UNIVERSITÀ
POLITECNICA
DELLE MARCHE

Master of Science in Biomedical Engineering

Individual estimation of physiological parameters of the glucose-insulin regulatory system: a modeling approach for reduced-sampling oral glucose tolerance test data

Supervisor:
Dott. Micaela MORETTINI

Candidate:
Libera Lucia DEL GIUDICE

Co-supervisor:
Dott. Andrea TURA

Dott. Agnese PIERSANTI

Academic Year 2022/2023

Abstract

Diabetes mellitus is a chronic metabolic disorder characterized by consistently elevated levels of blood glucose requiring daily insulin injections for affected individuals. Various diagnostic tests, both direct (such as the glucose clamp technique and insulin suppression test) and indirect (like the glucose tolerance test), are available for the diagnosis of diabetes mellitus. The gold standard for diagnosing diabetes is the Oral Glucose Tolerance Test (OGTT), which is less invasive and easier to perform than some other tests. However, it can be challenging to find OGTT data with more than 4 or 5 data points in research settings. Mathematical models within the realm of glucose metabolism have frequently been examined and applied. These models serve as an approximation of real-world phenomena by employing mathematical equations. In this domain, they elucidate various physiological processes by employing differential equations and logical deduction.

The primary objectives of this study were two-fold. Firstly, to systematically search for freely available datasets containing metabolic data. Overall, 16 studies were selected using a systematic process. These were later employed to validate a mathematical model describing the glucose-insulin system (G-I). Since this model involved a dense and intricate regularization procedure for estimating 13 parameters from 5-points OGTT, the present work aimed also to simplify the regularization process within the model while retaining the ability to estimate a vector comprising the 13 physiological parameters. The regularization was simplified by minimizing the sum of squared residuals, along with the second derivatives

of glucose and insulin weighted by appropriate values, denoted as *w1* and *w2*. Additionally, *w3* and *w4* were used to indicate the starting and ending points for the minimization process of the second derivatives.

The model was initially tested on two datasets with sufficient data for estimating the physiological parameter vector, using *w1*=2, *w2*=0.01, *w3*=3, and *w4*=ending point. The Root Mean Square Errors (RMSE) obtained from these initial estimates are 4.9717 and 1.1616, respectively. Subsequently, the model underwent validation using two datasets obtained through systematic research, each containing a limited number of data points. In one instance, the model was applied to a dataset of OGTT data from women with gestational diabetes, estimating all parameters while keeping w1, w2, and w4 values fixed and changing only the starting point for minimization (w3 = 1). In this case, the RMSE is 1.0258. The final validation involved a dataset with only 4 OGTT data points. In this case, there was the necessity of a reduction in the number of parameters to be estimated by fixing four of them. These four parameters were adjusted to match the values of parameters that were previously estimated using the most comprehensive OGTT datasets. Consequently, there are now two distinct sets of estimates. Using values of w1=0.068, w2=0.0015, w3=1, and w4=4, the RMSE are equal to 0.9678 and 1.3062, respectively. In conclusion, it is worth highlighting that the values obtained for the 13 parameters through this regularization method are in line with what is physiologically expected.

Summary

Introduction

Mathematical modeling in the field of glucose metabolism has a rich history. The mathematical model serves as an abstract representation of a real-world system, using mathematical concepts and language to approximate the reality. Typically, a model characterizes a system by defining a set of variables and equations that describe how these variables are interconnected. Given the increasing societal impact of Type 2 diabetes, the importance of developing and applying models in this context has grown significantly. There are various types of mathematical models employed in this domain, ranging from those based on *in vivo* methods to those utilizing *in vitro* approaches. An early method developed, used a tracer to study glucose kinetics [1]. Among the early studies that adopted this approach, Insel et al.[1] used a three-compartment model to analyze glucose kinetics. Subsequently, Radziuk et al.[2] examined the effectiveness of two-compartment models, and in 1979, Bergman et al.[3] introduced a minimal model that provides an indirect assessment of metabolic insulin sensitivity or resistance. The minimal model of glucose kinetics comprises a set of two interconnected differential equations involving four model parameters.

In the early 2000s, there was an expansion of the concept aimed at widening the application of the fundamental principle underlying the minimal model to include OGTT or Meal Glucose Tolerance Test (MGTT). Additionally, Caumo et al.[4] integrated the traditional minimal model for the rate at which glucose enters the bloodstream following an OGTT.

Within the field of glucose metabolism, mathematical models have also been developed to understand how incretin hormones influence insulin secretion. The models created by Dalla Man et al. [5],[6], represent extensions of conventional insulin secretion models, where they describe the impact of glucagon-like peptide 1 (GLP-1) as a multiplicative factor influencing the increase in insulin secretion.

It is possible to said that mathematical models play a crucial role in elucidating complex systems, assessing the effects of various components, and predicting their behaviour. They significantly contribute to researchers' understanding of physiological processes that otherwise are hardly measurable.

Indirect methods of measuring insulin resistance include OGTT and Intravenous Glucose Tolerance Tests (IVGTT). The OGTT is currently the standard test for diagnosing diabetes. It can take up to 2, 3 or 5 hours, and the blood sugar level is compared to a threshold to determine a diabetes diagnosis [7]. In a research context, it would be valuable to assess the glucose progression throughout the entire test [3]. This means collecting data at approximately 30-minute intervals. In clinical settings, only a few data points are usually collected, and the test typically lasts 2 hours, resulting in a data vector with no more than 4 to 5 points.

In the realm of research, having datasets with a larger number of samples would indeed be advantageous for implementing mathematical models capable of making precise estimates. However, this may not always align with the types of datasets typically available and consequently, regularization procedures are frequently employed.

Regularization is the practice of incorporating additional information into the solution of a potentially ill-posed problem, thereby mitigating the risk of overfitting [8]. By using regularization techniques, it is possible to strike a balance between the complexity of the mathematical models and the limited data available, helping to ensure that the models are both accurate and robust in making estimates.

In this study, the model proposed by Contreras et al.[9] was adopted, which delves into the kinetics of G-I system through the use of five differential equations. However, due to the limited data available in comparison to the number of physiological parameters that needed to be estimated, the model required the application of a regularization techniques. Given the intricate nature of the regularization procedure employed by [9], the objective of this study was to streamline their regularization process while still retaining the ability to estimate all the physiological parameters accurately.

The present work was conducted in two distinct phases. During the initial phase, the primary focus was on systematically searching for datasets containing metabolic data. The aim was to collect datasets that were freely available for use in the subsequent phase of the research. In the second phase, the study shifted its attention towards implementing the mathematical model of [9], simplifying the regularization process, and validating the model's performance. This validation was carried out using two of the datasets that had been acquired during the prior systematic data collection phase.

Chapter 1. The Glucose-Insulin regulatory system and the data for its study

The endocrine system is the primarily responsible of the control of metabolic pathways. While hormones play a crucial role in the ongoing regulation of metabolism, immediate adjustments fundamentally hinge on the balance between insulin and glucagon, both of which are released by the pancreatic endocrine cells. The pancreas is an elongated and nodular glandular organ situated in the abdominal region and connected to the digestive system. Apart from its acinar glandular tissue that produces pancreatic juice, it also contains endocrine tissue responsible for hormone production. This endocrine tissue comprises small, separate structures known as *Langerhans islets*, which are histologically different from the acinar tissue. These islands consist of two primary types of endocrine-active cells known as α and β cells, alongside a smaller population of various cell types called δ and φ cells. The β cells are responsible for the secretion of insulin, whereas α cells secretes glucagon. Both hormones, with a particular emphasis on insulin, play a crucial role as regulatory factors in the metabolism of all tissues [10].

1.1. The Glucose-Insulin regulatory system

Insulin is a polypeptide hormone, and it is primarily responsible for regulating the metabolism of glucose, lipids, and proteins, making it a pivotal hormone in maintaining the body's metabolic homeostasis. Insulin exerts a broad range of effects on cellular metabolism in nearly all tissues, with a particular emphasis on the liver, skeletal muscle, and adipose

tissue, which can be regarded as its primary target tissues. One of the most noticeable outcomes of insulin's influence on glucose metabolism is the rapid decrease in blood sugar levels (glycemia). Insulin secretion by β cells is controlled by the concentration of glucose in the bloodstream. This regulation mechanism involves the endocrine cells of the pancreas functioning as sensors for plasma glucose levels. Their response entails that a decline in blood sugar suppresses the release of the hormone, while an increase in blood sugar stimulates its secretion. For instance, during extended periods of fasting, the resulting low blood sugar levels lead to a roughly 50% reduction in average plasma insulin levels. A similar response is observed during intense and prolonged muscle exercise. In these situations, there is a counterregulatory regulation of insulin aimed at maintaining a consistent glycemic range between 70 and 110 mg/100 mL of plasma values. This ensures a steady delivery of glucose to the body's tissues, even though there are substantial fluctuations in glucose intake and utilization throughout the day, especially in relation to meal breaks. The effectiveness of this regulatory mechanism has its limitations when blood glucose levels exceed the range of 300-350 mg/100 mL of plasma or is below 45-50 mg/100 mL (this value represents the blood glucose threshold). Beyond this threshold, the body's regulatory mechanisms may struggle to maintain glucose homeostasis effectively.

After being secreted by β cells, insulin enters the abdominal portal vein and proceeds to the liver. In the liver, a portion (approximately half) of insulin is extracted by hepatocytes before it can enter the circulation outside the liver. Insulin clearance refers to its complete removal from the bloodstream. There have been suggestions that reduced blood clearance could raise the risk of developing diabetes mellitus, while hyperinsulinemia may be associated with

the development of cognitive impairments, including Alzheimer's disease, as well as certain types of cancer [10].

1.2. Diabetes

Diabetes is a persistent metabolic disorder marked by elevated levels of blood glucose, commonly known as blood sugar. Consequently, those afflicted with this ailment depend on one or more daily insulin injections. It implies a prolonged potential for severe complications, such as kidney disease, eye problems, nerve damage, and cardiovascular issues. Although it cannot be permanently cured, this form of diabetes can be effectively controlled to postpone its consequences [11]. Diabetes mellitus is characterized by noticeable symptoms, including heightened urination, increased thirst, and hunger, often accompanied by a reduction in body weight. However, the earliest and most distinctive indicator is a sustained elevation in blood sugar levels, which are already elevated during fasting and significantly rise after meals. The elevation in blood sugar levels results from the enhancement of glucogenic processes in the liver, which are no longer suppressed by the presence of insulin. Furthermore, due to the hormonal deficiency, glucose is unable to enter most tissue cells despite its high concentration in the bloodstream. Consequently, excess extracellular glucose coexists with a shortage of intracellular glucose. Additionally, the rise in glycemia, a hallmark of diabetes mellitus, has significant implications for kidney function [10].

The diagnosis of diabetes is established by identifying the presence of hyperglycemia. The revised criteria [12] include:

1. Fasting plasma glucose (FPG) level of ≥ 7.0 mmol/L;

2. A 2-hour postload glucose level > 11.1 mmol/L during an OGTT;

3. Presence of diabetes symptoms along with a casual (regardless of the time of the preceding meal) plasma glucose level of ≥ 11.1 mmol/L.

However, leading organizations such as the American Diabetes Association (ADA) [12], European Association for the Study of Diabetes (EASD) [13], International Diabetes Federation (IDF) [14], and World Health Organization (WHO) [15] have endorsed the use of glucated haemoglobin (HbA1c) levels for diabetes diagnosis. If any of these criteria are met, confirmation is necessary to establish the diagnosis. Confirmation can be achieved through various methods:

- Repeating the same test (either glucose or HbA1c) on a different blood sample taken on a subsequent day;
- Employing a different type of test as the confirmatory one compared to the initial assay. For example, if the initial measurement is glucose, HbA1c can serve as the confirmatory test in a subsequent sample, or vice versa;
- Measuring two different analytes, namely glucose and HbA1c, in samples collected on the same day.

It's important to note that repeat testing is not required in symptomatic individuals who have unequivocal hyperglycemia, meaning their glucose levels are consistently greater than 11.1 mmol/L (200 mg/dL) [16].

There are three types of diabetes: Type 1, Type 2, and gestational diabetes. The form of diabetes in which a primary deficiency in insulin predominates as a key pathogenic factor is

known as Type 1 diabetes, often referred to as juvenile diabetes. On the other hand, Type 2 diabetes, also known as adult diabetes, has a different pathogenesis and typically does not occur before the age of forty. In Type 2 diabetes, insulin levels are often reduced but can also be within the normal range or even elevated. Additionally, there is gestational diabetes, which affects women during the second or third trimester of pregnancy due to hormonal and physiological changes, including weight gain. This condition poses risks for both the child and the mother.

1.2.1 Type 1 diabetes

The Type 1 diabetes makes up just 5-10% of all diabetes cases and was formerly referred to as insulin-dependent diabetes or juvenile-onset diabetes [16]. The etiology of Type 1 diabetes can be explained by damage to the pancreatic cells due to environmental or infectious agents. In individuals who are susceptible to genetic alterations, the immune system is triggered to produce an immune response against altered β cells, or against molecules in β cells that are similar to viral proteins [17]. In this type of diabetes, the rate at which β cell destruction occurs varies considerably. It can be rapid in some individuals, especially infants and children, while in others, mainly adults, it progresses more slowly. Some patients, especially children and teenagers, may experience ketoacidosis as the initial symptom of the disease, others, may have mild fasting hyperglycemia that can quickly escalate to severe hyperglycemia or ketoacidosis, especially when they are dealing with an infection or other forms of stress. Additionally, some adults may retain a partial function of β cells, which is sufficient to prevent ketoacidosis for many years. However, over time, these individuals will eventually become reliant on insulin for survival and are at risk of developing

ketoacidosis [16]. Approximately the 80% of patients with Type 1 diabetes show circulating islet cell antibodies, and most of these patients have anti-insulin antibodies before receiving insulin therapy [17]. Since autoimmunity is considered the primary factor in the pathophysiology of Type 1 diabetes there is demonstrated that there is a significant association between Type 1 diabetes and other autoimmune conditions such as Graves' disease, Hashimoto's thyroiditis, and Addison's disease [18]. When these autoimmune disorders coexist, the prevalence of Type 1 diabetes tends to rise. Also, vitamin D plays a crucial role in both the development and prevention of Type 1 diabetes, as suggested by recent evidence[19]. Vitamin D deficiency independently predicts the onset of coronary artery disease in individuals with Type 1 diabetes. Moreover, another study has demonstrated that vitamin D deficiency in Type 1 diabetes can serve as a predictor for all-cause mortality [20].

1.2.2 Type 2 diabetes

The Type 2 diabetes, which constitutes 90-95% of all diabetes cases, includes individuals with insulin resistance and typically features a relative, rather than absolute, insulin deficiency. This type of diabetes often remains undiagnosed for extended periods because hyperglycemia develops slowly, and in its early stages, it may not be severe enough for patients to experience the typical symptoms of diabetes. Nonetheless, individuals with this condition are at an elevated risk of developing both macrovascular and microvascular complications [16].

Insulin resistance results in increased levels of fatty acids in the bloodstream, which, in turn, leads to reduced glucose transport into muscle cells and an increase in fat breakdown. This process subsequently leads to higher glucose production by the liver. For Type 2 diabetes to develop, both insulin resistance and dysfunction of the pancreatic β cells must occur simultaneously. Individuals who are overweight or obese typically experience some level of insulin resistance, but diabetes only manifests in those who do not produce enough insulin to match the degree of insulin resistance. In these individuals, insulin levels may be elevated, but they are still insufficient to normalize blood glucose levels [17].

1.2.3 Gestational diabetes

Gestational diabetes is a condition that develops during the second and third trimesters of pregnancy. It is characterized by significant insulin resistance, which is a result of hormonal changes triggered by the placenta. This activity emphasizes the etiology, epidemiology, pathophysiology, treatment, complications, and prognosis of gestational diabetes. The etiology of gestational diabetes appears to be linked to two main factors: the dysfunction of pancreatic β cells or a delayed response of these cells to changes in blood sugar levels, and a pronounced insulin resistance resulting from hormonal releases by the placenta during pregnancy. The primary hormone associated with increased insulin resistance in gestational diabetes is human placental lactogen. Other hormones contributing to insulin resistance and hyperglycemia during pregnancy include growth hormone, prolactin, corticotropin-releasing hormone, and progesterone. Several clinical risk factors have been identified for the development of gestational diabetes like an increased body weight, often indicated by a body mass index (BMI) greater than 25, educed physical activity levels, a family history of diabetes

mellitus in a first-degree relative, as well as hemoglobin A1C levels greater than 5.7, any significant marker of insulin resistance, such as acanthosis nigricans (a skin condition characterized by dark, thickened patches) [21]. Recommendations for screening the gestational diabetes typically involve conducting a screening test between 24 to 28 weeks of pregnancy. This test is usually a 50g, 1-hour oral glucose challenge test. If the results from this initial test are abnormal, meaning the glycemia is equal to or greater than 130 mg/dL (7.22 mmol/L), or equal to or greater than 140 mg/dL (7.77 mmol/L), a confirmatory test is necessary. The following criteria are used to diagnose gestational diabetes:

- Fasting blood glucose level ≥ 95 mg/dL;
- Blood glucose level after 1 hour ≥ 180 mg/dL;
- Blood glucose level after 2 hours ≥ 155 mg/dL;
- Blood glucose level after 3 hours ≥ 140 mg/dL;

The presence of two or more abnormal results in the 3-hour OGTT confirms the diagnosis of gestational diabetes [22].

