there are ongoing studies aimed at finding a way to substitute the lost β-cells, as of now, the process is irreversible and the only treatment consists in a balanced diet and administrations of insulin to maintain the glycemia in the desired range.

Type 2 Diabetes

Type 2 (or Non Insulin-Dependent) diabetes accounts for about 90 to 95% of the cases [2]. Rather than by the total absence of insulin, this form of diabetes is caused by the relative insufficiency of the hormone, whose quantity in the plasma is not adequate to maintain euglycemia. This does not imply that insulin levels in type 2 diabetic patients are lower than normal; on the contrary, sometimes insulinemia is much higher than in healthy subject. The defect in the metabolic system is due to the low responsivity to insulin of the body cells. The following effects are present: insufficient inhibition of hepatic glucose production, low glucose uptake by muscle and adipose tissue, delayed and weaker response of β-cells to a hyperglycemic stimulus. This induced low responsivity to insulin of the body cells has been shown to be strongly correlated with obesity, sedentary lifestyle and age; however, ongoing studies are unearthing a genetic predisposition to type 2 diabetes. Since the described process is gradual and the onset of severe hyperglycemia is slow, many cases of type 2 diabetes go undetected, to the point that it is estimated that about one third of cases is undiagnosed [19]. Insulin administration is normally not necessary with this form of diabetes, rather changes in dietary habits and lifestyle are prescribed to type 2 diabetic patients and have been proven to slow down if not partly reverse the process of loss of insulin sensitivity.

Insulin Sensitivity

As mentioned in the previous paragraphs, the capacity of insulin to regularize the level of glucose in plasma goes under the name of insulin sensitivity. However, this denomination encompasses a broad spectrum of phenomena and involves various organs in the body and many different biochemical pathways. On the one hand, glucose production is inhibited in the liver, so that both the processes of gluconeogenesis and glycogenolysis are slowed down and the storage of glucose into glycogen is stimulated. On the other, insulin binds to the receptors of muscle and adipose cells and initiates a cascade of biochemical events that increases their glucose uptake. It is not yet clear which of these metabolic pathways are mostly hampered in type 2 diabetes, but research is ongoing in this regard and aimed at understanding the causes underlying insulin resistance. So far, many factors have been found correlated with insulin resistance [41] are aging [20, 27], obesity [9] (and in particular body fat distribution [54, 57]), lack of exercise [36, 58] and pregnancy [16].

β-Cell Function

Although the actual process of insulin secretion is the result of a complex interaction of various factors, the ability of the pancreas to secrete insulin in an efficient fashion is denominated β-Cell Function or Responsivity. Many aspects are important in assessing the effectiveness of insulin secretion, the more relevant being the actual promptness with which β-cells react to a hyperglycemic stimulus and the amount of insulin secreted. In addition, it should be mentioned that liver extracts form the blood a great amount of insulin (~50%) even before it can reach the peripheral tissues, so this is another factor which must be taken into account. To circumvent this problem, scientists normally analyze the concentration in blood of C-peptide, a substance which is secreted equimolarly with insulin by the β-Cell, but is not subject to hepatic extraction and is therefore a better indicator of the pre-hepatic level of insulin. The secretion profile in response to an acute stimulus, such as an intravenous infusion of glucose, has been shown to be biphasic, with a first phase, in the first 5-10 minutes, characterized by an bolus-like secretion, followed by a second phase, in which a growing amount of insulin is released in a more continuous and regular fashion, until the glucose stimulus is removed. In the more physiologic context of a meal, where the glucose absorption in the blood is slightly delayed and more diluted in time, the insulin response still shows a similar trend, although the profile of secretion appears much smoother.

In case of type 1 diabetes, the population of β-cells is decimated or totally killed by an autoimmune reaction, but this is not the only case in which the response of the β-

cell becomes impaired. Degrade in β-cell function has been imputed to many different factors and research is ongoing in this respect. According to the glucotoxicity hypothesis, it is hyperglycemia itself the cause of the desensitization of β-cells [29], whereas the lipotoxicity hypothesis points towards high levels of Free Fatty Acids in blood as factors causing β-cell apoptosis [51]. Other explanations have been proposed, rising from the observation, in diabetic patients, of deposit of amylin (co-secreted with insulin) that seems affects the number of β-cells in the Islets of Langerhans [39] or defects in the metabolic pathway leading to the synthesis of insulin [38, 80].

The Disposition Index

The estimation of insulin sensitivity and β-cell function is very important and necessary to assess the overall efficiency of the glucose-insulin metabolic system, but the separate interpretation of just one of these two indices provides only a partial picture of the performance of the glycemic control mechanisms. The two parameters can assume different combinations of values and still produce a very similar promptness in the glycemic normalization response. In particular, the versatility of β-cells is such that they are able to respond, within a certain limit, to the lowering in insulin sensitivity (due e.g. to pregnancy or obesity), and increase the quantity of insulin secreted, compensating in this way for the low efficacy of insulin itself. Therefore, a more comprehensive paradigm is needed to quantify the actual performance of the glucose controlling mechanisms in an individual. First proposed in 1981 by Bergman et al [12], and then reviewed and updated in 1993 by Kahn et al. [37], the Disposition Index (DI) is the most commonly used method to jointly account for insulin sensitivity and secretion. The original formulation, used by Bergman and colleagues (whose original chart is reported in Figure 2.2), defines the DI simply as the product of the two indices, in formulae:

$$\Delta = \xi \cdot \Phi \tag{2.1}$$

where Δ indicates the DI, ξ is an insulin sensitivity and Φ a β-cell responsivity index. Subjects with decreasing values of DI have a growing tendency towards impaired glucose tolerance and diabetes. The relationship has been baptized "Hyperbolic Law" by Kahn et al. [37], who also introduced an additional parameter α, accounting for the different leverage of the sensitivity and secretion indices in the calculation of the DI. The new formulation proposed is the following

$$\Delta = \xi^{\alpha} \cdot \Phi \tag{2.2}$$

Research is still underway as about which pair of indices used to quantify insulin sensitivity and secretion are more suitable for the calculation of DI and if the value of α is actually significantly different from 1, but the DI paradigm is widely used to estimate the actual level of glucose tolerance in a subject and identify the possible defects either on the sensitivity or on the secretion side. A graphical representation of such use of the DI, with typical values found for different categories of subjects, is reported in Figure 2.3.

Protocols and Methods for the Study of the Glucose Metabolic System

In the previous paragraphs, we described the most important factors involved in the glycemic control system: insulin sensitivity and β-cell function. However, an objective quantification of these variables is not directly available via a simple measurement, as it is for blood pressure or white blood cell count. Although some researchers propose the use of mere measurements of basal glycemia and insulinemia, or indices calculated based on their values such as [79] and QUICKI [43], as indicators of the efficiency of the metabolic systems, these approaches provide rather limited information and, even if good correlation with other indices has been reported [79] they were not always found to be totally reliable [23]. The intrinsically dynamic nature of variables such as β-cell response, in fact, requires a more complex experiment to be analyzed, and that is the reason that led to the design of the so-called "Clamp" experiments".

The Euglycemic/Hyperglycemic Clamp

First proposed in 1979 [28], these techniques employ a very invasive experimental setup, requiring continuous and adaptive infusions of both glucose and insulin, aimed at controlling (or clamping) glycemia at a target level, until the reaching of a steady state. In the euglycemic/hyperinsulinemic clamp, the level of glucose is maintained at physiologic levels 90 mg/dL, via a constant infusion of insulin (which stimulates glucose uptake and inhibits production) and a variable glucose infusion, tailored to guarantee the stability of the glycemic level. Blood sampling is very frequent and, after the steady state is reached, the flux of glucose injected is assumed to be equivalent to the tissues total uptake and is used as an estimate of insulin sensitivity. The hyperinsulinemic clamp, instead, provides also a measure of β-cell response. In this case, glycemia is clamped to an over basal level: first a glucose bolus is injected, and then a continuous glucose infusion, aimed at maintaining the desired glycemic level. Average insulin concentrations in the first 10 minutes and between 10-120 minutes are used to estimate respectively the 1^{st} and 2^{nd} phase secretion indices,

whereas insulin sensitivity is estimated similarly to the other clamp protocol. Clearly these protocols are very expensive, invasive, require prepared technicians and are therefore not suitable for epidemiological studies.

Model-based Approaches

Also the clamp methods, though, even if they bring the system at a different steady state from the basal value, they just use a measurement or a derived quantity, as the estimate of insulin sensitivity or β-cell response. A different approach, instead, is used by the model-based methods. These methods rely on mathematical models of the system under study, and try to use the accessible information to measure the inaccessible portion [21, 74]. In the case of the glucose-insulin system, the accessible portion is the plasma, where glucose, insulin and C-peptide can be easily measured, but the aim of the study is to estimate some inaccessible information: the secretion of insulin and its effect on the tissues. Therefore an experiment is carried out to perturb the system or introduce a tracer, blood samples are collected and statistical assumptions are made for the noisy data. Then the model is used to explain the experimental data collected, and by doing these, the model parameter values are adapted to the observed system response, providing in this way the information about the inaccessible pool. In the study of glucose metabolism, the most popular approaches are represented by the minimal models, which will be used in this work and are explained in the following paragraphs.

Intra-Venous Glucose Tolerance Test (IVGTT)

Before explaining the models themselves, it is useful to give some details on the data on which they are used, and the experiment that produces them.

The Intra-Venous Glucose Tolerance Test (IVGTT) [11] consists in an injection of a bolus of glucose (~300 mg/kg of body weight), and subsequent drawing of blood samples for the following 4 hours. A modification to the original protocol has been proposed with either a bolus of tolbutamide [82] or an infusion of insulin [33] at minute 20, so that, even in those subjects with insufficient insulin response, the estimation of insulin sensitivity is possible. The models used for the interpretation of the data are mainly two: the model of glucose disappearance [11] and the one for C-peptide kinetics [67].

Minimal Model of Glucose Disappearance

A schematic representation of the model is provided in Figure 2.4. Glucose is present in a single compartment, the plasma (although this is a simplification of a two-

compartment dynamics), and the fluxes in and out of this compartment represent the tissue uptake and the Net Hepatic Glucose Balance (NHGB), both controlled by insulin. To guarantee the identifiability of all the parameters in the model, however, a reparameterization is necessary that leads to the following equations:

$$\begin{aligned} \dot{Q}(t) &= -\left[S_G + X(t)\right]\cdot Q(t) + S_G \cdot Q_b \\ \dot{X}(t) &= -p_2 \cdot X(t) + p_2 \cdot S_I\left[I(t) - I_b\right] \\ G(t) &= Q(t)/V \end{aligned} \qquad \begin{aligned} Q(0) &= D + G_b \cdot V \\ X(0) &= 0 \end{aligned} \qquad (2.3)$$

where D (mg/kg) denotes the glucose dose per unit of body mass, Q(t) (mg/kg) is glucose mass in plasma, G(t) (mg/dL) is plasma glucose concentration, I(t) (pmol/L) is insulin concentration, Q_b, G_b and I_b are their basal values, and X(t) is insulin action (min^{-1}). The model has then four uniquely identifiable parameters: S_G (min^{-1}), glucose effectiveness, S_I (min^{-1} $pmol^{-1}$ L), insulin sensitivity, p_2 (min^{-1}), the insulin action parameter, and V (dL/kg), the apparent glucose distribution volume per unit of body mass. S_G is defined as the ability of glucose concentration *per se* to inhibit glucose production and promote its uptake, whereas S_I is the ability of insulin to enhance the same processes elicited by S_G. The pool of insulin actually performing the control action is not the one in the plasma, I(t), which is assumed as a known, error-free input (forcing) function, but rather the insulin in a remote compartment, which mimics the insulin concentration in the interstitial fluid. A limitation of the model consists in the fact that the parameter of insulin sensitivity does not distinguish between the effect of insulin on the liver or the tissues, but this simplification is necessary to ensure model identifiability. Another shortcoming of the model consists in the fact glucose measurements prior to about 8 minutes must be excluded from the fit. This is due to the fact that the 1-compartment kinetics cannot account for the quickest phase of glucose kinetics. This is another modeling compromise [18], necessary to guarantee the correct parameter identification, and it could represent a drawback if the scope of the investigation is, e.g., the effect of a drug on the quickest phase of glucose disposal.

C-Peptide minimal model

This model is not directly used in this work, so a thorough description is beyond the scope of this section. However, some of the parameters used in the definition of the Disposition Index are obtained with this model, so a brief overview is provided.

A schematic picture of the model is provided in Figure 4.5. The model assumes the glucose profile as an error-free forcing function and fits the C-peptide data. C-peptide is used instead of insulin, because they are secreted equimolarly, but the former does

not undergo hepatic extraction as the latter. The two-compartmental kinetics is modeled using standardized parameters as proposed by Van Cauter et al. [72], while the purpose of the model is the estimation of the secretion profile. The model is studied on the above-basal state, meaning that all the system is considered at equilibrium before the beginning of the IVGTT experiment, and the effect of the disruption of this equilibrium on C-peptide (insulin) secretion is studied. The equations of the kinetics of C-peptide are:

$$\begin{aligned} \dot{CP_1}(t) &= -(k_{01}+k_{21})\cdot CP_1(t) + k_{12}\cdot CP_2(t) + SR(t) \qquad & CP_1(0)=0 \\ \dot{CP_2}(t) &= +k_{21}\cdot CP_1(t) - k_{12}\cdot CP_2(t) & CP_2(0)=0 \end{aligned} \tag{2.4}$$

where k_{ij} are the kinetics parameter (fixed at standard values, in min^{-1}), $CP_1(t)$ and $CP_2(t)$ are the above-basal C-peptide picomolar concentrations respectively in the accessible and peripheral compartment and $SR(t)$ represents the secretion rate (pmol L^{-1} min^{-1}):

$$\dot{SR}(t) = m\cdot X(t) \tag{2.5}$$

$$\dot{X}(t) = -m\cdot X(t) + Y(t) \qquad X(0)=X_0 \tag{2.6}$$

This is the amount of above-basal secretion reaching the accessible compartment. $X(t)$ (pmol/L) represents the releasable C-peptide in the β-cells, whereas $Y(t)$ (pmol L^{-1} min^{-1}) is the provision of new C-peptide. The initial amount of releasable C-peptide X_0 is secreted immediately after the glucose bolus and is responsible for the first-phase secretion, while the second and slower phase is caused by the provision $Y(t)$. The provision is a process stimulated by the above-basal concentration of glucose in blood $\left(G(t)-G_b\right)$ in mmol/L and obeys the following equation

$$\dot{Y}(t) = -\alpha\cdot\left[Y(t) - \beta\cdot\left(G(t)-G_b\right)\right] \qquad Y(0)=0 \tag{2.7}$$

The provision will therefore tend with a time constant of $1/\alpha$ (min) to a steady-state value linearly related to the above-basal concentration by the parameter β (min^{-1}).

The first- and second-phase secretion indices are then calculated from the model parameters. The first-phase responsivity, Φ_1 (dimensionless), is calculated as the ratio between the readily released C-peptide and the total increment of blood glucose that elicited the secretion:

$$\Phi_1 = \frac{X_0}{\Delta G} \tag{2.8}$$

The second-phase index, Φ_2 (min^{-1}) is already in the model as the parameter β, which modulates the stimulus of glucose concentration on the provision of new C-peptide

$$\Phi_2 = \beta \tag{2.9}$$

Finally the two indices can be integrated into a unique parameter Φ_{tot}, summarizing the overall β-cell responsivity

$$\Phi_{tot} = \frac{\Phi_1 \cdot \Delta G}{\int \left[G(t) - G_b \right] \cdot dt} + \Phi_2 \tag{2.10}$$

Oral Glucose/Meal Tolerance Test

Even though they are not directly treated in the present investigation, other protocols for the study of diabetes are used. The Oral Glucose Tolerance Test (OGTT) and the Meal Tolerance Test (MTT) are among those. The first consists in a oral dose (~75 g) of glucose assumed in the first 5 minutes of the test, whereas the second in a standard meal (10 kcal/kg of body weight) containing 45% carbohydrates, 15% protein and 40% lipids, consumed in 10 to 15 minutes. Following the dose, blood samples are normally collected at 0, 30, 60, 90 and 120 minutes, even though alternative protocols have been proposed [15]. Many studies [48-50, 63], included some revised versions of the minimal models [15, 24], are available in the literature approaching, either in a model-based or a non-parametrical fashion, the estimation of insulin sensitivity and beta-cell response.

Methodological Background

Traditional individual ML approach

Before exploring the advantages offered by the population approaches, it is interesting to review the traditional parameter estimation paradigm, which is used for fitting individual data.

The parameter estimation process takes place once the experiment has been completed, the data collected and a model of the physiologic system under evaluation has been proposed. The aim of the procedure is finding the model output that better describes the collected data. This implies tuning the parameter values to generate a profile fitting the experimental observations.

The most common approach is Maximum Likelihood (ML) estimation. With this approach, an assumption is made on the measurement error structure, and the observed data is interpreted as the noisy measurement of the model predicted value, which is a deterministic quantity depending only on the parameter values.

In formulae,

$$y(t_j) = f(t_j, \mathbf{p}) + \varepsilon(t_j) \tag{2.11}$$

where $y(t_j)$ represents the data sampled at time t_j, $f(t_j,\mathbf{p})$ is the model prediction produced by the set of parameter $\mathbf{p}$ and $\varepsilon(t_j)$, is the random variable representing the measurement error.

The probability distribution for the measurement noise can be freely chosen, but normally it is assumed as uncorrelated, normally distributed with zero mean and variance possibly dependent on the model prediction or the measured data. For example

$$\varepsilon(t_j) \sim N\left(0, \sigma^2 f^2(t_j,\mathbf{p})\right) \tag{2.12}$$

where σ is a parameter to optimize.

Under the assumptions of measurement noise in (2.12), the probability distribution for the observed data is Gaussian and thus

$$l(\mathbf{y}|\mathbf{p},\Sigma) = \frac{1}{\sqrt{2\pi|\Sigma|}}\exp\left(-\frac{1}{2}\left(\mathbf{y}-\mathbf{f}(\mathbf{t},\mathbf{p})\right)\Sigma^{-1}\left(\mathbf{y}-\mathbf{f}(\mathbf{t},\mathbf{p})\right)^T\right) \tag{2.13}$$

In addition, being the errors on the different samples uncorrelated, the overall likelihood can be obtained as the product

$$L(\mathbf{y}|\mathbf{p},\sigma) = \prod_{j=1}^{n}\left(\frac{1}{\sqrt{2\pi\cdot\sigma^2 f^2(t_j,\mathbf{p})}}\exp\left(-\frac{\left(y_j - f(t_j,\mathbf{p})\right)^2}{2\sigma^2\cdot f^2(t_j,\mathbf{p})}\right)\right) \tag{2.14}$$

If the data are considered as given, the probability distribution which arises depends only on the parameter values. Thus, the set of optimal parameter values is sought, that maximizes the probability (likelihood) of observing the data. After some simplifications, justified by the fact that points of maximum are not affected by monotone transformations the problem can be shown to be equivalent to

$$\hat{\mathbf{p}} = \operatorname{argmin}\left(-2\cdot\ln L(\mathbf{y}|\mathbf{p},\sigma)\right) = \sum_{j=1}^{n}\left(\left(\frac{y_j - f(t_j,\mathbf{p})}{\sigma\cdot f(t_j,\mathbf{p})}\right)^2 + \ln\left(\sigma^2\cdot f^2(t_j,\mathbf{p})\right)\right) \tag{2.15}$$

So, in general, it is preferred to minimize the negative log-likelihood.

Population approaches

In epidemiological studies, when experimental data is available for a large number of subjects, the investigator is often interested not only in the individual estimates, but also on information about the population distribution of the parameters.

Standard Two-Stage (STS)

The traditional approach to this problem consists in performing a basic statistical analysis on the individual results, obtained separately for each individual with the traditional ML approach. This methodology is often referred to as Standard Two-Stage (STS).

The first stage of the analysis estimates the individual parameter values, and in the second step, sample mean and covariance are calculated, as described in the following formulae

$$\begin{aligned} \theta_{STS} &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{p}_i \\ \mathbf{\Omega}_{STS} &= \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{p}_i - \boldsymbol{\theta}_{STS}\right)\left(\mathbf{p}_i - \theta_{STS}\right)^T \end{aligned} \tag{2.16}$$

where *i* denotes the different subjects.

STS is computationally inexpensive, but provides values which are known to be upwardly biased [25] and it does not take into account, for example, the precision of the individual estimates. In addition, the information about the population distribution does not play any role in the individual estimation procedure, and in this way a great amount of useful information is disregarded. In some cases, when individual experimental data is sparse of very noisy, the parameter evaluation may fail for several individuals, or provide imprecise estimates. An alternative, called "naive pooling", consists in treating all data as coming from a single individual, but this methodology does not take into account properly (and therefore does not provide any estimate of) between-subject variability (BSV), and thus can be very misleading when used with models highly non-linear in the parameters and datasets characterized by large BSV.

Global (GTS) and Iterative Two-Stage (ITS)

More sophisticated methods have been designed to take into account the population characteristics and use them to improve the robustness and reliability of the individual results. A first examples are GTS and ITS, proposed by Steimer et al. [62].

These approaches both rely on the STS results as a starting point: the population mean (θ_{STS}) and covariance (Ω_{STS}), and the individual parameter values $\overline{\mathrm{p}}_\mathrm{i}$ and their uncertainty $\overline{\mathrm{V}}_\mathrm{i}^{-1}$. Unlike STS, however, the population information is used to refine the individual estimates, which are consequently used to update the population parameters themselves, and the process is reiterated until a convergence criterion is met.

GTS employs only linear formulas. At each step, the individual parameter estimates are updated using the population parameters from the previous step ($\theta_{(k)}$ and $\Omega_{(k)}$), according to

$$\hat{\mathbf{p}}_{i,(k+1)} = \left(\overline{\mathbf{V}}_i^{-1} + \Omega_{(k)}^{-1}\right)^{-1}\left(\overline{\mathbf{V}}_i^{-1}\overline{\mathbf{p}}_i + \Omega_{(k)}^{-1}\theta_{(k)}\right) \tag{2.17}$$

where *i* denotes the individuals, so $\hat{p}_{i,(k+1)}$ represents the updated individual estimates for the parameters of the i^{th} individual. Once the individual estimates have been updated, new population values are obtained as follows

$$\theta_{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\hat{p}_{i,(k+1)} \tag{2.18}$$

$$\Omega_{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\left(\left(\overline{V}_i^{-1} + \Omega_{(k)}^{-1}\right)^{-1} + \left(\hat{p}_{i,(k+1)} - \theta_{(k+1)}\right)\left(\hat{p}_{i,(k+1)} - \theta_{(k+1)}\right)^T\right) \tag{2.19}$$

This procedure is repeated until convergence. An easy interpretation of the formulae can be seen by manipulating (2.19) and obtaining

$$\overline{V}_i^{-1}\left(\hat{p}_{i,(k+1)} - \overline{p}_i\right) = -\Omega_{(k)}^{-1}\left(\hat{p}_{i,(k+1)} - \theta_{(k+1)}\right) \tag{2.20}$$

From this notation, it can be seen how GTS balances the individual estimation uncertainty of the STS estimates, with the deviation from the population typical values.

GTS uses only linear operations, so the computation is very fast. Its main drawback is that the ML individual estimates and their precision must be obtainable for each subject, which might not be the case if data are sparse.

ITS, instead, uses a Maximum A Posteriori (MAP) Bayesian estimation to update the individual parameters. The population mean and variance are used to define a Gaussian prior probability distribution for the parameters, and the probability of the parameters belonging to this distribution is maximized jointly with the likelihood arising from the adherence of the model prediction to experimental data, (e.g. (2.14)). At each step, the population prior assumes the form

$$h\left(\mathbf{p}_{i,(k+1)}\middle|\theta,\Omega\right) = \frac{1}{\sqrt{2\pi\left|\Omega_{(k)}\right|}}\exp\left(-\frac{1}{2}\left(\mathbf{p}_{i,(k+1)} - \theta_{(k)}\right)\Omega_{(k)}^{-1}\left(\mathbf{p}_{i,(k+1)} - \theta_{(k)}\right)^T\right) \tag{2.21}$$

and therefore the new likelihood function taking into account the population prior and the experimental data can be written as

$$L\left(\mathbf{y}\middle|\mathbf{p}_{i,(k+1)},\theta_{(k)},\Omega_{(k)},\sigma\right) = h\left(\mathbf{p}_{i,(k+1)}\middle|\theta_{(k)},\Omega_{(k)}\right)l\left(\mathbf{y}\middle|\mathbf{p}_{i,(k+1)},\sigma\right) \tag{2.22}$$

In the individual step, the population parameters are fixed, so these likelihood functions (one for each individual) are maximized only in the individual parameters,

obtaining their estimates $\hat{\mathbf{p}}_{i,(k+1)}$ and their precision $V_{i,(k+1)}$. These values are then used to update the population mean and covariance

$$\theta_{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\hat{p}_{i,(k+1)}$$
$$\Omega_{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\left(V_{i,(k+1)} + (\hat{p}_{i,(k+1)} - \theta_{(k+1)})(\hat{p}_{i,(k+1)} - \theta_{(k+1)})^{T}\right) \quad (2.23)$$

ITS suffers initially from the same weaknesses as GTS, but the subjects for whom the individual ML values are not available can be reintroduced into the analysis in a second step. On the other hand, it is much more computationally cumbersome than GTS, as the MAP step is nonlinear and requires therefore much more calculations than GTS's linear formulas do.

Nonlinear Mixed-Effects Models (NLMEM)

Another approach is represented by the NLMEMs. These methods make some assumptions on the structure of the variability in the data, from which arises a population likelihood function that is maximized in order to obtain the optimal population parameter values. In other words, NLMEMs are very versatile and allow the designer to define various structures of the Between-Subject and Residual Unknown Variability; an appropriate probability distribution can be assumed for each parameter and for the concentrations' remaining unexplained variability (a mixture of measurement error and other, unforeseen error sources).

The individual parameters are assumed to be dependant on some features that do not vary across the population (the so-called *fixed effects* denoted θ) and some individual-specific factors (the *random effects* denoted η_i). This can be written as:

$$\mathbf{p}_i = \mathbf{d}(\theta, \eta_i) \quad (2.24)$$

The random effects are assumed normally distributed, with zero mean and variance Ω, but the modeler is granted flexibility in the definition of the function $\mathbf{d}(\theta, \eta_i)$, which can also incorporate other factors such as physiological information about the subject (e.g. body height and weight). This makes the NLMEMs very appealing and allows easy incorporation of covariates in the model.

Very commonly the model is nonlinear in the parameters, and one step of the likelihood evaluation requires a marginalization (an integral operation) along the random effects, which is in general not analytically solvable and therefore computationally very intensive. More specifically, the contribution of the i^{th} individual to the likelihood function assumes the following form

$$L_i\left(\mathbf{y}_i|\theta,\Omega,\Sigma\right)=-2\log\left(p\left(\mathbf{y}_i|\theta,\Omega,\Sigma\right)\right)=-2\log\left(\int_{-\infty}^{+\infty}l\left(\mathbf{y}_i|\eta,\Sigma\right)h\left(\eta|\theta,\Omega\right)d\eta\right) \qquad (2.25)$$

where $l\left(\mathbf{y}_i|\eta,\Sigma\right)$ is the likelihood of the individual parameter given the data, and $h\left(\eta|\theta,\Omega\right)$ is the population probability distribution of the parameters. All the individual contributions to the likelihood, homologous to (2.25), are multiplied together to obtain the overall population likelihood, which is generally nonlinear also for the population parameters. Thus, a numerical maximization approach like Newton-Raphson is normally used, which requires many evaluations of the likelihood function and its gradient. For each of these evaluations, the solution of the multidimensional integral is necessary, and this must be repeated for each of the subjects. Several approximations to the integral have been proposed to render the calculation more tractable. The most widely used approximations were developed by Beal and Sheiner at UCSF [7], and were made available in their software NONMEM [8]; they are the First-Order (FO), the First-Order Conditional Estimation (FOCE), and the Laplace approximation (LAP). The first is based on a first-order Taylor series linearization of the individual model around the population mean (all $\eta_i=0$), whereas in the second the first-order approximation is performed separately in each subject around a properly individualized parameter estimate ($\eta_i=\hat{\eta}_i$). LAP also performs an approximation around each subject's individual parameter estimate, but this is built on second-order numerical derivatives. More details on the approximation can be found in Bauer et al [6]

After the optimal population parameter values have been obtained via ML, they are used as a prior to perform a MAP Bayesian estimation of the individual parameters. This is the so-called *post-hoc* step.

Among the advantages of NLMEMs with respect to the Two-Stage methods are the greater flexibility in designing the hierarchical population variability and the fact that it is not necessary for the individual estimates to be obtainable in the traditional sense for all subjects. On the other hand, Nonlinear Mixed-Effects approaches require long computational times and a certain expertise for the implementation.

In chapter 3, the performance of these methods when applied to the glucose minimal model is assessed.

Monte Carlo profiling of the likelihood function

Alternative approaches to the solution of the integral step contained in the population likelihood function are represented from sampling techniques. Instead of

approximating the integral, this can be evaluated by means of sampling algorithms, which extract a large number of random samples from the distribution and provide an estimate of the non-approximated likelihood value. The necessary number of samples and computational time are very large, and this methodology is then unsuitable to be used inside the minimization algorithms, where a large number of likelihood evaluations are necessary while exploring the function domain to find the minimum. However, these sampling techniques can be used to assess the goodness of the solution reached by the approximated methods. The objective function can be evaluated at the approximated optimal value and in its surroundings, testing both the accuracy of the approximation, and the local precision of the parameter estimates.
In Chapter 4 this technique is used to test assess the goodness of the solutions provided by the different methods and evaluate the precision of the estimates of the population parameters.

Integration of covariates in the model

As already mentioned, one of the main advantages of the NLMEM approach consists in the possibility to incorporate covariates into the model. These are physiological characteristics of each subject that can be used to explain part of the Between-Subject Variability (BSV) in the model parameter values. Once the statistically significant correlations have been recognized and the nature of relationship between the model parameters and the covariates has been assumed, the regression coefficient can be inserted in the model along with the other the fixed effects and their value will be optimized together with the other ML population parameters.
In Chapter 5, we operate a covariate selection for the glucose minimal model, and then propose a model integrating several physiological variables.

Figure 2.1 Conceptualized scheme of the glucose-insulin system, adapted from [22].

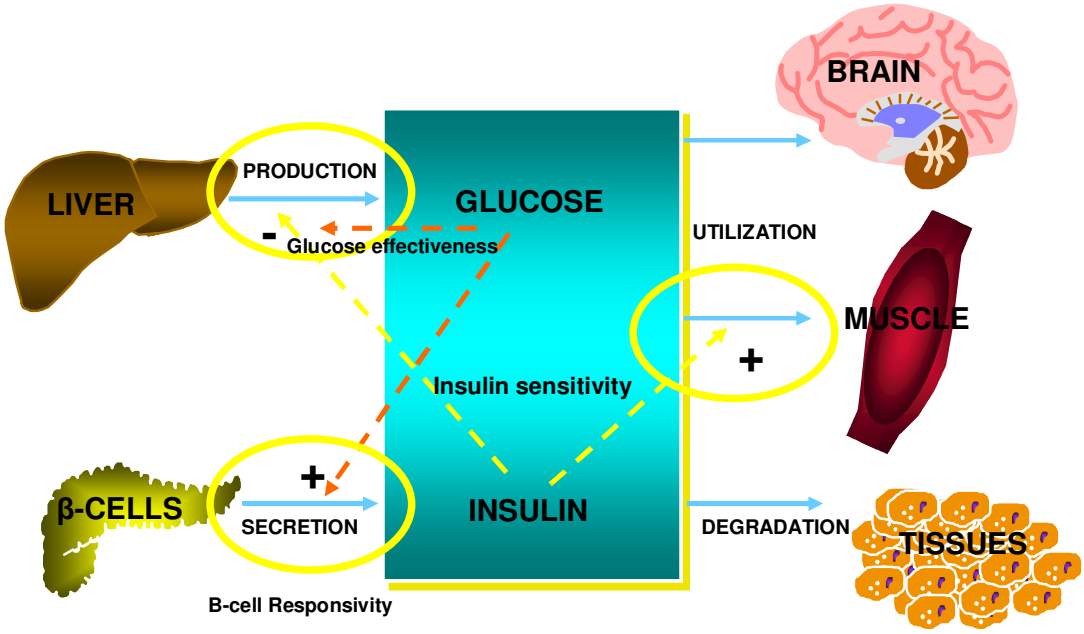


Figure 2.2 Original figure from [12], depicting the proposed hyperbolic relation between Insulin Sensitivity and Secretion.

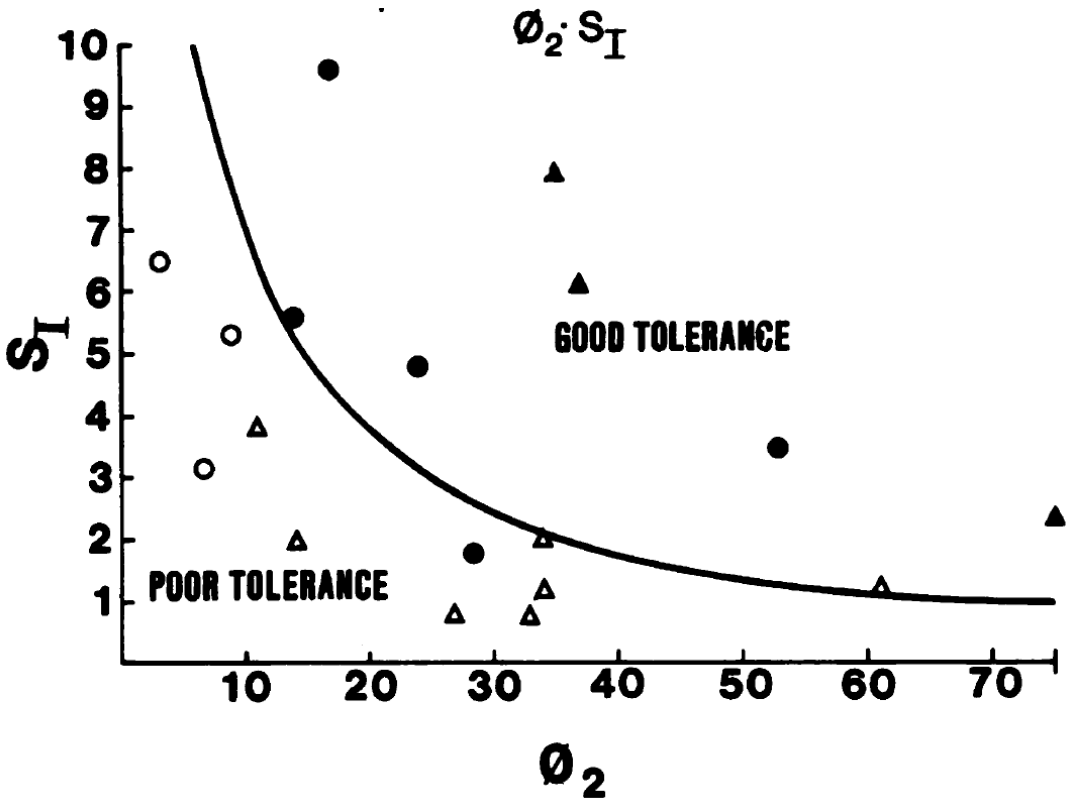


Figure 2.3 Example of the application of the DI paradigm, Figure from [40]. Different categories of subjects are ranked according to their DI value. The categories here represented are Relatives of Type 2 Diabetic Patients, women affected by Polycystic Ovarian Disease (PCO), subjects Impaired Glucose Tolerance (IGT), women formerly affected by Gestational Diabetes (GDM), Elderly Subjects and Type 2 Diabetic patients.

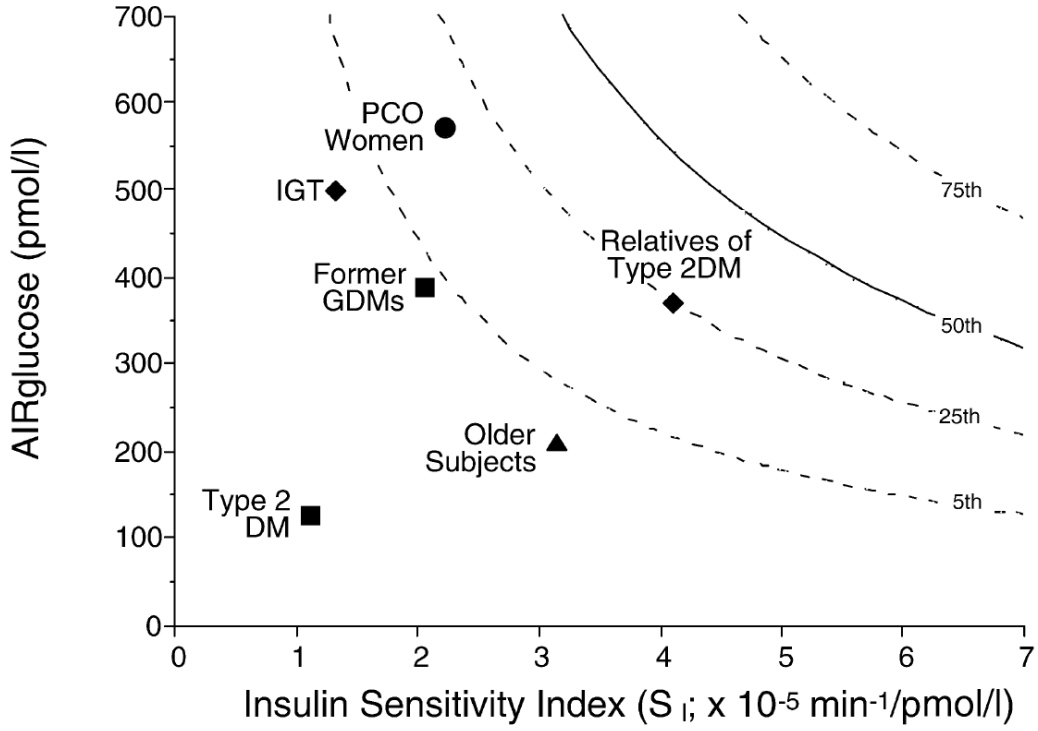


Figure 2.4 Representation of the glucose minimal model. This is the original set of parameters, which, once simplified, yields the identifiable parameterization described in the text. Figure from [18]. Insulin, from a remote compartment, influences glucose controlling both the Net Hepatic Glucose Balance (NHGB) and the tissue uptake.

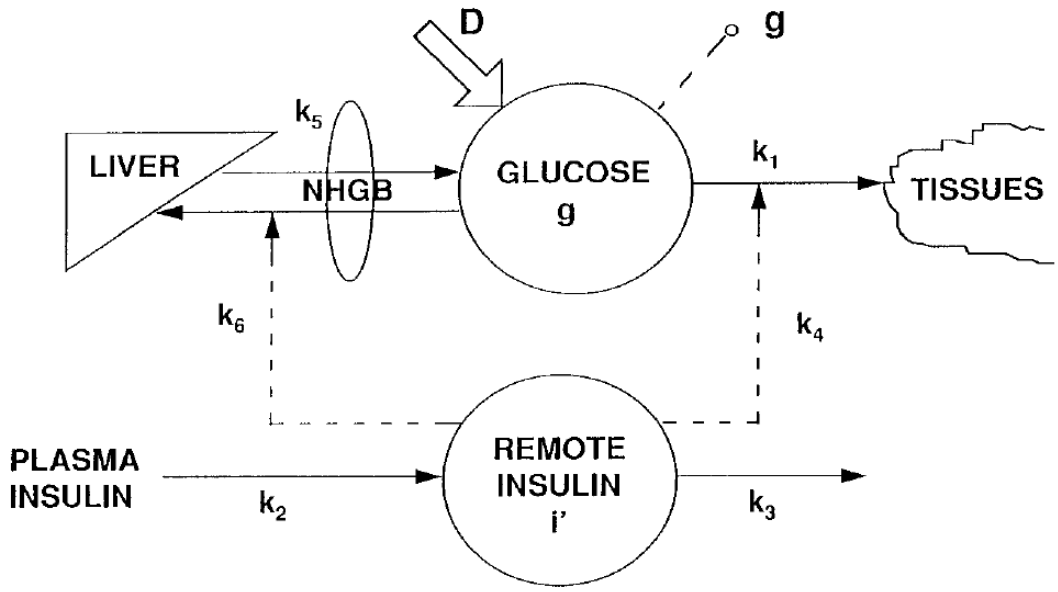
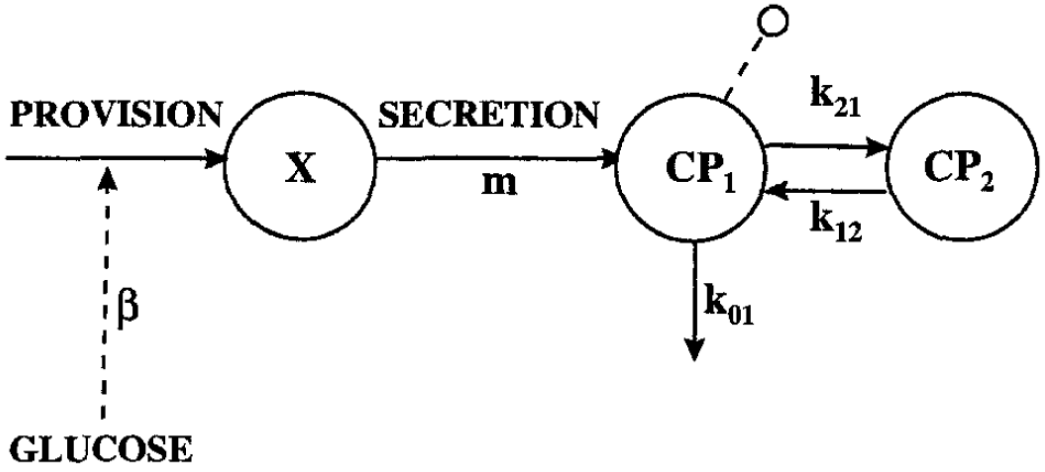


Figure 2.5 Schematic representation of the C-peptide minimal model, as represented in [68]. The section on the right corresponds to the C-peptide kinetics, whereas the section on the left represents the secretory component of the model.



Chapter 3

Population Approach to Improve IVGTT Glucose Minimal Model Parameter Estimation: a Simulation Study in Intensive and Sparse Sampling.

Overview

As mentioned in the introduction, IVGTT minimal model parameters are commonly estimated by Weighted Least Squares (WLS) on each single subject data. Sometimes, with sparse data, individual parameters cannot be satisfactorily obtained. This is what led us to probe the advantages of a population approach. However, several methodologies are available, and these approaches, even if very wide-spread in drug development studies, have not been applied in the context of metabolic models. Therefore, before starting our analysis, we decided to perform some tests, both to assess the reliability and advantages of the population approach vs. the traditional WLS, and to perform a comparison between the different methods, as to balance the benefits and drawbacks of each one, choosing the most suitable for the purposed of our analysis. The work described in this chapter is then a sort of a benchmark. We first performed the analysis in an intensive sampling situation, and then we reduced the number of samples in the dataset, with the purpose of testing the robustness of a population approach when dealing when sparse sampling. In order to perform an objective comparison (both among the different methods and among the results on the intensive and the sparse sampling situation) we needed to know the real values of the parameters, so that there was a gold standard to which the estimates could be compared. For this reason, we created and used a simulated dataset, but the set of parameters we used in the process was obtained from a real population, so that the data was fairly similar to real case-study. Our results reveal the advantages of a population approach with respect to the traditional individual WLS estimation paradigm even when the sampling scheme can be considered satisfactory. These benefits become more appreciable when the paucity of data becomes an issue: in these cases, WLS does not allow a proper estimation, whereas population paradigms make up for the scarcity of individual data by borrowing information from all the subjects. In particular, we found the NLMEMs to be more robust than ITS and GTS and less prone to the exclusion of very poorly sampled subjects, which affects TS. FO's approximation, however, seems to be too poor to cope with the nonlinearities of

the minimal model and with the large population variability which was characterizing some of the model's parameters. Both FOCE and LAP, instead, provided accurate results, but the latter proved computationally slower and required a tedious effort to tune the initial estimates, without really yielding much more accurate estimates. FOCE resulted therefore as the best choice. We then tested optimal setups of the population covariance matrix and concluded that neglecting the least significant off-diagonal terms, not only simplifies the population model, but also improves individual parameter estimation.

Introduction

Glucose effectiveness (S_G) and insulin sensitivity (S_I) are two important metabolic indices used in clinical and epidemiological studies of diabetes and hyperglycaemia. Estimates of S_G and S_I from an intravenous glucose tolerance test (IVGTT) are usually obtained by using the single-compartment minimal model method [11].

The traditional estimation paradigm is rooted in a Weighted Least Squares (WLS) approach applied on the experimental data of each single subject. Very often, in epidemiological studies, it is of interest to obtain also a description of the parameter distribution across the population. This information can then be either used as a Bayesian prior to facilitate the parameter estimation in further subjects, or for subsequent studies aimed at investigating the underlying biological reasons for parameter variability. Most frequently, this information is obtained by straightforward sample statistics (mean and covariance) on the individual parameter estimate set, following the Standard Two-Stage approach. This method does not account for individual parameter precision and is known to be prone to an upward bias in the estimation of population variance [25]. While this approach generates population statistics, the estimation process takes no advantage from the individuals' ensemble: no information is borrowed across individuals, and each subject is treated separately. The unintended consequences of this are that a potentially influential amount of information is discarded. The analysis of the parameter distribution in the population is carried out in a separate step, executed a posteriori, and the estimates of the population statistics play no role in the individual analysis.

Moreover, this traditional approach is feasible only in a data rich situation, i.e. a satisfactorily abundant sampling schedule for each subject. In case of more sparse or noisy data, the traditional individual approach may fail in some subjects, yielding unreliable or unrealistic (e.g. S_I virtually zero) estimates [55]. Bayesian approaches have been shown to be less prone to these kinds of estimation difficulties [45, 46, 56, 60], but they require independent (at least approximately) *a priori* statistical (i.e.,

mean, variance, covariance) knowledge on the model parameters, which is not always trivial to obtain.

Our purpose in this work is to evaluate the effective advantages of population approaches, which take into account the population structure when analyzing the data and simultaneously yield individual results and estimates of the population features. As already mentioned in Chapter 2, one family of these approaches is represented by iterative methods, such as ITS [6, 62, 73] and GTS [25, 62], which alternate recursively an individual and a population optimization phase. Both these Two-Stage (hereinafter TS) approaches need an individual estimate and its precision for each single subject to be available, but this might be difficult to achieve in case of sparse data. As we have seen in Chapter 2, another solution is provided by NonLinear Mixed-Effect Models (NLMEMs), which consist in maximizing a complex population likelihood. Due to the high computational price or even the unfeasibility of an exact solution, different approximation algorithms are used.

Population approaches to minimal model parameter estimation have been proposed before. To the best of our knowledge, the first report is by De Gaetano et al. [26], where NONMEM was used to fit a tolbutamide-modified IVGTT in 20 subjects. Vicini and Cobelli [73] used the ITS approach in 16 subjects studied with a standard IVGTT. Agbaje et al. [1] used a Bayesian hierarchical method on 65 insulin-modified IVGTTs. More recently Krudys et al. [44] tested both ITS and FOCE on 235 subjects given a tolbutamide-modified test. All these contributions focused on a single or few population methods applied to real data. Therefore, the true parameter values were unknown in all these studies and no unambiguous assessment could be done of the results provided by the population methods, except by comparing them to the traditional STS. This is a limitation, because STS results, even in a data rich situation, might sometimes carry a certain amount of uncertainty. The only other work to confront the problem with a simulated dataset is Erichsen et al. [31] which considers the performance of ITS, Bayesian hierarchical MCMC and FOCE on 40 simulated insulin-modified IVGTTs.

Our work develops on the concepts introduced by Erichsen et al., limiting our performance analysis to parametric methods. We believe this is a more unequivocal comparison, due to the fact that no hyper-priors are required in any case. In addition, we included other parametric methods such as GTS, FO and Laplace approximations; we explicitly addressed the assessment of the robustness of methods by simulating a data-poor environment; and we tested different setups for the population covariance matrix (full, diagonal and block matrix). All the algorithms have been run first on the full dataset to test their performance. Subsequently, the dataset was progressively

reduced (by 50% and 75%) and the analysis was repeated in order to asses the robustness of the estimates yielded by the different approaches.
For all our analyses we used the System for Population Kinetics [66]

Materials and methods

Synthetic Data

In order to ensure plausibility of data, we obtained our simulated dataset by using as a source data from a real IVGTT dataset. The dataset was previously described in [4, 5], a study in which insulin modified IVGTT (dose 330 mg/kg glucose at time 0, 0.02 units/kg of insulin at time 20) was performed on 204 healthy subjects, grouped in two main clusters according to their age. We chose to use the data from the group of young subjects, so 58 individuals (mean age 23±3 and mean BMI 24.5±2.9 kg/m^2) in the Clinical Research Center at the Mayo Clinic, Rochester, MN, USA. Blood samples were collected at -120, -30, -20, -10, 0, 2, 4, 6, 8, 10, 15, 20, 22, 25, 26, 28, 31, 35, 45, 60, 75, 90, 120, 180 and 240 min for measurement of glucose and insulin concentrations.
The process used to generate our simulated dataset was the following. First of all, the individual parameter values of each subject were estimated by means of the traditional WLS approach [3]. Once the individual values had been obtained, they were considered as true and used to generate new glucose concentration profiles using the individual insulin curves of the real dataset as forcing functions. Measurement noise equal to 2% or 5% of the simulated concentration value was added to the profiles. Hereinafter, these two simulated datasets will be referred to as Dataset A2 and Dataset A5. Because the simulated data are being fitted with the same model used to generate them, the simulations we are considering have no model error.

Undersampling of the dataset

In order to assess the robustness of estimates, the simulated datasets A2 and A5 were reduced by discarding glucose samples at random. In particular, approximately half of the samples were discarded from A2 and A5 (Datasets B2 and B5), then a quarter of the remaining samples was removed further (Datasets C2 and C5). The probability of each sample to be discarded was assumed to be 50%. Because removal was random, some individual data profiles were more affected than others. This choice leads to a situation common in practice in pharmacokinetics and pharmacodynamics studies: the information about some subjects might be very rich, while for others the sampling might be so unsatisfactory that individual estimates are impossible to obtain with the traditional estimation paradigm. The insulin concentration data are used in the model

as a forcing function and insulin was not undersampled. While this situation does not realistically reflect a practical experimental setting, where both glucose and insulin would be undersampled, here our purpose is the comparison of estimation methods and we focus our attention mainly on the effect of sparse glucose sampling. This is because glucose is the fitted dataset, while insulin is considered a forcing function in the minimal model. Moreover, because of the exogenous bolus of insulin injected during the IM-IVGTT protocol, the random discarding of insulin samples could yield very diverse outcomes (such as e.g. if in a subject the insulin infusion were totally missed by the sampling). Thus, it would be necessary to examine the individual results, as they would be subject-dependent. We decided to focus our analysis here on the method comparison. However, in order to gauge whether our results could be extended to a case where also the insulin samples are missing, we performed some test also with a Reduced Sampling Schedule (RSS) proposed by Steil et al. [61], discarding, in this case, also the insulin information.

Population Modeling Assumptions

We built the population model in the following way. Considering the ith individual, the model takes the general form:

$$\mathbf{y}_i = \mathbf{f}\left(\mathbf{d}\left(\theta, \eta_i\right)\right) + \boldsymbol{\varepsilon}_i \tag{3.1}$$

where $\mathbf{y}_i$ is the data profile of subject *i* and $\boldsymbol{\varepsilon}_i$ is the corresponding measurement error assumed to be normally distributed with zero mean and standard deviation described, in accordance with the standard minimal model identification approach

$$\varepsilon_{ij} \sim \mathrm{N}\left(0, \left(\sigma \cdot y_{ij}\right)^2\right) \tag{3.2}$$

with σ (proportional error variance) being an additional parameter to estimate. This stage of the population model deals with the variability in the data due to measurement and model error, in the population analysis jargon, this is referred to as Residual-Unknown-Variability (RUV). Parametric mixed-effect modeling requires to postulate at least some characteristics of the population probability distribution for the random effects (e.g. whether it is Gaussian or lognormal). We assume the random effects to be independent, with

$$\eta_i \sim N\left(0, \Omega\right) \tag{3.3}$$

with **Ω** being a positive definite covariance matrix. This second stage of variability, due to the differences among individuals, is called Between-Subject-Variability (BSV). In order to ensure the positivity and therefore the physiological plausibility of

parameter estimates, the parameters were described by the use of the following equation

$$\mathbf{p}_i = \exp(\theta + \eta_i) = \exp(\theta) \cdot \exp(\eta_i) \quad (3.4)$$

which implies a lognormal population distribution for the minimal model parameters and requires log-transformation of the original parameters. This assumption was also supported by the statistical analysis of the distribution of the individual parameters calculated with the traditional WLS methodology from the original dataset. In our dataset, potential outliers were detected among the WLS estimates. If these very few values (none for V, one for SG and five for SI parameters) are not considered, the Kolmogorov-Smirnoff test, run on the log-transformed parameters, detected departure from log normality only for p2 (p=0.016). In the simulations, they were intentionally left in the dataset to test the behavior of the population algorithms in presence of atypical individual values.

The standard interpretation of STS, ITS and GTS implies a Gaussian distribution for the parameters. Thus, in order to make the analysis coherent with our assumption of log-normality, we performed all the TS analyses on the natural logarithm of the parameters. This design choice yields results consistent with our assumption and allows us to compare the results produced by both TS and NLMEM.

Algorithms used

As a first step, we applied the traditional WLS estimation approach, referred to as STS. Then we used ITS and GTS, and finally, we applied the NLMEM approach testing the different approximations: First Order (FO), First Order Conditional Estimation (FOCE) and Laplace (LAP) In addition, for these last three algorithm implementations, SPK allows one to choose whether to neglect all or some of the off-diagonal terms of the parameter covariance matrix, assuming therefore no correlation between selected parameter pairs. This simplification reduces the number of parameters, at the expense of model flexibility. Thus, after comparing all the algorithms with a full covariance matrix setup, we moved on to test different population covariance matrix setups to assess the effective statistical significance of the parameters in the model. Details about the different algorithms employed can be found in Chapter 2.

Analysis of Results

In order to display and examine the *population* analysis results in a way that is compatible with previous literature and the original parameter units, we decided to transform the θ results for each parameter as follows:

$$\mu_{SG} = \exp(\theta_{SG}), \mu_{p2} = \exp(\theta_{p2}), \mu_{SI} = \exp(\theta_{SI}), \mu_V = \exp(\theta_V) \qquad (3.5)$$

and to calculate the following values

$$\delta_{SG} = \sqrt{\omega_{SG}}, \delta_{p2} = \sqrt{\omega_{p2}}, \delta_{SI} = \sqrt{\omega_{SI}}, \delta_V = \sqrt{\omega_V} \qquad (3.6)$$

where ω_p is the element on the diagonal of Ω corresponding to parameter p. These values, obtained with each of the methods, were then compared with their "true" value: respectively, the geometric sample mean and standard deviation of the true individual parameters. The choice to use these two parameters for the comparison is due to readability purposes: μ and δ correspond (and are therefore dimensionally similar) to the typical parameter values, and their CVs (since, conditional on a log-normal model for BSV, the elements of Ω relate to squared CVs, at least approximately). The percentages of discrepancy between the estimated and the true values are shown in Tables 3.2-3.4-3.6.

The goodness of the *individual* estimates was instead assessed by the square Root of the Mean Square Error (RMSE):

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(p_i - \hat{p}_i)^2}{N}} \qquad (3.7)$$

where p_i is the true parameter value for subject i, $\hat{p}_i$ its estimate, and N is the number of subjects in the analysis. This number was sometimes smaller than 58 due to the exclusion of some subjects (for example, when individual estimates would fail in ITS, GTS and LAP). For readability purposes, these values were indicated as percentages of the true population mean of each parameter.

In order to streamline the comparison across the different methods we defined a cumulative index ("discrepancy index of individual estimates", DIIE) to combine the results for each parameter together, and thus make them more readable. We decided to use a weighted mean which reflects the relative importance to the physiologist of each of the parameters: S_I was assigned a weight of 0.5, S_G 0.3, and p_2 and V 0.1. DIIE is used to compare in a more straightforward way the performance of the various algorithms and is reported in the last column of each table. Despite being heuristic, this proposed index allows us to summarize the results in a more effective way than reporting results for each single parameter.

Results

Full Dataset (A2 and A5)

The results obtained with dataset A2 and A5 are contained in Table 3.1. First of all, it should be mentioned that all the methods were successful on datasets A2 and A5. No subjects had to be excluded from the analysis and using the STS values as initial estimates for the other algorithms was almost always sufficient to get the desired procedure to run smoothly. The only exception was LAP, which required more care in the choice of the initial values, and for which the FOCE final values were used. Analyzing the estimates of the population typical values (i.e. μ, the exponential of the geometric mean of the fixed effects estimates), we can see that, in general, all methods provide results coherent with the ones that were used for the simulation. The only exceptions is FO, which overestimates μ_{SG} and μ_{p2} (respectively more than 7% and 16%) and underestimates S_I (about -36%). The results yielded by the other methods are all affected by a much smaller error, which is almost never over 5%. The parameters whose mean is estimated more precisely are V and S_I, whereas S_G and p_2 are affected by a slightly larger error. The population variability was estimated less precisely by the various algorithms in dataset A2. Almost all methods heavily underestimated δ_{SG} (by 20% or more) with the exception of FO, which surprisingly performs best with respect to this particular parameter, and STS, which produces an overestimate by 90%. δ_{p2} presents a very similar pattern, but less accentuated: all methods tend to underestimate its variability, with the exception of FO and STS, which respectively slightly and severely overestimate it. FO, on the other hand, is the only method to overestimate, and heavily (~38%), δ_{SI}, which is instead very well evaluated by all the other methods. This may be explained by the large BSV in S_I, as it is well known that the FO approximation performs poorly in the presence of large variation among subjects. The original parameter values from which the simulation was created belong to a reasonably "homogeneous" dataset of healthy young subjects; nonetheless, for some parameters like S_I and p2, remarkable population variability is present in the data. As a consequence, we suggest that the use of FO with the minimal model should be discouraged because unreliable even in presence of intensely sampled data. It is possible that the linearization carried out by this algorithm is too simple to cope with the complex non-linearities of the minimal model. These considerations are confirmed by the analysis of dataset A5, where STS and FO provide once again the worst performance. In particular, the expected amplification of population variability following higher noise in the data can be seen in the STS results. It is also interesting to note the increasing difficulty to estimate well the

minimal model fractional rate parameters, S_G and p_2. Because these are essentially exponential decay rates, one of glucose and the other of insulin action, it would be expected for them to be very sensitive to noise levels in the data. On the other hand, S_I, which is an "area" parameter, seems to be more robust to high noise, notwithstanding the poor performance of FO for this metabolic index.

From the analysis of the individual results, contained in Table 3.2, all methods demonstrate similar performances as far as individual estimates are concerned for dataset A2. This is not unexpected, since the noise level is low and the intensive sampling schedule obfuscates the effect of a misspecified population prior. All the methods estimate well V and S_I, whereas S_G and p_2 have a RMSE percent of more than 14%. This means that, in general, these two parameters are the ones whose estimates are on average more distant from the true values. It is worth mentioning that the parameters which are more poorly estimated in this analysis are the same whose estimates, obtained with the traditional WLS approach, are affected on average the largest estimation uncertainty (CVs about 18% and 22% for SG and p2 respectively, compared to about 3% for both SI and V). Therefore it is not surprising that also in this analysis they are the ones afflicted by the greatest error. Remarkably, anyway, the population approaches all performed slightly better than the mere traditional approach, represented by STS. The improvement is not substantial, but the gap due to the positive effect of the information borrowed across the population is expected to grow with the paucity of data. Dataset A5 shows that the impact of more than doubling the noise level is significant, and differences begin to widen. The TS methods (especially STS) appear more sensitive to high noise levels, and FOCE and LAP performances become measurably better than the other methods. As we mentioned, the poor performance of STS with richly sampled data casts fundamental doubts on STS as a gold standard procedure.

Undersampled Datasets (B-C)

From the analysis of the results obtained on dataset A2 and A5, it emerges that the benefit of population approaches can be already appreciated with rich sampling, although this may be due to a few very badly estimated subjects. However, the aim of this work is to assess the advantages of this technique in noisy, undersampled data. Let us then analyze the results from datasets B2 and B5 contained in Tables 3.3 and 3.4 respectively.

Running on dataset B2, the population methods perform much better than STS, both in estimating the population and the individual parameter values. The NLMEM methods, GTS and ITS yield more precise estimates of S_G and p_2's variability than

STS does. Based on dataset B5, the gap between STS and all the other methods broadens; also, FO performance is stable and comparable to other methods: as the number of individual samples decreases, a first-order linearization will be less dramatic than when substantial numbers of data are available for each individual. Also, at the individual level (Table 3.4), the STS estimates are much less accurate and, in this case, there also emerges a gap between NLMEMs and ITS or GTS. S_I is estimated with similar precision by NLMEM, ITS and GTS (with a RMSE of ~7%, whereas STS reaches 292%, even though this is partly due to several very poorly estimated subjects), but the values of S_G and p_2 produced by NLMEM appear more accurate than the ones obtained with any TS method (~20% and ~24% compared to ~28% and ~31% respectively).

If the paucity of data is further increased (25% of the original samples), the situation becomes very critical (Table 3.5 and 3.6). In the very data-poor framework depicted by datasets C2 and C5, some subjects were left with just 2 or 3 samples each, and this does not allow a proper individual estimation of both the parameter values and their precision with the traditional WLS paradigm. For this reason, some issues were encountered with STS and ITS, which were resolved by removing one subject, whereas a large amount of subjects had to be excluded from the GTS analysis. The reason for this exclusion was almost always the unavailability of the WLS precision estimates, which are necessary in the GTS algorithm. Looking at the results, it emerges that FOCE and LAP estimates prove definitely more robust and more precise than STS, both on a population and on an individual parameter level. The NLMEMs also prove more versatile than the TS methods, which encounter difficulties for the subjects whose sampling schedule is particularly poor. It is also interesting to have a visual representation of the behavior of STS and FOCE individual estimates when the data are very scarce. In Figure 3.1 and 2.3, the individual estimates for S_G and S_I are compared with the true values. Please note that for these data, another model for the covariance matrix was used, as explained below. The traditional approach (STS) produces a certain number of spurious outliers which are not present in the original dataset or in the solutions provided by the other methods. This happens because the discarding of random samples can drive the traditional WLS method to yield non-physiological values, which cannot be corrected unless additional information (such as from the population) is provided. On the other hand, when using a population method in a very data-poor situation, the population information plays a strong role and this can have the side-effect of shrinking the individual parameter values towards the population mean. This effect increases with growing data scarcity, so in general the population methods tend to absorb the outliers into the population main cluster.

This phenomenon is clearly shown in Figure 3.1-3.2: STS estimates tend to be scattered, whereas FOCE tends to keep the values clustered around the population mean, sometimes even excessively, like in the case of S_G (Figure 3.1). S_I seems instead to be reliably estimated also if the data are very scarce. Figure 3.3 offers an even clearer depiction of how the values of this parameter are biased towards the population mean as a consequence of data scarcity or poor quality. The displayed results refer to dataset C2 and C5, so a heavily undersampled situation, however it can be noticed that the regression of the values towards the mean is limited, suggesting that the population approach does not hinder the detection of outlier values for S_I. As already mentioned, the situation is different for S_G, with respect to which the model proves less sensitive. In the case of dataset C5, where the sampling schedule is very scarce and the level of noise in the measurements is high, the information in the data does not allow the proper estimation of S_G (as indicated by the negligible R^2 value). Therefore, if the only value of interest is S_I, the model can be simplified by neglecting the BSV for S_G, assuming that all the subjects share the same value of glucose effectiveness. The results are not reported here, but, with our data, this idea proved advantageous only with datasets C2 and C5, and providing only a moderate improvement in the overall %RMSE of S_I (~2% on 13% and 16% respectively). With the other datasets, the simplification did not show improvement in the S_I estimates, and with the full datasets A2 and A5 the detected effect on S_I was actually pejorative; p_2 is also poorly estimated in scarce sampling schedules situations, but this parameter is very strongly correlated with S_I and neglecting its BSV actually deteriorates its estimates.

Summarizing, we can say that the behavior of the different algorithms when dealing with decimated data confirms our previous findings: the NLMEMs prove more robust and are able to cope even with a situation of very scarce data, where the traditional approach fails and the TS methods are sometimes unable to run. It should be mentioned that, when the dataset is heavily under-sampled, it is often difficult to find proper initial values for methods such as Laplace and ITS, and the removal of several subjects from the analysis might be necessary, making these methods less appealing to the user.

Residual Unknown Variability

SPK additionally estimates the Residual Unknown Variability (RUV), i.e., the variability in the data which is not caused by differences among subjects. This is normally interpreted as the estimate of the measurement error, superimposed to model error and other sources of noise. In our case, we simulated data with a proportional

error structure (see Eq. (3.2)), and we introduced either a 2% or 5% CV. The parameter adjusting the RUV in the population model was not fixed in the computation, but the algorithms were free to optimize this value by estimating the scale parameter σ. It is interesting to inspect the results provided by the algorithms for the estimation of this parameter, shown in Table 3.7.

The size of the measurement error was on average evaluated quite well by FO, FOCE and LAP, whereas it is underestimated by the TS methods, by a growing extent as the data quantity and quality decreases. The method showing more explicitly this underestimation tendency is STS. An explanation could be the following: the estimation of the RUV is based on the analysis of the size of the model residuals. When the data points are very few, it is easy to accommodate the parameter values to fit the experimental data very well and, as STS does not use population information in the individual estimates, there are no constraints binding the parameters to remain close to a certain value. An effect of this phenomenon is a larger estimated BSV (greater values on the diagonal of Ω, see Table 3.1-3.3-3.5). Thus, there is a tradeoff between RUV and BSV. ITS and GTS show the same trend to a minor extent, whereas the NLMEMs seem able to cope well with the data-poor environment also from this point of view, and provide reliable estimates.

Optimal setup of the Ω matrix

So far, our analysis focused on the values of population mean and variance, but these are not the only population parameters. All the algorithms also produce as well estimates for the off-diagonal terms of $\boldsymbol{\Omega}$, which characterize the correlations between the parameters. For reasons of clarity, we do not go through an in-depth analysis of all these results, but some considerations of general validity can be made. The estimates of these second-order effects are in general very difficult to obtain; the various methods provide very dissimilar results and often completely in disagreement with the true values. This is particularly true for the terms representing a very weak correlation; in this case, the estimates may even have opposite sign with respect to the true value. This phenomenon becomes more and more relevant as the paucity of data increases. In fact, while the number of data points decreases, the number of parameters remains the same, and this leads inevitably to a degradation in the quality of the estimates; the covariance terms are the most prone to this phenomenon, whereas the mean is the most robust.

Therefore it might be a good choice, in a very data-poor framework, to reduce the number of parameters by neglecting these correlation terms. A possible option is then using a diagonal matrix, which excludes any correlation. This could be however a

strong simplification and lead to poor results; an intermediate choice would be keeping only the most relevant correlation terms. In our case significant correlations were detected between p_2 and S_I (~0.75) and between S_G and Volume (~-0.60), so a wise solution would be to include only these parameters in the optimization and neglect all the other off-diagonal elements. Therefore, we tested FOCE with different setups of the Ω matrix: full, diagonal, only with the P_2-S_I correlation term, or also the SG-Volume one. As far as the calculation of the population parameters is concerned (Table 3.8) no modeling strategy resulted as a clear winner. In general, it seems that excluding all the off-diagonal terms leads to an underestimation of the population variability, but further investigation would be needed to draw appropriate conclusions.

The individual parameter estimation results are included in Table 3.9. In this case the indices clearly show that neglecting some of the off-diagonal terms may actually help to simplify the model without deteriorating its performance. On the contrary, ignoring non-significant correlations seems to improve individual parameter estimation. In this case, when the only pairs of correlated parameters are S_I-p_2 and S_G-V, the individual parameters are estimated with a precision higher than that obtained with either the full BSV or only diagonal matrix. Also the model with only one off-diagonal term (S_I-p_2) performed well. This effect increases when the dataset is poorer, and therefore when the ratio between number of parameters and number of data points is very small and thus disadvantageous. Obviously, in a real case study the true values are not available, but a good decisional tool to test the level of significance of the covariance terms is the analysis of the objective function values provided by the algorithm. The results are not shown here for reasons of space, but the decision of dropping some correlation terms based on significance levels detected by the analysis of the objective function would have led to completely similar results, identifying the S_I-p_2 and S_G-V terms as the most relevant and worth introducing in the model.

Reduced Sampling Schedule (RSS)

In the analysis so far, we chose to undersample only the glucose profile while keeping the insulin information, so that we were able to probe not only the effect a sparse dataset, but also a situation in which several subjects were characterized by an extremely low number of samples. The same sample reduction for the insulin profile, even if followed by interpolation, in fact, would not in general be able to reconstruct the external insulin infusion which is part of the IM-IVGTT. However, in order to test the validity of our results in a real-case study, where both glucose and insulin samples would be missing, we decided to test also a 13-samples Reduced Sampling Schedule

similar to the one proposed by Steil et al. [61]. The samples included in this schedule were those taken at times 0-2-4-8-20-22-25-31-45-60-90-120-180, so a bit more than 50% of the samples were preserved, and the last sample is taken at time 180, thus shortening the experiment by 1 hour. The results obtained for these RSS datasets (both 2% and 5% CV) are contained in Table 3.10 and 3.11. For the sake of comparison, datasets B2 and B5 have a similar number of samples, but insulin was not discarded in that case. The use of a RSS, rather than random samples removal, has a positive effect on the traditional STS: both the population parameters and the RMSEs are more satisfactory, probably because the better undersampling operated by the RSS prevents some of the spurious outlier values caused by the extremely poor kinetic sampling in some subjects obtained with the random decimation.

The errors in the estimates yielded with STS, however, are still very large, in particular for several subjects. The same applies to GTS, which encounters estimation difficulties in some individuals. The advantages provided by the other population approaches, therefore, remain significant. In particular, the NLMEMs provide good population estimates and the size of the RMSEs is in line with the ones yielded by dataset B2 and B5. Overall, the undersampling of the insulin profile, if operated wisely, does not substantially deteriorate the parameter estimates obtained with the NLMEMs. This is encouraging with respect to the future design of experiments using a population method in conjunction with a RSS, resulting in less expensive, invasive and labor-intensive test.

Discussion and conclusions

We can draw several conclusions from the results collected in our analysis. Erichsen et al. [31], reported that, in a data rich situation no great advantage is introduced by the use of a population method. On the contrary, we detected a significant improvement both for the population and individual parameter estimates. Especially when the data become very sparse, some spurious outliers are identified by the traditional WLS approach (STS), whereas the hierarchical model structure of population methods prevents this. Moreover, as the experimental information decreases, the traditional approach shows its deficiencies even more, while the population methods prove very robust and are able to make up for the lack of individual information by borrowing it across the other subjects. S_I is estimated very accurately by the population methods, whereas S_G estimates are less precise, but nonetheless more reliable than the STS estimates.

Among the population methods, the most reliable and versatile are the NLMEMs, with a preference for FOCE. FO proved mostly inadequate for the MM, probably

because its approximation of the likelihood is too simple to cope with the complexity of the model and the amount of variability that characterizes the parameters. LAP, on the contrary, yields correct estimates, but not significantly better than FOCE, in comparison to which it is much more sensitive to starting values. The estimates we obtained with LAP were robust and the sensitivity to initial values purely implied a tedious process of tuning, however, we decided to choose FOCE as the reference method.

The TS population methods, ITS and GTS, performed well on the full and the 50% reduced datasets. GTS also has the advantage that all the formulas which are used are linear and this allows a very short computation time, much shorter than that of ITS or any of the NLMEMs. Anyway, these methods revealed unfeasible for some subjects in the dataset, in case of very sparsely sampled and noisy data. In addition, they failed with the heavily undersampled datasets (C2 and C5). This is probably due to the fact that they internally use the WLS individual estimates and their precision to start, which are obtained in a very unreliable way for some subjects, if only few samples are available.

Another effect which we experienced is that TS methods tend to overestimate the population variability, whereas NLMEMs tend to underestimate it. In addition, in a very data-poor situation the traditional estimation paradigm (STS) tends to produce outlying estimates. However, some subjects are truly different from the population mean: with population methods, the whole population, and therefore also these "true" outliers, become biased towards the population mean. The latter phenomenon occurs mostly with the parameters to which the model is less sensitive, in our case S_G, whereas very sensitive parameters, such as S_I, are normally less susceptible to this so-called "regression to the population mean" (shrinkage). This occurrence, well known in population modeling [42], was detected with the minimal model also by Erichsen et al. [31] and Krudys et al. [44]. This high sensitivity of the model with respect to S_I suggests that the application to different datasets, possibly including diabetic patients, characterized by a very low S_I and parameter estimation issues, would be feasible. Further research would be necessary to tune the proper setup for such datasets.

Another part of our analysis was aimed at investigating the optimal setup for the covariance random effect matrix $\boldsymbol{\Omega}$. We found out that, especially in data-poor situations, the BSV matrix should include only the most significant correlation terms; this excludes superfluous parameters from the optimization and improves the quality of the estimates of the remaining ones. In our analysis, we found out that the best choice is to include the correlation terms relative to S_I-p_2 and S_G-V, with the first pair being more tightly correlated than the second.

Further improvements can be obtained with the insertion into the model of physiologic information on the subjects, such as age, height, weight, body fat, etc. This information, obtainable non-invasively, can help in explaining some of the between-subject variability, if significantly correlated with the model parameters. A model that takes into account such clinical covariates is much more powerful, in that it not only uses information about the population, but also the individual characteristics of each subject. One of the main advantages of NLMEMs is that the regression coefficients for these physiologic variables are optimized by the algorithm together with the population parameters. Further studies will be therefore aimed at identifying the most significant covariates for the minimal model and incorporating them into the model.

In conclusion, our study confirmed that population approaches are a very appealing solution for estimating the minimal model parameters, mostly useful in a data-poor environment. In case of very under-sampled data for several subjects, TS methods are often unfeasible, whereas FOCE is able to cope even with this very critical situation. Our work also revealed the importance and advantages of properly shaping the random effect covariance matrix Ω, discarding the terms corresponding to very weak parameter correlations among subjects.

Acknowledgments – These results were partially presented at the IEEE/EMBS 2006 Annual Conference in New York City, NY, and published in the conference proceedings (Denti P, Bertoldo A, Vicini P, Cobelli C, "Identification of IVGTT minimal glucose model by nonlinear mixed-effects approaches", Conf Proc IEEE Eng Med Biol Soc. 2006;1:5049-52), and at the 6th IFAC Symposium on Modelling and Control in Biomedical Systems, 2006, Reims, France. A revised version of this chapter has been submitted to the IEEE transactions on biomedical engineering with the title "Nonlinear Mixed Effects to Improve Glucose Minimal Model Parameter Estimation: A Simulation Study in Intensive and Sparse Sampling".