1.3. Data for screening test

In the past decades, the connection between insulin resistance and type 2 diabetes has been widely acknowledged. Insulin resistance holds great significance in this context, not only as the most influential predictor for the future development of Type 2 diabetes, but it also becomes a target for treatment once hyperglycemia is present [23]. The global epidemic of Type 2 diabetes and impaired glucose tolerance is a significant contributor to illness and death worldwide. In these conditions, tissues like muscles, fat, and the liver become less

responsive or resistant to insulin. It was revealed that numerous other hormones and signaling events can dampen insulin's effects and play a crucial role in the development of Type 2 diabetes [24]. However, insulin resistance is a challenging problem that frequently exacerbates metabolic syndrome. It is commonly described as a reduced sensitivity or reduced response to the metabolic effects of insulin, such as insulin's ability to facilitate glucose disposal and inhibit hepatic glucose production (HGP) [25].

There are various techniques and measures accessible for assessing insulin resistance. It is imperative to examine and confirm their accuracy before employing them as investigative tools in patient assessments [26]. It is possible to distinguish between direct measures and indirect measures for estimating insulin resistance. Consequently, there exist various types of outcome data.

1.3.1 Direct measures

1.3.1.1 Glucose Clamp Technique

Glucose Clamp technique is a method for directly quantifying insulin secretion and insulin resistance. Two types of clamps are commonly used: the hyperglycemic clamp and the hyperinsulinemic or euglycemic clamp. Nowadays the hyperglycemic clamp technique is referred as the “gold standard” test for this purpose. The procedure proposed by DeFronzo et al.[27] it is performed after a 12-h overnight fast and aims to increase the plasma glucose concentration at a steady plateau while keeping it for 2 hours by using an intravenous glucose infusion. The entire process consists of two phases named “priming dose” and “maintenance dose”. During the first 15 min phase the dose required for increasing the plasma glucose

concentration by 125 mg/dL is computed per square meter of body surface area. Furthermore, the maintenance dose is given every 5 minutes, until the end of the study, and involves periodic adjustments of the glucose infusion based on the feedback mechanism [27]. After several hours of continuous insulin administration, steady-state conditions can be reached for insulin levels in the bloodstream, blood glucose levels, and the rate at which glucose is infused (referred to as glucose infusion rate, GIR). Assuming that the excessive presence of insulin effectively prevents the liver from releasing glucose and, given that there are no overall alterations in blood glucose levels during stable clamp conditions, the GIR must be equal to the rate at which glucose is utilized and removed from the body (M). The value of M is often standardized by the individual's body weight or fat-free mass in order to produce an estimate of insulin sensitivity [25]. Typical outcome of the test is described in Figure 1.

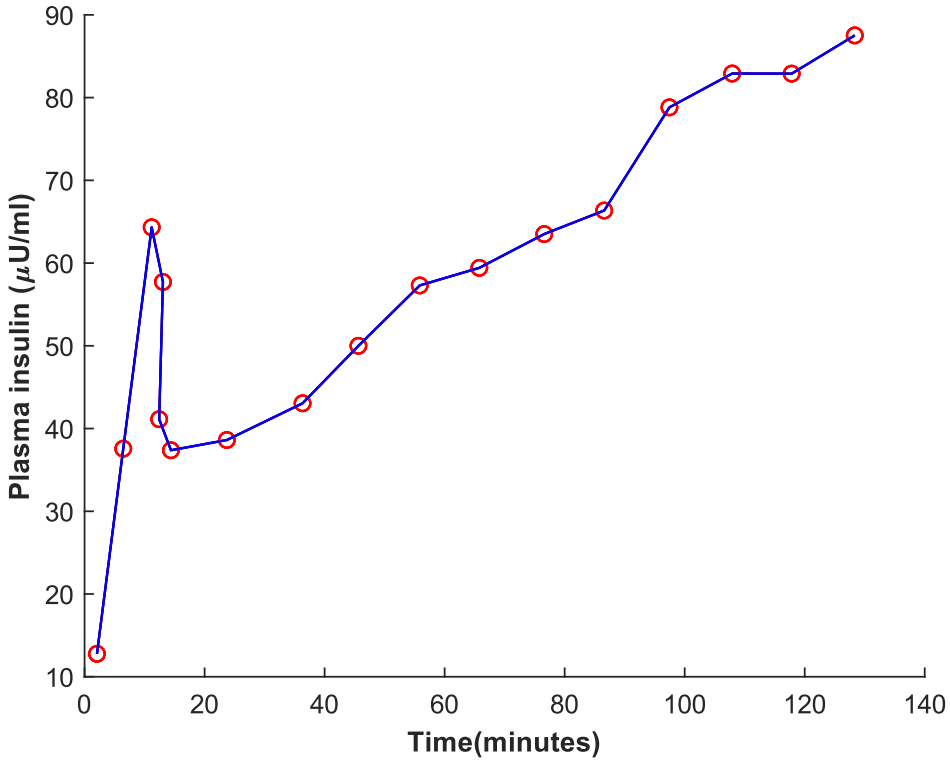

Figure 1: Plasma insulin response (μU/ml) during a hyperglycemic clamp test.

The plasma insulin response to the sustained hyperglycemia exhibited a two-phase pattern, with an initial rapid release of insulin followed by a phase of gradually increasing insulin concentration that persisted until the end of the study.

The hyperinsulinemic-euglycemic clamp test is infrequently conducted in clinical practice but is a valuable tool in medical research, particularly for evaluating the impact of various medications. This procedure typically spans about two hours. It involves administering insulin through a peripheral vein at a controlled rate ranging from 10 to 120 mU per m^2 per minute. To counterbalance the insulin infusion, a 20% of glucose solution is also administered to maintain blood sugar levels within the range of 5 to 5.5 mmol/L. The rate of glucose infusion is adjusted based on regular blood sugar checks conducted every five to ten minutes. Insulin sensitivity is determined by the rate of glucose infusion during the last half-hour of the test. High infusion rates (7.5 mg/min or higher) indicate that the patient is responsive to insulin (insulin-sensitive), while very low rates (4.0 mg/min or lower) suggest insulin resistance. Rates between 4.0 and 7.5 mg/min are inconclusive and may indicate "impaired glucose tolerance," an early sign of insulin resistance. This fundamental technique can be significantly improved by incorporating glucose tracers. Glucose can be labeled with stable or radioactive atoms. Before commencing the hyperinsulinemic phase, a 3-hour tracer infusion is employed to establish the basal rate of glucose production. During the clamp, the plasma tracer concentrations enable the calculation of whole-body insulin-stimulated glucose metabolism and the body's glucose production (endogenous glucose production) [25].

1.3.1.2 Insulin Suppression Test

Another method for the direct measure of insulin sensitivity/resistance is the Insulin Suppression Test (IST) consisting of an intravenous infusion of somatostatin or its analogue, the octreotide, employed following an overnight fast to inhibit the natural release of insulin and glucagon. At the same time, insulin and glucose are introduced into the bloodstream through the same vein and the process continues for a duration of 3 hours. Blood samples for measuring glucose and insulin levels are collected every 30 min for 2.5 hours from the opposite arm. Subsequently, samples are taken at 10-minute intervals between the 150th and 180th minute of the IST. The continuous infusion of insulin and glucose determines the steady-state plasma insulin (SSPI) and glucose concentrations (SSPG). As a result, individuals with insulin resistance will exhibit elevated levels of SSPG concentration, while those who are sensitive to insulin will demonstrate lower levels. Therefore, this test offers a direct assessment of how effectively external insulin facilitates the removal of an intravenously administered glucose load within stable conditions, during which the body's natural insulin release is inhibited [25].

1.3.2 Indirect measures

1.3.2.1 Oral Glucose Tolerance Test

The indirect measures of insulin sensitivity nowadays mostly used is the Glucose Tolerance Tests (GTT). The straightforward examination commonly employed in clinical settings to diagnose both type 2 diabetes and glucose intolerance is the OGTT. Following an overnight period of not eating and the consumption of either a standard glucose solution or a typical meal, blood samples are gathered at 0, 30, 60, and 120-minute intervals to analyze glucose and insulin levels in the bloodstream. It measures the body's capacity to process sugar [25]. The examination is conducted early in the morning (between 07:00 and 08:30) for practical reasons and because glucose metabolism follows a circadian rhythm, with its peak in the morning and lowest levels in the afternoon. The initial sample, referred to as the baseline, is collected to establish the starting point, which is considered as time 0'. At this point, the laboratory checks the blood sugar level. If the glucose concentration is below 126 mg/dL, glucose is administered. The patient is then given a solution containing 75g of glucose dissolved in 300-500 mL of water, which must be consumed within a maximum of 5 minutes. Following this ingestion, blood samples are collected at various time intervals. A "two-stage" OGTT involves taking baseline samples (at 0') and another at 120 minutes. Alternatively, a "five-stage" OGTT can be conducted, involving baseline sampling and additional samples every 30 minutes up to 2 hours: 0', 30', 60', 90', and 120'. In some cases, a "six-stage" sampling may be performed, extending the test up to 180 minutes: 0', 30', 60', 90', 120', and 180'. It's important to note that this test typically includes measuring insulin levels

(insulinemia) alongside blood sugar levels, with two blood tubes collected at each sampling time [7]. There are studies that have used glucose and C-peptide minimal models for measuring insulin action, β cell function, and the rate of meal glucose appearance using a 300-min OGTT and 420-min meal protocols. However, it was demonstrated by Chiara Dalla Man et al. [28] that is possible to use a two-hours-seven-sample OGTT (taken at 0,10,20,30,60,90,120) to have a good measurement of insulin sensitivity.

Administering glucose through the oral route is evidently more natural compared to intravenous glucose injection or the continuous infusion of insulin during hyperinsulinemic clamp procedures. Nonetheless, evaluating insulin's effects following the intake of glucose or a combination meal is more challenging than administering glucose intravenously [29]. The OGTT replicates the natural patterns of glucose and insulin in the body's physiological state more accurately than the glucose clamp, and IST procedures [25]. A typical shape of the of graph regulating the glucose disappearance during an OGTT is described in Figure 2.

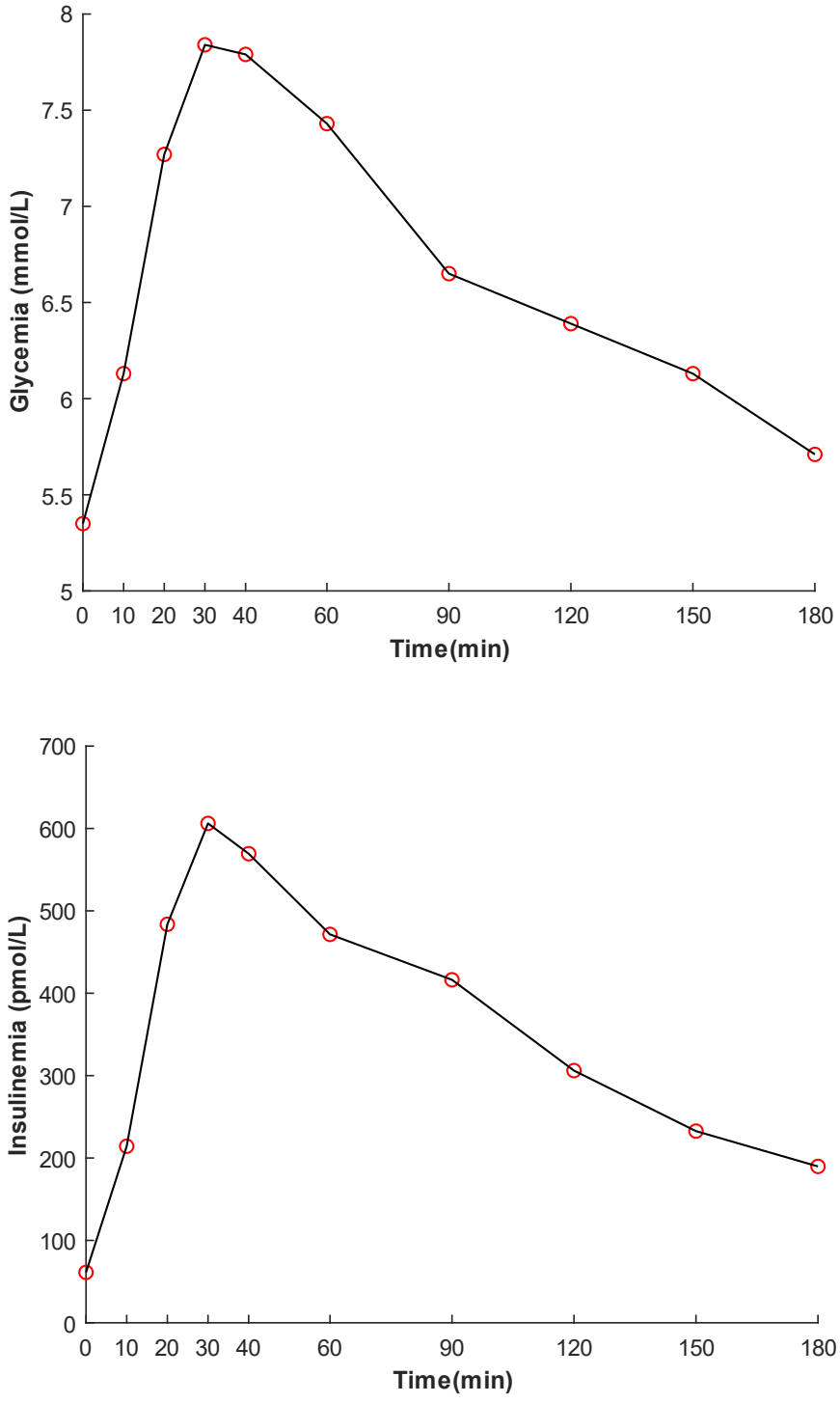

Figure 2: Top: Shape of the Glycemia (mmol/L) following a 3-OGTT. Bottom: Shape of the Insulinemia (pmol/L) following a 3-OGTT.

It is possible to distinguish four conditions:

- Basal condition (Fasting Glucose): the starting point of the graph represents the level of glucose in the fasting blood, that is, before the intake of the glucose solution;
- Glucose administration (Acute Glucose Response): immediately after taking the glucose solution, blood glucose levels will begin to rise. This initial peak reflects the body's response to glucose intake;
- Insulin Response: in response to increased glucose levels, the pancreas releases insulin to help cells absorb glucose from the blood. This leads to a decrease in blood glucose levels. The insulin spike should generally occur shortly after the glucose spike;
- Long-Term Response: over time, blood glucose levels should return to baseline values. However, in some people with glucose regulation problems, levels may remain high for a longer period.

In addition, the shape of the glucose curve in an OGTT was labeled as "monophasic" when the plasma glucose levels increased following an oral glucose load, reaching the highest point between 30 to 90 minutes, and then decreased until 120 minutes with a final decrease of at least 0.25 mmol/L between 90 and 120 minutes. Glucose curves that initially increased, then dropped to a minimum, and increased again by at least 0.25 mmol/L until 120 minutes were categorized as "biphasic" [30]. Additionally, Ismail et al.[31] further divided the biphasic group into two subcategories based on when the second rise in glucose levels occurred after the initial decrease:

- Biphasic90: this subgroup is characterized by a plasma glucose drop of ≥ 0.25 mmol/L at 60 minutes following the initial increase at 30 minutes, followed by a subsequent increase from the 60 to 90-minute time points by ≥ 0.25 mmol/L.
- Biphasic120: in this subgroup, the plasma glucose at 60 minutes dropped by ≥ 0.25 mmol/L after the initial increase at 30 minutes, and then it increased again from the 90 to 120-minute time points by ≥ 0.25 mmol/L.