Table 3.1 Comparison between the true population parameter values and the estimates provided by the different methods analyzing dataset A2 and A5 (full datasets). The distance of the estimated values for both the μ's and the δ's from the true values are reported as average percent differences, normalized to the true values.

dataset A2	$\Delta\mu_{SG}$ [%] ($\Delta\delta_{SG}$) [%]	$\Delta\mu_{SI}$ [%] ($\Delta\delta_{SI}$) [%]	$\Delta\mu_{p2}$ [%] ($\Delta\delta_{p2}$) [%]	$\Delta\mu_{V}$ [%] ($\Delta\delta_{V}$) [%]
STS	-7 (90)	3 (-1)	-2 (38)	0 (-1)
ITS	1 (-19)	-1 (5)	1 (-3)	0 (-3)
GTS	1 (-25)	-1 (5)	5 (-10)	0 (-2)
FO	7 (1)	-36 (38)	16 (8)	-1 (-1)
FOCE	1 (-26)	-1 (6)	2 (-5)	0 (-4)
LAP	1 (-20)	-1 (5)	2 (-6)	0 (-4)
dataset A5	$\Delta\mu_{SG}$ [%] ($\Delta\delta_{SG}$) [%]	$\Delta\mu_{SI}$ [%] ($\Delta\delta_{SI}$) [%]	$\Delta\mu_{p2}$ [%] ($\Delta\delta_{p2}$) [%]	$\Delta\mu_{V}$ [%] ($\Delta\delta_{V}$) [%]
STS	-26 (217)	10 (4)	-7 (-116)	0 (24)
ITS	-4 (25)	1 (4)	6 (-5)	1 (9)
GTS	-1 (11)	3 (-2)	7 (9)	1 (15)
FO	35 (-30)	-16 (31)	-6 (9)	-4 (-2)
FOCE	10 (-26)	-3 (9)	0 (-36)	0 (0)
LAP	5 (-12)	2 (8)	3 (-19)	0 (9)

Table 3.2 Square root of the mean square error (RMSE) of the individual parameter estimates obtained from the analysis of datasets A2 and A5, expressed as percentage of the true population mean. DIIE is an index which combines in a weighted average the precision on the four parameters (see text for detail). The last column contains the number of warnings indicating that one individual parameter value has hit the estimation method's upper or lower bounds; this happens when the individual random effect (η) assumes a value greater than 3 standard deviations.

dataset A2	S_G	S_I	p_2	V	DIIE	Warnings
STS	17.54	11.53	22.50	2.94	13.57	1/232
ITS	15.17	4.46	18.33	2.83	8.90	
GTS	15.81	4.28	19.19	2.79	9.08	
FO	14.45	4.00	18.65	2.74	8.48	1/232
FOCE	14.72	4.48	17.53	2.90	8.70	1/232
LAP	14.69	4.44	17.52	2.89	8.67	
dataset A5	S_G	S_I	p_2	V	DIIE	Warnings
STS	46.60	36.09	135.27	9.36	46.49	4/232
ITS	31.87	8.22	33.20	7.56	17.75	
GTS	31.59	9.57	37.48	7.89	18.80	
FO	24.63	9.74	23.95	6.61	15.32	
FOCE	20.84	7.93	25.40	6.12	13.37	
LAP	21.57	7.80	22.85	6.51	13.31	

Table 3.3 Comparison between the true population parameter values and the estimates provided by the different methods analyzing dataset B2 and B5 (50% decimated datasets). The distance of the estimated values for both the □'s and the □'s from the true values are reported as average percent differences, normalized to the true values.

dataset B2	$\Delta\mu_{SG}$ [%] ($\Delta\delta_{SG}$) [%]	$\Delta\mu_{SI}$ [%] ($\Delta\delta_{SI}$) [%]	$\Delta\mu_{p2}$ [%] ($\Delta\delta_{p2}$) [%]	$\Delta\mu_{V}$ [%] ($\Delta\delta_{V}$) [%]
STS	-8 (114)	17 (26)	-22 (134)	-1 (8)
ITS	0 (14)	2 (3)	-8 (16)	0 (1)
GTS	-0 (26)	3 (1)	-6 (17)	0 (5)
FO	19 (-6)	-19 (18)	5 (-17)	-2 (-4)
FOCE	10 (-13)	-1 (5)	-11 (3)	-1 (-5)
LAP	2 (-7)	2 (1)	-2 (4)	1 (-6)
dataset B5	$\Delta\mu_{SG}$ [%] ($\Delta\delta_{SG}$) [%]	$\Delta\mu_{SI}$ [%] ($\Delta\delta_{SI}$) [%]	$\Delta\mu_{p2}$ [%] ($\Delta\delta_{p2}$) [%]	$\Delta\mu_{V}$ [%] ($\Delta\delta_{V}$) [%]
STS	-14 (221)	36 (97)	-51 (234)	-1 (45)
ITS	5 (14)	3 (-2)	-11 (29)	1 (9)
GTS	16 (-4)	0 (9)	-16 (43)	0 (26)
FO	39 (-2)	-6 (14)	-12 (39)	-3 (-5)
FOCE	27 (-25)	-2 (5)	-23 (-7)	-2 (-15)
LAP	17 (-19)	4 (4)	-24 (-1)	0 (-11)

Table 3.4 Square root of the mean square error (RMSE) of the individual parameter estimates obtained from the analysis of datasets B2 and B5, expressed as percentage of the true population mean. DIIE is an index which combines in a weighted average the precision on the four parameters (see text for detail). The last column contains the number of warnings indicating that one individual parameter value has hit the estimation method's upper or lower bounds.

dataset B2	S_G	S_I	p_2	V	DIIE	Warnings
STS	36.65	292.04	46.93	8.08	162.52	4/232
ITS	28.16	7.36	30.55	6.72	15.85	
GTS	30.65	7.29	34.49	7.49	17.04	
FO	16.21	6.03	20.53	4.38	10.37	
FOCE	20.35	6.19	23.25	5.04	12.03	
LAP	21.16	6.40	23.79	5.50	12.48	1/232
dataset B5	S_G	S_I	p_2	V	DIIE	Warnings
STS	59.77	288.57	60.37	13.68	169.62	10/232
ITS	36.50	11.45	44.73	8.84	22.03	
GTS	39.47	15.70	47.00	10.21	25.41	
FO	33.86	13.07	36.00	8.42	21.14	
FOCE	33.97	12.02	40.29	8.40	21.07	
LAP	29.70	11.32	38.15	7.76	19.16	

Table 3.5 Comparison between the true population parameter values and the estimates provided by the different methods analyzing dataset C2 and C5 (75% decimated datasets). The distance of the estimated values for both the □'s and the □'s from the true values are reported as average percent differences, normalized to the true values. The symbols * and ** indicate the exclusion of respectively a negligible or consistent number of subjects (for details, please refer to Table 6).

dataset C2	$\Delta\mu_{SG}$ [%] ($\Delta\delta_{SG}$) [%]	$\Delta\mu_{SI}$ [%] ($\Delta\delta_{SI}$) [%]	$\Delta\mu_{p2}$ [%] ($\Delta\delta_{p2}$) [%]	$\Delta\mu_{V}$ [%] ($\Delta\delta_{V}$) [%]
STS*	-36 (300)	2 (115)	5 (215)	-3 (128)
ITS	-19 (55)	8 (-19)	22 (27)	3 (21)
GTS**	-21 (57)	3 (-6)	42 (78)	2 (-8)
FO	31 (-51)	-10 (3)	-9 (-10)	-3 (-40)
FOCE	16 (-17)	0 (-9)	-9 (14)	-2 (-31)
LAP	1 (-1)	4 (-15)	8 (26)	0 (-17)
dataset C5	$\Delta\mu_{SG}$ [%] ($\Delta\delta_{SG}$) [%]	$\Delta\mu_{SI}$ [%] ($\Delta\delta_{SI}$) [%]	$\Delta\mu_{p2}$ [%] ($\Delta\delta_{p2}$) [%]	$\Delta\mu_{V}$ [%] ($\Delta\delta_{V}$) [%]
STS	-40 (402)	29 (204)	-37 (335)	-8 (437)
ITS*	-2 (39)	7 (-13)	-4 (40)	1 (13)
GTS**	7 (23)	-8 (19)	-8 (80)	1 (17)
FO	46 (1)	-3 (-5)	-24 (55)	-4 (-24)
FOCE	35 (-11)	0 (-6)	-32 (6)	-4 (-37)
LAP	23 (29)	9 (-1)	-24 (23)	-3 (34)

Table 3.6 Square root of the mean square error (RMSE) of the individual parameter estimates obtained from the analysis of datasets B2 and B5, expressed as percentage of the true population mean. DIIE is an index which combines in a weighted average the precision on the four parameters (see text for detail). The column "Warnings" contains the number of warnings indicating that one individual parameter value has hit the estimation method's upper or lower bounds, whereas the last column indicates how many subjects were removed from the analysis.

dataset C2	S_G	S_I	p_2	V	DIIE	Warnings	Removed
STS	53.76	247.61	1182.57	15.67	259.76	12/228	1/58
ITS	38.52	12.95	69.48	10.84	26.07		
GTS	42.72	18.80	254.63	8.00	48.48	4/120	28/58
FO	31.61	13.42	29.22	9.98	20.11		
FOCE	34.94	13.10	38.20	10.65	21.92		
LAP	29.88	11.85	41.81	9.09	19.98		
dataset C5	S_G	S_I	p_2	V	DIIE	Warnings	Removed
STS	84.54	1257.53	1696.39	60.94	829.86	23/232	
ITS	41.57	19.35	49.03	10.78	28.12		1/58
GTS	49.21	25.76	107.97	13.67	39.81	6/144	22/58
FO	50.15	20.34	39.90	13.29	30.53		
FOCE	47.58	16.29	45.78	12.79	28.28		
LAP	44.76	17.64	41.93	12.41	27.68		

Table 3.7 Measurement %CVs estimated by the various methods on the simulated datasets. The symbols * and ** indicate the exclusion of respectively a negligible or a consistent number of subjects (for details, please refer to Table 6).

Estimated CV	dataset A2	dataset B2	dataset C2
True	2.0%	2.0%	2.0%
STS	1.7%	1.4%	* 0.9%
ITS	1.7%	1.4%	1.0%
GTS	1.7%	1.3%	** 0.8%
FO	2.1%	2.2%	2.4%
FOCE	2.0%	2.0%	2.1%
LAP	2.0%	1.9%	2.0%
Estimated CV	dataset A5	dataset B5	dataset C5
True	5.0%	5.0%	5.0%
STS	4.6%	3.9%	2.3%
ITS	4.7%	4.1%	* 2.9%
GTS	4.5%	3.7%	** 2.1%
FO	5.4%	5.3%	5.2%
FOCE	5.3%	5.3%	5.0%
LAP	5.3%	5.2%	5.1%

Table 3.8 Population parameter estimates provided by FOCE with datasets C2 and C5 and different setups of the **Ω** matrix. As in the previous tables, the percentage of discrepancy for μ and δ (in brackets) is displayed.

dataset C2	$\Delta\mu_{SG}$ [%] ($\Delta\delta_{SG}$) [%]	$\Delta\mu_{SI}$ [%] ($\Delta\delta_{SI}$) [%]	$\Delta\mu_{p2}$ [%] ($\Delta\delta_{p2}$) [%]	$\Delta\mu_{V}$ [%] ($\Delta\delta_{V}$) [%]
FOCE, Full Matrix	16 (-17)	0 (-9)	-9 (14)	-2 (-31)
FOCE, Corr. P2-SI and SG-VOL	12 (-54)	-1 (-8)	-3 (-23)	-1 (-25)
FOCE, Corr. P2-SI	10 (-76)	-1 (-7)	-1 (-15)	0 (-32)
FOCE, Diagonal Matrix	11 (-59)	-1 (-4)	1 (-17)	0 (-34)
dataset C5	$\Delta\mu_{SG}$ [%] ($\Delta\delta_{SG}$) [%]	$\Delta\mu_{SI}$ [%] ($\Delta\delta_{SI}$) [%]	$\Delta\mu_{p2}$ [%] ($\Delta\delta_{p2}$) [%]	$\Delta\mu_{V}$ [%] ($\Delta\delta_{V}$) [%]
FOCE, Full Matrix	35 (-11)	0 (-6)	-32 (6)	-4 (-37)
FOCE, Corr. P2-SI and SG-VOL	32 (-30)	-1 (-8)	-25 (-33)	-3 (-30)
FOCE, Corr. P2-SI	33 (-49)	-1 (-8)	-26 (-27)	-3 (-40)
FOCE, Diagonal Matrix	36 (-44)	-1 (2)	-27 (-12)	-4 (-41)

Table 3.9 DIIE values (calculated by combining the square root of the mean square errors on the four parameters) obtained with the different setups for the Ω matrix (rows) and the different datasets (columns). * indicates that few (1 or 2) individual parameters hit the estimation upper or lower bounds.

Ω matrix definitions	dataset A2	dataset B2	dataset C2
FOCE, Full Matrix	* 8.70	12.03	21.92
FOCE, Corr. P2-SI and SG-VOL	* 8.07	* 10.40	15.87
FOCE, Corr. P2-SI	* 8.34	* 10.49	16.43
FOCE, Diagonal Matrix	* 8.86	* 12.16	17.22
Ω matrix definitions	**dataset A5**	**dataset B5**	**dataset C5**
FOCE, Full Matrix	13.37	21.07	28.28
FOCE, Corr. P2-SI and SG-VOL	13.66	17.87	25.81
FOCE, Corr. P2-SI	13.82	18.90	24.79
FOCE, Diagonal Matrix	14.97	20.00	27.76

Table 3.10 Comparison between the true population parameter values and the estimates provided by the different methods analyzing dataset RSS2 and RSS5. The distance of the estimated values for both the μ's and the δ's from the true values are reported as average percent differences, normalized to the true values. * is a reminder of the exclusion of one of subject (for details, please refer to Table 11).

dataset RSS2	$\Delta\mu_{SG}$ [%] ($\Delta\delta_{SG}$) [%]	$\Delta\mu_{SI}$ [%] ($\Delta\delta_{SI}$) [%]	$\Delta\mu_{p2}$ [%] ($\Delta\delta_{p2}$) [%]	$\Delta\mu_{V}$ [%] ($\Delta\delta_{V}$) [%]
STS	-7 (56)	1 (2)	-4 (76)	0 (6)
ITS	-1 (-11)	-5 (3)	-3 (0)	0 (1)
GTS*	-1 (-15)	-5 (6)	9 (-6)	20 (4)
FO	9 (-15)	-35 (35)	17 (-5)	-1 (-4)
FOCE	2 (-23)	-6 (5)	5 (-7)	0 (-2)
LAP	2 (-22)	-6 (5)	3 (-5)	0 (-2)
dataset RSS5	$\Delta\mu_{SG}$ [%] ($\Delta\delta_{SG}$) [%]	$\Delta\mu_{SI}$ [%] ($\Delta\delta_{SI}$) [%]	$\Delta\mu_{p2}$ [%] ($\Delta\delta_{p2}$) [%]	$\Delta\mu_{V}$ [%] ($\Delta\delta_{V}$) [%]
STS	-31 (274)	4 (4)	8 (85)	0 (32)
ITS	-9 (44)	0 (-5)	14 (-10)	2 (10)
GTS	1 (11)	0 (-6)	14 (-2)	1 (18)
FO	31 (-36)	-12 (22)	-2 (-31)	-2 (-18)
FOCE	8 (-21)	-6 (1)	5 (-54)	0 (-24)
LAP	7 (-10)	-5 (10)	3 (-21)	-1 (5)

Table 3.11 Square root of the mean square error (RMSE) of the individual parameter estimates obtained from the analysis of the datasets RSS2 and RSS5, expressed as percentage of the true population mean. DIIE is an index which combines in a weighted average the precision on the four parameters (see text for details). The column "Warnings" contains the number of warnings indicating that one individual parameter value has hit the estimation method's upper or lower bounds, whereas the last column indicates how many subjects were removed from the analysis.

dataset RSS2	S_G	S_I	p_2	V	DIIE	Warnings	Removed
STS	25.37	49.70	28.64	5.13	35.84	1/232	
ITS	20.99	8.81	23.13	4.51	13.47		
GTS	21.16	49.63	23.34	4.36	33.93	1/228	1/58
FOF	18.25	8.63	20.19	4.40	12.25		
FOCEF	19.00	8.89	20.08	4.32	12.59		
LAPF	19.14	8.82	20.13	4.34	12.60		
dataset RSS5	S_G	S_I	p_2	V	DIIE	Warnings	Removed
STS	50.14	34.90	274.09	11.66	61.06	2/232	
ITS	34.62	9.39	40.71	9.73	20.13		
GTS	31.63	13.09	50.94	10.02	22.13		
FOF	27.14	13.28	25.61	9.17	18.26		
FOCEF	26.48	11.26	30.76	9.06	17.56		
LAPF	21.40	10.00	26.62	7.98	14.88		

Figure 3.1 Scatterplot of the S_G individual estimates provided by FOCE (black dots) and STS (white dots) based on dataset C5 vs. the true values. The correlation terms S_I-p_2 and S_G-V were used. The dashed line is the unity slope.

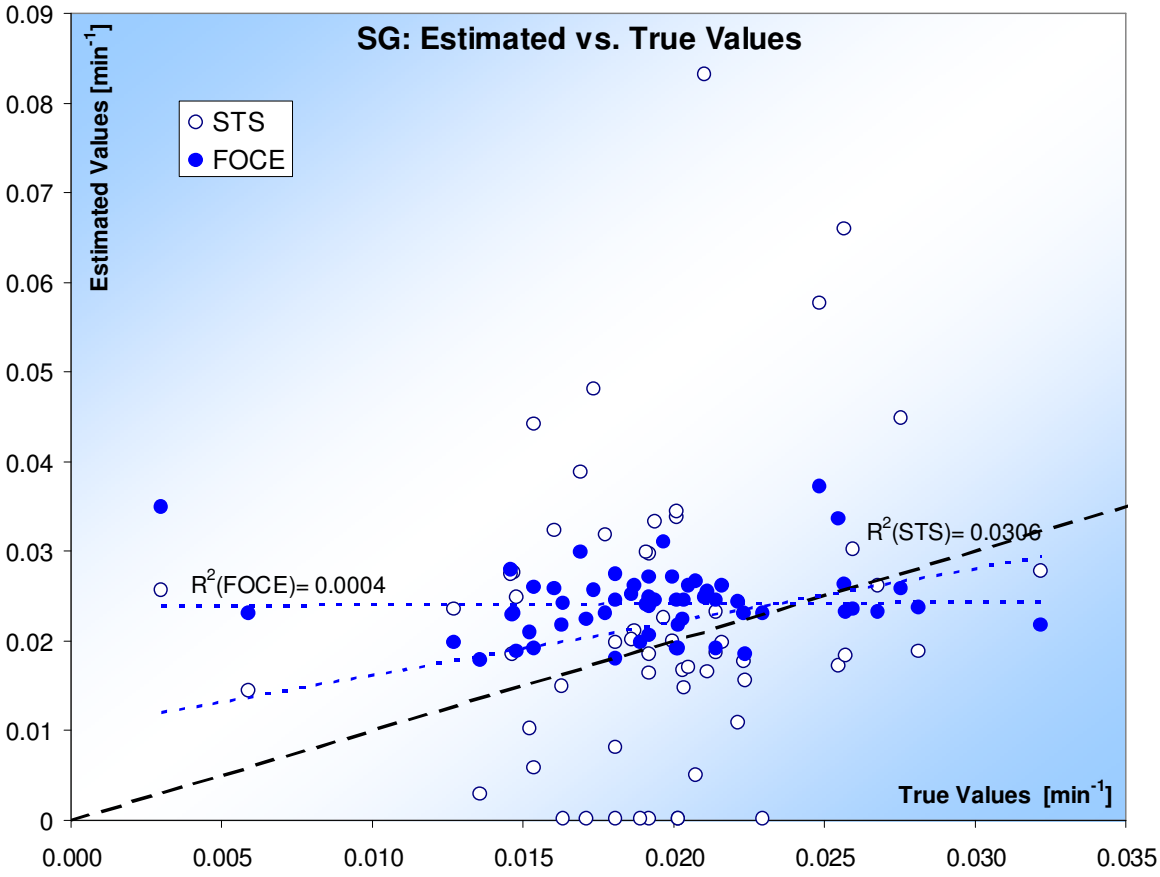


Figure 3.2 Scatterplot of the SI individual estimates provided by FOCE (black dots) and STS (white dots) based on dataset C5 vs. the true values. The correlation terms S_I-p_2 and S_G-V were used. The dashed line is the unity slope. For the sake of clarity, the big panel shows an enlarged portion of the chart, excluding some outliers yielded by STS. The small panel shows the entire dataset.

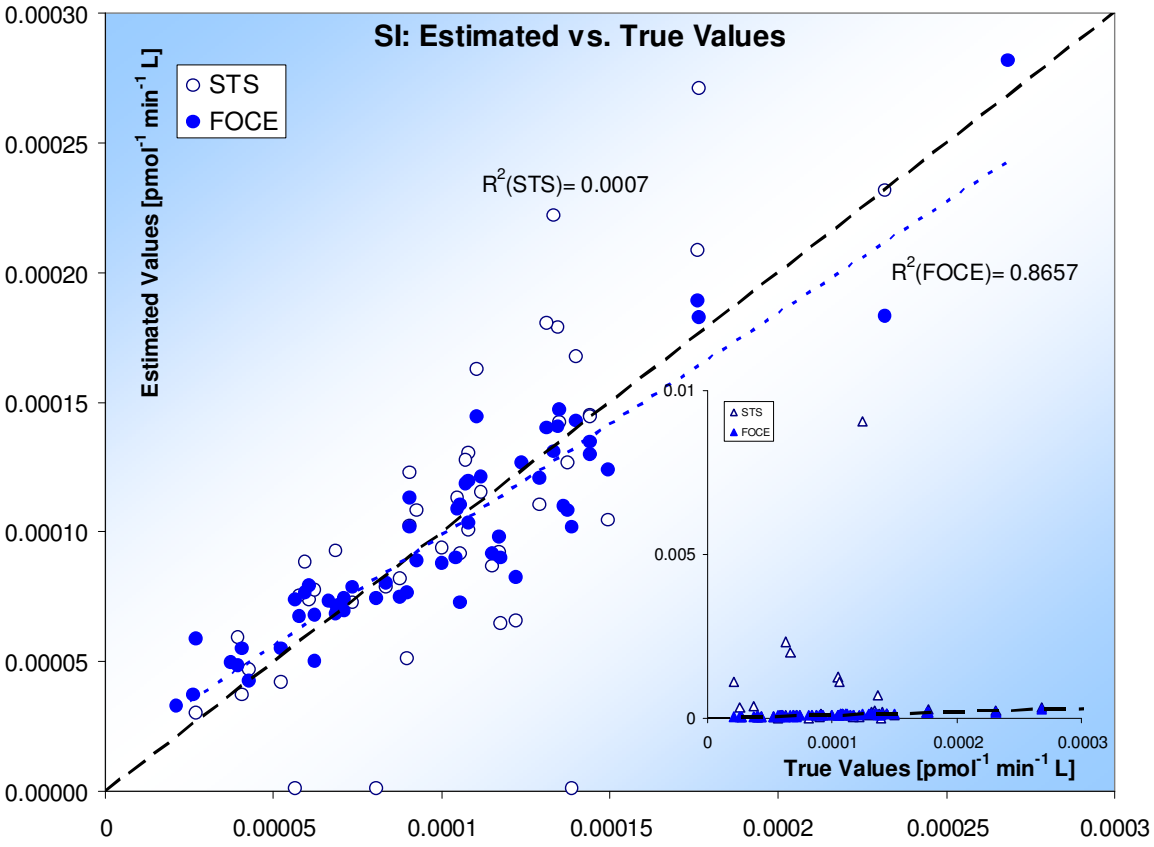
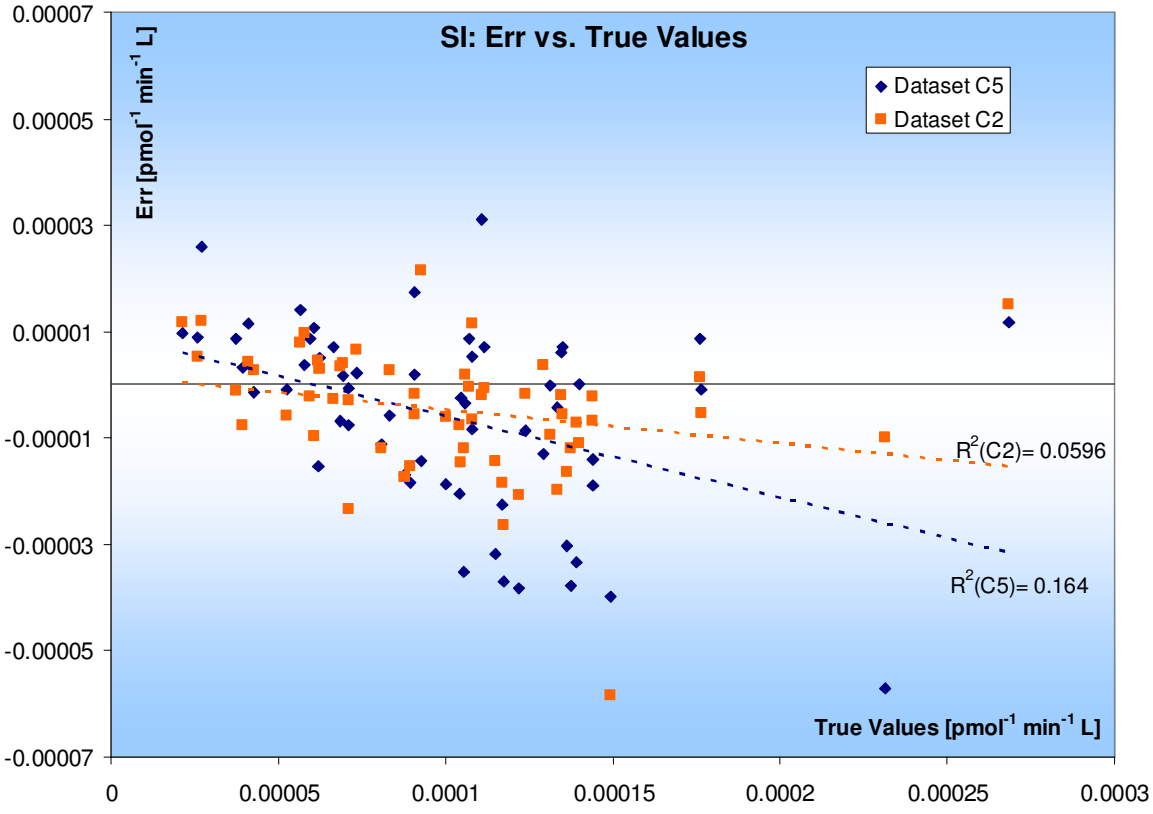


Figure 3.3 Scatterplot of the estimation error vs. the true values of SI. The values were obtained with FOCE based on dataset C2 (square-shaped) and C5 (diamond-shaped). Only the correlation terms SI-p2 and S_G-V were used. The estimation error is defined as the difference between the estimated and the true value. Regression lines and the R^2 coefficient are also displayed.



Chapter 4

Population Approach to Improve IVGTT Glucose Minimal Model Parameter Estimation: a Real-Case Study with Likelihood Function Profiling via Monte Carlo Sampling

Overview

The purpose of the work exposed in this chapter is the design of an appropriate population model of the glucose minimal model, serving as basis for the covariate analysis which is the subject of the following chapter. In order to select the suitable parameterization, we rely on the value of the objective function that can be employed for inference on the statistical significance of the parameters and the parameterization, and we also use some profiling techniques, which are useful to assess the precision of the estimates and the accuracy of the employed population method. The profiling of the true likelihood offers therefore an independent validation of the results obtained in the previous chapter, which were obtained from simulated data. In this case, as the data are real, there is no "true value" to which the different results can be compared, that is why we recur to a true likelihood profiling implemented via Monte Carlo sampling. Moreover, in this case, the analysis will be applied to the whole dataset, consisting of 204 subjects and encompassing both young and elderly individuals and characterized therefore by a broader range of variability. So, in a first section, we apply both the Two-Stage methods (STS, ITS and GTS) and the NLMEMs (FO, FOCE and LAP) to the new dataset and evaluate the quality of the results using the likelihood profiling. As a first step, as the implementation of the TS methods only allows a full covariance matrix, we performed a first evaluation to compare the accuracy of the solutions provided by the different methods. Confirming our finding from the previous chapter, FOCE appears to offer the best approximation, possibly even more reliable than LAP. FO dramatically fails in estimating even the typical values of the parameters presenting the largest BSV, namely SI and P2, whereas the TS approaches seem to provide underestimates of the RUV. Once we have confirmed that FOCE is the method of choice, we apply it to tune the appropriate model for our dataset. We tried different setups for the model: an approach using an error weighting scheme proportional to the model prediction rather

than the data, normal versus lognormal distribution for the parameters. Then we inspected the optimal setup for the population covariance matrix.

Introduction

Population kinetic modeling approaches are attracting growing interest in many fields of biomedicine thanks to their value in estimating population features from sparsely sampled data. However, their application often entails approximations of the original model function, whose effect is difficult to gauge. One way of assessing the quality of the provided estimates, is to calculate the value of their likelihood. However, some methods, like the TS approaches, are based on an iterative algorithm and not a likelihood minimization, so they do not provide such value. On the other hand, NLMEMs use an approximated version of the likelihood function and, as the approximation is different from method to method, the value provided by these algorithms is not an objective tool for a comparison among methods. Therefore, to perform an objective comparison, the value of the non-approximated likelihood was calculated. Since this implies the solution of an analytically intractable integral, so we have recurred to a MC sampling technique. We then focused on model refinement, selecting the most suitable parameterization, error model structure and setup for the population covariance matrix.

Materials and Methods

Dataset

A cohort of 204 healthy subjects was tested with insulin-modified IVGTT (IM-IVGTT) with full sampling schedule (240 min, 21 samples) at Mayo Clinic, Rochester, MN. More details on the experimental design and the dataset are available in the literature [4, 5]. The individuals had very wide ranges of age (mean 56 yrs, range 18-87) and BMI (27 kg/m^2, range 20-35) and were clustered into two main groups, one (59 individuals) encompassing young subjects (average age 23), and the other (145 individuals) elderly individuals (average age 69). At the time of the IVGTT experiment, also some relevant demographic information were collected about the subjects, and these will be the used in the next chapter for the covariate analysis.

Population statistical assumptions

To perform a population analysis, it is necessary to define a hierarchical model as well as establish key assumptions on the parameter probability distribution and the statistical structure of the measurement error. In fact, as explained in Chapter 2, when

encompassed in a population framework, each subject's individual parameters p_i are assumed as extracted from a population probability distribution which is defined by some hyper-parameters called fixed effects (that are therefore shared by all subjects). Each subject is then characterized by an individual-specific deviation from the population typical values. These features, which depend on each subject's distinctive attributes, are called random effects.

For the first step of this analysis, we chose to assume a lognormal distribution for all parameters, guaranteeing in this way the positivity of the individual values. We also chose to model the random effects with a full population covariance matrix, including all the possible correlations between the minimal model parameters. In formulae:

$$\left.\begin{array}{l}\mathbf{p}_i = \exp\left(\boldsymbol{\theta} + \boldsymbol{\eta}_i\right) \\ \boldsymbol{\eta}_i \sim N\left(0, \boldsymbol{\Omega}\right)\end{array}\right\} \rightarrow \mathbf{p}_i \sim LN\left(\boldsymbol{\theta}, \boldsymbol{\Omega}\right) \qquad (4.1)$$

where $\mathbf{p}_i$ are the parameter values of subject *i*, the elements of $\boldsymbol{\theta}$ are the population typical values, the elements of $\boldsymbol{\eta}$ are the individual random effects and $\boldsymbol{\Omega}$ their covariance matrix.

Concerning the part of the model accounting for measurement uncertainty, the model explaining the glucose profile for the i^{th} individual takes the general form

$$\mathbf{y}_i = \mathbf{f}\left(\mathbf{d}\left(\theta, \eta_i\right)\right) + \boldsymbol{\varepsilon}_i \qquad (4.2)$$

where $\mathbf{y}_i$ is the data profile of subject *i* and $\boldsymbol{\varepsilon}_i$ is the corresponding measurement error, assumed to be normally distributed with zero mean and standard deviation proportional to the measured experimental data

$$\varepsilon_{ij} \sim \mathrm{N}\left(0, \left(\sigma \cdot y_{ij}\right)^2\right) \qquad (4.3)$$

with σ being a further unknown parameter common to the whole population and considered therefore a fixed effect.

Methods

As mentioned before, the estimation of population fixed effects is in general not trivial and many approaches have been proposed in the literature [7, 25, 62]. In Chapter 3, we have already carried out a test on a simulated dataset to explore which is the most reliable method. In this chapter, the analysis is performed on a real data, so the true values of the parameters are not available and another method to assess the quality of the results is necessary.

Thus, we decided to analyze the value of the likelihood function calculated at the optimal values provided by the various algorithms. As previously described in the Background section, the evaluation of the likelihood function implies the

marginalization (an integral operation) along the random effects, which is in general not analytically solvable. For clarity, we report the formula of the i[th] individual's contribution to the log-likelihood:

$$L_i\left(\mathbf{y}_i\middle|\theta,\Omega,\Sigma\right)=-2\log\left(p\left(\mathbf{y}_i\middle|\theta,\Omega,\Sigma\right)\right)=-2\log\left(\int_{-\infty}^{+\infty} l\left(\mathbf{y}_i\middle|\eta,\Sigma\right)h\left(\eta\middle|\theta,\Omega\right)d\eta\right) \tag{4.4}$$

where $l\left(\mathbf{y}_i\middle|\eta,\Sigma\right)$ is the likelihood of the individual parameter given the data, and $h\left(\eta\middle|\theta,\Omega\right)$ is the population probability distribution of the parameters. This term and all the homologous terms corresponding to the other individuals, are summed together to obtain the overall population log-likelihood. To overcome this step, all the NLMEM approaches, during the function minimization recur to some approximation, but these approximations make the estimated value inappropriate for an objective comparison among the solutions provided by different methods. Therefore, we calculate the value of the true likelihood by solving the integral via a Monte Carlo (MC) sampling technique. For each set of fixed effects' optimal values provided by the methods, and for each subject, equation (4.4) is sampled 30,000 times. In addition to the optimal values, the likelihood is also evaluated at two other points near the supposed minimum along each fixed effect, so as to obtain a profile of the function itself along all the population parameters, allowing a visual and numerical assessment of the degree of optimization reached by the different algorithms [75].

For all the population methods included in the comparison, the implementations we used are the ones available in the SPK software [10, 66], Resource Facility for Population Kinetics (RFPK), a NIH/NIBIB research resource in the Department of Bioengineering at the University of Washington. When starting values were necessary for the NLMEMs, the estimates provided by the STS analysis were used.

It should be mentioned that SPK internally transforms the parameters to improve the estimation process. With the purpose of ensuring that the matrices Ω and Σ remain legitimate covariance matrices (and thus positive definite) throughout the optimization process, SPK employs the Cholesky factor of these matrices. More precisely, the elements on the diagonal (variance terms) are log-transformed and then the Cholesky factor of this transformed matrix is employed. This transformation, however, makes the interpretation of the results a bit more challenging, since there is in general no one-to-one correspondence between the elements of the covariance matrix and its Cholesky factor. The example for the terms of the **Ω** matrix is reported below. To interpret the notation, SPK internal convention for parameter names must be considered. Parameters are renamed as α_1 to α_{15}, of which the θs

come first (α_1 to α_4), then the terms of Cholesky factor of $\boldsymbol{\Omega}$, in row major order (α_5 to α_{14}) as indicated in (4.5), and finally $\boldsymbol{\Sigma}$ (α_{15}). This notation will be used also for the figures.

$$\Omega=\begin{bmatrix} \exp(2\alpha_5) & \alpha_6\exp(\alpha_5) & \alpha_8\exp(\alpha_5) & \alpha_{11}\exp(\alpha_5) \\ \alpha_6\exp(\alpha_5) & \alpha_6^2+\exp(2\alpha_7) & \alpha_6\alpha_8+\alpha_9\exp(\alpha_7) & \alpha_6\alpha_{12}+\alpha_{13}\exp(\alpha_7) \\ \alpha_8\exp(\alpha_5) & \alpha_6\alpha_8+\alpha_9\exp(\alpha_7) & \alpha_8^2+\alpha_9^2+\exp(2\alpha_{10}) & \alpha_8\alpha_{11}+\alpha_9\alpha_{12}+\alpha_{13}\exp(\alpha_{10}) \\ \alpha_{11}\exp(\alpha_5) & \alpha_6\alpha_{12}+\alpha_{13}\exp(\alpha_7) & \alpha_8\alpha_{11}+\alpha_9\alpha_{12}+\alpha_{13}\exp(\alpha_{10}) & \alpha_{11}^2+\alpha_{12}^2+\alpha_{13}^2+\exp(2\alpha_{14}) \end{bmatrix} \quad (4.5)$$

In our analysis, then, for the parameters used in the optimization of Ω, the profiling simply offers a way to evaluate with which accuracy the minimum of the objective function was reached, but inference on the precision of the terms of the covariance matrix is not trivial.

The population model was refined after the use of the likelihood profiling. The likelihood function test was employed as a decisional tool to rank the different models, as commonly done [77]. However, it is important to point out that, in SPK, the value of the objective function calculated is the negative log-likelihood (NLL), and therefore, in order to be comparable to a Chi-Squared distribution, the values should be doubled (i.e. the drop in the NLL value consequent to the introduction of one parameter should be at least 1.92 for the parameter to be considered statistically significant, with a p-value of 5%).

Results

Preliminary results

Before proceeding to the results of the profiler, it is interesting to see how the optimal parameter values provided by the various methods differ. The results are in Table 4.1. For each method and parameter, the values of θ and the square root of the corresponding element on the diagonal of Ω are displayed. Given the assumption of a log-normal distribution, θ is the geometric mean of the parameters and the square root of the diagonal elements of Ω can be assumed, in first approximation, as the CV of the population variability. The estimate of the square root of Σ yielded by the different approaches is contained in the last column.

Even at first glance, the estimates provided by FO appear very distant from the ones yielded by all the other methods. Moreover, this happens also for the values of θ, whose estimates are very stable across all the other algorithms. This phenomenon is particularly evident for SI, which is severely underestimated. Conversely, the variance term associated with SI is very large. The same happens with p2, a parameter strongly correlated to SI.

With the exception of FO, population means are estimated coherently by all methods and no significant difference can be detected among them. The variance terms, instead, present some differences. As already pointed out in the literature [25], the STS analysis tends to detect a larger population variability (likely overestimated) than the other methods. This is especially visible in our case for the insulin action parameter p2 (78% vs. an average of ~50%) and for glucose effectiveness SG (31% vs. ~17%). In general, the TS approaches provide a value ~65% for SI's variability as opposed to ~70% provided by FOCE and LAP. Another discrepancy is represented by LAP, which provides a values of 11% for SG's variability, versus an average of ~17% for ITS, GTS and FOCE. Finally, the RUV is estimated ~4%, larger than the CV normally assumed for glucose concentration measurements, ~2%. In addition, two-stage approaches tend to yield a slightly smaller value, ~3.8%, as opposed to NLMEMs, ~4.3%.

Likelihood Profiling Results

The MC evaluation of the Objective Function (Negative Log-Likelihood, or NLL) at the parameter optimal values provided by the different methods is shown in Table 4.2. For the NLMEMs, the value of the approximated NLL is also provided; the TS methods do not optimize a likelihood function, so those values are not available. First of all it must be reported that we were unable to profile the NLL around the values provided as optimal by FO, whose results are very distant from those provided by all the other methods. The approximated NLL value returned by FO was the lowest among the NLMEMs, but apparently it was only a bad estimate due to the poor performance of the approximation; in fact, the failure of the profiler was due to the fact that FO's proposed solution was generating an infinitive value for the NLL. The method which provided the "most likely" solution was FOCE, followed by LAP, then ITS, GTS and lastly STS. So, in general, it appears that the solution yielded by the FOCE and LAP are characterized by a better value for the NLL.

Analyzing more in details the profiles along each of the parameters, it appear that the minimum was reached almost along each direction by FOCE, with the exception of α_3 (the typical value of SI, discussed below) and α_5 (the term corresponding to the variance of SG), which seems slightly overestimated, in which also the performance of the other methods was poor. The estimates for most of the parameters are very similar, with the exception of α_{15}, the RUV. The graphical results in Figure 4.4 show that the RUV estimate provided by NLMEMs is more accurate than the one provided by TS approaches, which tend instead to provide an underestimate.

Taking a look at the precision in parameter estimates, Figure 4.5 contains some perhaps counterintuitive results for the θ estimates. Despite the large sample size of the population in our analysis, V is the only parameter that seems to have been estimated at a well-defined likelihood maximum. The same seems to be true for SG, although to a minor extent. On the other hand, SI and p2, which are also the parameters affected by the largest BSV, are not as accurately estimated, potentially pointing to the need for very large sample sizes to determine typical values of parameters affected by substantial biological variability. Examining Figure 4.3, in addition, some observations about the significance of the terms of the $\boldsymbol{\Omega}$ matrix can be done. Bearing in mind that a step in the NLL of 1.92 is the threshold to consider a parameter statistically significant (p-value 0.05), it can be observed that, for some of the αs, the step of the function between the optimal value and zero is lower or close to the threshold. This happens for α_8 - α_{11} - α_{12}, meaning that these parameters may not be significant in the model. We will analyze the setup of the $\boldsymbol{\Omega}$ matrix in the following paragraphs, but the results from the profiling indicate that the model can be simplified.

Optimal setup for the model

After confirming that FOCE provides the most accurate approximation, this method was used for all the following studies. All the results are summarized in Table 4.4. Even though the physiological interpretation of the parameters implies their positivity, we decided to test the effect of assuming all the parameters as normal; the worsening in the objective function value indicated that the lognormal distribution was more appropriate. Then we implemented different structures for the $\boldsymbol{\Omega}$ matrix, trying to simplify the model by removing the least significant terms. From the analysis of the parameter estimates and their precision (Table 4.3), it was possible to infer that the most significant correlations between the parameters are SG-VOL, VOL-P2 and SI-P2. In addition VOL-P2 does not appear very strong, and the other terms correspond to a very low correlation and their estimates are affected by a large uncertainty. The current implementation of SPK allowed multiple-blocks matrices, but an implementation preserving all the three terms was not available. So we tested all the available combinations comprising the most significant terms.. Results are presented in Table 4.4.

The worsening in the NLL between the model with full Ω matrix and its competitors indicates that some of the removed parameters were actually significant in the model; nonetheless, as no implementation was possible comprising all the three most

significant correlations (likely the best model), we decided that is was better to have a slightly under-parameterized model but only with statistically significant parameters, rather than an over-parameterization with some serious estimation issues, such as CVs>100%. This decision is also due to fact that the final scope of the analysis is the integration of covariates in the model, which will inevitably further increase the number of parameters. In addition, the model with a 3x3 block for VOL, SI and P2 was affected by the same problem, since the correlation term VOL-P2 was not significantly different from zero. For this reason, it was not excluded from the comparison.

Other models tested for the $\boldsymbol{\Omega}$ matrix presented no parameters whose CI was including zero, so they were all considered in the analysis. Comparing the objective function values and, bearing in mind that the statistical significance threshold for one parameter in SPK is 1.92, the model resulting as most significant was the one including the correlations VOL-SG and SI-P2.

After selecting this model, we also tested the effect of error measurement as proportional to the model prediction rather than the experimental data, but the results were unsatisfactory.

The results on the final model are displayed in Figure 4.8. The estimates of the population typical values exhibit the same trend as with the model with full population covariance matrix. θ_{VOL}, seems both well centered on the minimum and precisely estimated, and, to a minor extent, also θ_{SG} exhibits the same behavior. θ_{SI} and θ_{p2}, however, are not as accurately estimated, and also the precision is lower. The analysis of the variance terms reveals that the estimates of variability of SG and P2 are the less precise and possibly slightly overestimated, whereas the RUV seems underestimated. Finally, the off-diagonal terms preserved in this final model, are estimated with good precision; the other terms of the $\boldsymbol{\Omega}$ matrix, which did not results significant in the full model, have been removed.