1.3.2.2 Intravenous Glucose Tolerance Test

The IVGTT, as suggested by the name, requires intravenous infusion of the bolus of glucose and it is an alternative to the clamp techniques. In this test, the observed changes in plasma insulin levels over time are considered the "input," while the variations in plasma glucose levels are seen as the "output" of the system that manages the removal of glucose from the body [3].

1.4. Continuous Glucose Monitoring

Continuous Glucose Monitoring (CGM) systems offer a comprehensive overview of blood sugar levels in individuals dealing with diabetes. Informed patients can utilize this information to appropriately respond to their glucose levels, thus preventing instances of both low and high blood sugar. The sensors within all presently accessible CGM systems monitor glucose levels within the subcutaneous interstitial fluid for a period ranging from 6 to 14 days. An implantable sensor can extend this measurement duration to as long as 6 months. The CGM systems are typically composed of three primary elements, as shown in Figure 3:

- A glucose oxidase (GOD) based glucose sensor is inserted into the subcutaneous adipose tissue. This sensor continuously gauges the glucose concentration in the interstitial fluid.
- A transmitter is affixed to the sensor. Its role is to relay the collected data.
- A receiver or smartphone is employed to display the resulting information.

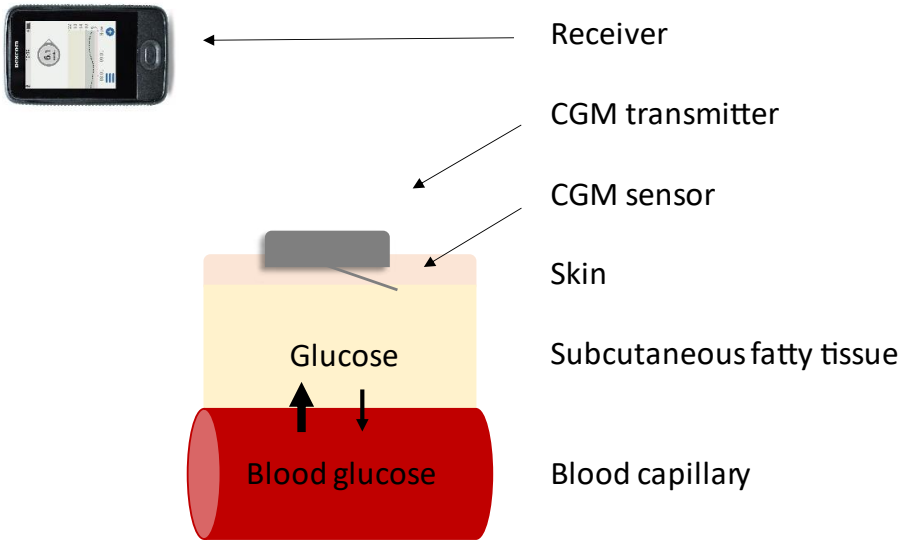

Figure 3: Redesigned schema of the components of a CGM system, from [32].

The estimation of glucose concentration hinges on the mechanism involving the production of hydrogen peroxide by GOD. This process is coupled with the release of an electric current, which is directly proportional to the glucose concentration within the interstitial fluid. Depending on the specific system being used, glucose levels are displayed either in real-time (referred to as rtCGM systems) or upon scanning (known as iscCGM systems). These systems offer various features including alerts, alarms, trend arrows, and data visualization, all of which encourage individuals to independently manage their diabetes treatment. The frequent stream of glucose data and the assortment of functions available necessitate thorough training for individuals with diabetes and their caregivers [32]. The distinctions between the two system types lie in the fact that rtCGM systems oversee glucose levels and promptly exhibit a recent reading, typically at 5-minute intervals. Conversely, the sensors of iscCGM systems record glucose levels every minute and store one measurement every 15 minutes. To retrieve glucose data from iscCGM systems and showcase it on the device screen, active scans are necessary. These scans should be conducted at least once every 8 hours to preserve a comprehensive record of daily glycemic data. The glucose values obtained through scanning iscCGM systems can be either transferred to a personal computer or uploaded to a cloud-based platform [32]. In today's context, there is a clear acknowledgment of the role played by usability and the human interface in the world of medical devices. Many studies have been conducted to enhance the usability of various medical devices, and manufacturers have taken significant steps to improve this aspect. However, when it comes CGM devices and sensors, there is a noticeable dearth of published studies addressing usability and human interface concerns.

What is required are some traditional time-and-motion studies to delve into several key aspects:

- The amount of time taking for clinicians to learn the intricacies of CGM device operation and how to instruct patients effectively in its use;
- The time and training investment required for patients to become proficient in using the device, including tasks like sensor insertion and removal, daily usage routines, and the transmission of data;
- The time and training needed by physicians to perform data analysis and interpret the results accurately;
- An assessment of how effectively the information gathered from CGM devices translates into actions and behaviours that lead to measurable improvements in clinical outcomes;

By conducting these time-and-motion studies, it is possible to gain a better understanding of the practical challenges and training needs associated with CGM devices, ultimately leading to more effective use and better patient outcomes.

The original implantable sensor developed by Dexcom was not brought to the commercial market, primarily due to significant limitations in the accuracy and precision of CGM technology at that time. However, there has been a remarkable improvement in the accuracy and precision of CGM technology over the years. The CGM data is now considered accurate enough across a wide range of glucose values to support tasks such as self-adjustment of insulin dosage, the detection of hypoglycemia, and the assessment of a patient's response

to therapy. It is important to note that the accuracy of CGM readings is closely linked to the specific glucose level being measured and the rate of change in glucose levels. In the present day, CGM technology boasts greater accuracy than what was available with blood glucose meters when they were initially introduced approximately 35 years ago, even though those meters were already being utilized for self-adjustment of insulin dosages. CGM devices offer additional benefits, including graphical displays, information on the rate of glucose level changes, and alarms, all of which can significantly enhance the overall effectiveness of CGM in managing diabetes and improving patient outcomes. In Figure 4 there is a typical Dexcom sensor now used with its specific app [33].

Figure 4: Example of the DexcomG6 CGM device with its app.

Chapter 2. Mathematical models for the study of the Glucose-Insulin regulatory system

Mathematical modeling in the realm of glucose metabolism has a rich history. To begin with, it is important to understand that a mathematical model serves as an abstract representation of a real-world system, employing mathematical concepts and language. It serves as a depiction of reality, but in an approximate form. Typically, a mathematical model characterizes a system through a collection of variables and a set of equations that define the interconnections among these variables. The variables can take various forms and signify specific attributes or characteristics of the system. The existing model consists of a series of functions that articulate the associations between these diverse variables [34]. The employment of models is driven by various factors. They have been utilized to compute parameters of physiological significance from experimental data, either indirectly or to offer a precise quantitative portrayal of the physiopathological mechanisms. Additionally, they have been employed to determine clinical utility indices derived from straightforward experimental tests.

With the growing societal impact of Type 2 diabetes, a condition involving disruptions in the glucose homeostasis system, the importance of developing and utilizing models in this context has increased substantially. There are various types of mathematical models applied in this domain, ranging from those rooted in *in vivo* approaches to those employing *in vitro* methods. While most mathematical models rely on ordinary differential equations, the process of translating physiological mechanisms into mathematical equations can follow

various logical paths [35]. One of the initial methods employed, involved utilizing a tracer to investigate glucose kinetics [1]. Among the early studies adopting this method, a three-compartment model was used by Insel et al.[1] to analyse the glucose kinetics. This model of glucose was then merged with another three-compartment model, specifically designed to describe insulin kinetics. The outcome was a comprehensive model that mathematically encapsulated both insulin regulation and glucose utilization. Significantly, this model incorporated both insulin-dependent and insulin-independent processes. It introduced a noteworthy concept: the impact of insulin on glucose utilization is characterized by a delay compared to the profile of plasma insulin concentration [1]. Nevertheless, Radziuk et al.[2] subsequently examined the efficacy of two-compartment models. These models did not presuppose a structural connection between glucose utilization and insulin concentration but were instead employed to compute the temporal pattern of glucose utilization based on glucose and tracer concentrations. The exploration of glucose utilization is closely associated with the measurement of insulin sensitivity, which refers to the capacity of insulin to enhance glucose utilization and inhibit glucose production.

Bergman et al. [3], in 1979, introduced a minimal model that offers an indirect assessment of metabolic insulin sensitivity or resistance. This analysis relies on data related to glucose and insulin collected during an IVGTT. The minimal model of glucose kinetics consists of a set of two interconnected differential equations encompassing four model parameters. The first equation characterizes the dynamics of plasma glucose within a solitary compartment, while the subsequent equation delineates the dynamics of insulin within a "remote compartment". This structured configuration of the minimal allows to ascertain the model

parameters that define the optimal alignment with the process of glucose disappearance observed throughout the IVGTT [25]. In contrast to the reliance on steady-state conditions seen in the glucose clamp and insulin sensitivity test, the minimal model approach operates with dynamic data. Moreover, it offers the possibility to estimate another parameter, the "glucose effectiveness" which is defined as the ability of glucose *per se* to enhance its own metabolism independently from any change in insulin concentration, suppressing the hepatic glucose production and stimulating glucose uptake by the peripheral tissues [3]. The minimal model has garnered significant interest for a variety of reasons, capturing the attention of both the modeling and experimental communities and achieving iconic status [35].

However, during the early 2000s, there was a concept expansion aimed at broadening the scope of the fundamental principle underlying the minimal model to apply it to an OGTT or a MTT. Specifically, Caumo et al.[4] integrated the traditional minimal model, that explains glucose kinetics, with an equation that accounts for the rate at which glucose enters the bloodstream, following an OGTT. In this specific research, they compared the insulin sensitivity index derived from the MTT of 10 normal individuals with the values obtained from the same individuals through an insulin-modified, frequently sampled intravenous glucose test (FSIGT). In summary, this research demonstrates that in individuals without any underlying health conditions, it is feasible to employ the minimal model in the context of a MTT/OGTT to calculate the insulin sensitivity index that closely aligns with the one determined through the FSIGT [4].

Progress in understanding how insulin impacts glucose utilization, as initially outlined by [1], has been achieved through the development of various models aimed at explaining two well-established phenomena: the non-linear relationship between insulin concentration and glucose utilization, as well as the influence of glucose concentration. Given that the underlying causes of these outcomes remain uncertain and there is limited empirical data supporting the mathematical models, these phenomena are frequently omitted from representations of insulin action, or when included, they are based on conceptual reasoning rather than direct experimental validation. To capture the non-linear relationship between glucose utilization and glucose concentration, a model by Bizzotto et al.[36] was employed. This model examined the connection between insulin concentration and glucose utilization, integrating it with a Michaelis-Menten equation that represents glucose utilization as a function of glucose concentration [36]. This modeling approach successfully explained numerous tracer-based studies encompassing a broad spectrum of glucose and insulin concentrations. It underscored the significance of the non-linear correlation between glucose utilization and glucose concentration as a quantitatively significant phenomenon in glucose homeostasis.

Within the realm of glucose metabolism, mathematical models have also been devised to elucidate how incretin hormones affect insulin secretion. The models developed by Dalla Man et al.[5],[6], represent extensions of conventional insulin secretion models, where they characterized the impact of GLP-1 as a multiplicative factor influencing the increase in insulin secretion. They put forward four different models to depict how this factor's reliance on GLP-1 concentration evolves during a hyperglycemic clamp study accompanied by GLP-1 infusion.

Their conclusion was that a linear relationship with both GLP-1 concentration and its derivative proved to be the most suitable model for describing this phenomenon.

In summary, a model can be instrumental in elucidating a system, assessing the influence of diverse components, and forecasting its behavior. In the context of the previously discussed mathematical models, they contribute significantly to researchers' understanding of insulin sensitivity and diabetes by simulating intricate physiological processes.

Chapter 3. The problem of data richness in the real-world scenario: a systematic review of the available datasets

Even though, the gold standard to identify diabetes and prediabetes, predict perinatal morbidity in pregnancy and to diagnose gestational diabetes remains the OGTT, it is possible to talk about limitations in finding available datasets for research purposes. A 5-hour Oral Glucose Tolerance Test (5-OGTT) is a laboratory investigation that can be useful in many ways such as in assessing and diagnosing various glucose metabolism disorders, including postprandial hypoglycemia. In this test, the blood samples are collected typically every 30 minutes for a total of 10 samples. It is also possible to have 5-OGTT with a collection of 13-15 samples and so on. However, the 2-hour Oral Glucose Tolerance Test (2-OGTT) is more commonly used in clinical practice and is the standard for diagnosing conditions like diabetes. In the 2-OGTT, blood samples are collected at multiple time points, including baseline (fasting), and then at 30 minutes, 60 minutes, 90 minutes, and 120 minutes after consuming a standardized glucose solution. The 2-OGTT is considered a more practical and widely accepted test because it reflects how the body responds to glucose after a meal, which is a critical aspect of glucose metabolism. While the 5-OGTT mentioned earlier has been used in specific research or in specific clinical contexts, it is not typically the first-line test for diagnosing diabetes or glucose metabolism disorders in routine clinical practice. Instead, the 2-OGTT is more commonly employed for these purposes [46]. Due to the reasons clarified earlier, it is significantly easier to locate and utilize data from a 2-OGTT than from an OGTT involving multiple samples collections. This implies that when working with a

model that utilizes data from a 2-OGTT and encounters missing or unspecified information, it may be necessary to implement complex regularization techniques. Nevertheless, data obtained from such tests, whether in clinical settings or research contexts, are usually subject to privacy protections or necessitate consent from the individuals involved. Therefore, it is relatively uncommon to come across datasets containing a substantial volume of this type of data.

This review is focused on identifying scientific datasets that contain metabolic data from human subjects. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework was used to guide the systematic review process [37].

3.1. Literature search strategy

The literature search was conducted from February 24th to May 24th, 2023, on three levels: Google search, journal search, and search in the public clinical trial, for a complete overview of three different search fields and with the objective of acquiring the maximum amount of data available.

The roots “Repository”, “Biobank”, “Platform” and “Dataset” were related to the type of publication and/or sites of interest; instead, the words “Metabolism”,” Diabetes”, “Glucose”, “Glycemia”, “Glycaemia”,” Physiology” concern the topics of interest.

The search terms were organized into two concepts:

1. Repository, Biobank, Platform, Dataset
2. Metabolism, Diabetes, Glucose, Glycemia, Glycaemia, Physiology

Within each concept were linked by the Boolean operator ‘OR’ and then combined with the Boolean operator ‘AND’ across them. The choice of using both “Glycemia” (with a modern root) and “Glycaemia” (with the Latin root) can be justified by their great usage in today’s scientific literature. The English language was set as a filter.

3.1.1 Google search

The aim of the Google search was to find suitable websites for a deeper and more detailed investigation regarding human metabolism or cellular metabolism. Six different searches were performed by keeping the string of the first concept fixed and using a single word of the second concept for each search. For instance, the following query was used for the initial search: “Repository OR Biobank OR Platform OR Dataset” AND “Metabolism”. Subsequently, for the remaining searches, only the last term was updated.

For each website obtained by the Google search, a new query was formulated matching the terms previously mentioned. For portals having the possibility to choose the type of results (i.e., insert the filter “Dataset” or “Data paper”) the query was constructed simply utilizing the terms of the second concept connected with the comma or the ‘OR’ (the choice depends on the website’s instructions). If this option was not available, the terms "Dataset," "Data paper," and "Repository" were incorporated using the operator 'AND’. The first 50 results of each website ranked in order of relevance were evaluated by viewing the description of the dataset.

3.1.2 Journal search

This search was conducted in “Scimago” [38] because this is one of the best tools for the evaluation of the degree of scientific influence of the journals. The general area of interest was first set to “Medicine” and then to “Engineering”. “Endocrinology, Diabetes and Metabolism” and “Biomedical engineering” were selected as sub-categories for each one of the areas of interest.

3.1.3 Search in the public clinical trials

For search in the public clinical trials, the website www.clinicaltrials.gov was employed. The domain “Condition of Disease” was set at “Diabetes” and in the field “Other Terms” the words “glucose, glycaemia, insulin” were inserted since they are parameters intended for the analysis. The filter “Study with results” was applied.