Final results

The results obtained with the finally selected model are reported in Table 4.5. Despite the relative homogeneity of the subjects in our dataset (all the subjects were healthy and had no previous history of glucose intolerance), a large BSV variability is detected for SI (70%) and P2 (~51%). Much smaller is the BSV for SG and VOL, which does not even reach 20%.

The RUV is estimated at 4.4%, larger than the degree of error normally assumed for glucose measurements (2%).

The histograms of the individual parameters' distribution seem to be compliant with the lognormality hypothesis; only one outlier is detected, as its optimal value for VOL lies at more than 3 Standard Deviations from the population typical values.
The scatterplots clearly indicate a strong correlation between log(SI) and log(P2), accounted for by the appropriate term in the Ω matrix, which estimates this correlation at about 0.87. Another relevant correlation, although less apparent from the scatterplots, is detected between log(SG) and log(VOL), r=-0.40.

Conclusions and Discussion

In this work we have implemented a population model for the glucose minimal model, which will be the starting point for the covariate analysis performed in Chapter 5. Even if the subjects in the dataset are all healthy and with no previous history of glucose intolerance, a certain amount of population variability is present for SI and P2, and the purpose of the covariates analysis will be trying to explain in a deterministic way this variability.
The true likelihood profiling via MC sampling proved a useful tool to probe the results provided by the different algorithms, and the results obtained in the previous chapter found in this work an independent validation. The unsatisfactory quality of the approximation provided by FO is more evident with this real dataset than in the previous study (Chapter 3), in which a smaller and simulated dataset was employed. The profiling also helped in assessing the accuracy and precision with which the parameters are estimated, and it confirmed that FOCE and LAP provide more accurate estimates of the RUV. A phenomenon revealed by the profiler might result a bit counterintuitive: even if our dataset comprised a large number of subjects and samples, some difficulties were encountered in the estimation of the population typical values for SI and P2. This is most likely due to the large BSV characterizing these parameters in our dataset, however, it is interesting to acknowledge this relatively low sensitivity of the OF to some of the population typical parameters, normally estimated with very satisfactory precision The profiling also allowed to inspect the overall quality of the estimates and statistical significance of the terms in the Ω matrix, even though the transformation internally operated by SPK made direct inference on specific parameters unfeasible. Possible improvements to this approach and therefore ideas for further investigation can imply the use of more sophisticated sampling algorithms (e.g. Markov Chain Monte Carlo), to gauge whether it is possible to obtain reliable results also with a smaller computational effort.

Table 4.1 Values for θ, root square of the diagonal terms of Ω and Σ, as provided by each of the tested methods. The units for BSV and RUV are % (dimensionless).

	$\log(S_G)$	$\log(V)$	$\log(S_I)$	$\log(p_2)$	σ
θ_{STS}	-4.02	0.477	-9.7	-3.6	3.89%
ω_{STS}	31%	13%	65%	79%	
θ_{ITS}	-3.99	0.478	-9.75	-3.51	3.89%
ω_{ITS}	16%	12%	65%	54%	
θ_{GTS}	-3.99	0.481	-9.75	-3.46	3.73%
ω_{GTS}	17%	12%	66%	48%	
θ_{FO}	-3.73	0.386	-11.7	-3.36	3.55%
ω_{FO}	27%	15%	823%	515%	
θ_{FOCE}	-3.96	0.471	-9.8	-3.52	4.39%
ω_{FOCE}	18%	12%	70%	52%	
θ_{LAP}	-3.97	0.474	-9.8	-3.5	4.44%
ω_{LAP}	11%	11%	70%	46%	

Table 4.2 The estimated values of the likelihood function, calculated at the maximum provided by the different methods. The first line contains, when available (TS methods do not provide it), the estimate based on the approximated Likelihood. The second line contains the NLL from MC and, in brackets, its standard deviation (SD). FO's likelihood profiling failed.

	STS	ITS	GTS	FO	FOCE	LAP
Approx	N/A	N/A	N/A	12492.40	12853.7	12873.7
MC	13023.1	12943.9	12980.3	N/A	12884.3	12912.2
(SD)	(5.8)	(4.7)	(5.0)	(N/A)	(4.1)	(6.7)

Table 4.3 Estimates and relative precision for the elements on the Ω matrix, as provided by the FOCE algorithm. Given the lognormal distribution for the parameters, the correlations are to be interpreted between the logarithms of the parameters. The "*" symbol, indicates that the confidence intervals include 0, indicating that the parameter might not be statistically significant.

Parameter	Covariance	CV%	LB	UB	Correlation
VAR(log(SG))	0.0310	25.4	0.0156	0.0465	
SG-VOL	-0.0086	33.5	-0.0142	-0.0030	-0.42
SG-SI*	-0.0081	146.0	-0.0313	0.0151	-0.07
SG-P2*	-0.0069	174.0	-0.0303	0.0166	-0.08
VAR(log(VOL))	0.0133	11.6	0.0103	0.0164	
VOL-SI	0.0233	26.4	0.0113	0.0354	0.29
VOL-P2	0.0114	46.8	0.0009	0.0219	0.19
VAR(log(SI))	0.4950	8.9	0.4090	0.5810	
SI-P2	0.3160	10.2	0.2530	0.3790	0.87
VAR(log(P2))	0.2660	12.7	0.2000	0.3330	

Table 4.4 NLL values provided by all the different models and parameterization attempted. The first column indicates the distinctive features of the model under evaluation, the second contains the NLL value, the third the number of parameters included in the model, whereas the last one indicates whether problems of statistical significance were detected in the model.

Model	NLL value	Number of Parameters	Confidence intervals including 0?
FULL Gaussian parameters	12919.5	15	YES, 3
FULL block	12853.7	15	YES, 2
FULL block for VOL-SI-P2	12861.5	12	YES, 1
only SG-VOL SI-P2	12862.2	11	NO
only VOL-SI SI-P2	12862.3	11	NO
only SG-VOL VOL-SI	12930.3	11	NO
only SG-VOL	12938.5	10	NO
only VOL-SI	12938.7	10	NO
only SI-P2	12869.3	10	NO
DIAGONAL	12943.3	9	NO
only SG-VOL SI-P2 Weight on model	12888.8	11	NO

Table 4.5 Population parameter values for the final model. The typical values are reported as exp(θ), so that the results are dimensionally compatible with the minimal model parameters. The terms of the Ω and Σ matrices are reported as CV% (approximated in the case of Ω) and correlations.

SG		VOL		SI		P2		Ω - Correlations		Σ
exp(θ)	Ω CV %	exp(θ)	Ω CV %	exp(θ)	Ω CV %	exp(θ)	Ω CV %	corr (SG-VOL)	corr (SI-P2)	CV %
0.0191 (0.0184 0.0196)	17.1 (13.2 20.1)	1.60 (1.57 1.63)	11.4 (10.0 12.6)	5.58E-5 (5.01E-5 6.15E-5)	69.8 (63.3 75.7)	0.0298 (0.0271 0.0326)	51.3 (43.8 57.7)	-0.40 (-0.65 -0.14)	0.87 (0.69 1.00)	4.4 (4.2 4.6)

Figure 4.1 Profiling of the objective function along the θs around the optimal values provided by the different population algorithms. The units in abscissa and ordinate are the same for all the panels, so that a graphical comparison of the parameter precision can be easily performed. As the parameters are the logarithms of the typical population values, the curvature in the minimum can be interpreted as proportional to the estimation CV.

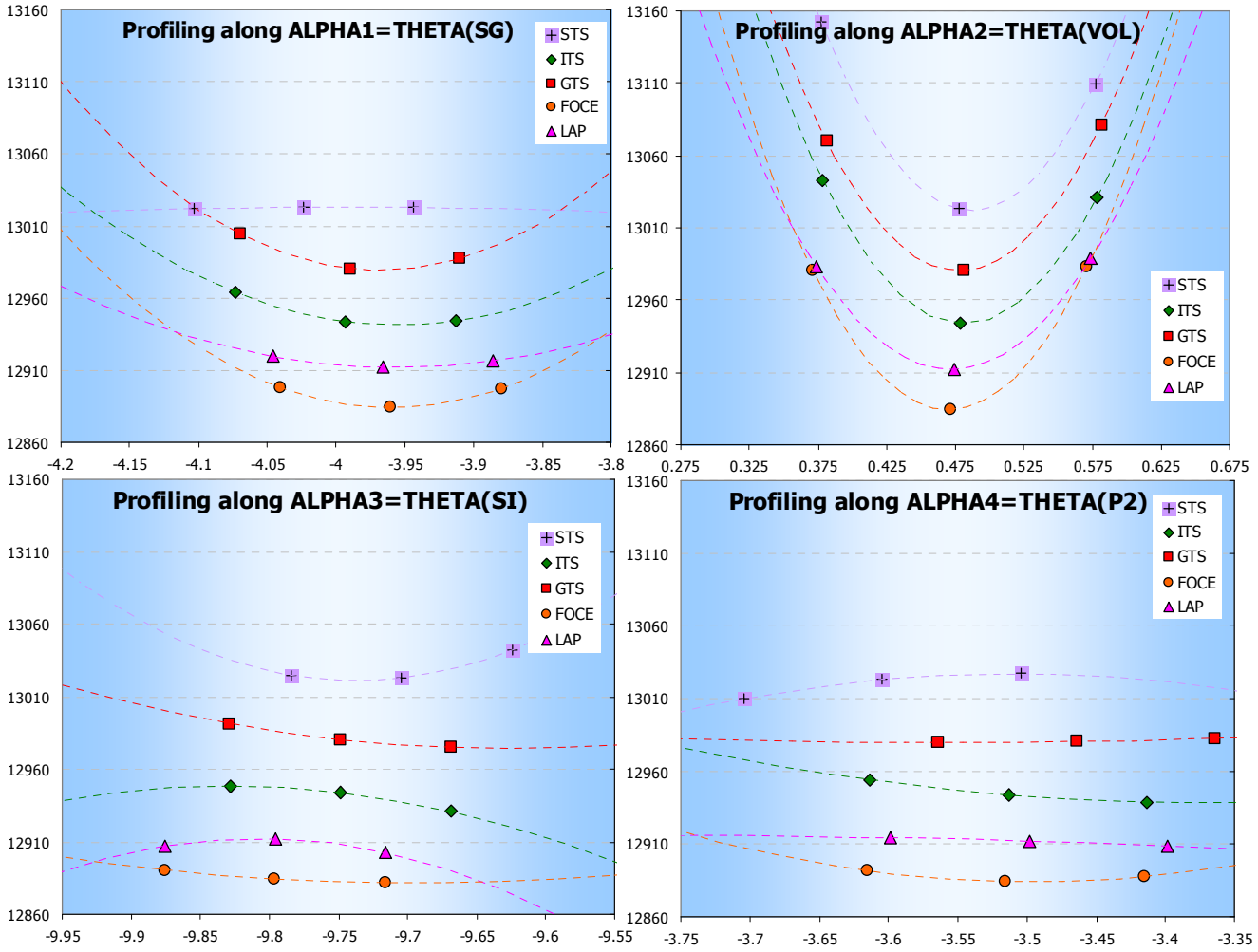


Figure 4.2 NLL profiling along the diagonal elements of the Cholesky factor of the Ω matrix. As explained in the text, inference on the precision of the parameters of the original Ω matrix is very hard, so the charts are exposed only as a qualitative evaluation of the optimal values provided by the different methods.

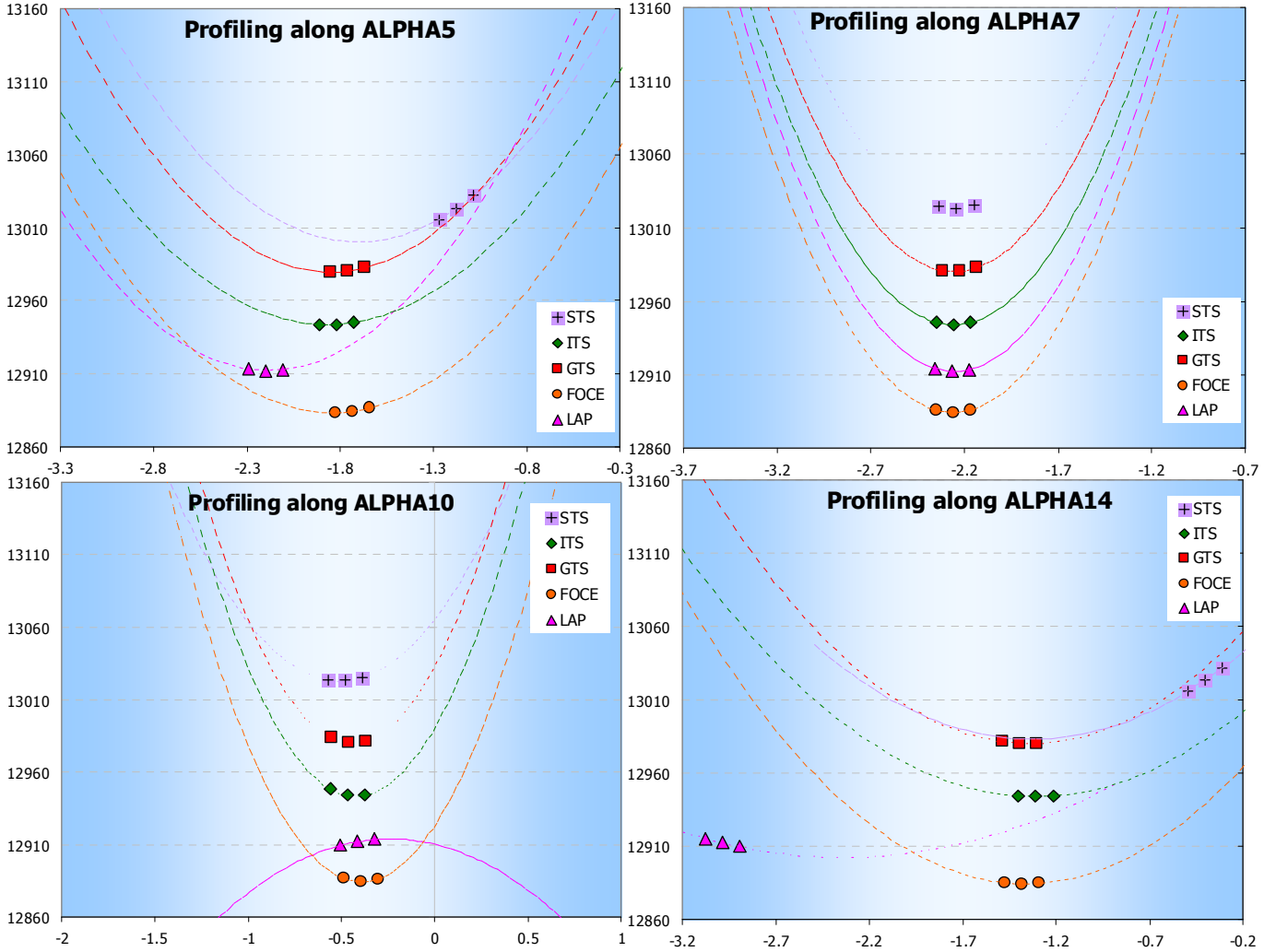


Figure 4.3 NLL profiling the off-diagonal elements of the Cholesky factor of the Ω matrix. As explained in the text, inference on the precision of the parameters of the original Ω matrix is very hard, so the charts are exposed (and without sharing the same scaling) only as a qualitative evaluation of the optimal values reached by the different methods.

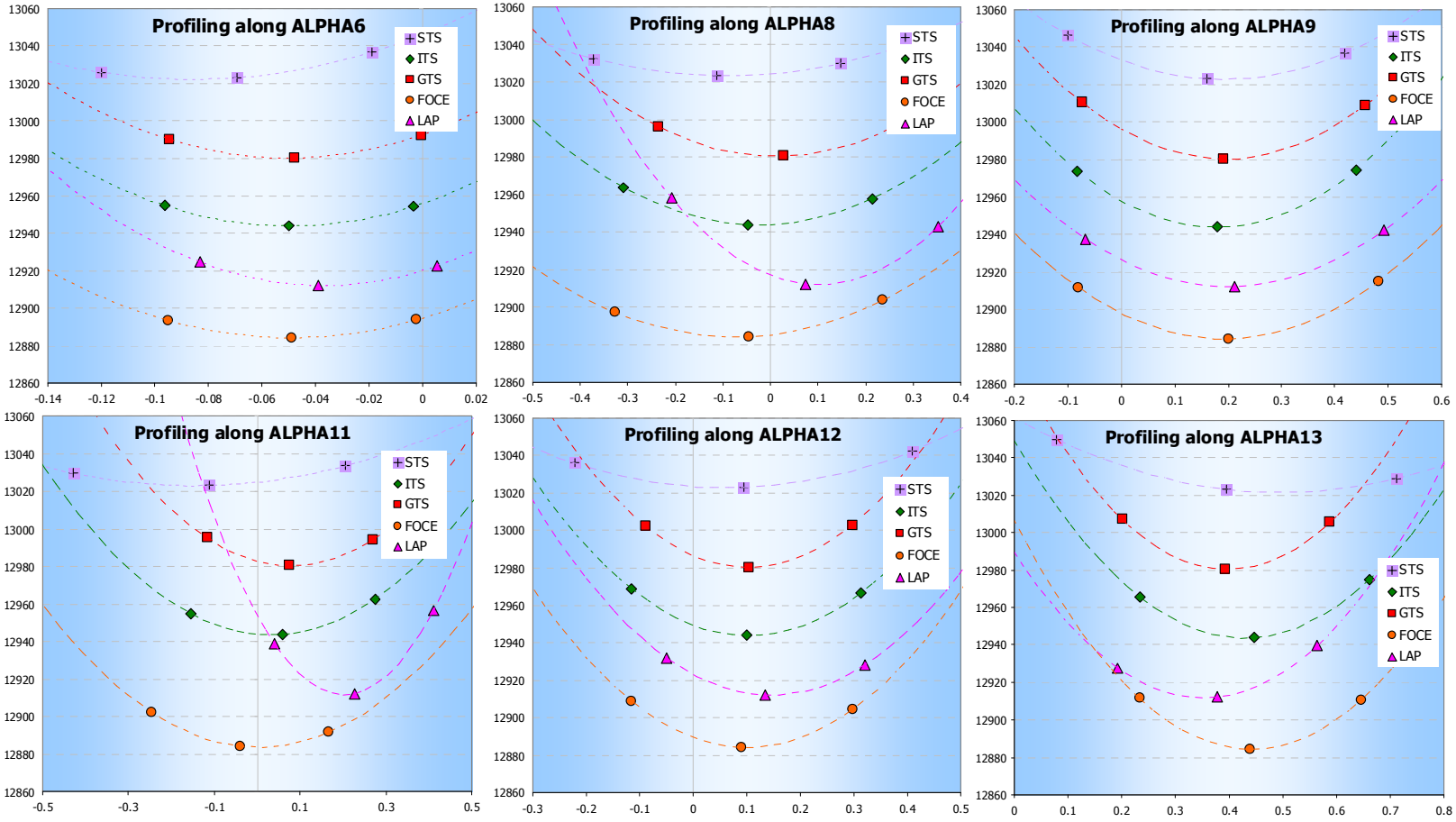


Figure 4.4 NLL profiling for the parameter Σ (log(σ) is actually used). For each method (STS, ITS, GTS, FOCE and LAP), the NLL is evaluated at the estimated optimal value (the central colored point) and two values around it. Note the biased estimate provided by the TS methods.

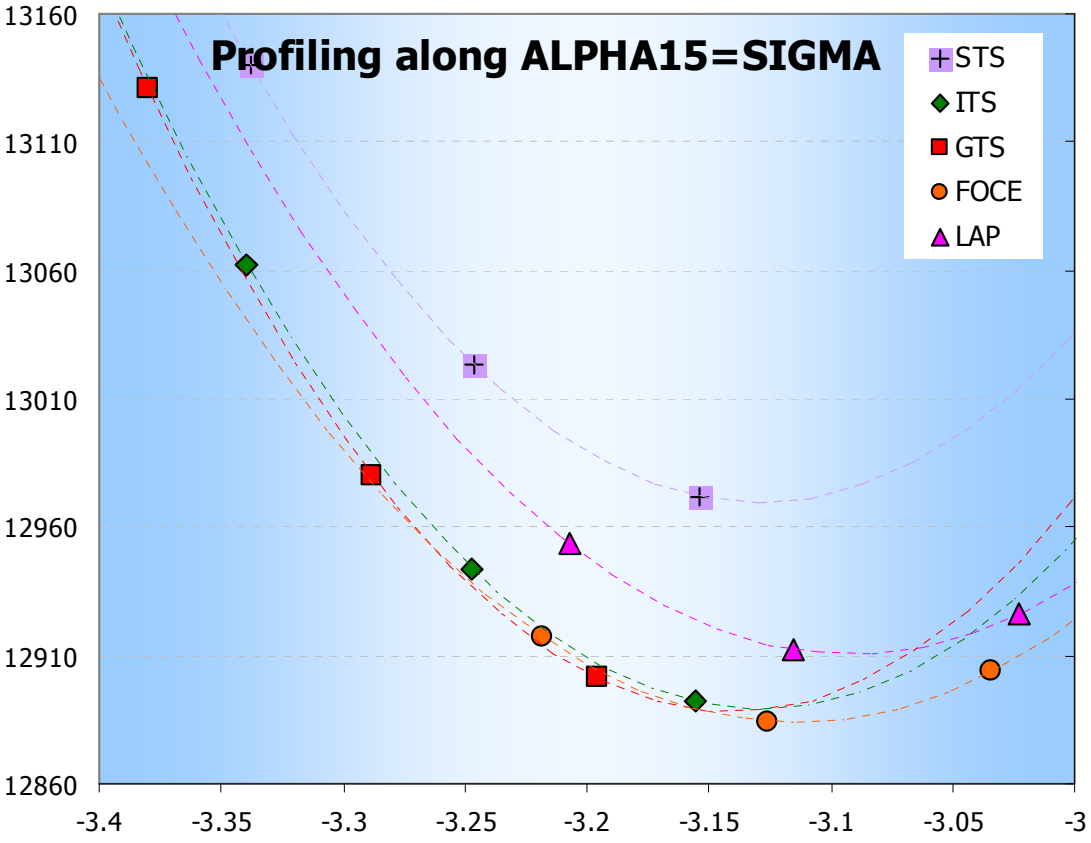


Figure 4.5 Profiling of the θs centered on their optimal value as provided by FOCE with full covariance matrix. The curvature of the objective function around the minimum provides a graphical evaluation of the precision of the estimates. Bearing in mind that in our model the θs are log-transformed with respect to the minimal model parameters, in first approximation the curvature can be considered proportional to the CV and the entries in abscissa can be approximately read as percentages of the mean value.

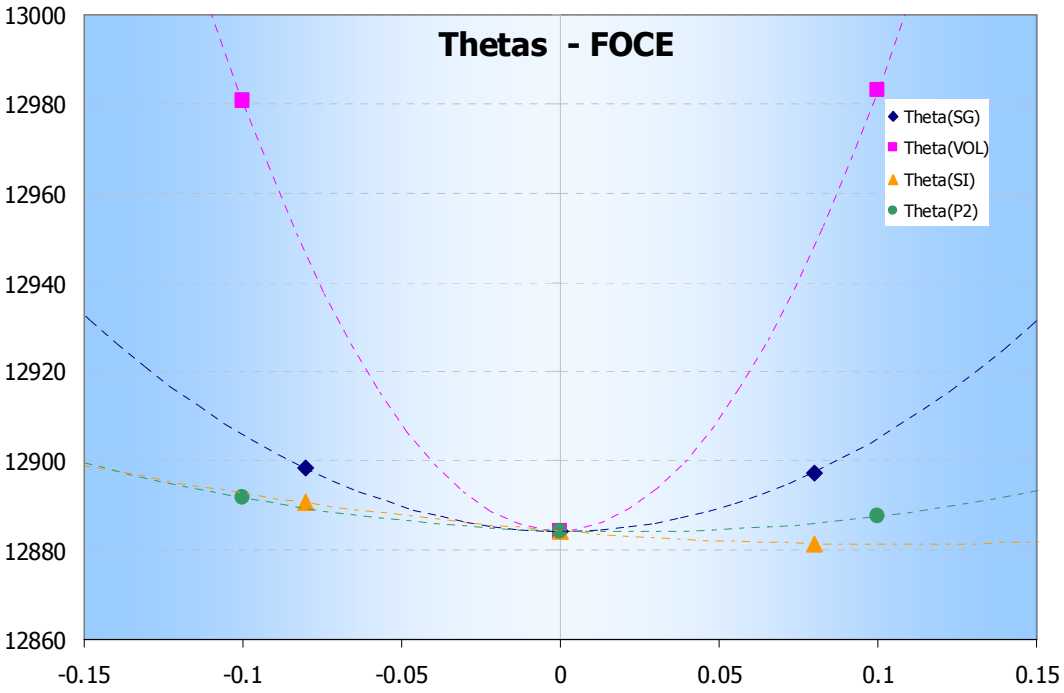


Figure 4.6 Histogram plots and smoothed densities for the log-transformed minimal model individual parameter values obtained with the final model.

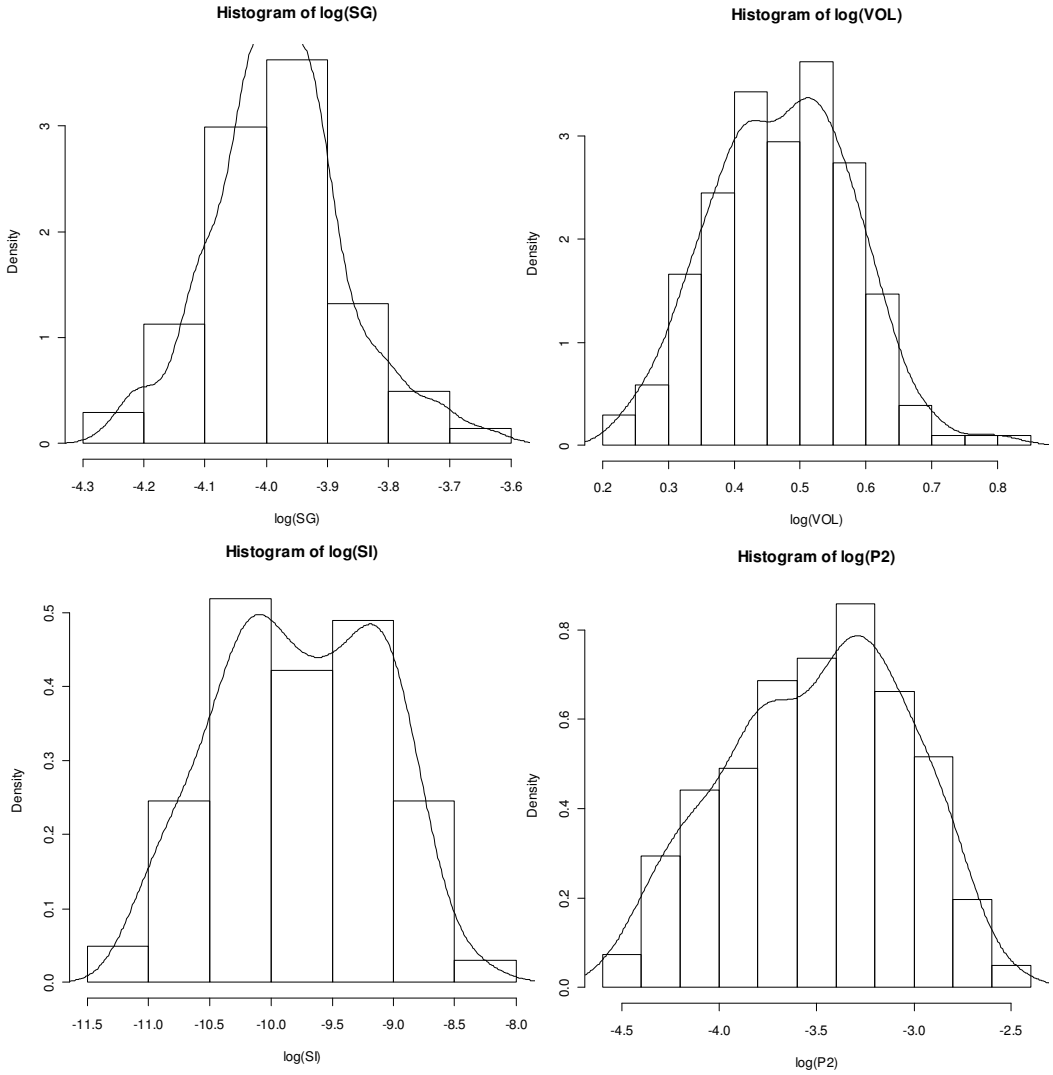


Figure 4.7 Scatterplot of the logarithm of the parameter values, as obtained with final model. A very strong correlation is apparent between log(SI) and log(P2), as accounted for by the relative term in the covariance matrix Ω.

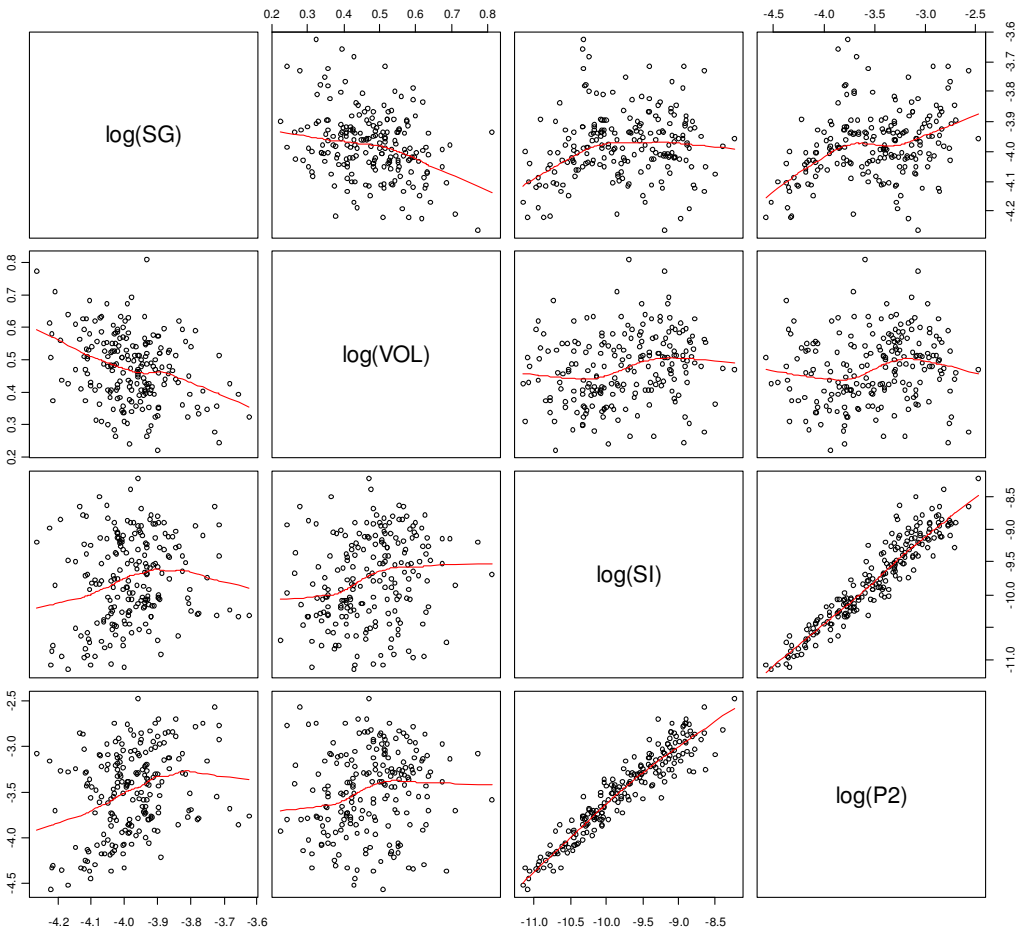
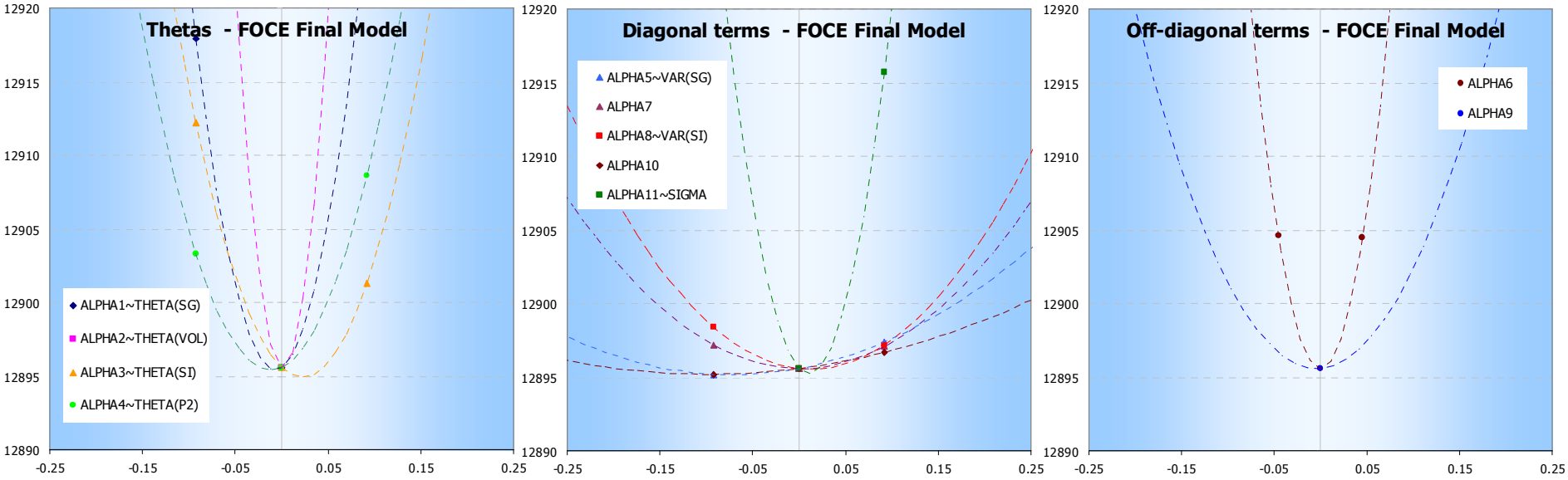


Figure 4.8 Profiling of the along the parameter estimates for the final model. As the population covariance matrix Ω has been optimized and the number of parameters reduced, the new meaning of the α values is reported in the following equations (the θs remain the same).

$$\boldsymbol{\Omega} = \begin{bmatrix} \exp(2\alpha_5) & \alpha_6 \exp(\alpha_5) & 0 & 0 \\ \alpha_6 \exp(\alpha_5) & \alpha_6^2 + \exp(2\alpha_7) & 0 & 0 \\ 0 & 0 & \exp(2\alpha_8) & \alpha_9 \exp(\alpha_8) \\ 0 & 0 & \alpha_9 \exp(\alpha_8) & \alpha_9^2 + \exp(2\alpha_{10}) \end{bmatrix} \qquad \boldsymbol{\Sigma} = \left[\exp(2\alpha_{11})\right]$$

In the following charts, all the parameters have been centered on their optimal value as provided by FOCE. The leftmost panel shows the θs, the central one the elements on the diagonal of the Cholesky factor of Ω and Σ, whereas the rightmost displays the off-diagonal elements. Because of the nature of the parameters analyzed (mean, variances and covariances) and the internal transformations operated by SPK, comparison of the precision based on the curvature is possible only among elements in the same chart.



Chapter 5

IVGTT Glucose Minimal Model Covariate Selection by NonLinear Mixed-Effects Approach.

Overview

The implementation of a population framework for the glucose minimal model carried out in the previous chapter was the first step towards the creation of a more sophisticated model including covariates. At the time of the IVGTT experiment, in fact, demographic information about the subjects were collected and are therefore available for our analysis. The aim is to inspect the presence of significant correlation between the individual parameter values and the physiological characteristics, and explore the possibility to employ the latter as predictors of response. One of the great advantages of NLMEMs consists in the flexibility to define the population distribution. It is possible to parameterize the population typical values as a function of some of the covariates, which are considered as fixed effects and their values optimized along with the other parameters. Covariate selection is a very complex problem and no method proposed so far has the guarantee to provide the "best model". Normally researchers rely on the Likelihood Ratio Test, inspecting whether the introduction of an additional parameter in the model is statistically significant. Anyway, the number of possible models is at least (considering, for example, only linear models) exponential relative to the number of covariates, and it is unfeasible to explore all the possibilities. What is normally done is to recur to stepwise techniques, which imply the insertion (and/or removal) of the most (least) significant parameter at each step, until no further modification to the model results into statistically significant changes in the objective function. However, also this methodology implies, at each step, the exploration of all the possible additions (subtractions), and, in our case, this would have spawned a huge number of NLMEM regressions. This is the reason which urged us to design an alternative approach, based on a preliminary skimming of the pool of possible covariate models, carried out by regression analysis on the individual parameter values provided by the base model (without covariates). After the group of candidate models has been narrowed down, a procedure similar to the stepwise is executed. Our method selected age, visceral abdominal fat and basal insulinemia as predictors for SI, and age, total abdominal fat and basal insulinemia for P2. The volume of distribution was found to be correlated to gender, age, the

percentage of total body fat and basal glycemia, whereas SG was best explained by height, weight and body surface area. This last model, however, might be due to statistical artifacts, rather than actual physiologic significance. Our analysis finds as alternatives the use of age or basal glycemia as predictors. For the other parameters, and especially for SI and P2, the incorporation of covariates resulted in a significant shrinking of the BSV. The overall predictive power of the model was considerably increased with the incorporation of easily, inexpensively and non-invasively collectible physiological information; a significant portion of the inter-individual variability is explained in a deterministic fashion, and a smaller part is left to a stochastic component, the random effects. This can not only offer a starting point for speculation about the significance of the relationships detected, but can also provide a tool to allow the design of less invasive and less expensive protocols for epidemiological studies of the glucose disposal metabolic system.

Introduction

Variability in biological systems is a widespread trait and poses an extremely challenging problem. Issues related to variability impinge on the design and execution of experiments on biological systems, especially in the clinical setting, where measurements are few and far between each other. The challenge posed by variability is compounded by the fact that, often, the traits to be measured cannot be quantified directly, but are amenable only to indirect quantification. It can be argued that, when such quantification is only possible through identification of a mathematical model, new, state of the art tools are mandatory to both quantify and explain sources of in vivo variability. An example of model-based biomedical investigation is provided by the so-called minimal model of glucose kinetics. The model, when fitted to individual responses from an intravenous glucose tolerance test (IVGTT), provides useful indices of an individual subject's metabolic state, including measures of insulin sensitivity and glucose effectiveness. The traditional paradigm to estimate the IVGTT glucose minimal model parameters is by means of Weighted Least Squares (WLS), performed on each single subject experimental data. However, in case of an epidemiologic study, encompassing large numbers of subjects, the use of a population approach has been shown to both improve the quality of the individual parameter estimates and to provide an accurate insight on the between-subject variability of the parameter in the population. Generally, in fact, the analysis of the population variability is performed only as an "a posteriori" step, by means of a statistical analysis (namely sample mean and covariance) on the individual results previously obtained by the traditional WLS approach. This method, however, does not normally

take into account the uncertainty in the estimates of the individual parameters and is therefore prone to provide biased values.

Population approaches, instead, are based on the assumption that all the subjects are realizations of the population and they share some common features typical to the population itself, called fixed effects. The population information is employed to support the individual parameter estimation process, because each subject is characterized by a random effect, which represents its deviation from the population typical values. In particular, parametric population approaches assume a probability distribution for the different subjects and then estimate the parameter values characterizing this distribution, i.e. means, variances and correlations.