3.2. Screening strategy and eligibility criteria

In the case of the Google search, the first 10 pages of search results were considered. This limitation was put in place to manage the data volume effectively and is justified by the ranking of results based on relevance, which led to the exclusion of less pertinent records. These records encompass a variety of sources, including websites and articles. A preliminary screening of the found websites enabled us to limit the results to only those concerning metabolic characteristics or cellular metabolism. This first stage was performed simply looking the home page of the websites and evaluating if it suitable for the settled goal. In this manner, the subsequent screening stages were simplified, commencing with the title

screening, thus enhancing the process's efficiency. Then, the findings were collected into Excel and the titles were examined to narrow down websites related to human metabolism that might have accessible datasets at no cost. After, the records were evaluated by full-text following some eligibility criteria, in particular were excluded:

- Websites and/or articles which do not contain any type of dataset;
- Websites and/or articles that contain an unavailable dataset;
- Websites and/or articles containing a dataset that necessitate permission from the dataset creators;
- Websites and/or articles in which the variables of interest (diabetes, glucose, insulin, etc.) are not part of the main goal of the study, instead, they are associated with a dataset whose aim doesn't concern the evaluation of metabolic disorders.

Approximately the same approach was adopted for the journal screening. The only difference in that part is in the choice to discard journals with a quartile inferior to Q2 since the Scimago assesses publications that are listed in the Scopus database, which is made available by the publisher Elsevier, spanning from 1997 to the current date. Each subject group of journals is divided into four quartiles: Q1, Q2, Q3, Q4. Q1 but the most esteemed journals within a particular subject area are those found in the first quartile, Q1, and so on. Therefore, it is decided to discard journals not so prestigious (Q3 and Q4). The journal's description of the topics that concerned them was examined and journals that do not deal with topics related to metabolic disorders or cellular metabolism and, those that do not have the dataset, have been excluded.

In the end, for the clinical trials search, as the generated results of the search were a perfect match for the type of data desired, the first 20 results obtained were directly exported into Excel. All studies underwent an eligibility assessment, with the dataset being evaluated first, followed by an analysis of the tolerance test.

For each analysed record of the three searches, a score in terms of colour is assigned (red=discarded, yellow=doubt, green= accepted) and the rule was that only the accepted and the doubt have been evaluated in the next step.

3.3. Data analysis

Each study was characterized according to its study design, the number of subjects participating in the study and their characteristics, including gender and age. Additionally, the type of tolerance test (IVGTT, OGTT, hyperglycemic clamp) was identified with their duration, and it was indicated whether CGM was used. A concise description of the collected dataset was also provided.

3.4. Results' dataset review

Overall, a systematic process yielded 16 datasets. A total of 367 websites were initially obtained from a Google search, and after the preliminary screening, 59 of them were assessed based on eligibility criteria and their titles. As a result, the 48 screened websites were evaluated in full-text, and only 18 of them were ultimately included in the review.

From the Scimago search, 58 journals were initially identified and assessed by reviewing their topic descriptions. Following the exclusion of 46 journals, the remaining 12 were incorporated into the review.

Additionally, in the search on www.clinicaltrials.gov, 20 studies were considered, but 12 of them were eliminated during the dataset evaluation phase. After the screening process, there were 8 studies remaining; however, 2 of them were subsequently excluded due to tolerance test evaluation. Ultimately, 6 studies were included in the review.

In the final stage of the process, a total of 428 records were gathered, but 357 of them were excluded because the datasets were not publicly available. The remaining records were assessed for eligibility criteria, and after excluding 55 more, only 16 records were ultimately included in the review. These 16 records were analyzed to generate the results. The Figure 5 shown the process of literature search and study selection while the Table 1 depicts a description of the studies included in the review.

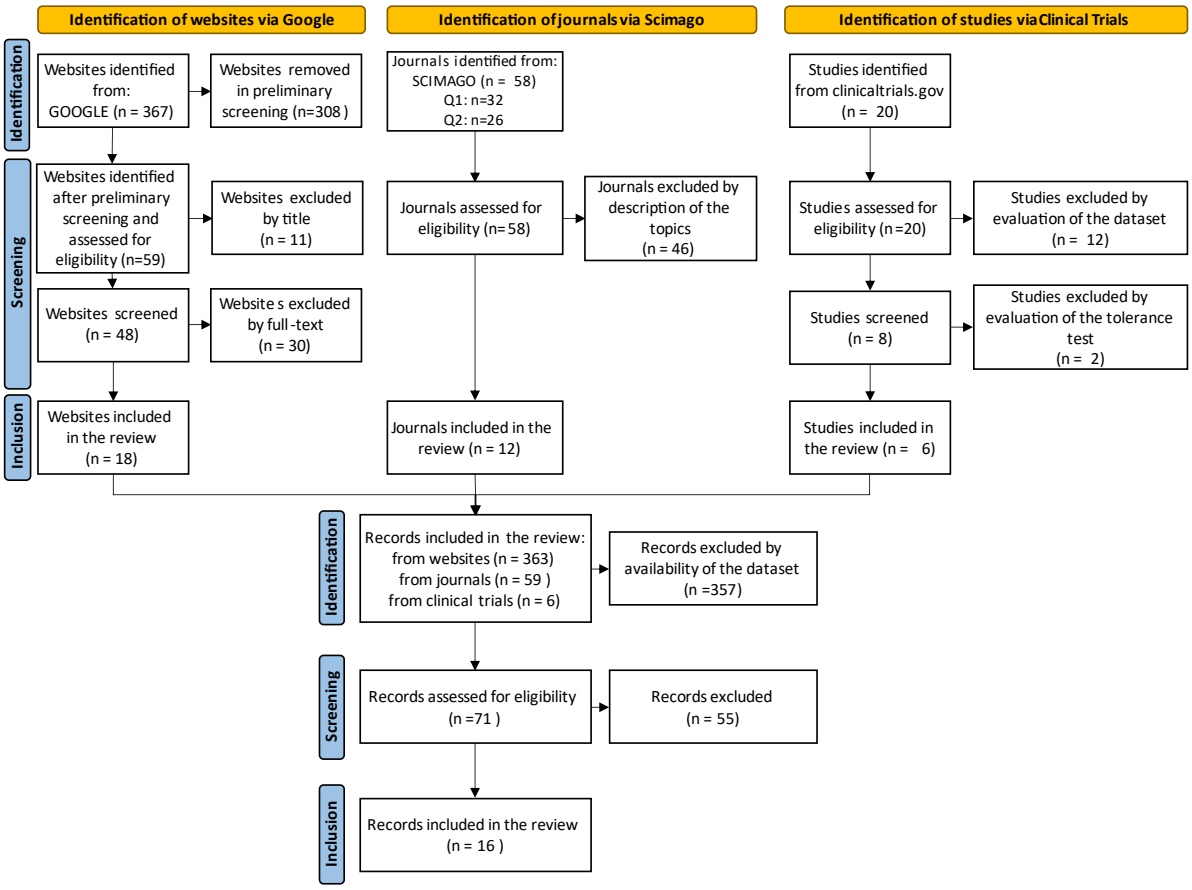

Figure 5: Flowchart of the performed screening process for the evaluation of the collected results.

Table 1. Summary of the analysed studies. There are reported the number of subjects included in the population (N.subjects) their gender (Sex(M/F)) and their age (years) in terms of mean and standard deviation (means(std)). Additionally, the sixth column show the type of the tolerance test and, next, there are: a column which contain information of if it is performed a continuous glucose monitoring and, lastl,y a brief description of the datasets.

Ref.	Study design	N.subjects	Sex (M/F)	Age	Tolerance Test	CGM	Brief description
[39]	Observational		F	21	OGTT (2h)		This dataset originates from the National Institute of Diabetes and Digestive and Kidney Diseases. Its primary objective is to predict, diagnostically, whether a patient has diabetes or not, utilizing specific diagnostic measurements provided in the dataset. All the patients included in this dataset are females who are at least 21 years old and of Pima Indian heritage.
[40]	Observational				OGTT (4h)		A Minimal Model developed by Bergman et al. is introduced, highlighting specific challenges in the assessment of the Insulin Sensitivity Index (SI) and Acute Insulin Response (AIR). Given these discoveries, it is advisable to exercise caution when making comparisons of SI estimates among different racial groups employing the Minimal Model.
[41]	Observational	10	M		OGTT (2h)		This study aimed to determine if measures of OGTT and IGTT-derived insulin sensitivity (ISI) exhibit variations when calculated using blood samples from venous versus arterialized sources. The study involved ten healthy men who participated in two trials randomly, each encompassing a 2h-OGTT, either conducted at rest or after exercise. Blood samples were collected simultaneously from a heated hand (arterialized) and an antecubital vein in the opposite arm (venous).
[42]	Observational	18 (pigs)	9M/9F		OGTT (2h) IVGTT (3h)		This study aimed to refine an OGTT model in young growing pigs and describe IVGTT in the same age group. Eighteen pigs were acquired one week after weaning and were trained to bottle-feed a glucose solution for two weeks, simulating the human OGTT. Subsequently, the pigs underwent both an OGTT (1.75 g/kg body weight) and an IVGTT (0.5 g/kg body weight). Blood samples were collected from indwelling vein catheters to measure glucose levels

							and diabetes-related hormones, including insulin, glucagon, and active glucagon-like peptide-1.
[43]	Observational	30	F		OGTT (2h)		This dataset introduces an Ordinary Differential Equations System for simulating Glucose-Insulin dynamics in the context of OGTT.
[44]	Observational	20			OGTT (2h)		This study has data from 20 healthy human subjects who underwent glucose ingestion, resulting in measurements of 83 metabolites and 7 hormones. The temporal patterns of these blood molecules are assessed using four key features to characterize differences among individuals and various molecules.
[45]	Observational	1492			OGTT (2h)		Data derived from an OGTT was employed to construct a predictive model utilizing the Support Vector Machine (SVM) algorithm. The model was trained and validated using OGTT data and demographic information obtained from 1,492 healthy individuals as part of the San Antonio Heart Study. This study involved the collection of plasma glucose and insulin concentrations before glucose intake and at three subsequent time-points (30, 60, and 120 minutes).
[46]	Cohort	1031	F		OGTT (2h)		The study aimed to investigate whether a 75 g OGTT administered between the 14th and 16th weeks of gestation could effectively identify two outcomes in a cohort of 1,031 pregnant women from the STORK study: 1. The development of Gestational Diabetes Mellitus (GDM). 2. The likelihood of giving birth to babies classified as large-for-gestational-age (LGA).
[47]	Cohort	16		35/65	OGTT	yes	The objective of this study was to assess the viability and efficiency of wearable devices in detecting early physiological changes that occur before the onset of prediabetes. Participants were equipped with a Dexcom 6 continuous glucose monitor (CGM) and an Empatica E4 wristband for a duration of 10 days. During this period, they received a standardized breakfast meal every other day. Following

							the 10-day monitoring period, participants returned to the clinic for an OGTT.
[48]	observational	12				yes	This dataset is accessible to researchers with an interest in enhancing the health and quality of life for individuals with type 1 diabetes. This dataset comprises 8 weeks' worth of data for each of the 12 individuals with type 1 diabetes included in the study. All participants were utilizing insulin pump therapy combined with continuous glucose monitoring (CGM). The dataset encompasses a wide range of information, including CGM, blood glucose levels from periodic self-monitoring, insulin doses, bolus and basal insulin, self-reported meal times, self-reported timings of activities, physiological data obtained from fitness bands.
[49]	Observational	17	16M/1F	58.1(6.6)	OGTT (3h) Hyperglycemic clamp		The aim of this project is to investigate the deficiencies in insulin secretion that play a role in the abnormal glucose metabolism observed in diabetes patients. Specifically, the study will examine the impact of signaling molecules released from the intestine, which stimulate insulin secretion. Participants with type 2 diabetes will undergo assessments of insulin secretion in response to both glucose and intestinal factors before and after receiving insulin treatment to lower their blood glucose levels.
[50]	Cohort	60;55	34M/26F; 30M/25F	28.8(16.7); 28.1(15.1)		yes	In this research study, individuals who were experiencing inadequate metabolic control despite being on optimized basal-bolus injection regimens were randomly assigned to one of two groups. One group received the Mini-Med Paradigm REAL-Time insulin pump (PRT), which is an insulin pump capable of receiving and displaying continuous glucose monitoring (CGM) data from a separate subcutaneous glucose sensor. The other group received conventional continuous subcutaneous insulin infusion (CSII). After a duration of 6 months, the study compared the glycemic outcomes between these two groups.
[51]	Cohort	50;50	26M/24F;	56.9(7.11);		yes	This study aimed to assess the impact of Dapagliflozin on 24-hour blood glucose

			25M/25F	56.8(9.71)			levels in patients with Type 2 Diabetes who had inadequate control of their condition, either through the use of Metformin or Insulin.
[52]	Cohort	93;98;98	47M/46F; 57M/41F; 49M/48F	60.74(7.94); 58.57(9.99); 59.23(9.28)		yes	This study sought to compare fasting blood glucose levels in patients diagnosed with Type 2 diabetes following 12 weeks of treatment with a new basal insulin analog as opposed to treatment with insulin glargine.
[53]	Cohort	11;11	5M/6F; 7M/4F	34.1(9.1); 45(13)	OGTT (2h)		In this study, participants will be divided into groups, and each group will take either 0g, 2g, or 4g of capsules or tablets in the morning following an overnight fast. After 40 minutes, they will ingest 75g of glucose dissolved in 300ml of water. Blood glucose, insulin, and triglyceride levels will be measured while fasting and at various intervals over the course of 2 hours after consuming the glucose solution. This protocol aims to assess the impact of different doses of capsules or tablets on metabolic responses to glucose intake.
[54]	Cohort	70;68	41M/28F; 45M/23F	36.82(11.34); 39.53(12.28)		yes	This clinical trial is taking place in Asia and has the objective of examining the ability of biphasic insulin aspart 30, in combination with metformin, to lower blood sugar levels and its safety profile in Chinese subjects with type 2 diabetes who have not previously used insulin. This study aims to compare the effects of biphasic insulin aspart 30 to biphasic human insulin 30 when used alongside metformin in individuals with type 2 diabetes who have not responded adequately to oral antidiabetic drug (OAD) therapy.

It's crucial to emphasize that in cases where there are empty cells in the table, this indicates that the information regarding that characteristic of the studies is not specified or available.

As evident from the Table 1, the majority of the studies employ an OGTT as a method for gathering glucose or insulin data. This choice is attributable to the fact that, as previously mentioned, in addition to being less invasive compared to other tests, it also stands as one of the gold standard diagnostic procedures for identifying Type 2 diabetes.

Chapter 4. Mathematical model for the study of the Glucose-Insulin regularoty system: small size datasets

The study of the glucose system has received considerable attention in recent decades. Given the complex nature of this condition, it's not surprising that it is investigated using a diverse range of tools that span various fields of study [55].

Specifically, regarding models, imbalances in the dynamics of G-I, as observed in conditions like diabetes, insulin resistance, and glucose intolerance, are prevalent issues in modern society [55]. As a result, the mathematical modeling of the G-I control system has been a topic of frequent exploration and study [56].

4.1. Model formulation

The objective of the present work is to simplify the regularization procedure for a model describing G-I kinetics so that it is possible to estimate all the parameters effectively, even with limited OGTT samples. For this purpose, the model of Sebastián Contreras et al.[9] was chosen since it clearly explain the G-I dynamics involving five main variables. The variable *I* pertains to insulin levels while the other four variables quantify the glucose concentration in different compartments: the stomach (*S*), the upper gastrointestinal tract in both the jejunum (*J*) and ileum (*L*), and in the bloodstream (*G*). These parameters were depicted in the box diagram in Figure 6.

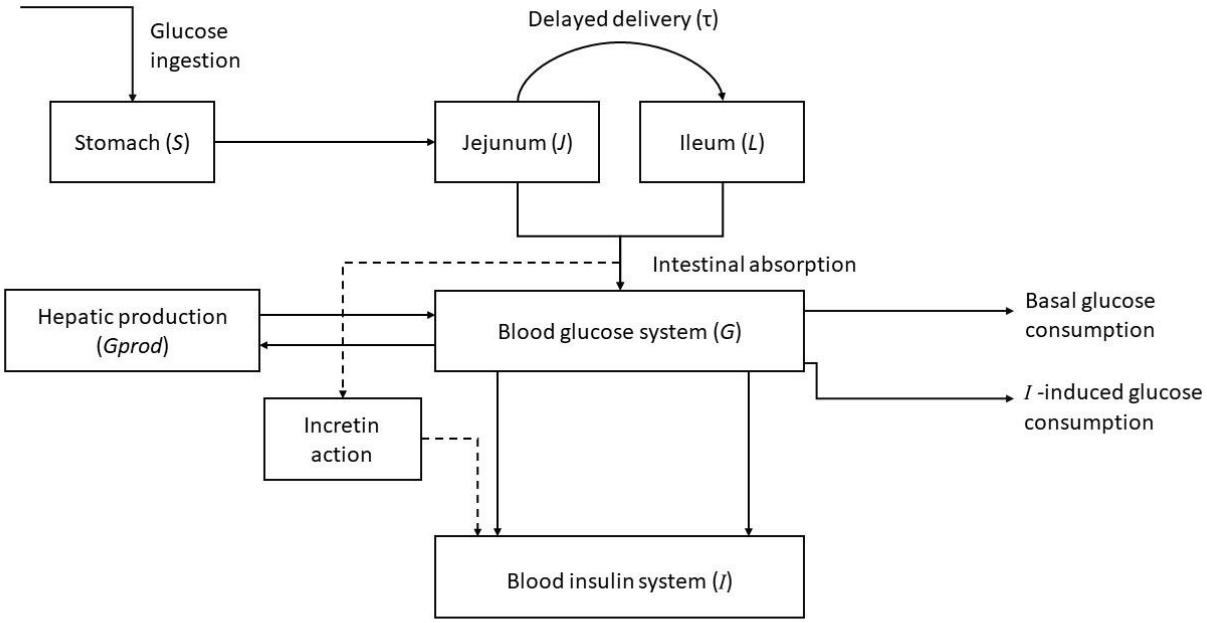

Figure 6: Redesigned box diagram illustrating the proposed model and how different compartments and variables interact.

To mathematically describe the emptying of the stomach for a liquid glucose bolus, it was assumed that its rate is directly proportional to the glucose content at any given time, *S(t)* is showed in Equation (1).

$$\frac{dS}{dt} = -k_{js}S \qquad S(0) = D \qquad (1)$$

Here, k_{js} represents a first-order kinetic constant describing the stomach emptying, *D* is the ingested glucose bolus in an OGTT (75g), and the negative sign is used to account for the decrease in glucose concentration, which corresponds to the emptying of the stomach.

The glucose exiting the stomach is transferred to the jejunum (*J*), acting as a source term that affects the rate of jejunum emptying. Since glucose absorption can occur in this segment of

the small intestine, it is possible to maintain the mass balance by recognizing that any glucose that is not absorbed will proceed further and be transported to the ileum (*L*) as it is possible to see in the Equation (2).

$$\frac{dJ}{dt} = k_{js}S - k_{gj}J - k_{jl}J \qquad J(0) = 0 \qquad (2)$$

Where, k_{gj} is a kinetic constant for glucose absorption in the jejune and, k_{jl} accounts for the glucose delivery from jejune to ilium. To depict the distribution of glucose transporters along the small intestine, it is posited that absorption takes place in two distinct regions of the intestine: the jejunum and the ileum, which are separated by a distance l. As a result, the time it requires for glucose to travel from the jejunum to the ileum is denoted as τ := l /U (where U represents the rate of transit). Consequently, there are two differential equations: one for the jejunum (2) and another for the ileum, Equation (3). The latter equation features a forcing function that is delayed by τ units of time.

$$\frac{dL}{dt} = k_{jl}\varphi(t) - k_{gl}L(t) \qquad \varphi(t) = \begin{cases} 0 & if\ t < \tau \\ J(t-\tau) & if\ t \geq \tau \end{cases} \qquad (3)$$

Thus, k_{gl} is the kinetic constant for glucose absorption in the bloodstream. To mathematically model the dynamics of blood glucose, source terms were incorporated into the glycemia equation. These include the intestinal absorption of glucose, which is adjusted by glucose bioavailability (η), and the hepatic contribution to glucose homeostasis (G_{prod}). The latter term indirectly represents the influence of glucagon on blood glucose levels as it possible to see in the Equation (4):

$$\frac{dG}{dT} = -k_{xg}G - k_{xgi}GI + G_{prod} + \eta\left(k_{gj}J + k_{gl}L\right) \qquad G(0) = G_b \qquad (4)$$

In this context, k_{xg} represents the rate of insulin-independent glucose uptake, k_{xgi} the uptake rate of insulin-sensitive tissues, η is the bioavailability of glucose absorbed in the intestines and G_{prod} represents the rate of hepatic glucose production. These parameters are used in the mathematical model to describe the different processes involved in glucose metabolism and regulation within the body. It is important to emphasize that the model implicitly takes into account the influence of glucagon through a hepatic glucose production function. This contributes a positive term to the equation governing blood glucose dynamics, representing the role of glucagon in promoting the release of glucose from the liver into the bloodstream. Therefore, G_{prod} also constitutes the resolution of the Equation (5):

$$\frac{dG_{prod}}{dG} = -\frac{G_{prod}}{G}\left(1 - \frac{k_2}{k_\lambda}\, G_{prod}\right) \qquad G_{prod}(G_b) = G_{prod}^0 \qquad (5)$$

By imposing the steady-state condition it is possible write the Equation (6):

$$G_{prod} = \frac{k_\lambda}{\frac{k_\lambda}{G_{prod}} + (G - G_b)} \qquad (6)$$

At the end, for the blood-insulin system the dynamics is expressed in the Equation (7):

$$\frac{dI}{dt} = k_{xi} I_b \left(\frac{\beta^\gamma + 1}{\beta^\gamma \left(\frac{G_b}{\tilde{G}}\right)^\gamma + 1} - \frac{I}{I_b}\right) \qquad (7)$$

In the previous equation k_{xi} represents a constant governing the rate of insulin breakdown in specific tissues, while β and γ are parameters linked to insulin production's half-saturation and acceleration aspects, respectively. They are responsible for regulating the initial and subsequent phases of insulin secretion by the pancreas. Additionally, $\tilde{G}$ signifies the modified G (glucose levels), influenced by the action of incretin hormones, while G_b and I_b represent

the stable-state values for these variables. For the parameter $\tilde{G}$, Sebastián Contreras et al.[9], at first suggested employing Hill's dynamics to model pancreatic secretion, along with a degradation term that is proportional to insulin. However, because these dynamics demonstrate saturation behaviour and are primarily designed for OGTT scenarios, they adjusted the mathematical representation of incretin action. The correction was made to reflect that incretin secretion is directly proportional to glucose levels within the intestinal lumen. This meaning is explained in the Equation (8):

$$\tilde{G} = G + f_{gj}\left(k_{gj}J + k_{gl}L\right) \qquad (8)$$

This aligns with physiological principles because intestinal epithelial cells lack glucose sensor proteins, making it impossible for them to directly sense the absolute amount of glucose in the intestine. Instead, their metabolic processes are intricately tied to the steady-state cytoplasmic concentration of glucose, which is in turn influenced by the rate of glucose transport across the cell membrane. In the (8), f_{gj} serves as a conversion factor that indirectly connects the rate of glucose absorption to insulin secretion through incretin action. It effectively quantifies the relative influence of incretin action compared to the direct glycemic effect on the pancreas.

Here's a summary of the differential equations characterizing the G-I model:

$$\frac{dS}{dt} = -k_{js}S \qquad S(0) = D \qquad (1)$$

$$\frac{dJ}{dt} = k_{js}S - k_{gj}J - k_{jl}J \qquad J(0) = 0 \qquad (2)$$

$$\frac{dL}{dt} = k_{jl}\varphi(t) - k_{gl}L(t) \qquad \varphi(t) = \begin{cases} 0 & if\ t < \tau \\ J(t-\tau) & if\ t \geq \tau \end{cases} \tag{3}$$

$$\frac{dG}{dT} = -k_{xg}G - k_{xgi}GI + G_{prod} + \eta\left(k_{gj}J + k_{gl}L\right) \qquad G(0) = G_b \tag{4}$$

$$\frac{dI}{dt} = k_{xi}I_b\left(\frac{\beta^{\gamma}+1}{\beta^{\gamma}\left(\frac{G_b}{G}\right)^{\gamma}+1} - \frac{I}{I_b}\right) \tag{7}$$

4.2. Model regularization

In the [9], the model detailed above was tested over 5 points OGTT of 407 volunteers who have ingested a 75g of glucose' bolus. However, due to the relatively high number of parameters (thirteen) that required estimation, compared to the number of OGTT samples available, the authors opted to create a dense and intricate regularization procedure that allows to increase the data density helping to prevent the overfitting and improve the model's generalizability. The regularization implemented by Contreras et al.[9] is showed in Equation (9).

$$min_{\vec{\theta}\in F_0 | F(\vec{\theta})=\vec{0}} \quad \lambda_1 J_{exp}(\vec{\theta},\alpha) + \lambda_2 J_{spline}(\vec{\theta},\alpha) + \lambda_3 J_{error}(\vec{\theta},\alpha) + \in ||\vec{\theta} - \vec{\theta}_j^*|| \tag{9}$$

Here, the first term is the conventional approach defining the parametric fitting problem, involving the minimization of a cost function, denoted as J_{exp}, which takes into consideration the disparities between the modeled curve and the experimental measurements. The second term (J_{spline}) was added to increase the data density of presumed measured data points by employing a soft interpolator to connect the experimental data points. Without losing the general applicability, it was used a combination of a cubic spline and a low-degree

polynomial interpolator. Moreover, the term J_{error} was integrated to consider errors stemming from the experiment, including variations in sampling time and laboratory techniques, and anticipated extreme values, derived from the experimental data and guided by clinical criteria, such as maximum and minimum values. The last term represents a local regularization.

4.3. Model simplified regularization

Starting from (9), the present work is focused on the simplification of the previous regularization technique facilitating its procedure and improving parameter estimation for the model, even when dealing with a limited number of OGTT samples. Therefore, the new proposed regularization aims to minimize the second derivative of the glucose and the insulin vectors retrieved from the model. The goal of this estimation is to find values for the 13 physiological parameters characterizing the models and collecting in a *v* vector, in order to minimize the sum of squared differences between the observed data and the model predictions. To do that, the following Equation (10) was used:

$$\min[\mathrm{f(v)}] = \min[\mathrm{RSS} + w1\ \mathrm{Glu''(w3{:}}\, w4) + w2\ \mathrm{Ins''}(w3{:}\, w4)] \qquad (10)$$

where, the first term is the residual sum of squares, the second term (Glu'') represents the second derivative of the output vector of the glucose levels and the term Ins'' describes the second derivative of the output vector of the insulin levels in the model. The choice of minimizing the second derivative of both vectors is justified by the aim to reduce rapid changes in concavity, which may not align with physiological patterns. Before doing that, both derivatives were normalized of an array to [0,1] since they exhibited two distinct scaling

factors. In addition, the last two terms were assigned weights, *w1* and *w2*, while *w3* and *w4* denote the initial and final points for minimizing the second derivatives of glucose and insulin. Various combinations of *w1*, *w2*, *w3*, and *w4* were evaluated using an iterative process, in which five distinct values were randomly generated for *w1* and *w2*, and all feasible combinations were tested to attain the lowest Root Mean Square Error (RMSE) and achieve the best possible fit. The implementation of the algorithm is illustrated in the following diagram:

Algorithm: Determine the optimal weights that result in the most accurate estimate

Data:
w1: consisting of five random values;
w2: consisting of five random values;
w3: containing all the points of the OGTT;
w4: containing all the points of the OGTT;
M: matrix containing all possible combinations of the previous data (each row is a possible combination).
Result: S matrix containing all values of RMSE, in descending order, with its combination of weights.
foreach *row of M* **do**
 The estimation of the 13 parameters in the *v* vector;
 The RMSE values;
 Compare the RMSE value with that of the top row:
 if it is greater than the previous one, swap the positions of the items to arrange them in descending order;
 if it is smaller, proceed to the next iteration.
end

The initial values for the parameters, composing the *v* vector, to be estimated are expressed in the Table 2.

Table 2: Initial values for the parameters on the reference of Contreras et al. [9]

Initial parameters	Values
k_{js}	0.1523
k_{gj}	0.1022
k_{jl}	0.0414
k_{gl}	0.1941
k_{xg}	0.0165
k_{xgi}	1.27e-07
k_{xi}	0.0177
γ	1.1786
β	82.2658
f_{gj}	2.8657
η	0.0186
τ	58.5458
k_{λ}	0.0363

4.4. Model implementation

The model G-I described in the [9] was implemented in Simulink 2023a and the procedure of parameters estimation is performed in Matlab 2023a. The input vectors of glucose and insulin were called with the terms G_{exp} and I_{exp} and the model parameter vector *v*, containing the 13 parameters to be estimated, is computed by solving a nonlinear least-squares curve fitting problem using the *lsqnonlin* MATLAB function. The lower and the upper bounds for the function are set to (0;20) for all the parameters except for β (0;100) and τ (0;80). The algorithm is set to *trust-region-reflective* and the function and the step size tolerances have been both set to 10-10.

4.5. Model testing

Keeping in mind the goal of using the model with a limited number of OGTT samples while still being able to estimate the physiological parameters, the described procedure was initially tested on two datasets with a large amounts of samples obtained from Pepino et al.[57] and Muscelli et al.[58]. On these datasets, it was possible to assess the model's effectiveness in estimating all the physiological parameters.

4.5.1 Pepino et al. dataset

This dataset originates from a study involving 17 non-obese individuals who do not use non-nutritive sweeteners (NNS) and are insulin sensitive. These participants underwent a 5-hour OGTT in which they consumed either sucralose or water approximately 10 minutes before the test. During the test, β cell function, insulin sensitivity, and insulin clearance were

assessed using a minimal model of glucose, insulin, and c-peptide kinetics. The dataset includes measurements of glucose, insulin, and c-peptide levels taken at 11 different time points: 0, 10, 20, 30, 60, 90, 120, 150, 180, 240, and 300 minutes after ingesting the bolus of sucralose. These measurements were likely collected to analyze how these parameters change over time in response to the sucralose ingestion, providing valuable insights into insulin sensitivity and glucose metabolism in the study participants.

4.5.2 Muscelli et al. dataset

Twenty-one subjects were volunteered for this study. None of these subjects had experienced weight loss or dietary changes in the three months leading up to the study, and none were taking any medications. The study involved a 3-hour OGTT with a 75g glucose load, during which plasma glucose concentrations were measured at 10-minute intervals. Insulin sensitivity was assessed by analyzing insulin and plasma glucose responses to the oral glucose load, and the oral glucose insulin sensitivity index (OGIS) was calculated as an indicator. The modeling approach used in this study comprises three main components: a model for the regulation and interpolation of plasma glucose concentrations, a model describing the relationship between insulin secretion (or C-peptide) and glucose concentration and, a model for C-peptide kinetics. The dataset from this study, which aims to quantify the impact of fluctuations on β cell function, includes 10 data points collected at the following time intervals: 0, 10, 20, 30, 40, 60, 90, 120, 150, and 180 minutes. These data points likely serve as the basis for analyzing how β cells respond to glucose levels over time.

4.6. Model validation

Subsequently, the model was validated on two datasets found in the previous chapter, coming from the systematic search, having less samples, the [46] and [45].

4.6.1 Lekva et al. dataset

This dataset was obtained from an OGTT conducted during the 14-16 week gestation period to determine which women developed gestational diabetes. The study examined data from the STORK study, a prospective longitudinal cohort study that tracked 1031 women with a low risk of Scandinavian inheritance throughout their pregnancies and during childbirth. For all participants, a 75g OGTT was administered in the morning following an overnight fast, and glucose levels were measured at various time points. Gestational Diabetes Mellitus (GDM) was diagnosed using four distinct sets of criteria:

1. The WHO 1999 criteria;
2. The IADPSG 2010 criteria;
3. The WHO 2013 criteria;
4. The Norway 2017 criteria.

Specifically, the dataset used for the validation contains OGTT data for women who developed gestational diabetes, with measurements taken at 0, 30, 60, 90, and 120 minutes during the test.

4.6.2 Abbast et al. dataset

This one was collected from a study that included OGTT data from both men and not pregnant women. The study utilized data generated from an OGTT to construct a predictive model employing the support vector machine (SVM) technique. The dataset includes measurements of plasma glucose and insulin concentrations taken both before and after the ingestion of glucose at specific time intervals, at 0,30,60, and 120 minutes. The goal of this research is to leverage these data points, along with demographic information, to build a predictive model that can potentially estimate various health-related outcomes or provide insights into glucose and insulin responses in healthy individuals.

For the dataset obtained from [46], the model successfully completed the parameter estimation process without requiring a reduction in complexity.

While processing data from [45], it became evident that the number of samples available was insufficient to accurately estimate all the physiological parameters. To address these issues, four parameters (k_{js},k_{gj},β,f_{gj}) were kept fixed at values that had been previously estimated using the datasets referenced in [57] and in [58], in order to reduce the number of parameters that needed to be estimated, resulting in a smaller v vector reducing the complexity of the problem. The four parameters that were set as fixed were selected because they were considered less influential on the dynamics of the model compared to the others. Additionally, they were the parameters that remained unestimated when the problem was not simplified or reduced. After implementing this process, the model was able

to proceed with its estimation task effectively and accurately. This approach was employed to enhance the assessment of the effectiveness of the simplified regularization procedure on a dataset characterized by a very limited number of data points.