A further step in epidemiological studies consists in trying to inspect the causes underlying the variability present in the data, in an attempt to identify whether some physiological characteristics of the subject significantly correlate with the model parameter values. These features, normally referred to in population analysis as *covariates*, can be used not only for speculation about the system being tested, but they can also be integrated into the population model itself to improve its predictive power. The coefficients driving the relationships between the individual parameter values and the covariates, in fact, can be introduced in the model and therefore optimized together with the remaining population fixed effects. In this way, a part of the population variability, previously accounted for by the individual random effects, is explained in a deterministic fashion. Moreover, the equations linking the model parameters to the independently measured covariates can be used to predict individual parameter values, thus potentially reducing the need for blood samples and invasive trials.

In this work, we apply methods for covariate selection to a large database of IVGTTs. We show that it is feasible to use independently measured, relatively easy to obtain covariates to explain observed variation in the metabolic parameters provided by the glucose minimal model. This work lays the foundation for further research aimed at investigating reduced protocols, not only based on blood sampling but also on individual-specific covariate values, for the identification of the glucose minimal model.

Materials and methods

In our study, we analyzed a heavily sampled dataset encompassing 204 healthy subjects (mean age 56 yrs, range 18-87; mean BMI 27 kg/m^2, range 20-35) who underwent a full sampling schedule insulin-modified [33] IVGTT (240 min, 21 samples, insulin infusion at 20 minutes lasting 5 minutes) [4, 5]. A description of the

minimal model of glucose disappearance and its parameters is provided in the introduction (Chapter 2). At the time of the experiments, additional information about the patients was collected, with the purpose of investigating which physiological characteristics were more significant predictors of the individual parameter values. More in detail, this information included age (AGE), Gender (SEX), Height (BH), Weight (BW) and basal levels of glucose (GBSL) and insulin (IBSL). In addition, body composition was measured using X-ray absorptiometry and computed tomographic scans, as described in [4, 5]; in this way the following quantities were obtained: total body fat (TBF, or expressed as a percentage %TBF), visceral (VAF) and total abdominal body fat (TAF), lean body mass (LBM). Lastly, some derived covariates were calculated: Body Mass Index (BMI), Body Surface Area (BSA). All these covariates are listed in Table 5.1 and graphical summaries of the covariates' distributions can be seen in Figure 5.1. Also some regressions between the most correlated covariates are reported in Figure 5.3. Three subjects had some missing values for VAF, TAF and %TBF, therefore their value was assumed to be the average of the population. This seemingly naïve approach is justified by the fact that, when the model was fit, regression was carried out on the values of the covariates centered each one on its mean, and in this way the effect of that covariate was totally null for the subjects with missing data.

The procedure for covariate selection is described below. As a first step, a population model without covariates was implemented (*base model*, BM), as explained in Chapter 4, and estimates obtained for all the population parameters and the individual random effects. The statistics of the individual parameter estimates (in log-transformed units) are shown in Table 5.2, while a histogram representation is shown in Figure 5.2. Regressions between the parameters highlighting correlations are available in Figure 5.4. Based on previous literature [1, 31, 44] and on these results, the parameters $\mathbf{p}_i$ were assumed lognormal. In addition, the measurement error ε_{ij} proportional to the experimental data y_{ij} and the covariance matrix of the random effects Ω comprised only the correlation terms between SI-P2 and VOL-SG.

In formulae

$$\left.\begin{array}{l}\mathbf{p}_i = \exp\left(\theta + \eta_i\right) \\ \eta_i \sim N\left(0, \Omega\right)\end{array}\right\} \rightarrow \mathbf{p}_i \sim LN\left(\theta, \Omega\right) \tag{5.1}$$

$$\Omega = \begin{bmatrix} \omega_{SG}^2 & \omega_{SG-VOL} & 0 & 0 \\ \omega_{SG-VOL} & \omega_{VOL}^2 & 0 & 0 \\ 0 & 0 & \omega_{SI}^2 & \omega_{SI-P2} \\ 0 & 0 & \omega_{SI-P2} & \omega_{P2}^2 \end{bmatrix} \tag{5.2}$$

$$\varepsilon_{ij} \sim N\left(0, \left(\sigma \cdot y_{ij}\right)^2\right) \tag{5.3}$$

In our reports, for the sake of easiness of interpretation of the results, we reported the values of $\exp(\boldsymbol{\theta})$ for each parameter typical value, so that the figures are dimensionally similar to reasonable parameters values and more easily interpretable. The same reasons led us to report the square root of the elements on the diagonal of Ω, which can be interpreted in first approximation in terms of %CV, and the off diagonal elements were transformed into the corresponding correlations, rather than covariances.

The Expected Hessian algorithm (FOCE with interaction) as implemented in SPK [66] has been employed for all computations.

The individual random effects obtained from BM were used for the regression analysis. The regression analysis could potentially be performed directly with SPK, but the long computation time required to fit each single model, makes this approach unfeasible. So we decided to adopt the strategy of performing an exploratory correlation analysis of parameter estimates and covariates in the statistical software R [65] and narrow down this way the pool of potential candidate models to test with a full-fledged nonlinear mixed-effects model analysis. Even though the algorithms for linear regression analysis are very fast, the number of possible models grows exponentially with the number of covariates, in our case 14, and this makes impractical the exhaustive exploration of all the possibilities. Therefore, we recurred to heuristic procedures which use branch-and-bound techniques to dramatically reduce the number of tested models. The criteria which were employed to rank the models and select the most significant, are the Adjusted R^2 (R_a^2) and Mallows' Cp [47]. The Adjusted R^2 measures the strength of the correlation existing between each parameter and a group of predictors (covariates), at the same time adjusting the result for the number of predictors. Its mathematical formulation is the following:

$$R_a^2 = 1 - \frac{VAR_{reg}}{VAR_{tot}} = 1 - \frac{SS_{reg} / df_{reg}}{SS_{tot} / df_{tot}} \tag{5.4}$$

where VAR, SS and df denote respectively the (unbiased) variance, the sum of squares and the degrees of freedom (# of samples - # of predictors). These quantities refer either to the data (subscripted *tot*) or to the regression model (subscripted *reg*).

Mallows' Cp compares each possible submodel with the full model (i.e. the model including all predictors) by calculating the ratio of the submodel's and the full model's residual unknown variance, further correcting for the number of parameters and available samples (degrees of freedom). Mathematically

$$Cp = \frac{SS_{reg}}{VAR_{all}} - N + 2P \tag{5.5}$$

where N indicates the number of data points, P the number of regressors used in the submodel, and subscripted *all* refers to the model using with all predictors.

We started by determining the range of model order (number of predictors included), based on Mallows' Cp. For each order in the range, the best model(s) were chosen for implementation in SPK. Then, Adjusted R-squared were computed for all these models: when the models were close or discrimination was difficult, the Adjusted R-squared was used to further rank models. These models were implemented as nonlinear mixed-effects models in SPK, and for the optimization, the initial values were the ones provided by the regression analysis of parameters vs. covariates performed in R. Whenever the most promising covariate model order was at the bounds of the range which was being analyzed, the range itself was broadened to explore the effects of choosing the best fitting models with greater or lower order. Once the mixed-effects optimal parameter values were obtained, the objective function minima were used to assess the actual significance of the covariates in the models, as it is commonly done [77]. In SPK, the value of the objective function reported to the user is the negative log-likelihood (NLL), and therefore the significance levels are halved with respect to a chi-squared distribution, as performed in [77] (i.e. the drop in the NLL value consequent to the introduction of a single parameter should be at least 1.92 for the parameter to be statistically significant).

Results

Results from the exploratory covariate ranking using the Adjusted R^2 and Mallows' Cp are reported in Table 5.3 and 5.4. Selection of covariate models to be tested with nonlinear mixed effects analysis proceeded as follows. A first selection of the order of the model (i.e. the number of predictors included) was performed by visual inspection of the Mallows' Cp charts for each parameter, by highlighting the portion of the chart where the best models congregated, and thus determining a preliminary range of models orders (i.e., number of covariates to be included). Subsequently, for each model order, if models had similar Mallows' Cp, they were further ranked by inspection of Adjusted R-squared values. Models were then run in SPK and the NLL value was obtained for every model. Covariates (with the exception of SEX) were centered on the mean, i.e. the difference of the individual value from the mean was entered in the model, as opposed to the raw value of the covariate. This does not affect the covariate selection process or the minimum of the objective function which is reached, but we introduced it mainly for two reasons. The first is that it allows us to

impute the values of missing covariates to the mean value for several subjects (in our case only 3); in this way those specific subjects will not influence the effect of the covariate. In addition, centering the regressors on their mean confers more stability to the value of the fixed effects modeling each parameter's typical values, making their interpretation easier.

The covariate selection process is summarized in Table 5.5. All the models which were tested in the SPK environment were numbered from 0 to 70, and they are grouped according to the "selection step" to which they belong. For each model, the value of the NLL and its difference with the value provided for the BM is reported.

Alternative covariate models were tested, following the sequence VOL-SG-SI-P2, and differences in NLL were recorded. We proceeded by steps, choosing the best covariate model for each parameter, and then using it as a starting point to further advance in the selection for the remaining parameters. For each parameter, all the candidates models selected via the explorative regression in R were implemented, and then ranked according to their Chi-Squared p-value with respect to the base model. The model proving more statistically significant was selected for improvement in the following step. It should be borne in mind that the objective function values provided by SPK are the NLL, and were therefore doubled before the use of the Chi-Squared test. In addition, during the covariate selection process, care was taken in analyzing the precision of the parameter estimates. Furthermore, whenever estimation produced a 95% confidence interval for a covariate coefficient which included zero, the validity of the model was questioned. It is interesting to notice that this additional rule did not interfere with the p-value ranking criterion. In other words, no model that would have been selected with the analysis of the NLL value was rejected in this way, and the presence of zero inside one of the confidence intervals always coincided with the model's p-value being smaller than those characterized by higher precision in the coefficient estimates.

Our covariate selection roadmap proceeded as follows. As a first step, the candidate models for Volume were tested. The improvement in the objective function value is very pronounced (-76 NLL, +4 parameters), and all the candidate models presented a very low p-value, much lower than the cutoff value of 5%. This can partly be attributed to the presence of an outlier in the base model, which, when incorporating the covariates, reduces considerably the size of its random effect. At this step, the model with the lowest p-value included SEX, AGE, %TBF and GBSL. This temporarily optimal model for VOL was selected and, in the following step, it was expanded to include the candidate models for SG. In this case, the improvement of the NLL value was much smaller (-8.2 NLL, +3 parameters). The selected model for

SG consisted in the insertion of BH, BW and BSA. Consequently, the models for SI were introduced, a large decrease in the NLL was detected (-60.2 NLL, +2 parameters) and our criterion selected, in this step, VAF and IBSL as predictors. The next candidates tested were the models for P2, and the best option was including AGE, TAF and IBSL (-22.8 NLL, +3 parameters). At this point, some of the previously rejected models for each of the parameters were evaluated once again, to test whether the inclusion of covariates for the other model parameters had modified the statistical significance level of the other predictors. The models whose performance had been very poor in the previous step (whose p-value was 100X or higher than those of the competitor models) were excluded from this reevaluation. This is also indicated in Table 5.5 with a # symbol following the model number. Models for SI were tested once again, and, at this stage, our criterion selected the addition of AGE into the model. After this inclusion, once again the performance of all the covariate models for other parameters in combination with the newly selected one for SI was assessed. However, no other model provided a lower p-value. As a further validation of the optimality of the chosen model, some hybrid models were implemented by selecting features from the alternative parameterizations providing the lowest p-values beside the optimal model. No other model proved better with respect to the selection criterion.

The final model was then number 38

$$\begin{aligned}
SG&=\theta_1\cdot\exp\left(\theta_9\cdot\left(BH-\overline{BH}\right)+\theta_{10}\cdot\left(BW-\overline{BW}\right)+\theta_{11}\cdot\left(BSA-\overline{BSA}\right)+\eta_1\right)\\
VOL&=\theta_2\cdot\exp\left(\theta_5\cdot SEXM+\theta_6\cdot\left(AGE-\overline{AGE}\right)+\theta_7\cdot\left(\%TBF-\overline{\%TBF}\right)+\theta_8\cdot\left(GBSL-\overline{GBSL}\right)+\eta_2\right)\\
SI&=\theta_3\cdot\exp\left(\theta_{12}\cdot\left(AGE-\overline{AGE}\right)+\theta_{13}\cdot\left(VAF-\overline{VAF}\right)+\theta_{14}\cdot\left(IBSL-\overline{IBSL}\right)+\eta_3\right)\\
P2&=\theta_4\cdot\exp\left(\theta_{15}\cdot\left(AGE-\overline{AGE}\right)+\theta_{16}\cdot\left(TAF-\overline{TAF}\right)+\theta_{17}\cdot\left(IBSL-\overline{IBSL}\right)+\eta_4\right)
\end{aligned}\qquad (5.6)$$

The optimal parameters values for the base and the final covariate model are reported in Table 5.6, along with their confidence intervals, and scatterplots of the individual parameter values used for the regression vs. the selected covariates are shown in Figure 5.5. Our analysis detects a negative correlation between log(SI) and AGE, VAF and IBSL, and similarly log(P2) is found to be negatively correlated with AGE, TAF and IBSL. All these relationship are apparent also from the scatterplots in Figure 5.5, and indeed the population variability decreases remarkably from 69.8% to 44.5% and from 51.3% to 39.5% for SI and P2 respectively. This means that the model has increased its predictive capability and is able to explain part of the inter-subject variability in a deterministic way. The drop in population variability for VOL is much more contained (from an initial 11.4% to 9.1%), but still statistically significant. A

negative correlation is detected with SEXM %TBF and GBSL, and a positive correlation with AGE; however, the relationships with %TBF and GBSL are apparent also in the scatterplots, whereas, interestingly, the correlation is not evident for AGE and SEX. Glucose effectiveness, instead, does not benefit much from the introduction of the covariates; in fact, the variability decreases only from 17.1% to 16.5%. Moreover, the inspection of the confidence intervals for the estimates of the variance terms reveals that this shrinking is not statistically significant.

The population covariance matrix used in our analysis included, besides the parameter variance terms, also the correlations SI-P2 and SG-VOL. Both these off-diagonal terms of the Ω are to be considered model parameters and both in the base and in the final model they result statistically significant, in that their estimation confidence interval does not include 0. Even though the results are not published for reasons of space, this holds true also for all the other models which were tested. The same can be said about the estimate of the RUV, which was not perturbed by the incorporation of the covariates. More in general, it can be pointed out that the precision of all the population parameters included in the base model (θ_1, θ_2, θ_3, θ_4 and the Ω terms) did not change significantly with the introduction of the covariates.

Besides model 38, which proved the best, there are others which, even if they did not result as winners, might be worth some considerations. These are models 44, 56 and 62, and in addition, if the confidence interval criterion is disregarded, also models 50, 55, 57, 65 and 67. They might indicate, in addition to the relationships included in the optimal model, also the influence of SEX on P2, VAF, TAF or BH on VOL, and the model BW-BMI-BSA for SG.

Discussion

It is interesting to analyze the discrepancies in the results provided by the nonlinear mixed-effects analysis and compare them with the initial explorative regression carried out on the individual estimates provided by the base model. The best model indicated by both Mallows' Cp and adjusted R^2 are larger than the ones finally selected by the population ML approach. Adjusted R^2 in particular was indicating as optimal the inclusion of 5 covariates per parameter, and Cp about 4. Apparently the penalization for the number of parameters employed in the model is more marked using the Chi-Squared test than the one applied by criteria like Cp or adjusted R^2. However, it is relevant to notice that the exploratory regression tests allowed narrowing down dramatically the number of necessary runs in SPK. The degree of accordance between the two regressions (the exploratory one in R and the ML in

SPK) has proved to be fairly satisfactory, and even if the exploratory analysis did not recognize the correct order and therefore did not single out the optimal model, it still provided a very good guess, and the model selected as final for each of the glucose minimal model parameters resulted, in the exploratory analysis, the best candidate among the ones with the same number of predictors. Our criterion allows to dramatically reduce the duration and improve the efficacy of the covariate selection process, in that not only the number of nonlinear mixed-effects runs is remarkably decreased with respect to a traditional stepwise approach, but also, in contrast to this methodology, it permits the evaluation of models in which two or more covariates are jointly inserted into the model, a possibility which is normally not explored with a tout-court stepwise approach [78]. Moreover, the choice of calculating the Chi-Squared statistics always with respect to the base model grants the modeler the possibility to carry out a comparison among all the models created in the various step of the analysis, without having to choose a new base for the comparison at each step.

In addition, it is interesting to observe how the criterion based on the p-value obtained by Chi-Squared test on the NLL, is always in concordance with the statistical significance of the regression parameter estimates. In other words, the models selected in the steps of our analysis were always characterized by regression parameters significantly different from 0. Moreover, if the fit of a model of a larger order (with more covariates) was attempted, most of the times the results were characterized by some of the parameter's confidence intervals including 0. This offers support to the criterion we have used for the covariate selection and corroborates the results about the precision estimates provided by SPK, which, because of the fact that they are obtained with a minimization algorithm, are based on the linearization of the Objective Function around the minimum.

Now let us take a look more in specific at the optimal model. Our dataset is composed only of healthy patients (meaning they did not present or where ever diagnosed before glucose metabolic disorders) and, even if a large span of age and BMI is covered, the variability between the subjects is limited and so might the capability of inference. However, some results are definitely very interesting.

The covariates included in the final model as ***predictors of SI*** do not come as a surprise; they have been reported previously in the literature and have a very convincing physiological explanation. It is widely known that insulin sensitivity decreases with ageing, and high level of basal insulin are normally used as a marker for insulin resistance, and incipient impaired glucose tolerance. Already in a pioneering publication by Bergman et al. [12], in Figure 5.6 is evident an inverse relationship among basal insulin and the logarithm of SI. And also Godsland and

colleagues [35] had already detected a similar correlation, even thought, in this case, the log-transformation was not employed. In addition, many studies in the literature have pointed out at the correlation between VAF and insulin resistance [34, 52, 76]. Basu and coworkers have already recognized the role of VAF when performing a regression analysis on part of the same dataset [4, 5]. Besides SI, a very analogous ***model*** was ***selected for p2***, measuring the inverse of the delay in the insulin action with respect to plasma insulin; in fact, with the only difference of TAF which substitutes VAF, the behavior of the predictors is at all similar. It is not clear if it is indeed the same predictors explaining the two different phenomena (namely the insulin sensitivity and its latency in turning into active insulin) or a limitation of the model which does not manage to separate these two aspects of the insulin metabolic system. In any case, the term in the model expressing the correlation of the random effects of SI and p2 does not decrease in value or statistical significance with the introduction of the covariates, meaning that also the residual inter-individual variability of the two parameters is highly correlated. It should be mentioned that the use of TAF instead of VAF as a predictor for P2 might not carry a strong physiological meaning, but may simply be caused by the very strong collinearity (r=0.874) between these two predictors. As it appears also from the selection process in Table 5.5, very often TAF is a good substitute for VAF and vice versa, so further validation on other datasets would be useful to investigate if the relationships we have found are corroborated or if our findings are just to be considered a proof of the correlation between insulin metabolism and abdominal fat, regardless of its specific distribution. In any case, all the relationships encompassed in the final model for SI and P2 are apparent also from the scatterplots with the individual values from the base model displayed in Figure 5.5. The same applies to a certain extent to the ***relationships found for VOL***; the inverse correlation is evident in the scatterplots in Figure 5.5 for PERTF and GBSL, whereas the relationship with SEX and AGE is not encountered graphically. This probably means that these correlations are present only after the volume has been corrected for %TBF and GBSL. We have no knowledge of studies in the literature regarding regression analysis of the apparent volume of distribution for the glucose minimal model, so no comparison with other data is possible. However, it should be borne in mind that that the apparent volume of distribution provided by the minimal model parameterization, and therefore used for the regression analysis, is intended per kg of the subject's body mass. If this is taken into account, the negative correlation with %TBF might find an explanation in the fact that the volume is proportional to the lean body mass of the subject, and

therefore, subjects with a greater percentage of body fat have a relatively lower volume per kg of body mass. In formulae

$$V \propto LBM \rightarrow \frac{V}{BW} \propto \frac{LBM}{BW} = \frac{BW\text{-}TBF}{BW} = 1 - \frac{TBF}{BW} = 1 - \frac{\%TBF}{100} \qquad (5.7)$$

Slightly different comments are due about the ***model proposed for SG***. From the scatterplots, no remarkable relationship is graphically detectable between log(SG) and the selected covariates separately. The correlation, in fact, is present only when the whole set BH-BW-BSA is used for the regression; if a model with fewer predictors is selected, AGE or GBSL (both inversely correlated to log(SG)) result more significant, whereas the explanatory power of BH, BW and BSA seem to be inconsistent when they are considered separately. This might point towards a more complicated nonlinear relationship between SG and these covariates, but it should be also considered that the covariates themselves are strongly correlated. In particular, BSA is obtained as a function, although non-linear, of BH and BW. Bonate has warned about the danger of collinearity in nonlinear mixed-effects models [14], reporting biased and very low precision estimates of the correlation coefficients. However, in our case, the estimates are affected on average by a 30% CV, which is not much larger than the uncertainty affecting the coefficients of the other covariates in the model. In addition, all the other lower order alternative models tested for SG (models 59-60-61) were affected by even greater problems of statistical significance, in that, not only the overall p-value was less satisfactory, but the regression parameters were estimated with very low precision and did not result significantly different from 0. Another candidate model for SG is BW-BMI-BSA, but, on the one hand, it provides a slightly worse p-value, and, on the other, it is prone to the same, if not stronger, collinearity issue. In any case, it should be borne in mind that SG is affected by model simplification from a two-compartmental into a one-compartment glucose kinetics [17, 18], and this can have a confounding effect that hinders the proper identification of physiologically plausible predictors.

Conclusions

In this work we propose a population model for the glucose minimal model, incorporating physiological information such as sex and age, easily measurable anthropometric data such as height and weight, body fat amount and distribution, basal levels of glucose and insulin. It is important to mention that many studies encompassing a correlation analysis of glucose-insulin model parameters are present in the literature, but we are not aware of any other study actually implementing a "dynamic" regression model such as the one we are proposing in this work. Our

model does not simply verify correlations on already calculated parameter values; instead the regression coefficients are integrated in the model and in this way physiological information plays an active role in the parameter estimation process. The introduction of the covariates in the model, in fact, helps in explaining a substantial fraction of the population variability for the parameters SI and p2, even though the inter-subject variability is not too large in the dataset under test, which is encompassing only healthy subjects, rather than individuals with different degrees of glucose intolerance disorders. The predictive power of the model is then considerably increased with the incorporation of easily, inexpensively and non-invasively collectible physiological information; a significant portion of the inter-individual variability is explained in a deterministic fashion, and only a smaller part involves a stochastic component, the random effects. This can not only offer a starting point for speculation about the significance of the relationships detected, but can also provide a tool to allow the design of less invasive and expensive protocols for epidemiological studies of the glucose disposal metabolic system. The advantages of a population approach to parameter estimates has been shown extensively in the literature, and even so far it has been mostly employed in drug development studies of pharmacokinetics and pharmacodynamics, the modeling of metabolic systems can greatly benefit from the use of these techniques. Nonlinear mixed-effects modeling, in particular, allows not only a more accurate evaluation of the population features, but also enhances the individual parameter accuracy and precision, in that the population information is used as a prior which supports the individual parameter estimation process. In addition, thanks to the introduction of the individual values of the significant covariates, this population prior is tailored to each one of the subjects, further improving the estimate precision. Our covariate model, being a first proposal, suffers from several limitations and gives therefore rise to possible speculation. The existence of more complex and possibly nonlinear relationships should be investigated, as our analysis was limited simply at linear models. Moreover, different parameterization could be tested to see, for example, if aggregating the parameters (e.g. using SG*VOL or SI*VOL) allows to unveil additional or stronger relationships. Furthermore, additional research is necessary to clarify the relevance of the covariates detected for SG, establishing whether the present results are driven by covariate collinearity and, in this case, investigate to propose a more satisfactory solution, possibly employing more physiologically meaningful predictors, such as age or basal glycemia, which our analysis indicates as potential candidates. Finally, since the amount of variability in the data is limited and therefore the possibility of statistical artifacts and model misspecification can not be excluded, it would be important to

validate and corroborate our results by repeating the analysis on different datasets, possibly encompassing subjects with a broader range of variation concerning glucose disposal. Another idea to assess the robustness and statistical significance of the results is the use of a bootstrap analysis, to test whether the covariates identified with the full dataset are identifiable also in replicate sub-datasets.

Table 5.1 Covariates for the glucose-insulin system measured in our 204 subject database. Statistics include: minimum and maximum value, 1st and 3rd quartile, mean and median. Covariate 1 is SEX (Gender).

ID	Covariate	Name	Units	Min	1stQ	Median	Mean	3rdQ	Max
2	AGE	Age	years	18	27	65	55.53	71	87
3	BH	Body height	cm	145	163	171	170.9	178	194
4	BW	Body weight	kg	53	68.9	79	77.94	87	127
5	BMI	Body mass index	kg/m^2	19.6	24.23	26.76	26.61	29.06	34.85
6	BSA	Body surface area	m^2	1.505	1.771	1.937	1.917	2.047	2.596
7	LBM	Lean body mass	kg	30.1	38.5	51.84	49.53	58.68	74.58
8	VAF	Visceral abdominal fat	cm^2/CT slice	11.86	62.62	127.5	141.8	204.7	478.2
9	TAF	Total abdominal fat	cm^2/CT slice	43.94	195.1	294.5	301.8	404.4	837.5
10	TBF	Total body fat	grams	4884	17370	22570	23410	28420	46990
11	%TBF	Percent total body fat	%	7.3	25.85	31.55	32.39	39.68	56.7
12	GBSL	Basal glucose (fasting, pre-dose)	mg/dL	72.96	86.74	90.31	91.34	94.72	123.8
13	IBSL	Basal insulin (fasting, pre-dose)	pmol/mL	5.4	18.71	23.85	27.25	32.29	80.25

Table 5.2 Minimal model parameters for the glucose-insulin system estimated in our 204 subject database using BM-1. Note that all parameters have been log-transformed to comply with the normality assumption on the random effects required by parametric nonlinear mixed effects modeling.

Parameter	Name	Units	Min.	1stQ	Median	Mean	3rdQ	Max.
log(SG)	Glucose effectiveness	log(min^{-1})	-4.263	-4.041	-3.984	-3.979	-3.921	-3.624
log(VOL)	Glucose volume of distribution	log(dL/kg)	0.2224	0.4028	0.4795	0.4742	0.5496	0.8124
log(SI)	Insulin sensitivity	log(min^{-1} per pmol/mL)	-11.16	-10.28	-9.772	-9.767	-9.208	-8.225
log(p2)	Insulin action parameter	log(min^{-1})	-4.576	-3.806	-3.441	-3.498	-3.136	-2.477

Table 5.3 Table containing the best candidate models selected by Mallows' Cp. The columns indicate the number of covariates included in the model (order of the model), whereas the rows contain the minimal model parameters. For each parameter and for each order, the model indicated as "best" by Mallows' Cp is reported. The overall (regardless of the order) "best" model for each parameter is indicated in boldface. A * indicates that this criterion seemed to exclude a model of that particular order.

# of Covariates	2	3	4	5	6
log(SG)	AGE, GBSL	BH, BW, BSA	**BH, BW, BSA, GBSL**	BH, BW, BSA, LBM, GBSL	*
log(VOL)	*	*	SEX, AGE, %TBF, GBSL	**SEX, AGE, VAF, %TBF, GBSL**	SEX, AGE, BH, VAF, %TBF, GBSL
log(SI)	*	AGE, VAF, IBSL	**SEX, AGE, VAF, IBSL**	SEX, AGE, BMI, VAF, IBSL	*
log(P2)	*	AGE, TAF, IBSL	**AGE, VAF, %TBF, IBSL**	SEX, AGE, BSA, VAF, IBSL	*

Table 5.4 Values of R-squared and adjusted R-squared for the candidate models in the range selected by Mallows' Cp, listed by parameter. The "best" model based on exploratory analysis is suggested to be the one with the highest adjusted R-squared (boldface).

Log(SG) (best covariates based on adjusted R-squared: BH, BW, BSA, LBM, GBSL)

Covariates	R^2	adjR^2
GBSL	0.03544	0.03066
AGE	0.03371	0.02892
AGE GBSL	0.05097	0.04153
BH BW BSA	0.06960	0.05564
BW BMI BSA	0.05821	0.04409
BH BW BSA GBSL	0.08197	0.06352
BH BW BSA LBM GBSL	**0.08952**	**0.06653**

Log(VOL) (best covariates based on adjusted R-squared: SEX, AGE, VAF, %TBF, GBSL)

Covariates	R^2	adjR^2
AGE TBF GBSL	0.3995	0.3905
SEX AGE %TBF GBSL	0.4264	0.4149
AGE VAF %TBF GBSL	0.4196	0.4080
SEX AGE VAF %TBF GBSL	**0.4365**	**0.4223**
AGE BH VAF %TBF GBSL	0.4329	0.4186
SEX AGE TAF %TBF GBSL	0.4324	0.4180
SEX AGE BH VAF %TBF GBSL	0.4382	0.4211

Log(SI) (best covariates based on adjusted R-squared: SEX, AGE, BMI, VAF, IBSL)

Covariates	R^2	adjR^2
TAF	0.3761	0.373
VAF	0.3418	0.3385
VAF IBSL	0.4836	0.4784
TAF IBSL	0.4779	0.4727
AGE IBSL	0.4734	0.4682
AGE VAF IBSL	0.5265	0.5193
AGE TAF IBSL	0.5186	0.5113
VAF %TBF IBSL	0.5123	0.5050
SEX AGE VAF IBSL	0.5376	0.5283
AGE VAF %TBF IBSL	0.5369	0.5276
SEX AGE BMI VAF IBSL	**0.5414**	**0.5299**

Log(P2) (best covariates based on adjusted R-squared: SEX, AGE, BSA, VAF, IBSL)

Covariates	R^2	adjR^2
TAF	0.3358	0.3325
AGE IBSL	0.4155	0.4097
TAF IBSL	0.3880	0.3819
AGE TAF IBSL	0.4512	0.4429
AGE VAF IBSL	0.4431	0.4347
AGE VAF %TBF IBSL	0.4636	0.4528
SEX AGE VAF IBSL	0.4625	0.4517
SEX AGE TAF IBSL	0.4583	0.4474
AGE VAF TBF IBSL	0.4574	0.4465
AGE TAF %TBF IBSL	0.4565	0.4456
SEX AGE BSA VAF IBSL	**0.4674**	**0.4539**
SEX AGE BW VAF IBSL	0.4667	0.4533

Table 5.5 Outline of the covariate selection process. The first column groups the models under test according to the minimal model parameter which was analyzed fro covariate inclusion at the current step of the selection process. The second column contains the code of the model: a * indicates that several of the regression parameters in the model did not result statistically different from zero according to their estimate's confidence intervals, whereas a # indicates that the performance of the model was so unsatisfactory to lead to its exclusion from being retested in the following steps. Columns 3 to 6 contain the codes of the covariate included for each of the parameters. NLL is the value of the Negative Log-Likelihood provided by SPK, whereas the number of θs in the model (typical values + regression coefficients) is reported in the following column. ΔNLL and ΔP are the difference in NLL value and number of parameters, with respect to the base model. Finally the last column displays the p-value as reported by the Chi-Squared test, in other words, it is the probability of committing an error if rejecting the null hypothesis that all the additional parameters added to the base model are insignificant.

Parameter under test	Model	SG	VOL	SI	P2	NLL	θs	□NLL	□P	p-value
BASE	0	0	0	0	0	12862.2	4			
VOL	1	0	AGE-TBF-GBSL	0	0	12791.9	7	-70.3	3	2.81E-30
	2	0	SEX-AGE-%TBF-GBSL	0	0	12786.2	8	-76.0	4	7.59E-32
	3#	0	AGE-VAF-%TBF-GBSL	0	0	12791.5	8	-70.7	4	1.42E-29
	4*	0	SEX-AGE-VAF*-%TBF-GBSL	0	0	12785.2	9	-77.0	5	1.88E-31
	5	0	AGE-BH-VAF-%TBF-GBSL	0	0	12785.9	9	-76.3	5	3.73E-31
	6*	0	SEX-AGE-TAF*-%TBF-GBSL	0	0	12785.5	9	-76.7	5	2.52E-31
	7*	0	SEX*-AGE-BH*-VAF*-%TBF-GBSL	0	0	12784.1	10	-78.1	6	3.78E-31
SG	8	GBSL	SEX-AGE-%TBF-GBSL	0	0	12783.8	9	-78.4	5	4.76E-32
	9	AGE	SEX-AGE-%TBF-GBSL	0	0	12783.4	9	-78.8	5	3.21E-32
	10*	AGE*-GBSL*	SEX-AGE-%TBF-GBSL	0	0	12782.3	10	-79.9	6	6.53E-32
	11	BH-BW-BSA	SEX-AGE-%TBF-GBSL	0	0	12778.0	11	-84.2	7	5.46E-33
	12	BW-BMI-BSA	SEX-AGE-%TBF-GBSL	0	0	12780.0	11	-82.2	7	3.80E-32
	13*	BH-BW-BSA-GBSL*	SEX-AGE-%TBF-GBSL	0	0	12777.3	12	-84.9	8	1.42E-32
	14*	BH-BW-BSA-LBM*-GBSL*	SEX-AGE-%TBF-GBSL	0	0	12776.2	13	-86.0	9	2.36E-32

SI	15#	BH-BW-BSA	SEX-AGE-%TBF-GBSL	VAF	0	12747.8	12	-114.4	8	5.31E-45
	16#	BH-BW-BSA	SEX-AGE-%TBF-GBSL	TAF	0	12752.8	12	-109.4	8	6.90E-43
	17	BH-BW-BSA	SEX-AGE-%TBF-GBSL	VAF-IBSL	0	12717.8	13	-144.4	9	6.18E-57
	18#	BH-BW-BSA	SEX-AGE-%TBF-GBSL	TAF-IBSL	0	12726.1	13	-136.1	9	2.03E-53
	19#	BH-BW-BSA	SEX-AGE-%TBF-GBSL	AGE-IBSL	0	12731.0	13	-131.2	9	2.40E-51
	20*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	AGE*-VAF-IBSL	0	12717.2	14	-145.0	10	2.02E-56
	21#	BH-BW-BSA	SEX-AGE-%TBF-GBSL	AGE-TAF-IBSL	0	12724.3	14	-137.9	10	2.00E-53
	22*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	VAF-%TBF*-IBSL	0	12717.8	14	-144.4	10	3.61E-56
	23*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	AGE*-VAF-%TBF*-IBSL	0	12717.2	15	-145.0	11	1.12E-55
	24*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	SEX*-AGE*-VAF-IBSL	0	12717.2	15	-145.0	11	1.12E-55
	25*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	SEX*-AGE*-BMI*-VAF-IBSL	0	12716.7	16	-145.5	12	3.63E-55
P2	26	BH-BW-BSA	SEX-AGE-%TBF-GBSL	VAF-IBSL	TAF	12699.9	14	-162.3	10	9.68E-64
	27	BH-BW-BSA	SEX-AGE-%TBF-GBSL	VAF-IBSL	AGE-IBSL	12700.5	15	-161.7	11	1.02E-62
	28	BH-BW-BSA	SEX-AGE-%TBF-GBSL	VAF-IBSL	TAF-IBSL	12697.9	15	-164.3	11	8.11E-64
	29	BH-BW-BSA	SEX-AGE-%TBF-GBSL	VAF-IBSL	AGE-TAF-IBSL	12695.0	16	-167.2	12	2.73E-64
	30	BH-BW-BSA	SEX-AGE-%TBF-GBSL	VAF-IBSL	AGE-VAF-IBSL	12697.2	16	-165.0	12	2.31E-63
	31	BH-BW-BSA	SEX-AGE-%TBF-GBSL	VAF-IBSL	AGE-VAF-%TBF-IBSL	12695.3	17	-166.9	13	1.97E-63
	32	BH-BW-BSA	SEX-AGE-%TBF-GBSL	VAF-IBSL	SEX-AGE-VAF-IBSL	12695.4	17	-166.8	13	2.17E-63
	33*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	VAF-IBSL	AGE-VAF-TBF*-IBSL	12695.9	17	-166.3	13	3.52E-63
	34*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	VAF-IBSL	SEX*-AGE-TAF-IBSL	12694.5	17	-167.7	13	9.10E-64
	35*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	VAF-IBSL	AGE-TAF-%TBF*-IBSL	12694.8	17	-167.4	13	1.22E-63
	36*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	VAF-IBSL	SEX-AGE-BSA*-VAF-IBSL	12695.0	18	-167.2	14	7.65E-63
	37*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	VAF-IBSL	SEX-AGE-BW*-VAF-IBSL	12695.2	18	-167.0	14	9.28E-63
SI	38	BH-BW-BSA	SEX-AGE-%TBF-GBSL	AGE-VAF-IBSL	AGE-TAF-IBSL	12687.6	17	-174.6	13	1.14E-66
	39#	BH-BW-BSA	SEX-AGE-%TBF-GBSL	VAF-%TBF-IBSL	AGE-TAF-IBSL	12692.9	17	-169.3	13	1.93E-64
	40*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	AGE-VAF-%TBF*-IBSL	AGE-TAF-IBSL	12687.3	18	-174.9	14	4.53E-66
	41*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	SEX*-AGE-VAF-IBSL	AGE-TAF-IBSL	12687.4	18	-174.8	14	4.99E-66
	42*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	SEX*-AGE-BMI*-VAF-IBSL	AGE-TAF-IBSL	12686.1	19	-176.1	15	7.28E-66

P2	43*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	AGE*-VAF-IBSL	TAF	12698.6	15	-163.6	11	1.60E-63
	44	BH-BW-BSA	SEX-AGE-%TBF-GBSL	AGE-VAF-IBSL	AGE-IBSL	12690.0	16	-172.2	12	2.13E-66
	45*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	AGE*-VAF-IBSL	TAF-IBSL	12696.9	16	-165.3	12	1.72E-63
	46*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	AGE-VAF-IBSL	AGE-VAF*-IBSL	12689.5	17	-172.7	13	7.20E-66
	47*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	AGE-VAF-IBSL	AGE-VAF*-%TBF-IBSL	12687.5	18	-174.7	14	5.50E-66
	48*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	AGE-VAF-IBSL	SEX-AGE-VAF*-IBSL	12687.5	18	-174.7	14	5.50E-66
	49*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	AGE-VAF-IBSL	AGE-VAF*-TBF*-IBSL	12688.2	18	-174.0	14	1.08E-65
	50*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	AGE-VAF-IBSL	SEX*-AGE-TAF-IBSL	12686.4	18	-175.8	14	1.90E-66
	51*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	AGE-VAF-IBSL	AGE-TAF*-%TBF*-IBSL	12687.0	18	-175.2	14	3.39E-66
	52*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	AGE-VAF-IBSL	SEX-AGE-BSA*-VAF*-IBSL	12687.1	19	-175.1	15	1.91E-65
	53*	BH-BW-BSA	SEX-AGE-%TBF-GBSL	AGE-VAF-IBSL	SEX-AGE-BW*-VAF*-IBSL	12687.3	19	-174.9	15	2.31E-65
VOL	54#	BH-BW-BSA	AGE-TBF-GBSL	AGE-VAF-IBSL	AGE-TAF-IBSL	12694.5	16	-167.7	12	1.68E-64
	55*	BH-BW-BSA	SEX-AGE-VAF*-%TBF-GBSL	AGE-VAF-IBSL	AGE-TAF-IBSL	12686.1	18	-176.1	14	1.42E-66
	56	BH-BW-BSA	AGE-BH-VAF-%TBF-GBSL	AGE-VAF-IBSL	AGE-TAF-IBSL	12687.3	18	-174.9	14	4.53E-66
	57*	BH-BW-BSA	SEX-AGE-TAF*-%TBF-GBSL	AGE-VAF-IBSL	AGE-TAF-IBSL	12686.8	18	-175.4	14	2.80E-66
	58*	BH-BW-BSA	SEX*-AGE-BH*-VAF*-%TBF-GBSL	AGE-VAF-IBSL	AGE-TAF-IBSL	12685.8	19	-176.4	15	5.45E-66
SG	59*	GBSL*	SEX-AGE-%TBF-GBSL	AGE-VAF-IBSL	AGE-TAF-IBSL	12692.5	15	-169.7	11	4.23E-66
	60*	AGE*	SEX-AGE-%TBF-GBSL	AGE-VAF-IBSL	AGE-TAF-IBSL	12692.9	15	-169.3	11	6.25E-66
	61*	AGE*-GBSL*	SEX-AGE-%TBF-GBSL	AGE-VAF-IBSL	AGE-TAF-IBSL	12692.3	16	-169.9	12	1.99E-65
	62	BW-BMI-BSA	SEX-AGE-%TBF-GBSL	AGE-VAF-IBSL	AGE-TAF-IBSL	12688.6	17	-173.6	13	3.01E-66
	63*	BH-BW-BSA-GBSL*	SEX-AGE-%TBF-GBSL	AGE-VAF-IBSL	AGE-TAF-IBSL	12687.4	18	-174.8	14	4.99E-66
	64*	BH-BW-BSA-LBM*-GBSL*	SEX-AGE-%TBF-GBSL	AGE-VAF-IBSL	AGE-TAF-IBSL	12686.9	19	-175.3	15	1.57E-65
MIXED	65*	BH-BW-BSA	SEX-AGE-VAF*-%TBF-GBSL	AGE-VAF-IBSL	AGE-IBSL	12688.6	17	-173.6	13	3.01E-66
	66*	BH-BW-BSA	SEX-AGE-TAF*-%TBF-GBSL	AGE-VAF-IBSL	AGE-IBSL	12689.1	17	-173.1	13	4.89E-66
	67*	BH-BW-BSA	SEX-AGE-VAF*-%TBF-GBSL	AGE-VAF-IBSL	SEX*-AGE-TAF-IBSL	12685.0	19	-177.2	15	2.52E-66
	68*	BW-BMI-BSA	SEX-AGE-%TBF-GBSL	AGE-VAF-IBSL	SEX*-AGE-TAF-IBSL	12687.3	18	-174.9	14	4.53E-66
	69*	BH-BW-BSA	SEX-AGE-TAF*-%TBF-GBSL	AGE-VAF-IBSL	SEX*-AGE-TAF-IBSL	12685.6	19	-176.6	15	4.50E-66
	70*	BW-BMI-BSA	SEX-AGE-VAF-%TBF-GBSL	AGE-VAF-IBSL	SEX*-AGE-TAF-IBSL	12685.4	19	-176.8	15	3.71E-66

Table 5.6 Summary of nonlinear mixed-effects maximum likelihood regression for the base and the final covariate models for the minimal model of glucose kinetics. Values are reported as follows. Typical values for the parameters (θ) are in the original units. Given that the between-subject variability is modeled as lognormal, variance measures are reported as coefficients of variation, whereas the covariance terms (the off-diagonal element of Ω) are in terms of correlation. Coefficients for the covariates are in logarithmic units; therefore, as a first approximation, they can be interpreted as proportional changes in the minimal model parameter per unit change of the covariate. By way of explanation, SI = 5.58E-5*exp(-0.00703*ΔAGE), where ΔAGE is the deviation of AGE from the mean AGE. This model can be approximated as SI ~ 5.58E-5*(1-0.00703*ΔAGE), which predicts approximately a 0.7% drop in SI for every year of AGE above mean AGE. The Residual Unknown Variability is reported as CV%,

Model	SG			VOL			SI			P2			Off-diagonal Ω		Σ	NLL
	θ		CV %	θ		CV %	θ		CV %	θ		CV %	Ω (SG-VOL)	Ω (SI-P2)	CV %	
Base	0.0191 (0.0184 0.0196)		17.1 (13.2 20.1)	1.60 (1.57 1.63)		11.4 (10.0 12.6)	5.58E-5 (5.01E-5 6.15E-5)		69.8 (63.3 75.7)	0.0298 (0.0271 0.0326)		51.3 (43.8 57.7)	-0.40 (-0.65 -0.14)	0.87 (0.69 1.00)	4.4 (4.2 4.6)	12862.2
Final	0.0186 (0.018 0.0192)		16.5 (12.7 19.6)	1.68 (1.64 1.73)		9.1 (7.9 10.1)	5.68E-5 (5.3E-5 6.07E-5)		44.5 (39.6 48.8)	0.0299 (0.0276 0.0323)		39.5 (32.1 45.7)	-0.67 (-0.96 -0.39)	0.85 (0.63 1.00)	4.4 (4.2 4.6)	12687.6
				SEX	-0.0813 (-0.12 -0.0425)											
	BH	0.0214 (0.00732 0.0354)		AGE	0.00169 (0.00113 0.00226)		AGE	-0.00703 (-0.0108 -0.00329)		AGE	-0.00814 (-0.0117 -0.00457)					
	BW	0.0446 (0.0168 0.0724)		%TBF	-0.00981 (-0.012 -0.0076)		VAF	-0.00235 (-0.00332 -0.00137)		TAF	-0.000656 (-0.00125 -0.0000637)					
	BSA	-3.78 (-6.07 -1.49)		GBSL	-0.00473 (-0.00643 -0.00303)		IBSL	-0.0276 (-0.0343 -0.0209)		IBSL	-0.0101 (-0.016 -0.00411)					

Figure 5.1 Histogram plots and smoothed densities for the covariate set (see also Table 5.1). See text for abbreviations and details.

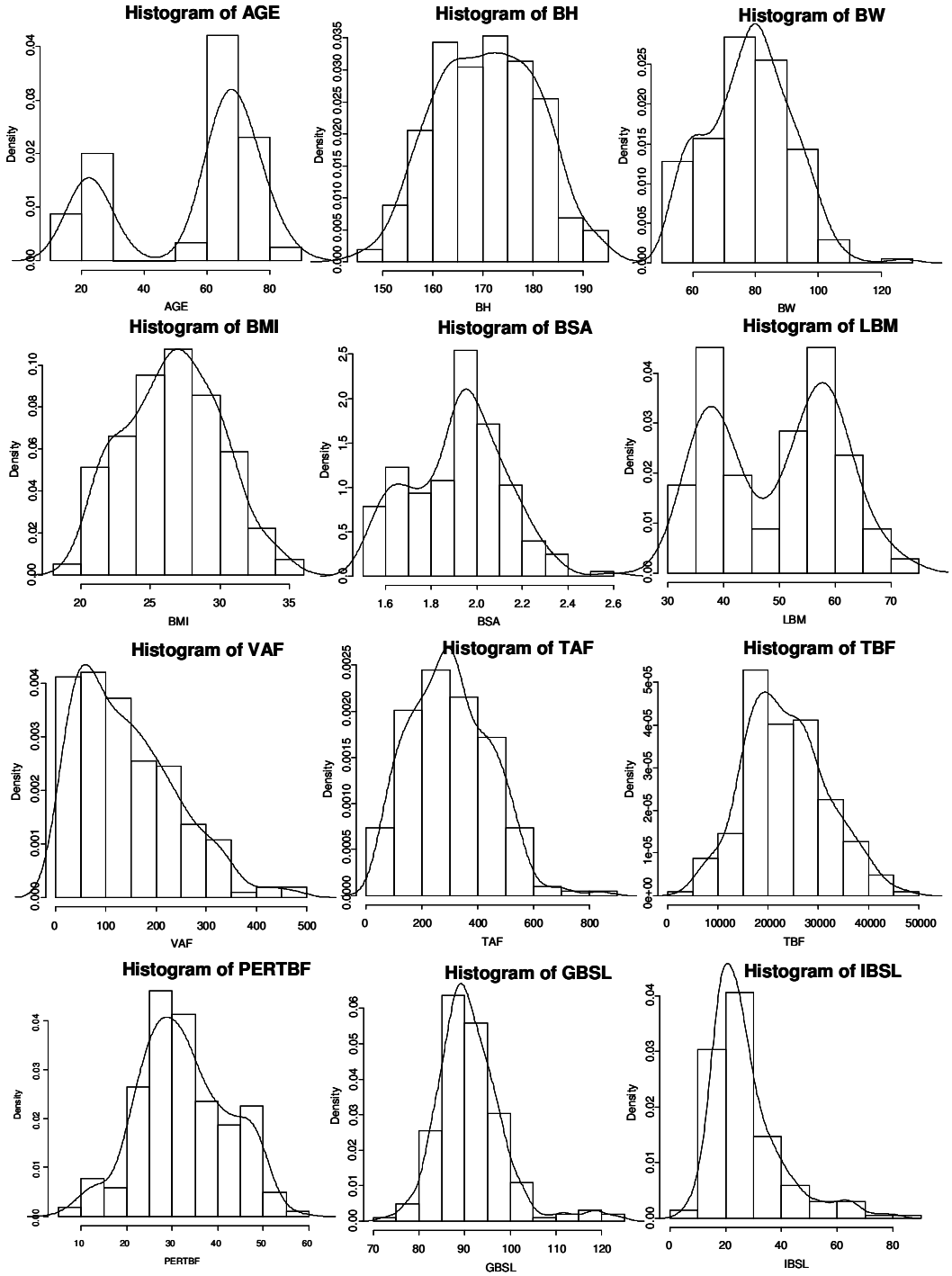


Figure 5.2 Histogram plots and smoothed densities for the minimal model parameter set (see also Table 5.2). See text for abbreviations and details.

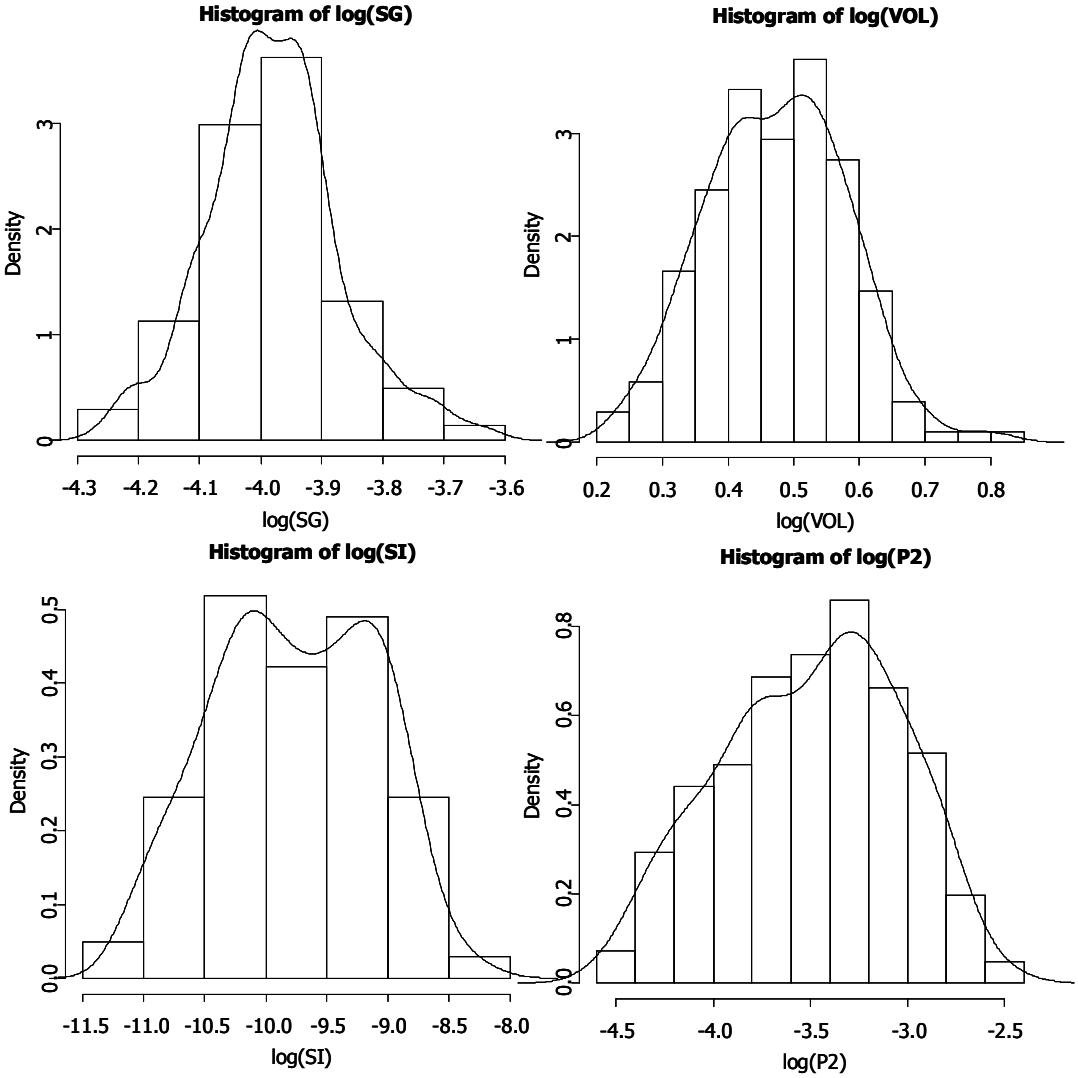


Figure 5.3 Scatterplot of the covariates most closely correlated. The anthropometric covariates (BH, BW, BMI and BSA) and the Body-Fat related covariates (TBF, %TBF, LBM, VAF and TAF) were grouped together as they are closely related or tied by some functional relation. A smoothed tendency line is superimposed to depict the trend of the relation.

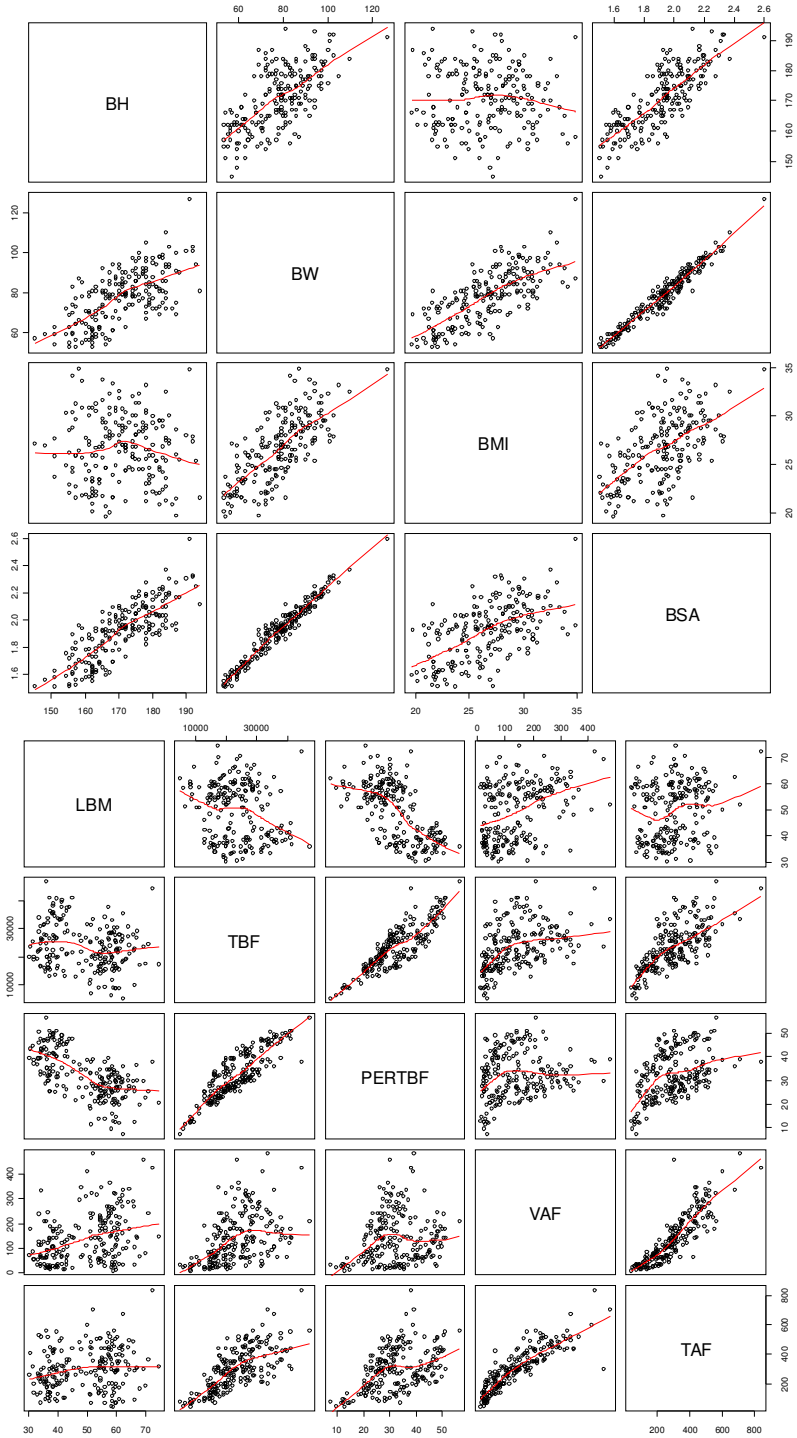


Figure 5.4 Scatterplot of the logarithm of the parameter values, as obtained with the base model. A very strong correlation is apparent between log(SI) and log(P2), as accounted for by the relative term in the covariance matrix Ω.

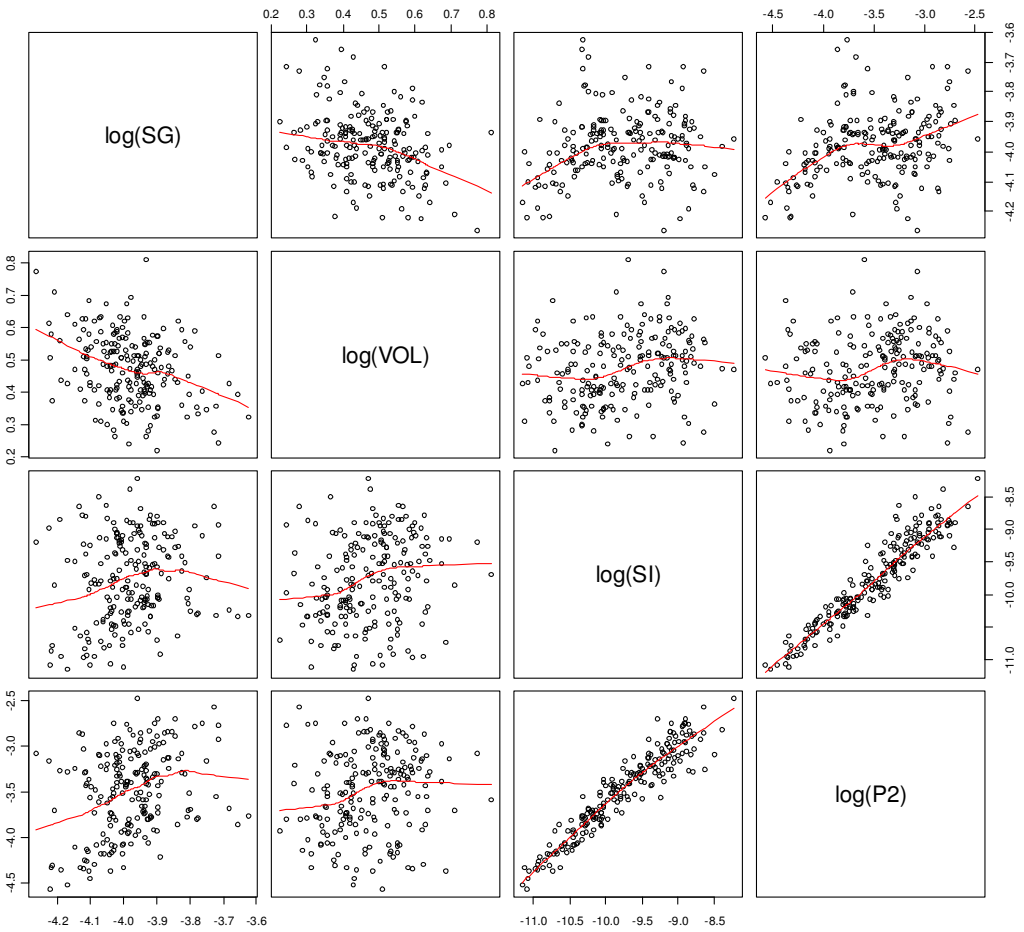
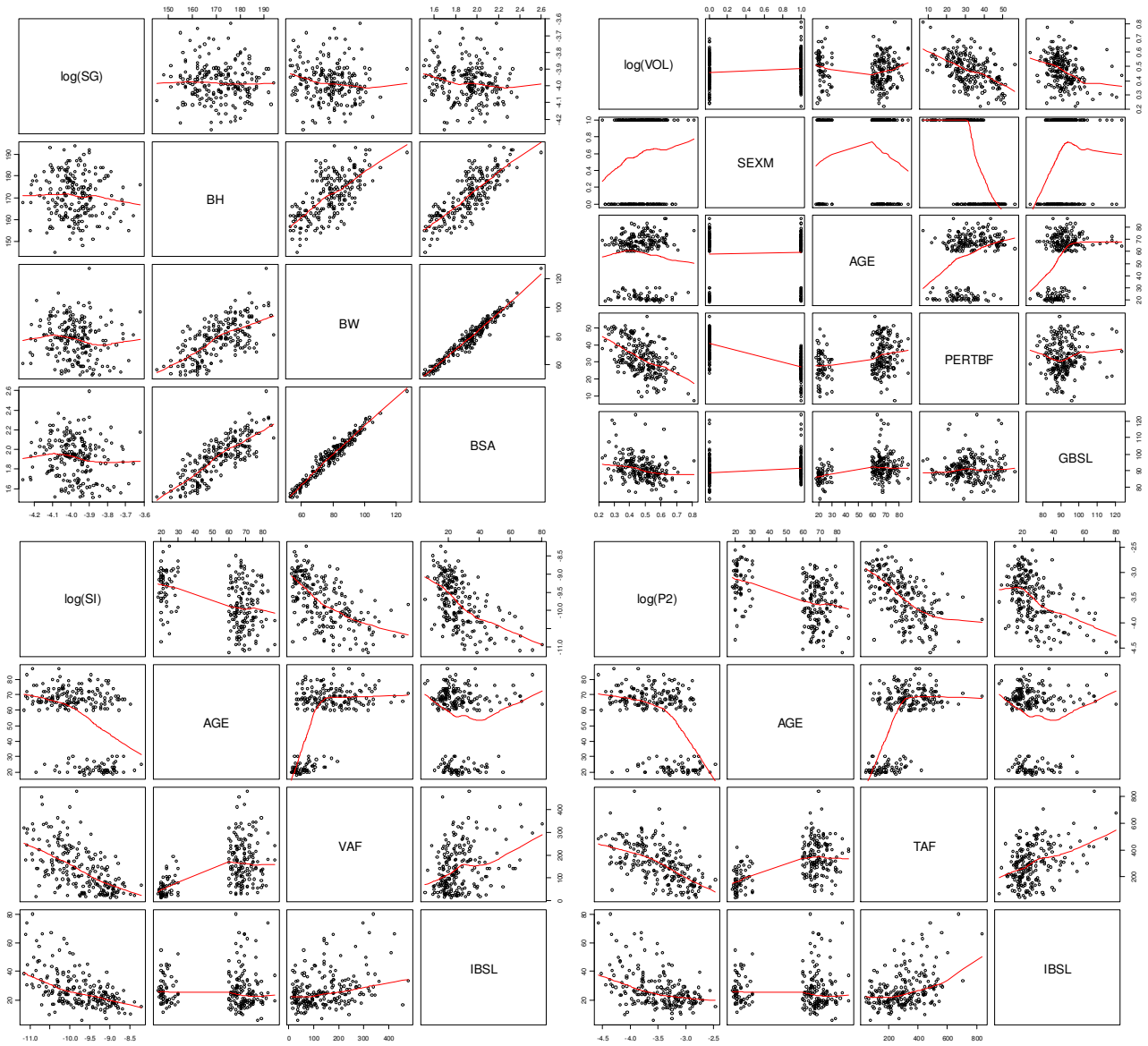


Figure 5.5 Scatterplot of the logarithm of the parameter values obtained with the base model and the respective covariates included in the final model.



Chapter 6

Assessment of Glucose Disposal Metabolic System. The Disposition Index: Total Least Squares VS Population Approaches

Abstract

It is widely recognized that in order to asses the actual performance of the glucose disposal metabolic system of a subject, it is important to interpret the data of insulin sensitivity together with information concerning insulin secretion. The disposition index, first proposed in Bergman et al. [12], is one of the most widely accepted methods to do so. This paradigm assumes that subjects with similar efficiency in disposing of glucose have a similar value of the product of insulin sensitivity and insulin secretion index. Due to the implied inverse relationship between the sensitivity and secretion indices, this paradigm is often referred to as the hyperbolic law. More recently the validity of this law has been questioned [37] and the hypothesis has been put forward that relationship is quasi-inverse, with the introduction of an additional parameter α as exponent of the sensitivity index. The method used so far to investigate the validity of the disposition index laws consists in analyzing a dataset from a population of subjects with similar glucose disposal efficiency and therefore supposedly sharing the same disposition index. For each individual, an insulin sensitivity and a secretion index are estimated, together with their precision, and a fit approach is normally applied to find the curve which best explains the data. In the literature many publications are found that attempt to address this problem, but the details of the fit algorithm implementation are often not clearly stated, if not overlooked. In the current investigation, we first show the crucial importance of the setup of the fitting algorithm and then propose a more statistically sound fit approach based on the Total Least Squares (TLS) technique. However, all the fit approaches rely on the hypothesis that all the subjects share the same value of the disposition index. In this way they fail to account for a very important factor which is inevitably intrinsic in biological data: the variability due to inter-individual differences. Thus, we propose a new approach to the disposition index estimation, which consists in obtaining its value based on population features estimated with a nonlinear mixed-effects model. After testing the performance of all the proposed methods on simulated datasets, we apply our newly designed tool to the analysis of

real data, to investigate the validity of the pseudo-hyperbolic law versus the traditional hyperbolic paradigm.

Introduction

An accurate and exhaustive evaluation of the efficiency of a subject's glucose metabolic system must take into account both information on insulin sensitivity and on insulin secretion (i.e. β-cell responsivity). These two characteristics must be interpreted jointly, because it is indeed the balance of the two that characterizes the actual efficiency of the system.

Even though other authors [69] had previously detected a similar relationship, it is widely acknowledged that the Disposition Index (hereinafter DI, Δ in the formulae), was first proposed in 1981 by Bergman and coworkers [12] and it is, to these days, one of the most widespread methods to account jointly for both insulin sensitivity (ξ) and β-cell responsivity (Φ). This paradigm assumes that subjects with similar efficiency of the glucose disposal metabolic system have a similar value of the product $\Delta=\xi\cdot\Phi$, where ξ is a sensitivity index (such as insulin sensitivity as provided by IVGTT glucose minimal model [11]) and Φ a β-cell responsivity index (such as Φ as provided by the C-peptide minimal model [67]). Due to the implied inverse proportionality among ξ and Φ, this is widely known as the hyperbolic law.

Recently a more comprehensive paradigm has been proposed, suggesting that the disposition index should be calculated as $\Delta=\xi^{\alpha}\cdot\Phi$ thus introducing an additional parameter, α. This further parameter is meant to accommodate for a different leverage of the secretion and sensitivity indices, with respect to the calculation of the DI. It is worth noting that, if the values of α is 1, this model would degenerate in the pure hyperbolic law. Hereinafter, we will refer to this newly proposed DI paradigm as pseudo-hyperbolic law.

In the original paper by Bergman and colleagues, the relationship between Φ_2 and SI was considered, but later many different indices (especially for insulin secretion) have been used for the same purpose. Many studies in the literature [13, 32, 37, 53, 70, 71] employ the DI, and address the problem of its determination in different ways, inspecting, for different combinations of sensitivity and secretion indices, also about whether the parameter α is significantly different from 1, thus providing evidence against the validity of the pure hyperbolic law to the advantage of the more complex pseudo-hyperbolic law. The most widely-used approach consists in trying to fit a hyperbola (or pseudo-hyperbola) on the data of insulin sensitivity and β-cell responsivity. However, the choice of the fit method used raises some issues and the

problem needs to be considered with the necessary care. As we show in the present work, the setup for the fit algorithm is crucial and what could be thought of as small details, and therefore possibly overlooked, might turn out to lead to extremely different outcomes.

In this work, we propose a novel nonlinear approach based on the Total Least Squares (TLS) technique, which will be outlined in the Materials and Methods section and we compare its performance on several simulated dataset, with that of the other methods previously used in the literature. Our algorithm proves more methodologically sound and our simulations show that the results provided are more reliable than the previously used methods. However, similarly to the all the other fit approaches, our algorithm implicitly relies on the hypothesis that all the subjects in the population have the same disposition index. So the only variability which is assumed to be present in the data is the estimation uncertainty of the parameters of insulin sensitivity and β-cell responsivity. This is might be a very strong simplification, and lead to very unreliable results.

Therefore we propose a new approach that considers the Disposition Index and the parameter α as features that characterize the joint population distribution of the parameters of insulin sensitivity and β-cell responsivity. The population analysis can be carried out with nonlinear mixed-effects modeling tools such as NONMEM [8] or SPK [66], and the values of DI and α can be obtained directly from the analysis of the population covariance matrix.

Finally, after testing also this new population approach on simulated data, we apply it on real data [4, 5] to test the validity of the hyperbolic law.

Background

In this section, we present a brief overview of the problem, introducing the solution we propose and a collection of the alternative approaches suggested and used so far in the literature to render the problem more easily tractable. The main issues raised by each of the proposed simplifications will be underlined to clarify the reasons which lead us to design our own approaches.

The standard scenario in which the analysis is normally carried out is the following: a population of subjects whose glucose metabolic system operates with the same degree of efficiency is considered and, for each one of the subjects, estimates of both insulin sensitivity and secretion indices (ξ_i and Φ_i), along with their precision (the variances are denoted $\sigma^2_{\xi i}$ and $\sigma^2_{\Phi i}$) are calculated. Therefore, assuming that the error

affecting the estimates is Gaussian, the estimates of the indices are characterized by the following probability distributions

$$\begin{aligned} \bar{\xi}_i &\sim N\left(\xi_i, \sigma^2_{\xi i}\right) \\ \bar{\Phi}_i &\sim N\left(\Phi_i, \sigma^2_{\Phi i}\right) \end{aligned} \tag{6.1}$$

where the barred values denote the true but unknown values of the indices. At the same time, in order to adhere to the DI paradigm, the true values are assumed to be obeying to the formula

$$\bar{\xi}_i^{\alpha} \cdot \bar{\Phi}_i = \Delta_i = \Delta \tag{6.2}$$

where *i* runs across the subjects and Δ is the DI characterizing the population. It is important to note that (6.2) emphasizes the assumption that Δ_i has the same value for all subjects. This hypothesis is possibly a strong simplification, as it appears quite unlikely that all subjects share exactly the same value of DI, however, this assumption is implicitly underlying all the fit approaches proposed in the literature. In this work, we first analyze the problem relying on assumption (6.2) to be true, and then we probe the consequences caused by the fallacy of this hypothesis (i.e., what if there is population variability in the DI values?).

Geometrical approaches (fit)

Total Least Squares Approach

If we interpret ξ_i and Φ_i as the coordinates x and y of a Cartesian coordinate system, a geometrical interpretation of the problem is possible. In a Cartesian plane, each subject is represented by a point and thus, the reason behind the name of the hyperbolic law is explained: (6.2) implies that all the points corresponding to the subjects' true parameter values lie on a curve, a pseudo-hyperbola. Therefore, finding the values of α and Δ characterizing the population, consists in identifying the curve that better fits the data; i.e., the curve lying at the minimum distance from the data points. The most common procedure to solve this kind of problem is tuning the parameters to minimize the sum of squares of the (weighted) residuals. For each point $\left(\xi_i, \Phi_i\right)$, the residual is defined as its distance to the closest point lying on the pseudo-hyperbola. This point, which in a sense can be considered as the projection on the curve, will be denoted $\left(\hat{\xi}_i, \hat{\Phi}_i\right)$. With this notation, the objective function that is minimized in the TLS fit approach can be written as follows:

$$F_{TLS} = \sum_i \frac{\left(\xi_i - \hat{\xi}_i(\alpha,\Delta)\right)^2}{\sigma^2_{\xi_i}} + \sum_i \frac{\left(\Phi_i - \hat{\Phi}_i(\alpha,\Delta)\right)^2}{\sigma^2_{\Phi_i}} \tag{6.3}$$

As it can be seen in (6.3), the difference of uncertainty along the x- and y-direction is properly taken into account by using the estimation uncertainties as weighting factors. The approach of using in the fit all the directions affected by error is generally known as Total Least Squares (TLS) or Error-in-variables, and efficient methods are available for the TLS fit of a straight line. However, in the case of a pseudo-hyperbola, the calculation of the closest point is not trivial to achieve. We were not able to identify a closed-form solution to the problem, so we had to resort to a numerical minimization procedure. We implemented the fit by nesting, inside the calculation of objective function, a further level of minimization to compute the projections of each data point on the pseudo-hyperbola. The Matlab [64] code is available upon request to the authors.

In order to approximate the distance from each data point to the curve, many different simplifications have been proposed in the literature. We briefly go through them in the next paragraphs.

1-variable fit

An easy way to bypass the problem consists in considering the error as affecting only one variable, so either the secretion or the sensitivity index. This naïve procedure raises several important issues: first of all, it is not clear upon which theoretical grounds one should choose to interpret one of the two indices as error-free and, secondly, which of the two variables should be chosen and why. The resulting objective functions are listed below

$$F_X = \sum_i \frac{\left(\xi_i - \sqrt[\alpha]{\frac{\Delta}{\Phi_i}}\right)^2}{\sigma_{\xi_i}^2} \tag{6.4}$$

$$F_Y = \sum_i \frac{\left(\Phi_i - \frac{\Delta}{\xi_i^\alpha}\right)^2}{\sigma_{\Phi_i}^2} \tag{6.5}$$

Log-transformation

The most widely used approach, inaugurated by Kahn et al. [37], consists in applying a log-transformation to the original data, so that, in the log-transformed space, the problem is simplified and becomes a linear fit. The log-transformation of both variables, ξ and Φ, turns the hyperbola into a straight line as shown in the following formulae:

$$\Delta = \xi_i^{\alpha} \cdot \Phi_i \xrightarrow{\log - transformation} \ln(\Delta) = \alpha \cdot \ln(\xi_i) + \ln(\Phi_i) \tag{6.6}$$

However, the transformation is nonlinear and does not preserve the Gaussian error structure of the model; in this way, the hypotheses supporting the classical fit approach are not satisfied. Therefore, to try to preserve information on the precision of the estimates of the individual indices, a first-order approximation can be used. This is equivalent to using the numerical values of the Coefficient of Variation (CV) of the original estimates as the Standard Error (SE) of the log-estimates, but this simplification is quite strong, especially if the uncertainty is high, as in this case the non-approximated probability distribution of the log-transformed data is heavily skewed to the right. On top of this, one should bear in mind that the optimal solution in the log-transformed space does not in general preserve its optimality under nonlinear transformations. Therefore, especially if a large variability is present, the two solutions might be very different. In addition, even if the log-transformation simplifies the problem to the fit of a straight line, the error is still present in both variables and a proper technique must be employed to take this into account. Many approaches are used in literature, but the most widespread is the Riggs Perpendicular Weighted (PW) regression [59]. This method is essentially an orthogonal fit that uses a correction factor λ (ratio of the variances) to correct for the difference of uncertainty along the two different directions. So technically a fit along just one direction is performed, but the correction factor λ, allows somehow to take into account also the second direction. Even though a linear method is available, so no minimization of the objective function is required, we propose the formula here for the sake of comparison with the other methods

$$F_{\log RIGGS} = \sum_i \left(\lambda \cdot \left(\ln \xi_i - \frac{\ln \Delta - \ln \Phi_i^{\perp *}}{\alpha} \right)^2 + \left(\ln \Phi_i - \ln \Delta + \alpha \cdot \ln \xi_i^{\perp *} \right)^2 \right) \tag{6.7}$$

The symbol $^{\perp *}$ denotes the orthogonal projection on the straight line, once the axes have been rescaled to accommodate for the heteroscedasticity along the different directions. However, a strong simplification introduced by this method is represented by the fact that it assumes each direction to be affected by a specific degree of uncertainty, and the same for all the subjects. In this way, all the data points will be considered equally reliable in the fit. In our case, as the uncertainty changed from direction to direction and from subject to subject, the following formula was used for λ

$$\lambda = median\left(\frac{\sigma^2_{\ln(\Phi_i)}}{\sigma^2_{\ln(\xi_i)}} \right) = median\left(\frac{CV^2_{\Phi_i}}{CV^2_{\xi_i}} \right) \tag{6.8}$$

An alternative simplistic approach to execute the fit for the straight line is once again considering the error in just one of the log-variables:

$$F_{\log X} = \sum_i \frac{\left(\ln \xi_i - \frac{\ln \Delta - \ln \Phi_i}{\alpha} \right)^2}{\sigma^2_{\ln \xi_i}} \tag{6.9}$$

$$F_{\log Y} = \sum_i \frac{\left(\ln \Phi_i - \ln \Delta + \alpha \cdot \ln \xi_i \right)^2}{\sigma^2_{\ln \Phi_i}} \tag{6.10}$$

or, a more consistent TLS approach can be attempted:

$$F_{\log TLS} = \sum_i \frac{\left(\ln \xi_i - \widehat{\ln \xi_i} \right)^2}{\sigma^2_{\ln \xi_i}} + \sum_i \frac{\left(\ln \Phi_i - \widehat{\ln \Phi_i} \right)^2}{\sigma^2_{\ln \Phi_i}} \tag{6.11}$$

where $\left(\widehat{\ln \xi_i}, \widehat{\ln \Phi_i} \right)$ is the prediction in the TLS sense, i.e. the point on the straight line lying at the shortest weighted distance from $\left(\ln \xi_i, \ln \Phi_i \right)$. As already mentioned, this linear case, the TLS projection can be obtained with a closed-form formula, and therefore it is computationally very rapid.

Population approach (NLMEM)

All the methods described so far are based on a geometrical approach and therefore implicitly rely on hypothesis (6.2). This makes them statistically inconsistent in case of population variability in the DI values. The methodology exposed in the following paragraphs, instead, relies on a hierarchical structure for the variability in the dataset: in addition to sharing with the previous models the assumptions on the measurement error, a different, broader hypothesis on the population probability distribution of the secretion and sensitivity indices is made. The indices $\bar{\xi}_i$ and $\bar{\Phi}_i$ are assumed as lognormally distributed in the population; this is a common assumption in the literature, that also guarantees the positivity and physiological plausibility of the parameters. The joint population distribution for the true but unknown values $\bar{\xi}_i$ and $\bar{\Phi}_i$ has the following general mathematical formulation:

$$\overbrace{\begin{bmatrix} \bar{\xi}_i \\ \bar{\Phi}_i \end{bmatrix}}^{z_i} \sim LN\left(\overbrace{\begin{bmatrix} \mu_\xi \\ \mu_\Phi \end{bmatrix}}^{\mu}, \overbrace{\begin{bmatrix} \omega^2_\xi & \rho\omega_\xi\omega_\Phi \\ \rho\omega_\xi\omega_\Phi & \omega^2_\Phi \end{bmatrix}}^{\Omega} \right) \tag{6.12}$$

where μ and ω^2 denote respectively the population geometric mean and variance of the subscripted index, and ρ is the correlation among the logarithms of the indices.