Chapter 5. Model results

The glycemia (mmol/L) and insulinemia (pmol/L) curves derived from the Pepino data [57] are displayed in the Figure 7 and in the Figure 8 while the trends of the data provided by [58] are respectively in Figure 9 and in Figure 10.

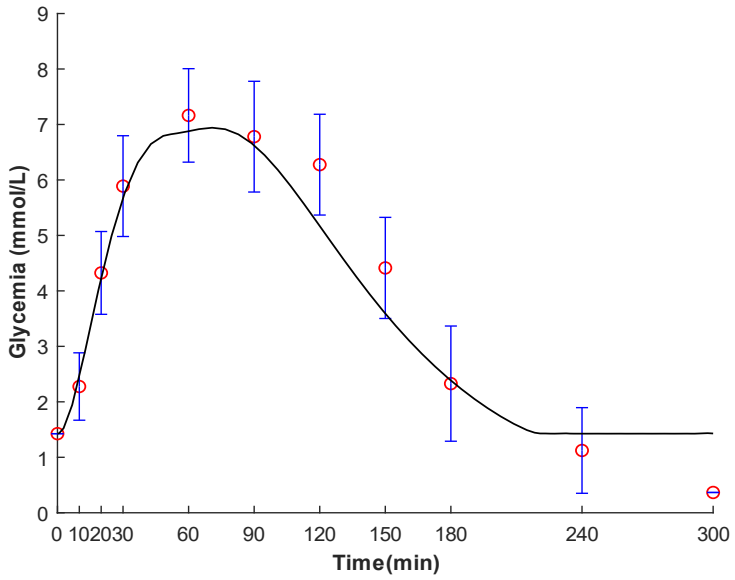

Figure 7: Glycemia (mmol/L) curve obtained by experimental data of Pepino et al. [57]. The red circles are the initial data, the black curve represents the estimated fit and the blue lines are the standard deviations.

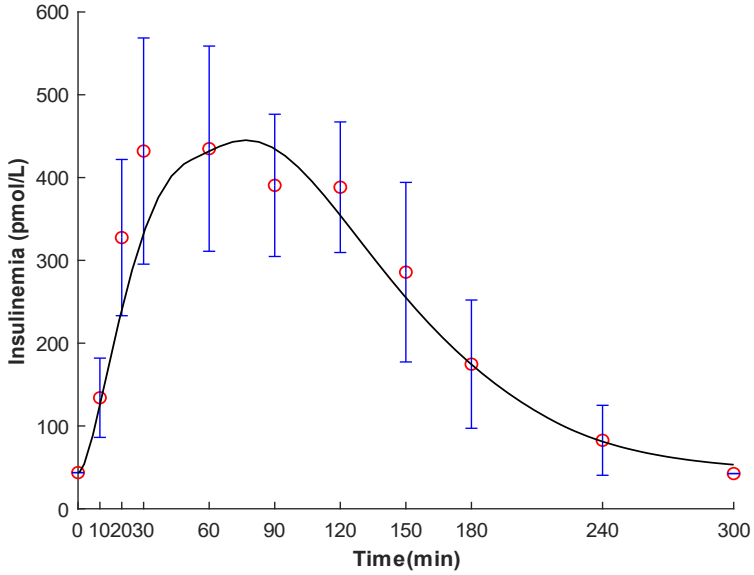

Figure 8: Insulinemia (pmol/L) curve obtained by experimental data of Pepino et al. [57]. The red circles are the initial data, the black curve represents the estimated fit and the blue lines are the standard deviations.

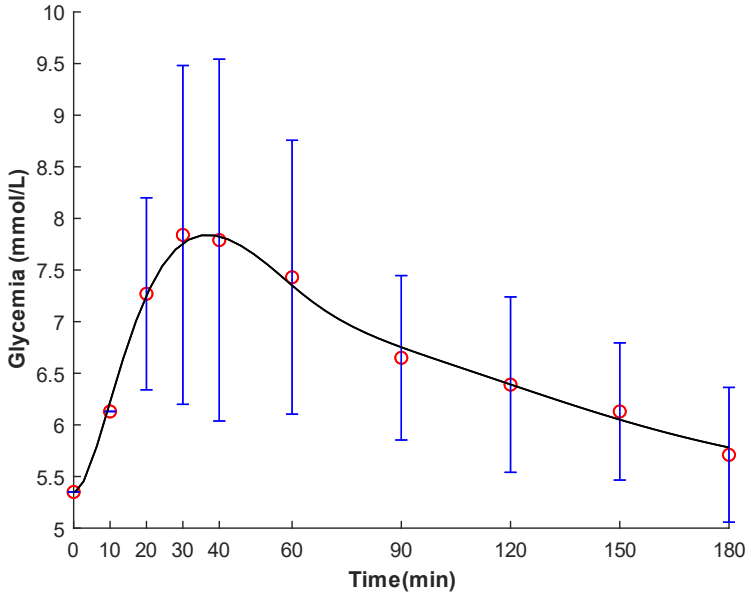

Figure 9: Glycemia (mmol/L) curve obtained by experimental data of Muscelli et al. [58]. The red circles are the initial data, the black curve represents the estimated fit and the blue lines are the standard deviations.

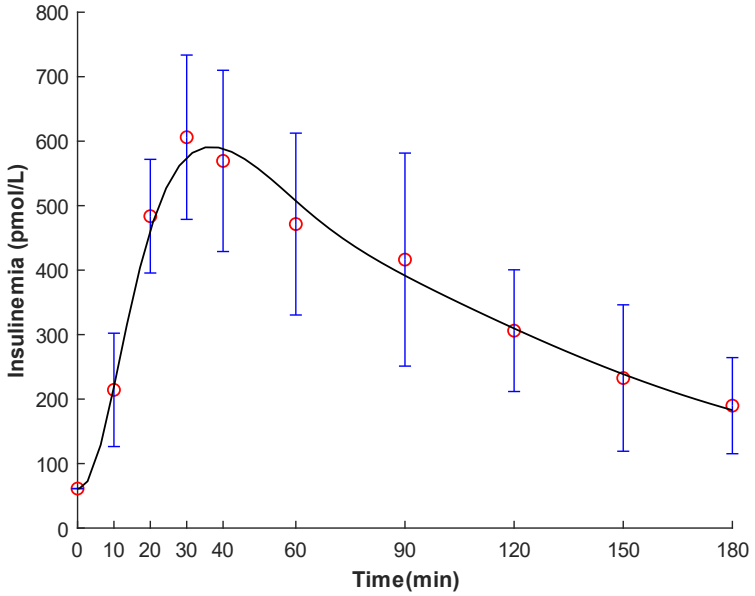

Figure 10: Insulinemia (pmol/L) curve obtained by experimental data of Muscelli et al. [58]. The red circles are the initial data, the black curve represents the estimated fit and the blue lines are the standard deviations.

It can be observed that both datasets exhibit a consistent fit utilizing the same weight and minimizing the second derivative of glucose and insulin from the third point onward. Furthermore, it is possible to highlight that all the trends fall within the ranges established by the standard deviation.

The Figure 11 and Figure 12 shown the estimate on reference experimental data by [46] of glycemia and insulinemia respectively.

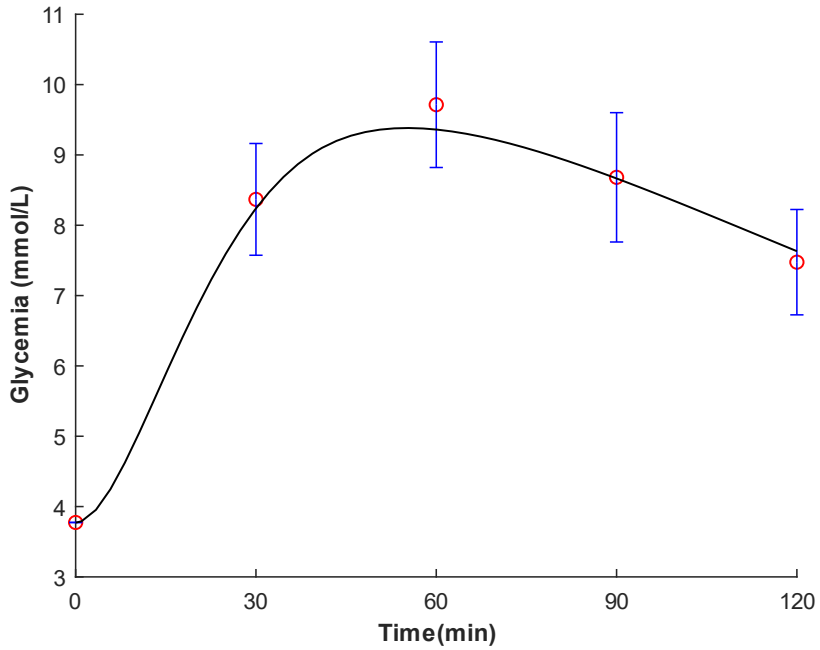

Figure 11: Glycemia (mmol/L) curve obtained by experimental data of Lekva et al. [46]. The red circles are the initial data, the black curve represents the estimated fit and the blue lines are the standard deviations.

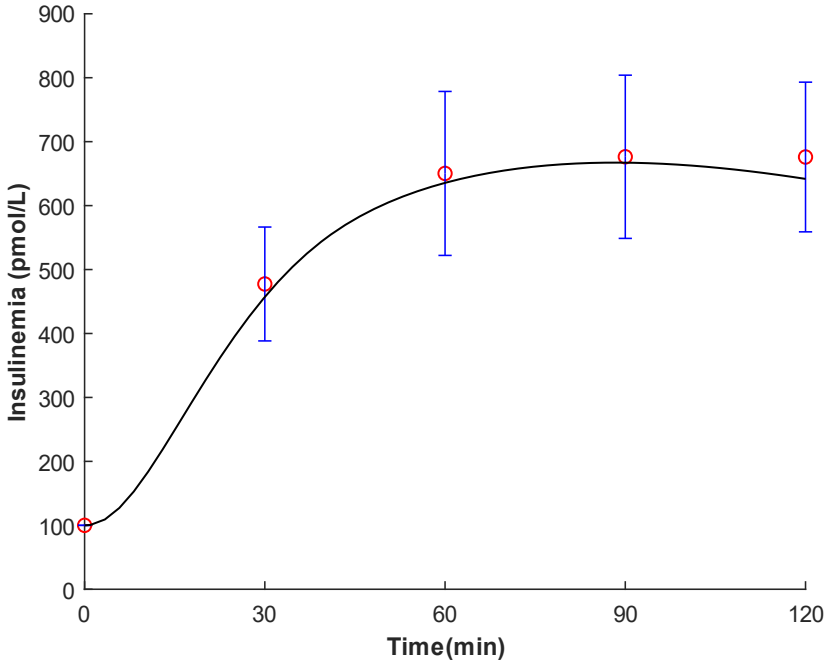

Figure 12: Insulinemia (pmol/L) curve obtained by experimental data of Lekva et al. [46]. The red circles are the initial data, the black curve represents the estimated fit and the blue lines are the standard deviations.

The curves derived from [45] are depicted in Figure 13 and in Figure 14. These trends were generated with the adjustment of four parameters using estimates derived from Pepino's data [57].

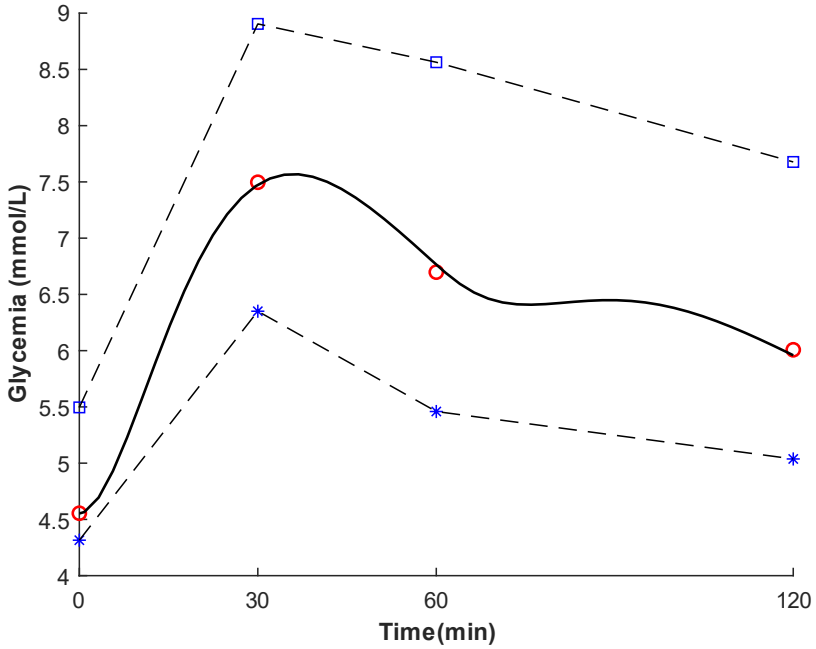

Figure 13: Glycemia (mmol/L) curve obtained by experimental data of Abbas et al. [45] with four physiological parameters set as those of Pepino et al. [57]. The red circles are the initial data, the black curve represents the estimated fit and the dashed lines are the first and the third interquartile.

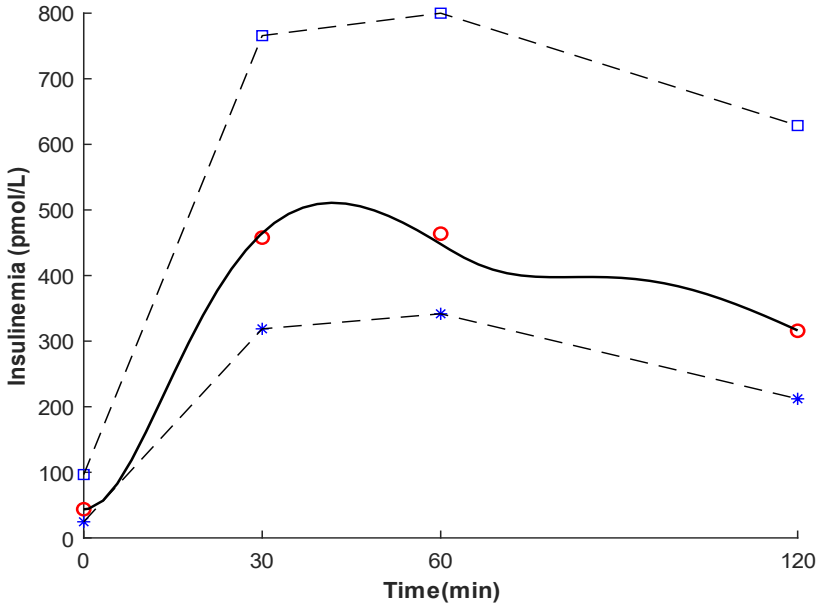

Figure 14: Insulinemia (pmol/L) curve obtained by experimental data of Abbas et al. [45] with four physiological parameters set as those of Pepino et al. [57]. The red circles are the initial data, the black curve represents the estimated fit and the dashed lines are the first and the third interquartile.

In Figures 15 and Figure 16, there are shown the trends of glycemia and insulinemia from the dataset of Abbas et al.[45], but this time, the four parameters were adjusted based on Muscelli's estimates [58].

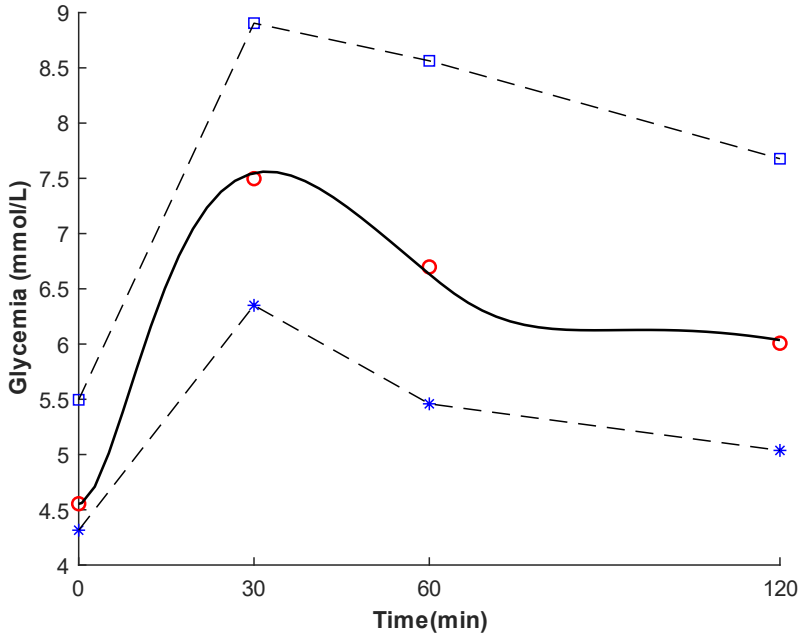

Figure 15: Glycemia (mmol/L) curve obtained by experimental data of Abbas et al. [45] with four physiological parameters set as those of Muscelli et al. [58]. The red circles are the initial data, the black curve represents the estimated fit and the blue lines are the dashed lines are the first and the third interquartile.