Under this hypothesis, also the probability distribution of the DI is lognormal, and more precisely

$$\bar{\xi}_i^{\alpha} \cdot \bar{\Phi}_i = \Delta_i \sim LN\left(\alpha\mu_{\xi} + \mu_{\Phi},\quad \alpha^2\omega_{\xi}^2 + \omega_{\Phi}^2 + 2\alpha\rho\omega_{\xi}\omega_{\Phi}\right) \tag{6.13}$$

If the whole population is characterized by one value of DI, shared by all the subjects, this implies that population variance of the DI is zero. In other words, the following constrain holds

$$\alpha^2\omega_{\xi}^2 + \omega_{\Phi}^2 + 2\alpha\rho\omega_{\xi}\omega_{\Phi} = 0 \tag{6.14}$$

Working equation (6.14) under the conditions

$$\begin{aligned} -1 \le \rho \le 1 \\ \omega_{\xi} \ge 0 \\ \omega_{\Phi} \ge 0 \\ \bar{\alpha} > 0 \end{aligned} \tag{6.15}$$

we obtain one uninteresting solution ($\omega_{\xi} = \omega_{\Phi} = 0$, i.e., all subjects have the same SI and PHI) and the more complex one

$$\begin{aligned} \rho &= -1 \\ \frac{\omega_{\Phi}}{\omega_{\xi}} &= \alpha \end{aligned} \tag{6.16}$$

So, if the correlation ρ among $\log\left(\bar{\xi}_i\right)$ and $\log\left(\bar{\Phi}_i\right)$ is -1, and the ratio between their population variances is α^2, then all the subject will share the same value of DI. This can be intuitively interpreted graphically by plotting the subjects on a Cartesian plane. If the correlation is -1, they all belong to the same pseudo-hyperbola. It is also interesting to notice that, once α is fixed, what controls the population variability of the DI is the parameter ρ: as it gets closer to zero, the points get more and more scattered around the pseudo-hyperbola.

So the joint population distribution can be rewritten as follows

$$\overbrace{\begin{bmatrix} \bar{\xi}_i \\ \bar{\Phi}_i \end{bmatrix}}^{\bar{z}_i} \sim LN\left(\overbrace{\begin{bmatrix} \mu_{\xi} \\ \mu_{\Phi} \end{bmatrix}}^{\mu}, \overbrace{\omega^2 \cdot \begin{bmatrix} \frac{1}{\alpha} & \rho\alpha \\ \rho\alpha & \alpha \end{bmatrix}}^{\Omega}\right) \tag{6.17}$$

This represents the first stage of the variability in the model, and more specifically the between-subject variability. On top of this, the within-subject variability (estimation uncertainty) is still present and equations (6.1) still hold.

The advantage of this model is that its hypotheses are much less restrictive, and it accommodates the TLS fit approach as a special case with $\rho = -1$.

In order to estimate the parameters of the joint population distribution, a population approach, such as NonLinear Mixed-Effects Models (NLMEMs) [7, 25], can be applied to analyze the insulin sensitivity and secretion data of a group of subjects consistently with the statistical model in Eq. (6.17). Once estimates of the population parameters are available, the information about the DI can be extracted from the joint population covariance matrix Ω. Another even more robust option consists in using a population approach to jointly fit the IVGTT glucose and C-peptide minimal models, estimating, at the same time, not only the individual sensitivity and secretion indices, but also the DI and α, which are features of the whole population.

Real and generated datasets

Before applying the new TLS and population approaches on a real dataset, a comparison of all the methods was carried out on a simulated data to assess their performance. With the purpose of making the simulation as realistic as possible, a real dataset [4, 5] was first analyzed to inspect the distribution of the indices across a population and have guesstimate of the real-life nature and size of their estimation errors.

The real dataset consists of 204 healthy subjects who underwent an insulin-modified IVGTT (IM-IVGTT) with full sampling schedule (240 min, 21 samples). The individuals belong to two large groups characterized by different age: 59 subjects were young (age 23 ± 3) and the remaining 145 were elderly (age 69 ± 6). For each subject, the IVGTT glucose minimal model [11] was used for identification and estimation of SI (used along with its uncertainty, whereas the IVGTT C-peptide minimal model [67] provided the estimates for the secretion indices Φ_1 and Φ_2, subsequently combined into Φ_{tot}. Both these models are described in Chapter 2. The fit of the models for each individual was performed with the modeling software SAAMII [3].

A lognormal distribution across the population was detected for both $\bar{\xi}_i$ and $\bar{\Phi}_i$ and therefore hypothesized for the generation of the simulated dataset, using the general structure in Eq. (6.12) and therefore, to respect the DI paradigm, Eq. (6.17). The typical values of $\bar{\xi}_i$ and $\bar{\Phi}_i$ ($\mu_{\bar{\xi}}$ and μ_{Φ}) were tuned so to obtain 100 as a symbolic expected value for Δ, while ω was fixed to 0.5, implying, for $\alpha = 1$, a variability of about 50%, in agreement with the real dataset. In addition, different values for the parameter α (0.5, 1 and 2) were used and, for each of these values, 100 sets, each one consisting of 1000 subjects, were generated. These datasets will be referred to in the paper as A, B and C respectively.

Analyzing the probability distribution of the estimation uncertainty detected in the real dataset, a constant CV pattern was recognized for Φ_i, whereas constant SD was more suitable for ξ_i. The error on Φ_i was characterized on average by 8% CV, and for ξ_i, where SD - and not CV - is constant, an SD corresponding to the 5% of ξ_i's population typical value was detected. However, this level of estimation error did not seem to account for all the variability in the dataset and, when performing the fit of a pseudo-hyperbola, a much larger uncertainty was found a posteriori, pointing towards a five-fold increase in the weighted residuals ($\hat{\sigma}^2 \cong 25$). Therefore, we decided, for our simulation, to use about 40% for Φ_i's CV and 25% of the typical population value for ξ_i's SD.

After generating the data, for each value of α, all the algorithms were tested on the 100 repetitions and the results were collected to provide an estimate of bias and precision for both Δ and α.

As a first step, the subjects were assumed to have all the same DI value, so no variability of Δ_i was simulated. This was obtained by fixing the correlation parameter ρ to -1. Visually, this means that all the subjects' parameter values, prior to the perturbation due to the simulation of the estimation error, superimpose exactly over the pseudo-hyperbola. In a further step of the analysis, population variability for DI was introduced in the simulation, and thus hypothesis (6.2) was not respected: in this way the possible consequences of using the fit approaches in presence of such a misspecification are investigated. Several datasets with progressively increasing levels of variability in the DI values were generated, and this was obtained by using different values for the parameter ρ (-0.8, -0.5 and -0.2). At the same time, the level of simulated estimation error in the individual values ξ_i and Φ_i was decreased, trying to maintain about the same level o overall variability. In this way the focus of the analysis was on the consequences of the different hierarchy underlying the variability, rather than its level. However, as stated in (6.13), the population variability in the DI depends both on ρ and α, so the overall level of variability was preserved for the case $\alpha = 1$. For this particular setting, the generated datasets were characterized respectively by an approximate population CV% of 32% ($\rho = -0.8$), 50% ($\rho = -0.5$) and 63% ($\rho = -0.2$) for the values of Δ_i.

Finally, the real dataset was analyzed with our newly designed approaches (both TLS and the population NLMEM method), to inspect the validity of the hyperbolic law and probe the statistical significance of the hypothesis $\alpha \neq 1$. As different values of

DI are expected to characterize the young and elderly population, the fit was also executed separately for the two clusters of subjects. We used the Insulin Sensitivity provided by the IVGTT glucose minimal model as ξ (on the x axis), whereas for the Φ we used the first-phase, second-phase and total beta-cell responsivity provided by the IVGTT C-peptide model (respectively Φ_1, Φ_2 and Φ_{tot}). The fit approach is very sensitive to outlying values and so, in order to detect and remove these subjects from the dataset, the jackknife technique [30] was employed. This approach consists in refitting each of the datasets several times, excluding each time one single different subject and calculating, *a posteriori*, the residual that the subject presents when not included in the parameter optimization process. In this way, the problem of outliers with leverage strong enough to "hide" themselves is circumvented. After the collection of all the so-called jackknife residuals, the subjects showing an abnormally large deviation from the model prediction were excluded.

Results

For each setting of the parameters used in the simulation (α and ρ), the estimates of the 100 repetitions were collected and displayed in boxplots, which show the size of the quartiles and the position of the median value. The median can be used to assess the accuracy of the estimation algorithm, whereas the width of the boxes and whiskers is an indicator of its precision.

The number of methods initially included in the comparison was rather large and some of them performed extremely poorly. In order to make the presentation of results more readable, in a first step all the methods were considered, but, after the first comparison, the results from the 1-variable fit approaches were not reported. First, the datasets including only estimation uncertainty of the individual parameters were tested, then the effect of population variability in the DI values was probed, and finally, the real dataset is analyzed.

No population variability in the DI

In this section, the results obtained on the datasets with no population variability in the DI values are reported. These data were generated respecting hypothesis (6.2), that is to say, the value of the correlation ρ was fixed to -1.

All methods were run on simulation Dataset B ($\alpha = 1$), but the results obtained with the 1-variable fit approaches were so numerically so different from the others, that, for readability purposes, they had to be displayed in separate panels. They are collected in Figure 6.2. Even at first glance, it is clear that attempts to estimate the

parameters with approaches considering only one variable produce extremely biased estimates. Numerically this seems more apparent when the error is assumed along the X-direction, but this is due only to the mathematical formulation of the model, and the quality of the estimates provided by the error-on-Y approach is not much more satisfactory. Figure 6.1 (upper panel) provides a graphical depiction of the results and allows one to realize that, even if the bias in the estimates is numerically much more remarkable for the X-algorithms, all the methods based on the fit in one variable provide comparably poor results. The 1-variable fit approach is unsatisfactory also when used to fit a straight line on the log-transformed data (Log-X and Log-Y). All these methods systematically under- or over-estimate the parameters. Since a very similar behavior characterizes the results obtained on datasets A and C and they do not add much to the present discussion, they are not reported.

The results provided by all the other approaches are displayed in Figure 6.3. When compared to the 1-variable methods, they all yield much more reasonable outcomes, but with some important differences with one another. The log-transformation, in fact, seems to affect the results, in particular the estimate of Δ: Log-Riggs tends to underestimate Δ (about -40% or more), whereas Log-TLS provides overestimates (about +40% or more). The precision and accuracy with which the other parameter, α, is estimated by the 2-variables log-transformation algorithms are more satisfactory; Log-TLS, however, performs slightly better. Also with the parameter α, in fact, Log-Riggs tends to provide an underestimate (about -15% to -25%) and Log-TLS an overestimate (about +10%). The newly proposed TLS and population (NLMEM) methods yield the best performance, both in terms of precision and accuracy, and for both parameters. The size of the bias for Δ is shrunk very significantly compared to the other 2-variables fit approaches, at most ~20% as opposed to al least ~40%, and a similar trend is evident also for the precision, in that the TLS and NLMEM results are generally less scattered than the other 2-variables' algorithms The improvement in the estimate of α is also significant: the bias is at most ~5%, instead of ~10% for Log-TLS and ~20% for Log-Riggs. The changes in the values of the parameter α used in the simulation (0.5, 1 and 2) do not seem to have a clear effect on the results; on the contrary, the methods generally seem to provide a similar performance with different values of α.

Population variability in the DI

In this section, then, the consequences of the introduction of population variability are investigated. The outcomes of the methods based on the error in only one variable,

whose very poor quality is apparent from the results presented in last section, are not reported for all the following tests, and the focus is moved on the 2-variables methods and NLMEM population approach. Figure 6.4-6.5-6.6 contains the boxplots regarding the datasets characterized by population variability in the DI. Figure 6.4 reports the results for a situation ($\rho = -0.8$) in which the effect of population variability is limited and comparable to the one due to individual uncertainty. Even if the NLMEM method seems to perform overall slightly better, also TLS and Log-TLS, for several settings of α, provide satisfactory results. However, while the other methods' performance seem to vary with the value of α, NLMEM seems more robust and immune to this phenomenon. As the level of DI variability increases, the estimates of α and Δ provided by the geometric fit approaches become less and less reliable. In the situation depicted in Figure 6.5 ($\rho = -0.5$), and even more in Figure 6.6 ($\rho = -0.2$), a large fraction of the variability is caused by differences across subjects in the population rather than estimation uncertainty of the individual parameters, and hypothesis (6.2) becomes therefore a strong simplification. In such cases, all the geometric fit approaches fail. The estimates they yield for both α and Δ are affected by a very large bias, and the direction of this bias (in the sense of over- or under-estimation) seems to be strongly dependent on the value of α used for the generation of the dataset. It is very interesting to notice that, even if the accuracy of the estimates is very poor, the relative level of precision is almost increased. It can be argued that, in this case, the parameter Δ assumes a slightly different meaning: in this new framework there is not anymore one single value of DI for the population, but rather a probability distribution centered around the value of Δ. Therefore, the deterioration of the geometric fit estimates of this parameter is somewhat expected, but the estimation problems involve also the estimates of α, whose quality degrades significantly, even though, at least for $\alpha = 1$, the overall level of variability in the data does not substantially change, what is different is the underlying hierarchy of the variability itself. The method that seems to suffer more from the introduction of population variability is Log-Riggs, whose bias in the estimate of both Δ and α is significantly bigger than the one obtained with the other methods.

Moreover, it is interesting to take a look at the values of $\hat{\sigma}^2$ obtained with the different datasets and contained in Table 6.1. The value of $\hat{\sigma}^2$ is a marker for goodness of fit and it can be roughly considered as the average squared residual: the average squared weighted distance between each data point and the model prediction, i.e., the closest point on the pseudo-hyperbola. If the *a priori* guesstimate of the level of error in the data is reasonable and the model is correct, a value of about 1 is

expected. As it can be seen in Table 6.1, the values are around 1 only for the datasets with no or relatively small population variability in the DI ($\rho = -1$ or -0.8). This does not hold true when the DI variability is larger ($\rho = -0.5$ or -0.2), where the values of $\hat{\sigma}^2$ clearly indicate lack of fit and so that, possibly, the *a priori* level of uncertainty is not large enough to account for all the variability detected in the data.

The real dataset

Finally, both the TLS and the NLMEM approaches are applied to the real datasets. Since the TLS approach is not able to cope with large variability in the population and the fit is very sensitive to outliers, a jackknife technique [30] was employed to detect and remove these values, as explained in the methods section. On average, one or two subjects were removed from each of the datasets. The analysis was executed both on the young and elderly populations separately and on the whole group. In this last case, there is population variability in the DI *by hypothesis*, so the TLS approach was used only as a mean of comparison, but the results are expected to be unreliable. Table 6.2 reports the optimal values provided by the TLS and NLMEM algorithms for Δ and α, for SI coupled with Φ_1, Φ_2 and Φ_{tot}. The graphical depiction of the TLS fits for the separate groups of young and elderly subjects is reported in Figures 6.5-6.6-6.7.

A first important result is given by the values of $\hat{\sigma}^2$ detected by TLS, which are always much bigger than 1, indicating that the level of variability in the data unexplained by the model is much greater than the level of estimation uncertainty indicated by the estimation CVs. This might either indicate that the estimation confidence intervals for the individual indices SI and Φ are heavily underestimated, or that does not manage o consistently explain the data. This could possibly indicate, for example, the presence of population variability. A further clue pointing towards the DI population variability is given by the values of ρ provided by the NLMEM approach. The correlation detected is never more significant than about -0.5, indicating a remarkable level of scattering of the DI values. A particularly low correlation is detected for SI- Φ_1, and particularly when fitting the cluster of elderly subjects or the whole population.

All these hints seem to support the presence of population variability in the DI values, so the NLMEM results are to be considered more accurate and are used to test the statistical significance of the parameter α. The values of α detected are smaller than 1 for all the datasets, but the difference is statistically significant only for the elderly cluster or when the entire population is employed for the analysis. It should be

pointed out, however, that the cluster of young subjects is smaller (59 vs 145) and therefore it's harder to reach statistical significance. The estimate of α is more uncertain when using Φ_1 as secretion index, and the value 1 cannot be excluded in this case, whereas for Φ_{tot} and Φ_2, the introduction of α in the model seems more statistically relevant.

Finally, it is interesting to observe the effect of using the TLS approach on the whole population, which, by hypothesis, is affected by population variability. The fit degenerates into a straight line ($\alpha = 0$) for **SI-**Φ_1 and provides very large values of $\hat{\sigma}^2$ in all cases. NLMEM, instead, even if it detects a large amount of variability, provides more robust estimates.

Discussion

The results clearly show that it is essential to consider the error in both variables; when only error along one direction is taken into account, the results are very biased. In particular, the X and Log-X methods heavily overestimate both Δ and α, creating in this way a curve "rotated clockwise" with respect to the real one. The Y and Log-Y methods, on the other hand, have the opposite effect, rotating anti-clockwise the pseudo-hyperbola. The explanation for this phenomenon is easy. When just the error on X is considered, the data points characterized by a very high ξ and a very low Φ lie in a section of the chart where the pseudo hyperbola is almost flat and therefore parallel to the X direction. This causes the residuals for these points to be very large and thus, these points have a very strong leverage in the fit. As a consequence, the curve rotates clockwise to fit better these data points. When instead, just the error on Y is considered, the high-leveraged points become the ones with very low ξ and very high Φ, so the opposite phenomenon is observed. This is shown clearly in Figure 1 (upper panel), depicting the prediction curves yielded by these 1-variable-fit methods on an example dataset.

On the other hand, when considering the error in both directions, caution should be used to calculate the distance between the data points and the curve. The use of a log-transformation, which significantly simplifies the problem, comes with a price: because of its nonlinearity, the logarithm stretches the axes and the distances among the points in an uneven way, and this can substantially change the leverage that the points have in the fit procedure. Moreover, as already mentioned, also the error structure is somehow modified and we had to employ an approximation to deal with this problem. It is hard to tell which of the two phenomena affects more severely the results. The nonlinear rescaling of the data is expected lead to underestimating Δ, as

the subjects with high ξ or Φ values will tend to clustered together, whereas the distance among the points close to the axes will increase and drive them farther from the real curve. An easy example to clarify this phenomenon is provided by the fact that, given any set of positive numbers z_i different form each other,

$$\exp\left(mean\left(\ln\left(z_i\right)\right)\right) < mean\left(z_i\right) \tag{6.18}$$

On the other hand, the first-order approximation we used for the error structure is expected to have the opposite effect and partially account for the uneven stretching, as it attributes a smaller uncertainty, and thus a bigger leverage, to the data points with high ξ and Φ values. In fact, the application of a first-order approximation to the error propagation is numerically the same as considering the values of the CVs in the original space as SDs for the log-transformed data. Thus, in case of an hypothetical dataset characterized by the same variance for all data points (constant SD), after the log-transformation the subjects with higher values of ξ and Φ, whose CVs in the original untransformed space are smaller, carry smaller SDs and therefore a bigger leverage in the log-transformed space.

Possibly, this is why Log-Riggs, which assumes the same degree of uncertainty for all the data points in the log-transformed plane and thus it is not prone to this latter effect, tends to underestimate Δ. Log-TLS, on the other hand, tends to provide overestimates of Δ even if the accuracy of the optimal values for α is not so poor. So, rather than being rotated like in the case of the 1-variable fits, these curves obtained with a log-transformation approach seem to be affected by a sort of offset. The example contained Figure 1 (lower panel) gives a visual representation of this phenomenon. TLS, even though computationally much more cumbersome, should then be considered as a more reliable alternative, because it is consistent with the ML estimator and does not perform approximations to calculate the distance between the data points and the curve. The NLMEM, on the other hand, even if it sits on more general hypothesis, still accommodates the situation $\rho = -1$ as a particular case, and indeed the results are similar to if not better than TLS. NLMEM is actually over-parameterized when there is no population variability in the DI, but it still performs very satisfactorily.

It is also very interesting to investigate the consequences of the presence of population variability. As already mentioned, assumption (6.2) is necessary for the geometric fit approaches to be statistically sound, and therefore the presence of other forms of variability besides the estimation uncertainty of the secretion and sensitivity indices undermines the validity of all the methods exposed here, except NLMEM.

The effects of the falsification of hypothesis (6.2) are clearly portrayed in our results. In fact, even if the overall amount variability in the data is about the same, the performance of all the geometry-based approaches degrades significantly when population variability in the DI values comes into play. The underlying reason is that, as the magnitude of population differences surpasses the effect of individual uncertainty, the estimates of precision for ξ_i and Φ_i, even if correct, become less and less informative about the actual distance of the data points from the average curve. The uncertainty on the parameters ($\sigma^2_{\xi_i}$ and $\sigma^2_{\Phi_i}$), in fact, does not have in general any relationship with how far their values are from the population mean. The geometric fit approaches are not designed to account for this kind of hierarchical variability, so they fail when different sources of variability come into play. Their inadequacy is stressed by two facts. First, the effect of the introduction of population variability does not cause the results of the geometric fit approaches to be less precise (i.e., larger scattering of the results obtained on 100 repetitions), possibly indicating more uncertainty in the estimates. On the contrary, the relative level of precision is even more satisfactory, but the bias becomes very large. Second, the size and direction of the bias of the estimates strongly depends on the value of α used in the simulation of the dataset. The different settings of α introduce asymmetries in the data and the fraction of variability which is not explainable as estimation uncertainty assumes patterns that lead to either over- or underestimation by the geometrical fit methods. The NLMEM, on the contrary, is designed to cope with the hierarchical structure of the variability and provides reliable results, regardless of the value of α.

When running the algorithms on the real datasets (Table 6.2), the TLS fit approach yields a high value of $\hat{\sigma}^2$ and NLMEM detects low level of correlation ρ among the individual indices. Both these markers indicate the presence of considerable population variability in the DI values, so the TLS results should be regarded as unreliable. Therefore, with the purpose of investigating the statistical significance of parameter α and thus the validity of the pseudo-hyperbolic versus the hyperbolic law, the NLMEM results are used. Even though a validation on other independently created datasets would be needed to strengthen the current findings, the estimates of α provided by the NLMEM algorithm seem to point towards a value of α smaller than 1, supporting in this way the pseudo-hyperbolic law with respect to the simple hyperbolic law. The statistical significance, however, is not reached for the cluster of young subjects, but this might be due to their small number. In addition, it seems that the values of α are not significantly changing with the age characterizing the population in analysis; rather they seem related to the combination of indices used to

calculate the DI (0.8~0.9 for Φ_1, 0.66~0.8 for Φ_2, and 0.61~0.8 for Φ_{tot}). Further research on different datasets would be necessary to corroborate this idea, but this phenomenon is encouraging and supports the use of the DI paradigm, also with the additional parameter α, as a broader classifier even among subjects belonging to heterogeneous populations (i.e. young, elderly, healthy, glucose intolerant, type I and I diabetes). In other words, if the value of α only depends on the parameters used in the calculation of the DI and not on the characteristics of the population of subjects in exam, then, once the value of α is determined with an extensive analysis comprising large datasets, the DI can be calculated for each subject separately and "off-line", without needing a fit procedure. In this way, the analysis of a population would be necessary only in a first phase, in order to determine the correct value of α, but consequently, the DI values of a subject can be simply determined by estimating insulin sensitivity and secretion, and then applying the formula using the appropriate value of α.

Consistently with the theory underlying the DI paradigm, the values of DI characterizing the elderly subjects are lower than the ones related to the young, and the graphical representations in Figure 6.7-6.8-6.9 (even though obtained with the TLS method) show that the curve for the elderly subjects is closer to the origin than the one for the young ones.

Further research should be aimed at implementing a joint population model, estimating the secretion and sensitivity indices directly from glucose and insulin profiles, and then extracting the information on the Δ and α from the population covariance matrix.

Acknowledgments

These results were partially presented at the EASD 2007 Annual Conference in Amsterdam, NL and included in the conference Proceedings (P. Denti, M. Campioni, G. Toffolo, R. Rizza, C. Cobelli "Is the disposition index law hyperbolic? Importance of the regression methodology"). I would like to thank Eng. Domenico Dei Tos, for performing the analysis on the real dataset as part of his Laurea Thesis project.

Table 6.1 The average $\hat{\sigma}^2$ obtained across the 100 runs for the TLS method is reported for datasets with increasing degrees of population variability. Only the results for $\alpha = 1$ are reported.

	$\rho = -1$	$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$
$\hat{\sigma}^2$	0.912	1.54	23.3	811

Table 6.2 Results obtained with the TLS and NLMEM aproach on the real dataset. The fit was performed both separately on the young and elderly subjects and then on the whole population. The columns contain the estimates for Δ, α yielded by both methods, for the population approach, also the confidence interval of α is reported. In addition, the value of $\hat{\sigma}^2$ provided by TLS and ρ given by NLMEM are contained in the last two columns. The symbol # indicates that, when TLS was used to examine the relation **SI**-Φ_1 in the whole dataset, the lower bound of parameter α was hit and the pseudo-hyperbola degenerated into a straight line. The symbol * indicates a value of α not statistically different from 1.

		Δ POP	Δ TLS	α POP	α TLS	$\hat{\sigma}^2$ TLS	ρ POP
SI-Φ_1	Y	1132	937	0.90* (0.67-1.13)	0.85	63	-0.303
	E	432	332	0.83 (0.68-0.99)	0.68	49	-0.222
	ALL	493	50	0.80 (0.67-0.92)	0.00#	113	-0.111
SI-Φ_2	Y	42	25	0.80* (0.60-1.01)	0.58	24	-0.536
	E	25	26	0.65 (0.56-0.74)	0.66	19	-0.524
	ALL	27	16	0.66 (0.58-0.74)	0.50	5080	-0.499
SI-Φ_{tot}	Y	90	40	0.80* (0.60-1.01)	0.51	13	-0.376
	E	35	32	0.58 (0.49-0.67)	0.56	17	-0.367
	ALL	42	14	0.61 (0.53-0.68)	0.19	1982	-0.198

Figure 6.1 Graphical example of the behavior of all the geometrical fit approaches on a simulated dataset with 300 subjects. The curves provided by the 1-variable (upper panel) and 2-variables fits (lower panel), together with the curve characterized by the true parameter values are displayed.

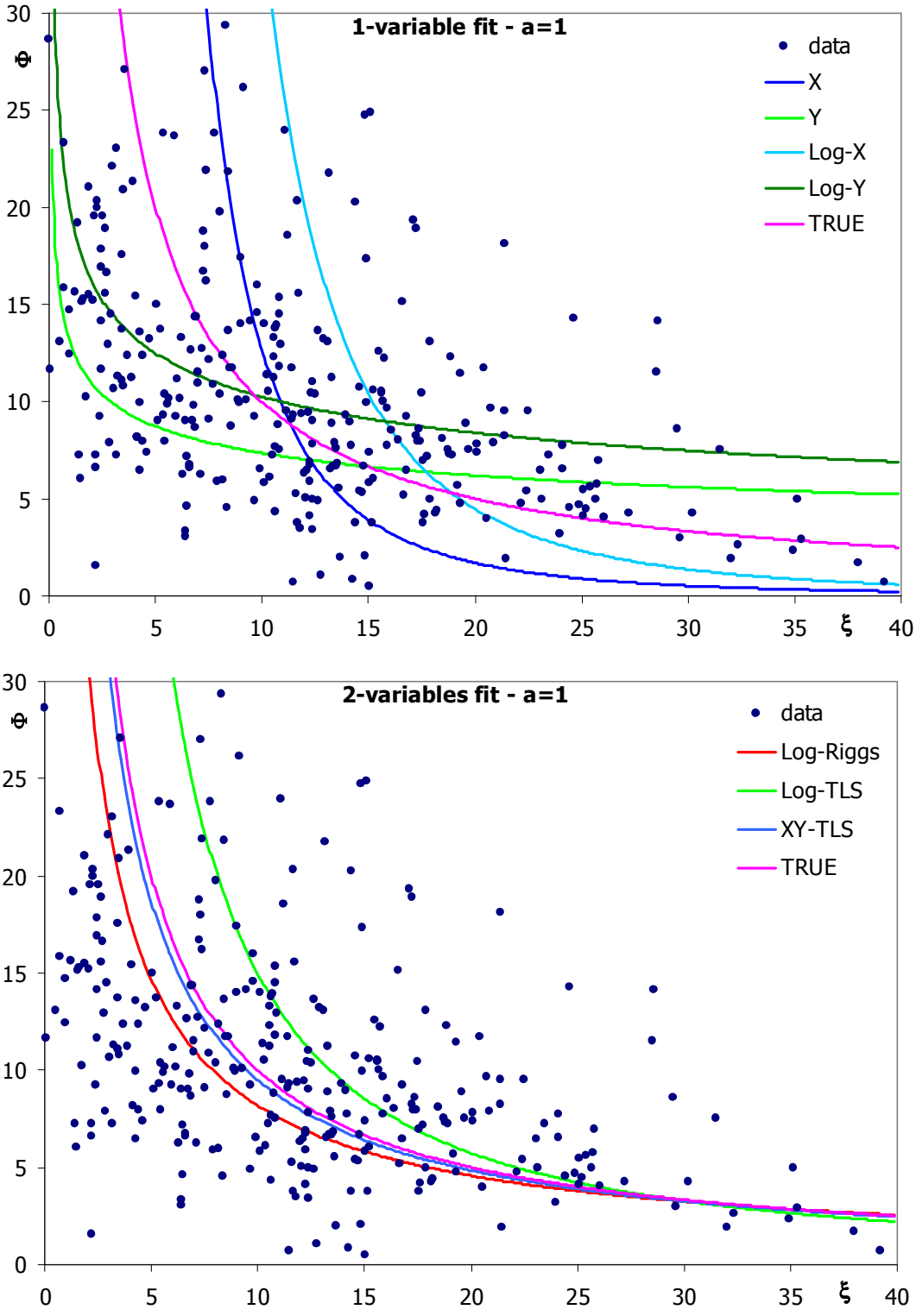


Figure 6.2 Boxplots displaying the parameter estimates (Δ in the upper panel and α in the lower one) obtained by running the 1-variable algorithms on Dataset B ($\alpha = 1$). Because of large difference among some of the values, the results have been split into separate panels. The dashed vertical line indicates the parameter value used in the simulation.

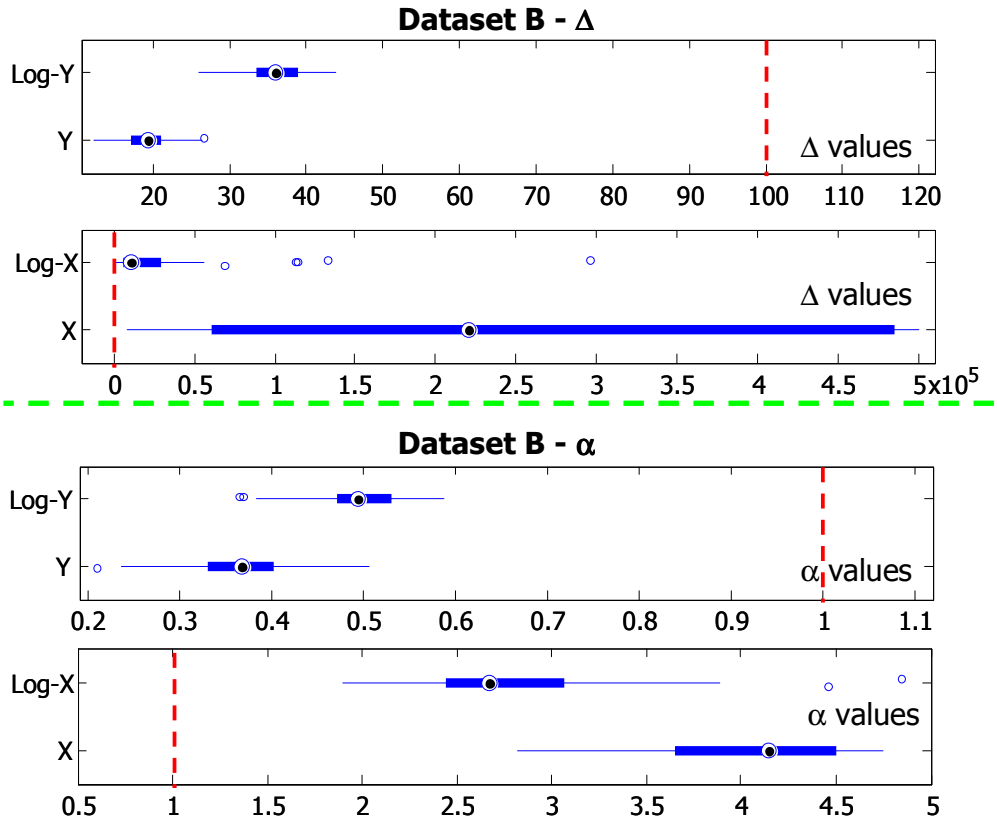


Figure 6.3 Boxplots displaying the parameter estimates (Δ in the upper box and α in the lower one) obtained by running the 2-variables geometric algorithms and the population method (denoted NLMEM) on Dataset A ($\alpha = 0.5$), B ($\alpha = 1$) and C ($\alpha = 2$), so with $\rho = -1$. The dashed line indicates the parameter value used in the simulation.

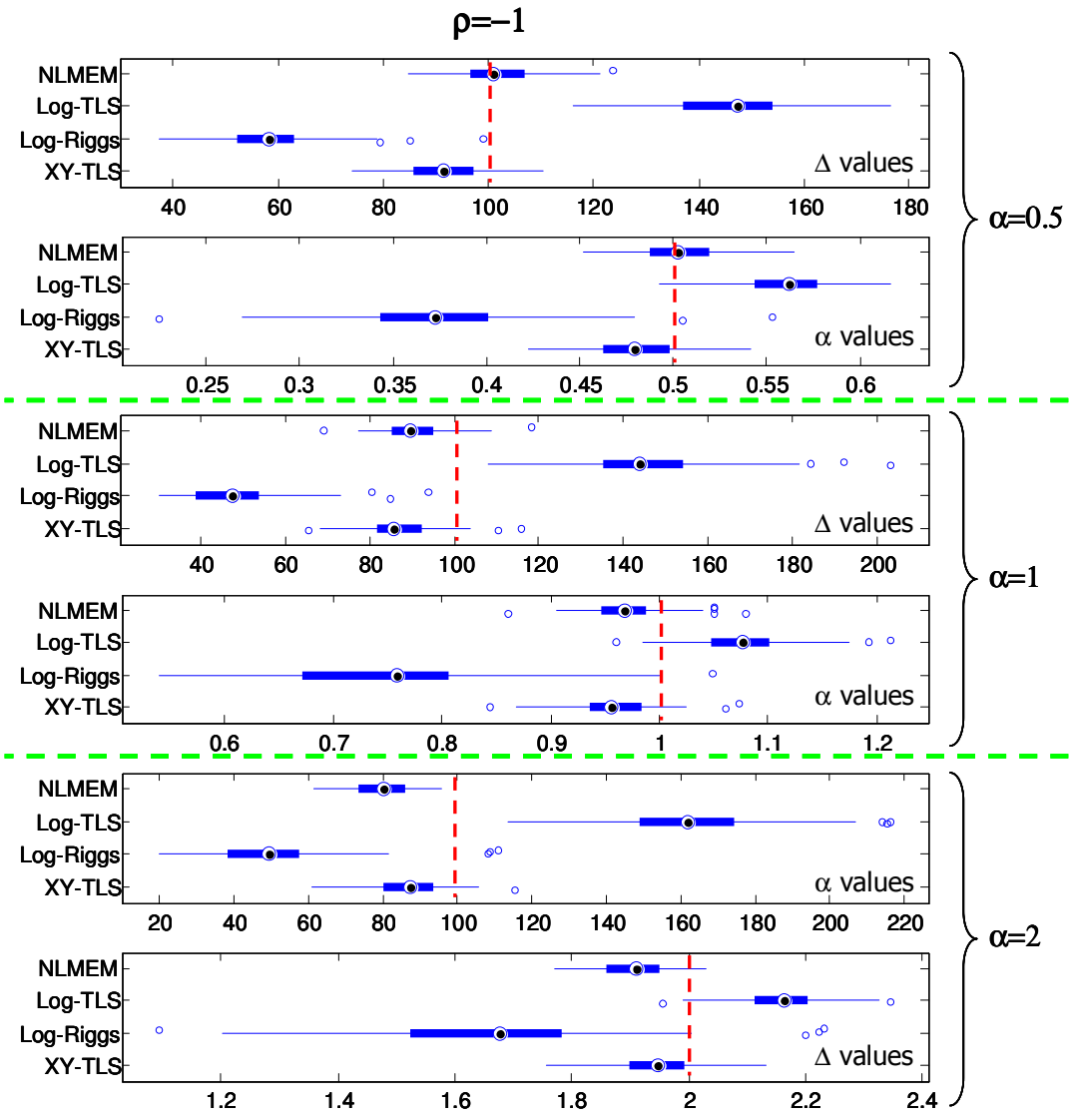


Figure 6.4 Boxplots displaying the parameter estimates obtained on the dataset including population variability with $\rho = -0.8$. Top to Bottom, in the three panels are respectively displayed the versions with $\alpha = 0.5$, $\alpha = 1$ and $\alpha = 2$. The dashed vertical lines indicate the real parameter value used in the simulation.

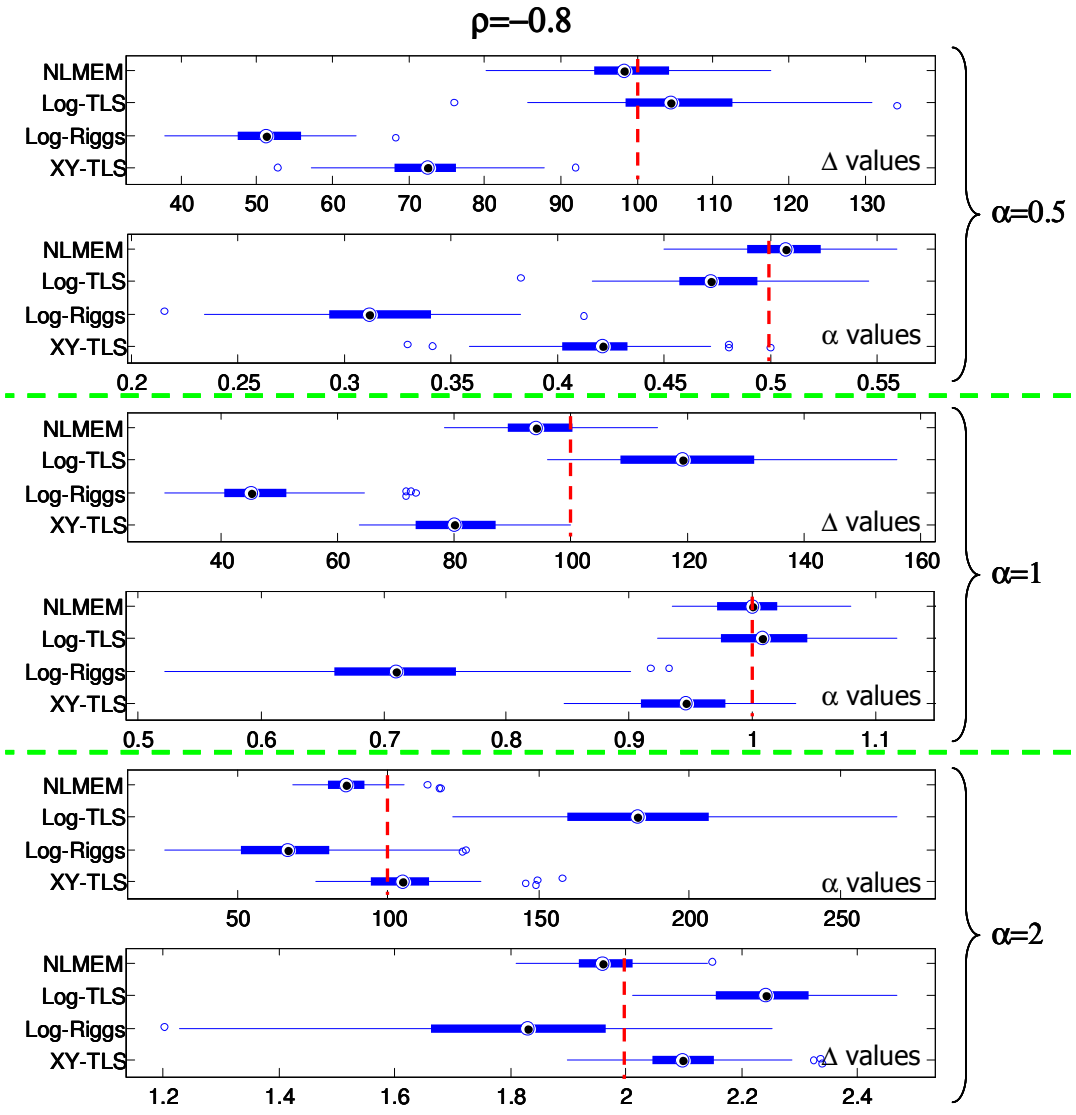


Figure 6.5 Boxplots displaying the parameter estimates obtained on the dataset including population variability with $\rho = -0.5$. Top to Bottom, in the three panels are respectively displayed the versions with $\alpha = 0.5$, $\alpha = 1$ and $\alpha = 2$. The dashed vertical lines indicate the real parameter value used in the simulation.