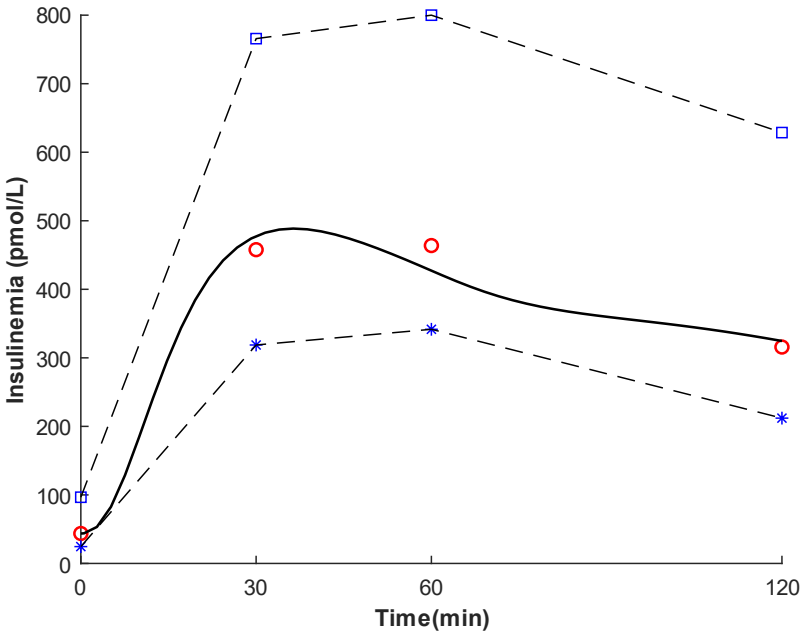

Figure 16: Insulinemia (pmol/L) curve obtained by experimental data of Abbas et al. [45] with four physiological parameters set as those of Muscelli et al. [58]. The red circles are the initial data, the black curve represents the estimated fit and the blue lines are the dashed lines are the first and the third interquartile.

In the Table 3 the estimates of physiological parameters are presented using the Contreras et al. [9] dataset. These are accompanied by their respective regularization procedures, the second and third columns of the table displaying results for the full regularization and simplified regularization methods, respectively. The last row of Table 3 features the calculation of RMSE using the simplified regularization procedure.

The Table 4 contains parameter estimates obtained from the Pepino [57] and Muscelli [58] datasets.

The Table 5 reports the estimated parameter values from the Lekva et al.[46] dataset while in Table 6, parameters are estimated using the Abbas et al.[45] dataset, with two sets of estimates: one with four parameters held constant based on previous Pepino's estimates and another utilizing the Muscelli estimates as a reference. Lastly, Table 7 displays the weights assigned to each dataset and showcases all the RMSE values obtained in the analysis.

Table 3: The parameters estimated using the dataset from Contreras et al.[9] are presented in three columns: Left Column: parameter names; Center Column: values of the estimates by Contreras et al.[9] with their regularization; Right Column: values of the parameters estimated with the [9] dataset using the simplified regularization, along with RMSE value.

Parameters	Values	
k_{js}	0.1523	0.0650
k_{gj}	0.1022	0.0680
k_{jl}	0.0414	0.0201
k_{gl}	0.1941	0.3639
k_{xg}	0.0165	0.0146
k_{xgi}	1.27e-07	5.5558e-08
k_{xi}	0.0177	0.0110
γ	1.1786	1.5934
β	82.2658	88.5255
f_{gj}	2.8657	2.8045
η	0.0186	0.0112
τ	58.5458	73.8226
k_{λ}	0.0363	1.9001
RMSE		1.2133

Table 4: Estimated values of the physiological parameters with the dataset from Pepino et al.[57] and Muscelli et al.[58], using the simplified regularization.

Parameters	Values	
	Pepino et al.[57]	Muscelli et al.[58]
k_{js}	0.0568	0.0855
k_{gj}	0.0451	0.0659
k_{jl}	0.0118	0.0193
k_{gl}	0.1696	0.0252
k_{xg}	0,0146	0.0125
k_{xgi}	2.99E-07	8.65E-11
k_{xi}	0.0229	0.0162
γ	0.9186	2.0858
β	89.2395	95.0421
f_{gj}	5.3162	2.1465
η	0.0341	0.014
τ	48.4072	55.8685
k_{λ}	4.76E-04	0.3524

Table 5: Estimated values of the physiological parameters using data from the Lekva et al.[46]

Parameters	Values
	Lekva et a.[46]
k_{js}	0.0479
k_{gj}	0.0644
k_{jl}	0.0178
k_{gl}	0.0193
k_{xg}	0.0047
k_{xgi}	8.70E-06
k_{xi}	0.0075
γ	1.5017
β	82.045
f_{gj}	3.3964
η	0.0285
τ	41.003
k_{λ}	4.3637

Table 6: The parameters estimated using different datasets containing a limited number of data points are presented in the columns. Second column: parameter estimates from the Abbas et al.[45] with 4 parameters set equal to the estimates of Pepino et al. [57]; third column: parameter estimates from the Abbas et al.[45] dataset with 4 parameters fixed to Muscelli's estimates[58].

Parameters	Values	
	Abbas et al.[45] with 4 parameters fixed by Pepino et al.[47].	Abbas et al.[45] with 4 parameters fixed by Muscelli et al.[48].
k_{js}	0.0560	0.0855
k_{gj}	0.0451	0.0659
k_{jl}	0.0194	0.0389
k_{gl}	0.1119	0.0229
k_{xg}	0.0138	0.0119
k_{xgi}	2.2753e-05	7.1117e-06
k_{xi}	0.0289	0.0129
γ	1.4094	2.2411
β	89.2395	95.0421
f_{gj}	5.3162	2.1465
η	0.0295	0.0210
τ	59.1427	59.4611
k_{λ}	2.3464	0.4684

Table 7: Values of the weights (w1,w2) and of the starting and ending point to regularize (w3,w4).In the las row, there is the value of the computed RMSE for each dataset.

Weights	Values				
	[57]	[58]	[46]	[45]	[45]
w1	2	2	2	0.068	0.068
w2	0.01	0.01	0.01	0.0015	0.0015
w3	3	3	1	1	1
w4	end point	end point	end point	end point	end point
RMSE	4.9717	1.1616	1.0258	0.4554	0.8851

Chapter 6. Discussion

Diabetes is a chronic metabolic disorder characterized by elevated levels of blood glucose. As a result of this condition, individuals affected by diabetes typically rely on one or more daily insulin injections for management and control. There are three primary types of diabetes: Type 1, often referred to as juvenile-onset diabetes; Type 2, commonly known as adult-onset diabetes; and gestational diabetes, which occurs during pregnancy. In recent decades, there has been a growing recognition of the strong connection between insulin resistance and the development of Type 2 diabetes. Insulin resistance plays a pivotal role in this context, serving not only as the most influential predictor for the future onset of Type 2 diabetes, but also as a target for treatment once hyperglycemia (high blood sugar) is already present. In the realm of research, the challenge often lies in locating datasets that include metabolic data with a sufficient number of data points. This is crucial for the successful implementation and the accurate validation of specific mathematical models. Mathematical models are frequently created and applied within the realm of glucose metabolism. Essentially, a mathematical model serves as a simplified portrayal of the real-world, utilizing variables and mathematical equations to elucidate the connections among these variables. The systematic research outlined in Chapter 3, primarily centers around the identification of scientific datasets containing metabolic data. The results come from among three different research, Google search, journal search and search in the public clinical trial. The primary objective of the Google search was to identify suitable websites and online sources where relevant datasets could potentially be located. This step was instrumental in pinpointing the right platforms and repositories for further investigation. The focus of the journal search was

on some of the most esteemed journals as per Scimago rankings (Q1 and Q2). The intention was to identify articles, within these journals, that contained pertinent information related to metabolic data. Journals of high repute are often reliable sources of scientific data. The exploration of public clinical trials aimed to provide a comprehensive overview of the available data in clinical settings. This step ensured that the research covered all potential sources of relevant information, especially data generated within the clinical context. By combining these three research methods, the study aimed to cast a wide net and gather a comprehensive dataset for the subsequent validation of the model describing the kinetics of glucose and insulin. The utilization of the English language filter, for all searches, is regarded as a practical approach, given that English is the predominant language for scientific publications and research worldwide. This choice is made to ensure that the search results are relevant and directly applicable to the research objectives. Additionally, the implementation of a preliminary screening for the websites search is considered a good measure since Google is a vast search engine housing an immense volume of data, and without some form of filtering, the search results can become overwhelming and require substantial time for examination. Through the execution of an initial screening, the results can be narrowed down to a more manageable set of potentially pertinent sources, thus enhancing the efficiency and focus of the research process. For the same reason, the decision was made to exclusively consider journals with Q1 and Q2 rankings, taking into account that, for each research, the results were analysed in order of relevance.

Overall, 16 studies were included in the review. The types of studies are categorized into two main groups: observational and cohort studies, with a slight majority favouring observational

studies. Specifically, there are 9 observational studies ([39]–[45], [49]) and 7 cohort studies ([46]–[48], [50]–[54]). It's noteworthy that most of the cohort studies are derived from clinical trials. The populations under analysis are predominantly mixed, comprising both male and female participants. However, there are 3 studies that exclusively focus on women ([39], [43], [46]), and one study involves only male participants, the [41]. Notably, one of the studies that exclusively analyses women ([46]) is aimed at diagnosing gestational diabetes, while the study involving only men is conducted both at rest or post-exercise ([41]). In this case the samples are taken simultaneously from the antecubital vein (venous blood) and from heated hand (arterialised). Nonetheless, there is one study that exclusively considers animals, specifically pigs ([42]). All the results were categorized based on the glucose tolerance test. Specifically, 9 out of the 16 studies only consider OGTT ([39]–[41], [43]–[47], [53]), while the populations in the 6 other studies were analyzed using CGM ([47], [48], [50]–[52], [54]). Furthermore, within the [42], the population involve pigs that undergo both OGTT and IVGTT. In the latter, the outcomes are not only glucose and insulin but also glucagone and c-peptide. In the [49], a combination of OGTT and hyperglycemic clamps is utilized for analysis. At the end, it is important to highlight the fact that, the majority of the studies involving OGTT performed in 2 hours ([39], [41]–[46], [53]) but with different number of samples, as well as the [39], with only one sample at the end of the test; the [41], [44], that have collected seven samples for each subject; the [43], [46], [53], considering 5 samples during the test. The [42] have considered 11 samples and the [45] has only 4 samples. Additionally,[49] have involved a 3-hours OGTT with the unique sample at the end of the test while, the duration of the OGTT in [40] in about 4 hours with 4 samples (one for

each hour). Essentially, the outcomes derived from this analysis indicate that the OGTT is the most commonly utilized test, albeit with varying durations and varying numbers of samples.

The advantage of this review lies in the discovery of real and readily usable datasets. Conversely, the disadvantage stems from the challenge of locating publicly accessible data that can be downloaded without requiring author permission. This limitation resulted in a significantly lower quantity of datasets being included in the review than initially anticipated, despite the multitude of websites and journals identified in the initial stages of the research.

Given the significant attention that has been directed towards studying the glucose system in recent decades, particularly the imbalances in the dynamics of glucose-insulin system observed in conditions such as diabetes, the mathematical modeling of the G-I control system has become a subject of frequent investigation and research. Contreras et al. [9], formulated a model that describe the G-I kinetics but require the estimation of 13 physiological parameters. Due to the scarcity of OGTT samples (only 5 samples available), they decided to incorporate a regularization technique to accurately estimate all the parameters in this scenario. Given the intricacy of Contreras' regularization method, the objective of this study was to streamline it while preserving the ability to estimate all the physiological parameters. The decision to implement this model was based on two primary factors: its relevance to the kinetics of the G-I system and its reliance on OGTT data. Therefore, it was the appropriate model to evaluate using the findings from the systematic review.

Specifically, the approach adopted in this work, involved minimizing the second derivatives of the vectors containing glucose and insulin data. Additionally, a set of weights (*w1* and *w2*)

was applied to the second derivative of each vector, while *w3* and *w4* specified the starting and the ending points of the vectors, delineating the range over which the minimization procedure was initiated and concluded. The decision to minimize the second derivative of both vectors was motivated by the goal of mitigating abrupt alterations in concavity, which might not correspond to physiological patterns. To perform this, both derivatives were normalized to a range of [0,1], as they exhibited two distinct scaling factors. Multiple combinations of *w1*, *w2*, *w3*, and *w4* were assessed through an iterative process in which five distinct values were randomly generated for each parameter, and all possible combinations were examined to attain the lowest RMSE and achieve the best possible fit. Since there were 13 parameters to be estimated, after the decision on what to minimize in the new regularization procedure, the model was then tested using the datasets [57] and [58] that contained an adequate number of data points for estimate all the 13 parameters. These are datasets coming out from 5-hours OGTT and 3-hours OGTT respectively. In the Fig.7 and Fig.8 there are the trend obtained with the [57], while in the Fig.9 and Figure there are the outputs obtained with the data of [58]. It is possible to see that all trends fall within the range defined by their respective standard deviations. Glycemia levels were calculated in millimoles per liter (mmol/L), while insulin levels were measured in picomoles per liter (pmol/L). The trends do not deviate significantly from the initial values, as can also be observed from the RMSE values that are equal to *4.9717* from [57] data, and equals to *1.1616* using [58] datasets.

After employing OGTT with a substantial amount of data, the model was subsequently validated using two OGTT datasets, which, in contrast, contained relatively few data points.

Specifically, these datasets were sourced from Abbas et al.[45] and from Lekva et al. [46], through the systematic review process. The dataset of [45] was generated from an OGTT which was utilized for constructing a predictive model employing the Support Vector Machine (SVM) algorithm. This model involved the measurement of plasma glucose and insulin levels both before and at three consecutive time points (30, 60, and 120 minutes) following glucose consumption. However, in the [46], the primary objective is to assess whether administering a 75g OGTT between the 14th and 16th weeks of gestation can reliably predict two outcomes within a group of 1,031 pregnant women participating in the STORK study.

These two datasets were selected due to their limited sample sizes, making them suitable for testing the model on very small datasets. The Fig.11 and Fig.12 display the trends that were derived from the dataset [46]. The curves observed in these figures still fall within the ranges defined by their standard deviation. It is also crucial to consider that the data from [46] are collected from diabetics. This is evident from the fact that the curves settle at high values rather than resembling basal values. Nevertheless, this trend doesn't deviate significantly from the initial values and boasts the lowest RMSE among all the trends, with a value of *1.0258*.

The implemented model encountered challenges when estimating values using dataset [45], because it contained only 4 samples. To address this limitation, 4 parameters were set. Therefore, the parameters were initially established using estimates derived from the Pepino et al.[57] and subsequently refined using estimates from Muscelli et al.[58]. As a result, the number of parameters to estimate was reduced from 13 to 9, which helped alleviate the

model's difficulties. Consequently, the results depicted in Fig.13, Fig.14, Fig.15 and Fig.16 indicate that Abbas' trends, with the assistance of this adjustment, exhibit a reasonable fit. It is also important to emphasize that, whether using Pepino's estimates or Muscelli's estimates, the trend initially rises until the 30th minute, then declines between the 30th and 60th minutes, and subsequently increases once more between the 60th and 90th minutes. The trend observed appears to consistently exhibit a biphasic nature; nevertheless, it is impossible to definitively confirm this conclusion due to the absence of data at point 90.

The estimates of the physiological parameters obtained using dataset [46] can be found in Table 5 while those obtained through the latest procedure using Abbas' data [45] are listed in Table 6. Across all these tables, it is evident that all the estimates exhibit plausible values, closely resembling the original values obtained by Contreras et al.[9] through their regularization method.

In Table 7, the values of *w1*, *w2*, *w3*, *w4*, and *RMSE* are collected for each test. It should be noted that the *w1* and *w2* values remain the same for data obtained from [46], [57], [58], however, adjustments were necessary for the [45] dataset. The *w3,* denotes the point at which the function began to minimize, and this varies among the four datasets. For [57] and [58], it is the third point, while for [46] and [45], it is the first point. This variation can be attributed to the smaller sample sizes in the [46] and [45] datasets, making it less meaningful to initiate the analysis from the third point onwards. The *w4* consistently represents the final point in all datasets.

Conclusions and future applications

In conclusion, it can be stated that when working with OGTT datasets containing five or more data points, the model implemented with simplified regularization successfully estimates all 13 physiological parameters. However, when dealing with datasets containing fewer than 5 points (in this case, 4 samples), it becomes necessary to reduce the parameter vector by fixing 4 out of the 13 parameters. Through the previously described procedure, it was observed that setting the parameters to the values estimated from the data of Muscelli et al. [58] yielded the best results, as these estimates closely matched those obtained from the Contreras [9] dataset.

It is important to acknowledge both the advantages and limitations of the procedure used in this study. In the systematic research phase, it was noted that the decision to restrict the procedure to freely available and downloadable datasets does introduce a potential *bias*. It may limit the diversity and representativeness of the data used for estimation, potentially biasing the results towards the characteristics of those specific datasets. This bias can be overcome by incorporating a more diverse range of data sources, including datasets that may not be freely available but are more representative of the broader population or research objectives. Expanding the dataset sources can help reduce *bias* and enhance the generalizability of the procedure's outcomes. On the positive side, this approach still yielded an acceptable number of usable datasets covering a wide range of tolerance tests. In future applications, it is advisable to expand the search to include also widely used scientific databases such as Scopus and PubMed.

Regarding the implementation and validation of the G-I model, a limitation was working with average population data instead of data from multiple individuals. However, this approach had the advantage of simplifying regularization resulted in plausible estimates. In future applications, it is recommended to apply the model to datasets that encompass multiple subjects to enhance the robustness of the analysis.

References

[1] P. A. Insel *et al.*, «Insulin control of glucose metabolism in man: a new kinetic analysis», *J Clin Invest*, vol. 55, fasc. 5, pp. 1057–1066, mag. 1975, doi: 10.1172/JCI108006.

[2] J. Radziuk, K. H. Norwich, e M. Vranic, «Experimental validation of measurements of glucose turnover in nonsteady state», *Am J Physiol*, vol. 234, fasc. 1, pp. E84-93, gen. 1978, doi: 10.1152/ajpendo.1978.234.1.E84.

[3] R. N. Bergman, Y. Z. Ider, C. R. Bowden, e C. Cobelli, «Quantitative estimation of insulin sensitivity», *Am J Physiol*, vol. 236, fasc. 6, pp. E667-677, giu. 1979, doi: 10.1152/ajpendo.1979.236.6.E667.

[4] A. Caumo, R. N. Bergman, e C. Cobelli, «Insulin sensitivity from meal tolerance tests in normal subjects: a minimal model index», *J Clin Endocrinol Metab*, vol. 85, fasc. 11, pp. 4396–4402, nov. 2000, doi: 10.1210/jcem.85.11.6982.

[5] C. Dalla Man, F. Micheletto, A. Sathananthan, R. A. Rizza, A. Vella, e C. Cobelli, «A model of GLP-1 action on insulin secretion in nondiabetic subjects», *Am J Physiol Endocrinol Metab*, vol. 298, fasc. 6, pp. E1115–E1121, giu. 2010, doi: 10.1152/ajpendo.00705.2009.

[6] C. Dalla Man, F. Micheletto, M. Sathananthan, A. Vella, e C. Cobelli, «Model-Based Quantification of Glucagon-Like Peptide-1-Induced Potentiation of Insulin Secretion in Response to a Mixed Meal Challenge», *Diabetes Technol Ther*, vol. 18, fasc. 1, pp. 39–46, gen. 2016, doi: 10.1089/dia.2015.0146.

[7] P. J. Phillips, «Oral glucose tolerance testing», *Aust Fam Physician*, vol. 41, fasc. 6, pp. 391–393, giu. 2012.

[8] M. Morettini, L. Burattini, C. Göbl, G. Pacini, B. Ahrén, e A. Tura, «Mathematical Model of Glucagon Kinetics for the Assessment of Insulin-Mediated Glucagon Inhibition During an Oral Glucose Tolerance Test», *Front Endocrinol (Lausanne)*, vol. 12, p. 611147, mar. 2021, doi: 10.3389/fendo.2021.611147.

[9] S. Contreras, D. Medina-Ortiz, C. Conca, e Á. Olivera-Nappa, «A Novel Synthetic Model of the Glucose-Insulin System for Patient-Wise Inference of Physiological Parameters From Small-Size OGTT Data», *Frontiers in Bioengineering and Biotechnology*, vol. 8, 2020, Consultato: 22 settembre 2023. [Online]. Disponibile su: https://www.frontiersin.org/articles/10.3389/fbioe.2020.00195

[10] C. Casella e V. Taglietti, *Principi di fisiologia*, vol. 2.

[11] A.-M. Aalto, A. Uutela, e A. R. Aro, «Health related quality of life among insulin-dependent diabetics: disease-related and psychosocial correlates», *Patient Education and Counseling*, vol. 30, fasc. 3, pp. 215–225, mar. 1997, doi: 10.1016/S0738-3991(96)00963-9.

[12] «American Diabetes Association», American Diabetes Association. Consultato: 7 ottobre 2023. [Online]. Disponibile su: https://diabetesjournals.org/

[13] *European Association For The Study Of Diabetes." Diabetologia 45.5 (2002): R23–R23. Web.*

[14] *International Diabetes Federation. IDF Diabetes Atlas, 10th edn. Brussels, Belgium: 2021. Available at: https://www.diabetesatlas.org*.

[15] «World Health Organization (WHO)». Consultato: 7 ottobre 2023. [Online]. Disponibile su: https://www.who.int

[16] «Diagnosis and Classification of Diabetes Mellitus», *Diabetes Care*, vol. 33, fasc. Suppl 1, pp. S62–S69, gen. 2010, doi: 10.2337/dc10-S062.

[17] «Mechanism linking diabetes mellitus and obesity».

[18] «Thyroid Dysfunction and Diabetes Mellitus: Two Closely Associated Disorders | Endocrine Reviews | Oxford Academic». Consultato: 20 settembre 2023. [Online]. Disponibile su: https://academic.oup.com/edrv/article/40/3/789/5288751

[19] C. Mathieu e K. Badenhoop, «Vitamin D and type 1 diabetes mellitus: state of the art», *Trends in Endocrinology & Metabolism*, vol. 16, fasc. 6, pp. 261–266, ago. 2005, doi: 10.1016/j.tem.2005.06.004.

[20] C. Joergensen, P. Hovind, A. Schmedes, H.-H. Parving, e P. Rossing, «Vitamin D Levels, Microvascular Complications, and Mortality in Type 1 Diabetes», *Diabetes Care*, vol. 34, fasc. 5, pp. 1081–1085, mag. 2011, doi: 10.2337/dc10-2459.

[21] C. Spaight, J. Gross, A. Horsch, e J. J. Puder, «Gestational Diabetes Mellitus», *Endocr Dev*, vol. 31, pp. 163–178, 2016, doi: 10.1159/000439413.
[22] D. R. Coustan, «Gestational Diabetes Mellitus», *Clinical Chemistry*, vol. 59, fasc. 9, pp. 1310–1321, set. 2013, doi: 10.1373/clinchem.2013.203331.
[23] R. Taylor, «Insulin Resistance and Type 2 Diabetes», *Diabetes*, vol. 61, fasc. 4, pp. 778–779, mar. 2012, doi: 10.2337/db12-0073.
[24] A. R. Saltiel e C. R. Kahn, «Insulin signalling and the regulation of glucose and lipid metabolism», *Nature*, vol. 414, fasc. 6865, Art. fasc. 6865, dic. 2001, doi: 10.1038/414799a.
[25] R. Muniyappa, S. Lee, H. Chen, e M. J. Quon, «Current approaches for assessing insulin sensitivity and resistance in vivo: advantages, limitations, and appropriate usage», *Am J Physiol Endocrinol Metab*, vol. 294, fasc. 1, pp. E15-26, gen. 2008, doi: 10.1152/ajpendo.00645.2007.
[26] M. Gutch, S. Kumar, S. M. Razi, K. K. Gupta, e A. Gupta, «Assessment of insulin sensitivity/resistance», *Indian J Endocrinol Metab*, vol. 19, fasc. 1, pp. 160–164, 2015, doi: 10.4103/2230-8210.146874.
[27] R. A. DeFronzo, J. D. Tobin, e R. Andres, «Glucose clamp technique: a method for quantifying insulin secretion and resistance», *Am J Physiol*, vol. 237, fasc. 3, pp. E214-223, set. 1979, doi: 10.1152/ajpendo.1979.237.3.E214.
[28] E. Breda, M. K. Cavaghan, G. Toffolo, K. S. Polonsky, e C. Cobelli, «Oral glucose tolerance test minimal model indexes of beta-cell function and insulin sensitivity», *Diabetes*, vol. 50, fasc. 1, pp. 150–158, gen. 2001, doi: 10.2337/diabetes.50.1.150.
[29] C. Dalla Man *et al.*, «Two-hour seven-sample oral glucose tolerance test and meal protocol: minimal model assessment of beta-cell responsivity and insulin sensitivity in nondiabetic individuals», *Diabetes*, vol. 54, fasc. 11, pp. 3265–3273, nov. 2005, doi: 10.2337/diabetes.54.11.3265.
[30] O. Tschritter, A. Fritsche, F. Shirkavand, F. Machicao, H. Häring, e M. Stumvoll, «Assessing the shape of the glucose curve during an oral glucose tolerance test», *Diabetes Care*, vol. 26, fasc. 4, pp. 1026–1033, apr. 2003, doi: 10.2337/diacare.26.4.1026.
[31] H. M. Ismail *et al.*, «The shape of the glucose concentration curve during an oral glucose tolerance test predicts risk for type 1 diabetes», *Diabetologia*, vol. 61, fasc. 1, pp. 84–92, gen. 2018, doi: 10.1007/s00125-017-4453-6.
[32] G. Freckmann, «Basics and use of continuous glucose monitoring (CGM) in diabetes therapy», *Journal of Laboratory Medicine*, vol. 44, fasc. 2, pp. 71–79, apr. 2020, doi: 10.1515/labmed-2019-0189.
[33] D. Rodbard, «Continuous Glucose Monitoring: A Review of Successes, Challenges, and Opportunities», *Diabetes Technol Ther*, vol. 18, fasc. Suppl 2, pp. S2-3-S2-13, feb. 2016, doi: 10.1089/dia.2015.0417.
[34] S. Dundar, B. Gokkurt, e Y. Soylu, «Mathematical Modelling at a Glance: A Theoretical Study», *Procedia - Social and Behavioral Sciences*, vol. 46, pp. 3465–3470, 2012, doi: 10.1016/j.sbspro.2012.06.086.
[35] A. Mari, A. Tura, E. Grespan, e R. Bizzotto, «Mathematical Modeling for the Physiological and Clinical Investigation of Glucose Homeostasis and Diabetes», *Frontiers in Physiology*, vol. 11, 2020, Consultato: 7 ottobre 2023. [Online]. Disponibile su: https://www.frontiersin.org/articles/10.3389/fphys.2020.575789
[36] R. Bizzotto *et al.*, «Glucose uptake saturation explains glucose kinetics profiles measured by different tests», *American Journal of Physiology-Endocrinology and Metabolism*, vol. 311, fasc. 2, pp. E346–E357, ago. 2016, doi: 10.1152/ajpendo.00045.2016.
[37] M. J. Page *et al.*, «The PRISMA 2020 statement: an updated guideline for reporting systematic reviews», *BMJ*, p. n71, mar. 2021, doi: 10.1136/bmj.n71.
[38] «Scimago Journal & Country Rank». Consultato: 8 settembre 2023. [Online]. Disponibile su: https://www.scimagojr.com/
[39] «Pima Indians Diabetes Database - dataset by data-society», data.world. Consultato: 26 settembre 2023. [Online]. Disponibile su: https://data.world/data-society/pima-indians-diabetes-database
[40] J. Ha, R. Muniyappa, A. S. Sherman, e M. J. Quon, «When MINMOD Artifactually Interprets Strong Insulin Secretion as Weak Insulin Action», *Frontiers in Physiology*, vol. 12, 2021, Consultato: 26 settembre 2023. [Online]. Disponibile su: https://www.frontiersin.org/articles/10.3389/fphys.2021.601894

[41] R. Edinburgh, A. Hengist, H. Smith, J.-P. Walhin, J. Betts, e J. Gonzalez, «Dataset for "Prior exercise alters the difference between arterialised and venous glycaemia – implications for blood sampling procedures"». Consultato: 26 settembre 2023. [Online]. Disponibile su: https://researchdata.bath.ac.uk/352/
[42] E. Manell, P. Hedenqvist, A. Svensson, e M. Jensen-Waern, «Establishment of a Refined Oral Glucose Tolerance Test in Pigs, and Assessment of Insulin, Glucagon and Glucagon-Like Peptide-1 Responses», *PLoS ONE*, vol. 11, fasc. 2, p. e0148896, feb. 2016, doi: 10.1371/journal.pone.0148896.
[43] «OGTT/Datos_OGTT.xlsx at main · hugofloresar/OGTT», GitHub. Consultato: 26 settembre 2023. [Online]. Disponibile su: https://github.com/hugofloresar/OGTT/blob/main/Datos_OGTT.xlsx
[44] «TDA_HumanOGTT/Data/Metabolites_and_hormones_data.xlsx at main · sfujita0601/TDA_HumanOGTT», GitHub. Consultato: 26 settembre 2023. [Online]. Disponibile su: https://github.com/sfujita0601/TDA_HumanOGTT/blob/main/Data/Metabolites_and_hormones_data.xlsx
[45] H. T. Abbas *et al.*, «Predicting long-term type 2 diabetes with support vector machine using oral glucose tolerance test», *PLoS One*, vol. 14, fasc. 12, p. e0219636, dic. 2019, doi: 10.1371/journal.pone.0219636.
[46] T. Lekva *et al.*, «Prediction of Gestational Diabetes Mellitus and Pre-diabetes 5 Years Postpartum using 75 g Oral Glucose Tolerance Test at 14–16 Weeks' Gestation», *Sci Rep*, vol. 8, fasc. 1, Art. fasc. 1, set. 2018, doi: 10.1038/s41598-018-31614-z.
[47] B. Bent *et al.*, «Engineering digital biomarkers of interstitial glucose from noninvasive smartwatches», *npj Digit. Med.*, vol. 4, fasc. 1, Art. fasc. 1, giu. 2021, doi: 10.1038/s41746-021-00465-w.
[48] «OhioT1DM Dataset». Consultato: 26 settembre 2023. [Online]. Disponibile su: http://smarthealth.cs.ohio.edu/OhioT1DM-dataset.html
[49] «ClinicalTrials.gov ID NCT00469833».
[50] «ClinicalTrials.gov ID NCT00441129».
[51] «ClinicalTrials.gov ID NCT02429258».
[52] «ClinicalTrials.gov ID NCT01027871».
[53] «ClinicalTrials.gov ID NCT03135015».
[54] «ClinicalTrials.gov ID NCT00807092».
[55] C. Cobelli, C. D. Man, M. G. Pedersen, A. Bertoldo, e G. Toffolo, «Advancing Our Understanding of the Glucose System via Modeling: A Perspective», *IEEE Transactions on Biomedical Engineering*, vol. 61, fasc. 5, pp. 1577–1592, mag. 2014, doi: 10.1109/TBME.2014.2310514.
[56] R. N. Bergman, «Minimal Model: Perspective from 2005», *Hormone Research*, vol. 64, fasc. Suppl. 3, pp. 8–15, gen. 2006, doi: 10.1159/000089312.
[57] M. Y. Pepino, C. D. Tiemann, B. W. Patterson, B. M. Wice, e S. Klein, «Sucralose Affects Glycemic and Hormonal Responses to an Oral Glucose Load», *Diabetes Care*, vol. 36, fasc. 9, pp. 2530–2535, set. 2013, doi: 10.2337/dc12-2221.
[58] E. Muscelli *et al.*, «Impact of incretin hormones on β-cell function in subjects with normal or impaired glucose tolerance», *American Journal of Physiology-Endocrinology and Metabolism*, vol. 291, fasc. 6, pp. E1144–E1150, dic. 2006, doi: 10.1152/ajpendo.00571.2005.