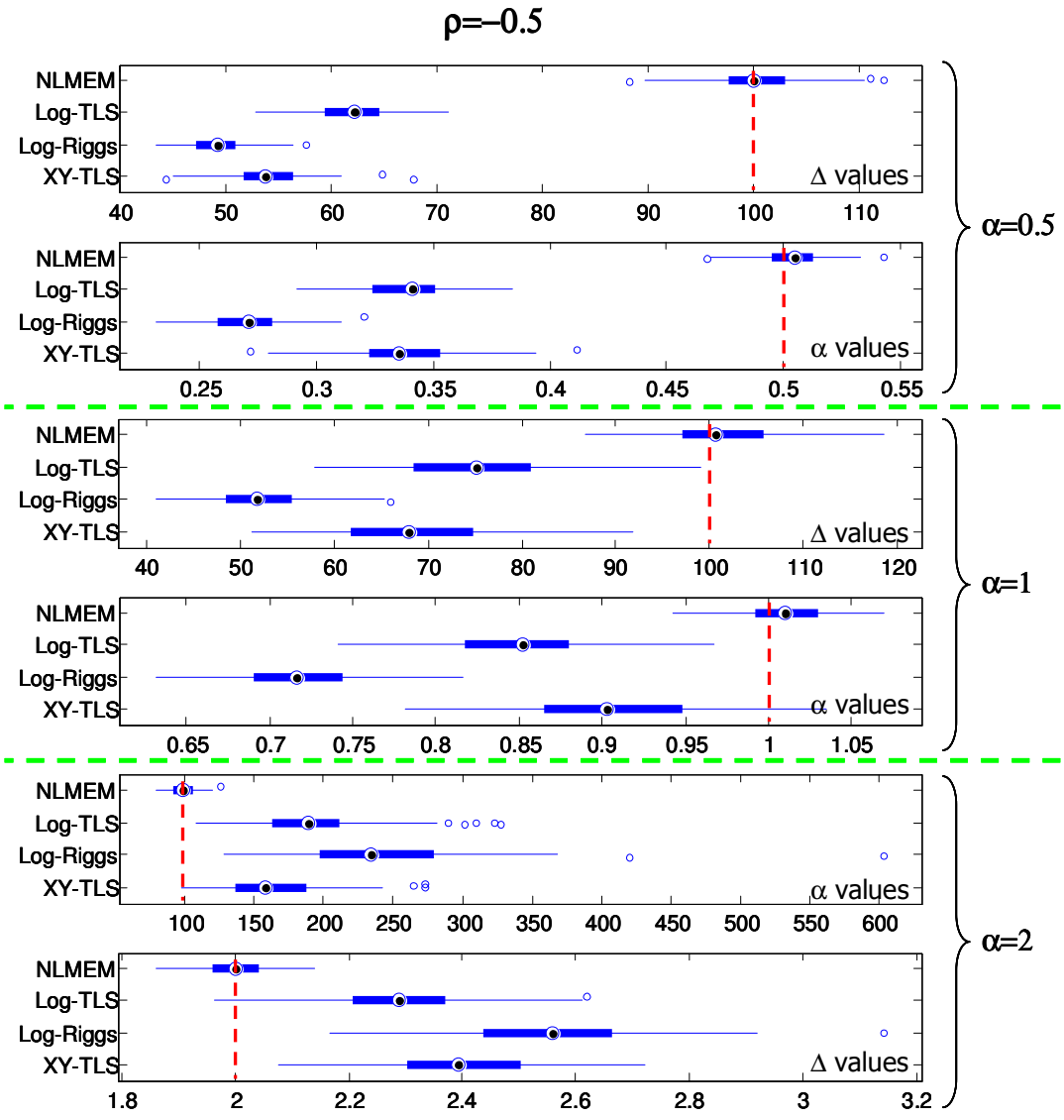


Figure 6.6 Boxplots displaying the parameter estimates obtained on the dataset including population variability with $\rho = -0.8$. Top to Bottom, in the three panels are respectively displayed the versions with $\alpha = 0.5$, $\alpha = 1$ and $\alpha = 2$. The dashed vertical lines indicate the real parameter value used in the simulation.

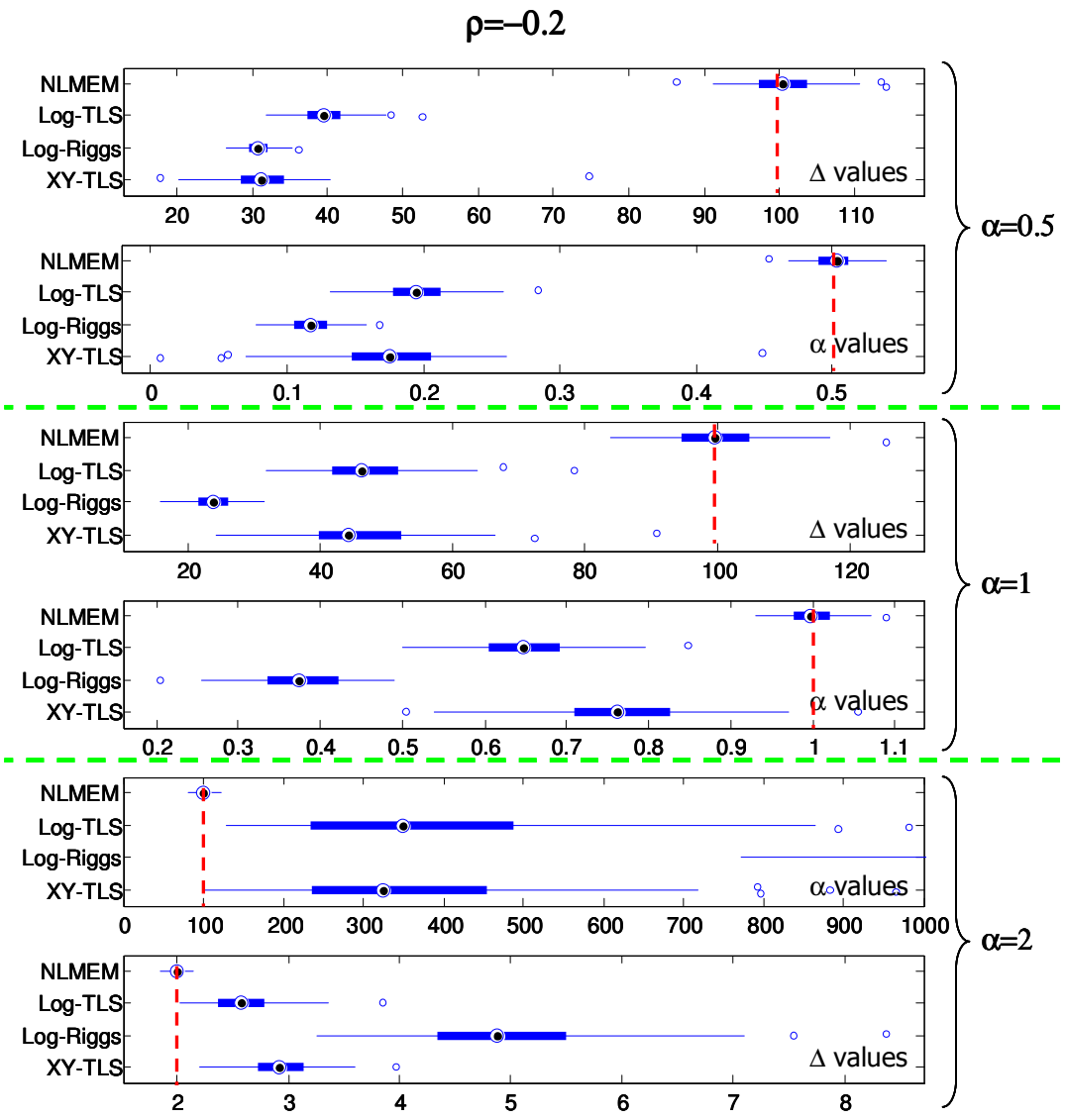


Figure 6.7 TLS results for the real dataset: Φ_{tot} vs SI. The young subjects are indicated with a full dot, and the corresponding fit is the solid line, whereas the elderly subjects are the hollow dots and the dashed line.

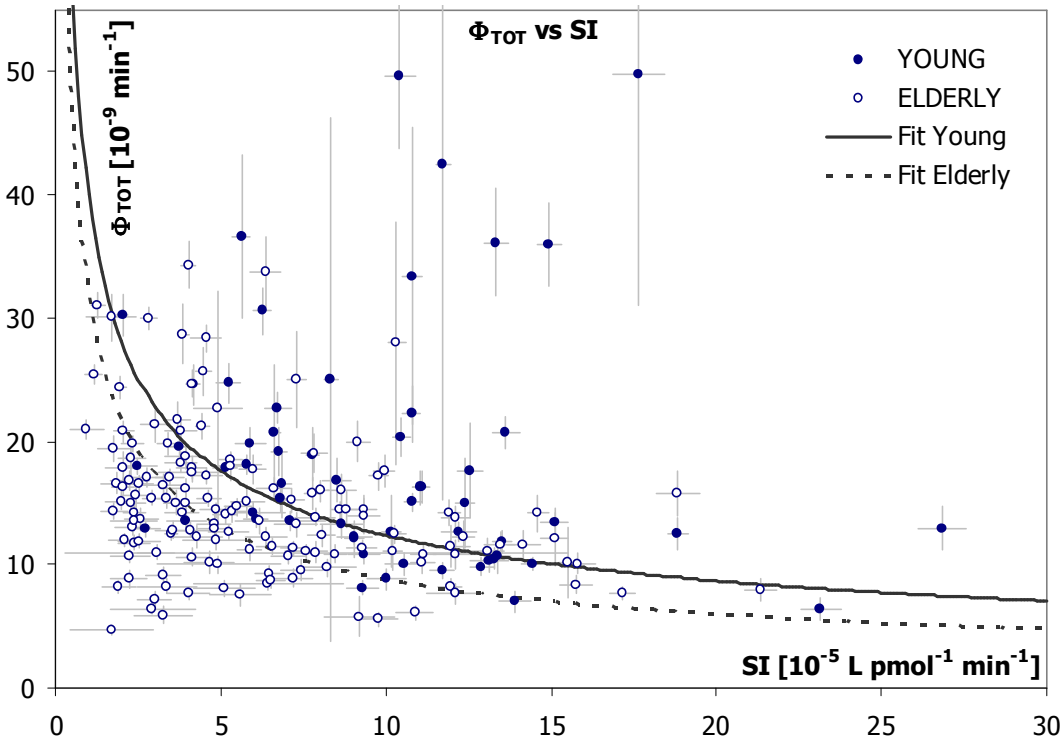


Figure 6.8 TLS results for the real dataset: Φ_1 vs SI. The young subjects are indicated with a full dot, and the corresponding fit is the solid line, whereas the elderly subjects are the hollow dots and the dashed line.

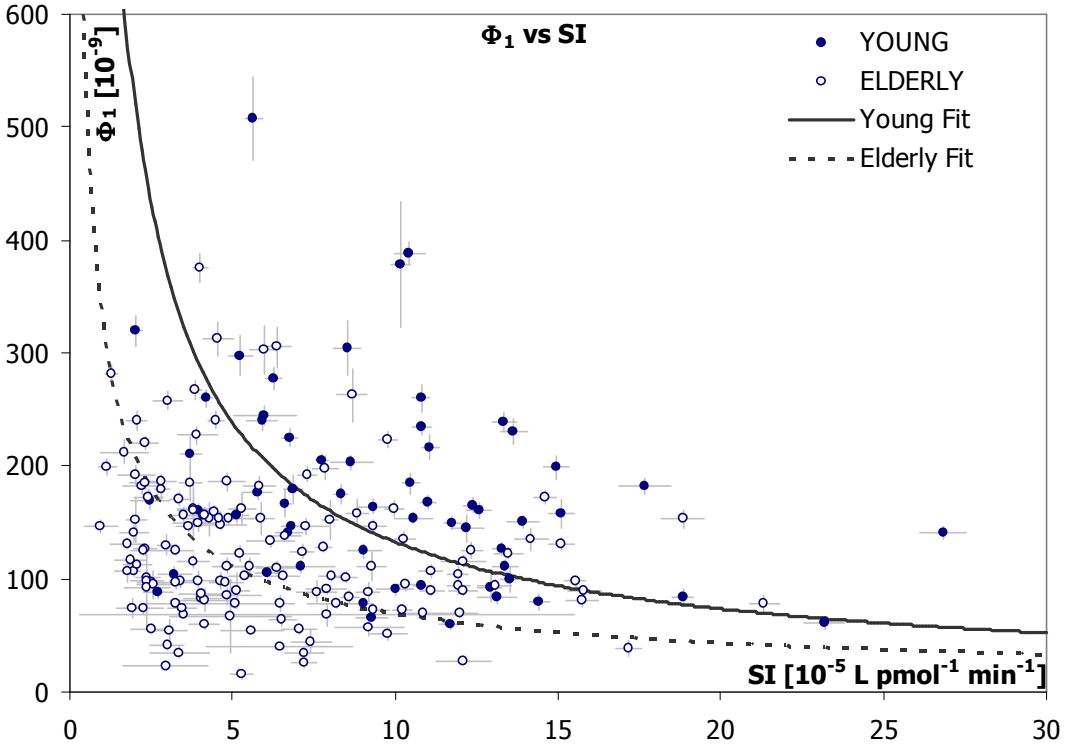
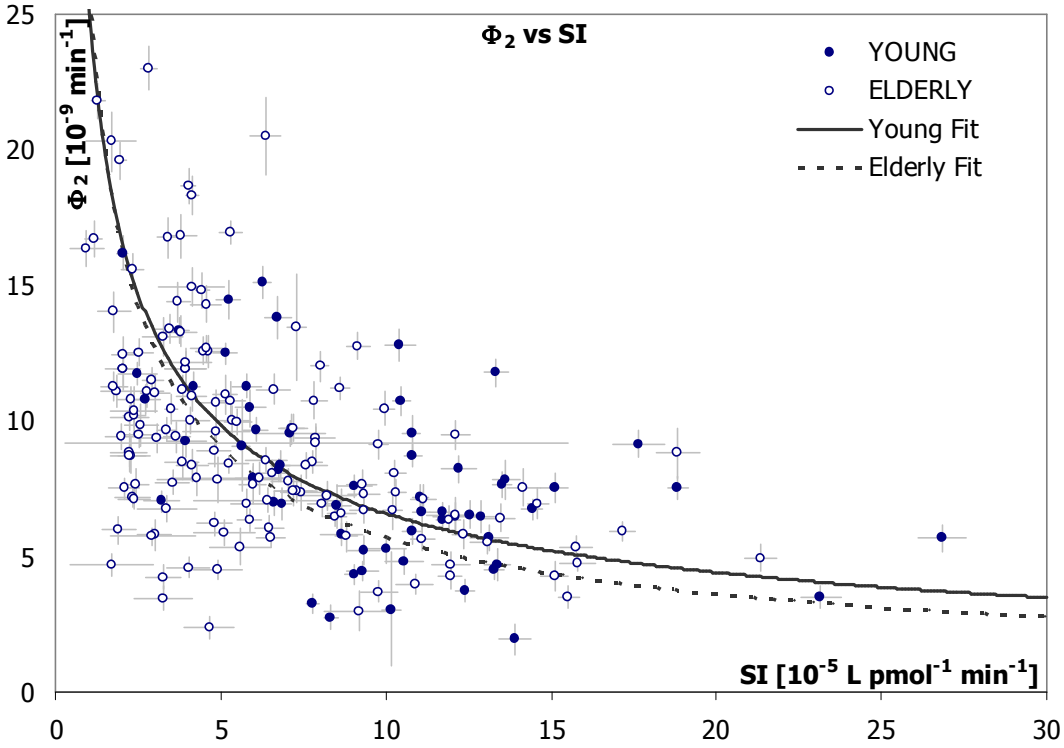


Figure 6.9 TLS results for the real dataset: Φ_2 vs SI. The young subjects are indicated with a full dot, and the corresponding fit is the solid line, whereas the elderly subjects are the hollow dots and the dashed line.



Chapter 7
Discussion and Conclusions

In this work, the advantages that population approaches can introduce in the study of glucose-insulin metabolism have been explored and some interesting results have been found.

First of all, in Chapter 3, the application of population methodologies to the traditional IVGTT model of glucose disappearance has been tested on simulated data, comparing a wide range of different methods and identifying in FOCE the algorithm of choice. In addition, the results obtained highlight the robustness of population approaches with respect to the traditional estimation paradigm when dealing with sparse sampling, and the importance of tuning the structural model by optimizing the covariance matrix, i.e., eliminating the most negligible correlation terms. In Chapter 4, the results obtained with the simulations have been confirmed on real data thanks to the likelihood function profiling via a Monte Carlo (MC) sampling technique. In first place, the MC sampling proved very useful in assessing the precision of the approximations, confirming the choice of FOCE as most accurate method. Moreover, the MC likelihood profiling allowed an assessment of the precision of the estimates: some estimation uncertainties were unexpectedly detected for the typical values of SI and P2 and, more predictably, for some elements of the population covariance matrix. In addition, the results about precision were used, together with the analysis of the objective function value, to detect over-parameterization and allow simplification of the model. In this analysis, the simple implementation of MC sampling was used, but further research could be aimed at optimizing and expediting the MC likelihood profiling tool by employing more sophisticated tools such as Markov Chain MC. In Chapter 5, the population model was further enriched with the integration of demographic information (called covariates) collected at the time of experiment. This so-called covariate model proved an important tool mainly for two reasons. First, it allowed a sophisticated statistical analysis between the minimal model parameters and the covariates. In particular, strong relations between SI and age, basal insulinemia and abdominal fat have been revealed. Second, the introduction of the covariates into the model augmented even further its explanatory power.

The most interesting advantage of the methodologies proposed consists in the fact that, since more information is obtained from the population and the covariates, less data is needed from the experiments. This allows the design of less invasive and less

expensive protocols, more suitable to a large-scale application. The dataset used in this analysis only comprised healthy subjects and was therefore characterized by a limited amount of variability; therefore the application of the same techniques on more extensive datasets would be very beneficial to confirm the present findings.
Finally, in Chapter 6, both a new non-approximated TLS geometrical fit approach and a NLMEM approach have been proposed for the estimation of the DI in a population, and tested on simulated data against simplified alternative algorithms. The TLS approach has been shown to provide more reliable results than the ones obtained with other geometrical methods previously used in the literature and based on simplifications of the fit. However, all geometrical fit approaches (and thus also the TLS one) are based on the assumption that all the subjects in the population in consideration share the same value of DI. Indeed, tests on simulated datasets revealed that in cases where a hierarchical structure of the variability is present in the data, the provided estimates are very unreliable. The novel population approach, instead, has been designed to deal with such a situation and yields accurate estimates also in this case. The inspection of a real dataset indicates the presence of population variability besides the uncertainty characterizing the estimates of the individual indices, therefore the newly-proposed population method was employed for the analysis. The results support the hypothesis of the pseudo-hyperbolic law with respect to the simple hyperbolic formula initially proposed by Bergman and colleagues, but further research on larger datasets would be needed to corroborate these results. In the study presented here, the values of secretion and sensitivity indices previously obtained in each individual with the traditional approach were used, but, since the proposed method extracts the information on the DI from the population covariance matrix, an alternative and possibly more powerful alternative might consist in interpreting directly the glucose, insulin and C-peptide profiles with a joint population model. In this way, both estimates of the individual and population parameters can be obtained and then the population covariance matrix can be used for inference on the DI.

Summarizing, this work shows the significant potential of population approaches in the study of glucose-insulin metabolism. The advantages range from the possibility of designing simplified and less invasive alternatives to the traditional IVGTT protocol, to the ability to interpret correctly data with an inherently hierarchical structure of the variability, such as in the case of the study of the DI.

References

1. Agbaje, O.F., Luzio, S.D., Albarrak, A.I., Lunn, D.J., Owens, D.R., and Hovorka, R., *Bayesian hierarchical approach to estimate insulin sensitivity by minimal model.* Clin Sci (Lond), 2003. **105**(5): p. 551-60.

2. American Diabetes Association (ADA). Available from: http://www.diabetes.org/about-diabetes.jsp.

3. Barrett, P.H., Bell, B.M., Cobelli, C., Golde, H., Schumitzky, A., Vicini, P., and Foster, D.M., *SAAM II: Simulation, Analysis, and Modeling Software for tracer and pharmacokinetic studies.* Metabolism, 1998. **47**(4): p. 484-92.

4. Basu, R., Breda, E., Oberg, A.L., Powell, C.C., Dalla Man, C., Basu, A., Vittone, J.L., Klee, G.G., Arora, P., Jensen, M.D., Toffolo, G., Cobelli, C., and Rizza, R.A., *Mechanisms of the age-associated deterioration in glucose tolerance: contribution of alterations in insulin secretion, action, and clearance.* Diabetes, 2003. **52**(7): p. 1738-48.

5. Basu, R., Dalla Man, C., Campioni, M., Basu, A., Klee, G., Toffolo, G., Cobelli, C., and Rizza, R.A., *Effects of age and sex on postprandial glucose metabolism: differences in glucose turnover, insulin secretion, insulin action, and hepatic insulin extraction.* Diabetes, 2006. **55**(7): p. 2001-14.

6. Bauer, R.J., Guzy, S., and Ng, C., *A survey of population analysis methods and software for complex pharmacokinetic and pharmacodynamic models with examples.* AAPS J, 2007. **9**(1): p. E60-83.

7. Beal, S.L. and Sheiner, L.B., *Estimating population kinetics.* Crit Rev Biomed Eng, 1982. **8**(3): p. 195-222.

8. Beal, S.L., Sheiner, L.B., and Boeckmann, A.J., *NONMEM Users Guides*. 1989-2006, Icon Development Solutions, Ellicott City, Maryland, USA.

9. Beard, J.C., Ward, W.K., Halter, J.B., Wallum, B.J., and Porte, D., Jr., *Relationship of islet function to insulin action in human obesity.* J Clin Endocrinol Metab, 1987. **65**(1): p. 59-64.

10. Bell, B.M., *Approximating the marginal likelihood estimate for models with random parameters.* Applied Mathematics and Computation, 2001. **119**(1): p. 57-75.

11. Bergman, R.N., Ider, Y.Z., Bowden, C.R., and Cobelli, C., *Quantitative estimation of insulin sensitivity.* Am J Physiol, 1979. **236**(6): p. E667-77.

12. Bergman, R.N., Phillips, L.S., and Cobelli, C., *Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose.* J Clin Invest, 1981. **68**(6): p. 1456-67.

13. Bergman, R.N., Ader, M., Huecking, K., and Van Citters, G., *Accurate assessment of beta-cell function: the hyperbolic correction.* Diabetes, 2002. **51 Suppl 1**: p. S212-20.

14. Bonate, P.L., *The effect of collinearity on parameter estimates in nonlinear mixed effect models.* Pharm Res, 1999. **16**(5): p. 709-17.

15. Breda, E., Cavaghan, M.K., Toffolo, G., Polonsky, K.S., and Cobelli, C., *Oral glucose tolerance test minimal model indexes of beta-cell function and insulin sensitivity.* Diabetes, 2001. **50**(1): p. 150-8.

16. Buchanan, T.A., Metzger, B.E., Freinkel, N., and Bergman, R.N., *Insulin sensitivity and B-cell responsiveness to glucose during late pregnancy in lean and moderately obese women with normal glucose tolerance or mild gestational diabetes.* Am J Obstet Gynecol, 1990. **162**(4): p. 1008-14.

17. Caumo, A., Vicini, P., and Cobelli, C., *Is the minimal model too minimal?* Diabetologia, 1996. **39**(8): p. 997-1000.

18. Caumo, A., Vicini, P., Zachwieja, J.J., Avogaro, A., Yarasheski, K., Bier, D.M., and Cobelli, C., *Undermodeling affects minimal model indexes: insights from a two-compartment model.* Am J Physiol, 1999. **276**(6 Pt 1): p. E1171-93.

19. Centers for Disease Control and Prevention (CDC). Available from: http://apps.nccd.cdc.gov/DDTSTRS/default.aspx.

20. Chen, M., Bergman, R.N., Pacini, G., and Porte, D., Jr., *Pathogenesis of age-related glucose intolerance in man: insulin resistance and decreased beta-cell function.* J Clin Endocrinol Metab, 1985. **60**(1): p. 13-20.

21. Cobelli, C. and Caumo, A., *Using what is accessible to measure that which is not: necessity of model of system.* Metabolism, 1998. **47**(8): p. 1009-35.

22. Cobelli, C., *Presentation at the University of Virginia: Measurement of Insulin Sensitivity and β-cell Function from Intravenous and Oral Glucose Tolerance Tests: Necessity of Models*. 2005.

23. Cutfield, W.S., Jefferies, C.A., Jackson, W.E., Robinson, E.M., and Hofman, P.L., *Evaluation of HOMA and QUICKI as measures of insulin sensitivity in prepubertal children.* Pediatr Diabetes, 2003. **4**(3): p. 119-25.

24. Dalla Man, C., Caumo, A., and Cobelli, C., *The oral glucose minimal model: estimation of insulin sensitivity from a meal test.* IEEE Trans Biomed Eng, 2002. **49**(5): p. 419-29.

25. Davidian, M. and Giltinan, D.M., *Nonlinear models for repeated measurement data*. Monographs on statistics and applied probability 62. 1998, Boca Raton, Fla.: Chapman & Hall/CRC. xv, 359 p.

26. De Gaetano, A., Mingrone, G., and Castageneto, M., *NONMEM improves group parameter estimation for the minimal model of glucose kinetics.* Am J Physiol, 1996. **271**(5 Pt 1): p. E932-7.

27. DeFronzo, R.A., *Glucose intolerance and aging: evidence for tissue insensitivity to insulin.* Diabetes, 1979. **28**(12): p. 1095-101.

28. DeFronzo, R.A., Tobin, J.D., and Andres, R., *Glucose clamp technique: a method for quantifying insulin secretion and resistance.* Am J Physiol, 1979. **237**(3): p. E214-23.

29. DeFronzo, R.A., *Pathogenesis of type 2 diabetes mellitus.* Med Clin North Am, 2004. **88**(4): p. 787-835, ix.

30. Efron, B., *Nonparametric Estimates of Standard Error - the Jackknife, the Bootstrap and Other Methods.* Biometrika, 1981. **68**(3): p. 589-599.

31. Erichsen, L., Agbaje, O.F., Luzio, S.D., Owens, D.R., and Hovorka, R., *Population and individual minimal modeling of the frequently sampled insulin-modified intravenous glucose tolerance test.* Metabolism, 2004. **53**(10): p. 1349-54.

32. Festa, A., Williams, K., D'Agostino, R., Jr., Wagenknecht, L.E., and Haffner, S.M., *The natural course of beta-cell function in nondiabetic and diabetic individuals: the Insulin Resistance Atherosclerosis Study.* Diabetes, 2006. **55**(4): p. 1114-20.

33. Finegood, D.T., Hramiak, I.M., and Dupre, J., *A modified protocol for estimation of insulin sensitivity with the minimal model of glucose kinetics in patients with insulin-dependent diabetes.* J Clin Endocrinol Metab, 1990. **70**(6): p. 1538-49.

34. Gastaldelli, A., Miyazaki, Y., Pettiti, M., Matsuda, M., Mahankali, S., Santini, E., DeFronzo, R.A., and Ferrannini, E., *Metabolic effects of visceral fat accumulation in type 2 diabetes.* J Clin Endocrinol Metab, 2002. **87**(11): p. 5098-103.

35. Godsland, I.F., Agbaje, O.F., and Hovorka, R., *Evaluation of nonlinear regression approaches to estimation of insulin sensitivity by the minimal model with reference to Bayesian hierarchical analysis.* Am J Physiol Endocrinol Metab, 2006. **291**(1): p. E167-74.

36. Kahn, S.E., Larson, V.G., Beard, J.C., Cain, K.C., Fellingham, G.W., Schwartz, R.S., Veith, R.C., Stratton, J.R., Cerqueira, M.D., and Abrass, I.B., *Effect of exercise on insulin action, glucose tolerance, and insulin secretion in aging.* Am J Physiol, 1990. **258**(6 Pt 1): p. E937-43.

37. Kahn, S.E., Prigeon, R.L., McCulloch, D.K., Boyko, E.J., Bergman, R.N., Schwartz, M.W., Neifing, J.L., Ward, W.K., Beard, J.C., Palmer, J.P., and et al., *Quantification of the relationship between insulin sensitivity and beta-cell function in human subjects. Evidence for a hyperbolic function.* Diabetes, 1993. **42**(11): p. 1663-72.

38. Kahn, S.E. and Halban, P.A., *Release of incompletely processed proinsulin is the cause of the disproportionate proinsulinemia of NIDDM.* Diabetes, 1997. **46**(11): p. 1725-32.

39. Kahn, S.E., Andrikopoulos, S., and Verchere, C.B., *Islet amyloid: a long-recognized but underappreciated pathological feature of type 2 diabetes.* Diabetes, 1999. **48**(2): p. 241-53.

40. Kahn, S.E., *Beta cell failure: causes and consequences.* Int J Clin Pract Suppl, 2001(123): p. 13-8.

41. Kahn, S.E., *The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of Type 2 diabetes.* Diabetologia, 2003. **46**(1): p. 3-19.

42. Karlsson, M.O. and Savic, R.M., *Diagnosing model diagnostics.* Clin Pharmacol Ther, 2007. **82**(1): p. 17-20.

43. Katz, A., Nambi, S.S., Mather, K., Baron, A.D., Follmann, D.A., Sullivan, G., and Quon, M.J., *Quantitative insulin sensitivity check index: a simple, accurate method for assessing insulin sensitivity in humans.* J Clin Endocrinol Metab, 2000. **85**(7): p. 2402-10.

44. Krudys, K.M., Kahn, S.E., and Vicini, P., *Population approaches to estimate minimal model indexes of insulin sensitivity and glucose effectiveness using full and reduced sampling schedules.* Am J Physiol Endocrinol Metab, 2006. **291**(4): p. E716-23.

45. Magni, P., Sparacino, G., Bellazzi, R., Toffolo, G.M., and Cobelli, C., *Insulin minimal model indexes and secretion: proper handling of uncertainty by a Bayesian approach.* Ann Biomed Eng, 2004. **32**(7): p. 1027-37.

46. Magni, P., Sparacino, G., Bellazzi, R., and Cobelli, C., *Reduced sampling schedule for the glucose minimal model: importance of Bayesian estimation.* Am J Physiol Endocrinol Metab, 2006. **290**(1): p. E177-E184.

47. Mallows, C.L., *Some Comments on Cp.* Technometrics, 1973. **15**(4): p. 661-675.

48. Mari, A., Pacini, G., Murphy, E., Ludvik, B., and Nolan, J.J., *A model-based method for assessing insulin sensitivity from the oral glucose tolerance test.* Diabetes Care, 2001. **24**(3): p. 539-48.

49. Mari, A., Schmitz, O., Gastaldelli, A., Oestergaard, T., Nyholm, B., and Ferrannini, E., *Meal and oral glucose tests for assessment of beta -cell function: modeling analysis in normal subjects.* Am J Physiol Endocrinol Metab, 2002. **283**(6): p. E1159-66.

50. Matsuda, M. and DeFronzo, R.A., *Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp.* Diabetes Care, 1999. **22**(9): p. 1462-70.

51. McGarry, J.D., *Banting lecture 2001: dysregulation of fatty acid metabolism in the etiology of type 2 diabetes.* Diabetes, 2002. **51**(1): p. 7-18.

52. Miyazaki, Y., Glass, L., Triplitt, C., Wajcberg, E., Mandarino, L.J., and DeFronzo, R.A., *Abdominal fat distribution and peripheral and hepatic insulin resistance in type 2 diabetes mellitus.* Am J Physiol Endocrinol Metab, 2002. **283**(6): p. E1135-43.

53. Pacini, G., Thomaseth, K., and Ahren, B., *Contribution to glucose tolerance of insulin-independent vs. insulin-dependent mechanisms in mice.* Am J Physiol Endocrinol Metab, 2001. **281**(4): p. E693-703.

54. Peiris, A.N., Sothmann, M.S., Hennes, M.I., Lee, M.B., Wilson, C.R., Gustafson, A.B., and Kissebah, A.H., *Relative contribution of obesity and body fat distribution to alterations in glucose insulin homeostasis: predictive values of selected indices in premenopausal women.* Am J Clin Nutr, 1989. **49**(5): p. 758-64.

55. Pillonetto, G., Sparacino, G., Magni, P., Bellazzi, R., and Cobelli, C., *Minimal model S(I)=0 problem in NIDDM subjects: nonzero Bayesian estimates with credible confidence intervals.* Am J Physiol Endocrinol Metab, 2002. **282**(3): p. E564-73.

56. Pillonetto, G., Sparacino, G., and Cobelli, C., *Numerical non-identifiability regions of the minimal model of glucose kinetics: superiority of Bayesian estimation.* Math Biosci, 2003. **184**(1): p. 53-67.

57. Pouliot, M.C., Despres, J.P., Nadeau, A., Moorjani, S., Prud'Homme, D., Lupien, P.J., Tremblay, A., and Bouchard, C., *Visceral obesity in men. Associations with glucose tolerance, plasma insulin, and lipoprotein levels.* Diabetes, 1992. **41**(7): p. 826-34.

58. Prigeon, R.L., Kahn, S.E., and Porte, D., Jr., *Changes in insulin sensitivity, glucose effectiveness, and B-cell function in regularly exercising subjects.* Metabolism, 1995. **44**(10): p. 1259-63.

59. Riggs, D.S., Guarnieri, J.A., and Addelman, S., *Fitting straight lines when both variables are subject to error.* Life Sci, 1978. **22**(13-15): p. 1305-60.

60. Sparacino, G., Tombolato, C., and Cobelli, C., *Maximum-likelihood versus maximum a posteriori parameter estimation of physiological system models: the C-peptide impulse response case study.* IEEE Trans Biomed Eng, 2000. **47**(6): p. 801-11.

61. Steil, G.M., Volund, A., Kahn, S.E., and Bergman, R.N., *Reduced sample number for calculation of insulin sensitivity and glucose effectiveness from the minimal model. Suitability for use in population studies.* Diabetes, 1993. **42**(2): p. 250-6.

62. Steimer, J.L., Mallet, A., Golmard, J.L., and Boisvieux, J.F., *Alternative approaches to estimation of population pharmacokinetic parameters: comparison with the nonlinear mixed-effect model.* Drug Metab Rev, 1984. **15**(1-2): p. 265-92.

63. Stumvoll, M., Mitrakou, A., Pimenta, W., Jenssen, T., Yki-Jarvinen, H., Van Haeften, T., Renn, W., and Gerich, J., *Use of the oral glucose tolerance test to assess insulin release and insulin sensitivity.* Diabetes Care, 2000. **23**(3): p. 295-301.

64. The MathWorks. *Matlab 2007b*. Available from: http://www.mathworks.com/.

65. The R foundation. *The R Project for Statistical Computing*. 2.7.0: Available from: http://www.r-project.org/.

66. The RFPK team - University of Washington. *System for Population Kinetics (SPK)*. Available from: http://spk.rfpk.washington.edu.

67. Toffolo, G., De Grandi, F., and Cobelli, C., *Estimation of beta-cell sensitivity from intravenous glucose tolerance test C-peptide data. Knowledge of the kinetics avoids errors in modeling the secretion.* Diabetes, 1995. **44**(7): p. 845-54.

68. Toffolo, G., Cefalu, W.T., and Cobelli, C., *Beta-cell function during insulin-modified intravenous glucose tolerance test successfully assessed by the C-peptide minimal model.* Metabolism, 1999. **48**(9): p. 1162-6.

69. Turner, R.C., Holman, R.R., Matthews, D., Hockaday, T.D., and Peto, J., *Insulin deficiency and insulin resistance interaction in diabetes: estimation of their relative contribution by feedback analysis from basal plasma insulin and glucose concentrations.* Metabolism, 1979. **28**(11): p. 1086-96.

70. Utzschneider, K.M., Carr, D.B., Hull, R.L., Kodama, K., Shofer, J.B., Retzlaff, B.M., Knopp, R.H., and Kahn, S.E., *Impact of intra-abdominal fat and age on insulin sensitivity and beta-cell function.* Diabetes, 2004. **53**(11): p. 2867-72.

71. Utzschneider, K.M., Prigeon, R.L., Carr, D.B., Hull, R.L., Tong, J., Shofer, J.B., Retzlaff, B.M., Knopp, R.H., and Kahn, S.E., *Impact of differences in fasting glucose and glucose tolerance on the hyperbolic relationship between insulin sensitivity and insulin responses.* Diabetes Care, 2006. **29**(2): p. 356-62.

72. Van Cauter, E., Mestrez, F., Sturis, J., and Polonsky, K.S., *Estimation of insulin secretion rates from C-peptide levels. Comparison of individual and standard kinetic parameters for C-peptide clearance.* Diabetes, 1992. **41**(3): p. 368-77.

73. Vicini, P. and Cobelli, C., *The iterative two-stage population approach to IVGTT minimal modeling: improved precision with reduced sampling. Intravenous glucose tolerance test.* Am J Physiol Endocrinol Metab, 2001. **280**(1): p. E179-86.

74. Vicini, P., Gastonguay, M.R., and Foster, D.M., *Model-based approaches to biomarker discovery and evaluation: a multidisciplinary integrated review.* Crit Rev Biomed Eng, 2002. **30**(4-6): p. 379-418.

75. Vicini, P., Bell, B.M., Burke, J.V., Salinger, D.H., and the RFPK. *Assessing Nonlinear Mixed-Effects Model Parameter Estimates via Profiling of the True Likelihood.* . in *ACOP Meeting*. 2008. Tucson, AZ.

76. Wagenknecht, L.E., Langefeld, C.D., Scherzinger, A.L., Norris, J.M., Haffner, S.M., Saad, M.F., and Bergman, R.N., *Insulin sensitivity, insulin*

secretion, and abdominal fat: the Insulin Resistance Atherosclerosis Study (IRAS) Family Study. Diabetes, 2003. **52**(10): p. 2490-6.

77. Wahlby, U., Jonsson, E.N., and Karlsson, M.O., *Assessment of actual significance levels for covariate effects in NONMEM.* J Pharmacokinet Pharmacodyn, 2001. **28**(3): p. 231-52.

78. Wahlby, U., Jonsson, E.N., and Karlsson, M.O., *Comparison of stepwise covariate model building strategies in population pharmacokinetic-pharmacodynamic analysis.* AAPS PharmSci, 2002. **4**(4): p. E27.

79. Wallace, T.M., Levy, J.C., and Matthews, D.R., *Use and abuse of HOMA modeling*. Diabetes Care, 2004. **27**(6): p. 1487-95.

80. Ward, W.K., LaCava, E.C., Paquette, T.L., Beard, J.C., Wallum, B.J., and Porte, D., Jr., *Disproportionate elevation of immunoreactive proinsulin in type 2 (non-insulin-dependent) diabetes mellitus and in experimental insulin resistance.* Diabetologia, 1987. **30**(9): p. 698-702.

81. World Health Organization (WHO). Available from: http://www.who.int/diabetes/en/.

82. Yang, Y.J., Youn, J.H., and Bergman, R.N., *Modified protocols improve insulin sensitivity estimation using the minimal model.* Am J Physiol, 1987. **253**(6 Pt 1): p. E595-602.