The University Of Sheffield.

Investigating the Potential Effect of Consanguinity on Type 2 Diabetes Susceptibility in a Saudi Population.

Thesis by

Ibrahim Metaan Gosadi

Supervisors: Dr. Dawn Teare, Professor Elizabeth Goyder

A dissertation submitted in fulfilment

of the requirements for the degree of Doctor of Philosophy

Faculty of Medicine, Dentistry and Health

School of Health and Related Research

The University of Sheffield

September 2013

Summary

Background: Several studies suggested association between consanguinity and risk of developing Type 2 Diabetes (T2D).

Aim: To examine mechanisms by which consanguinity might increase the risk of T2D in a Saudi population.

Methodology: 362 adult male participants were recruited, 179 were T2D patients and 183 healthy participants were siblings of recruited patients. Severity was assessed in patients by recording age at diagnosis. Diabetes risk in healthy subjects was inspected by measuring their body mass index (BMI), fasting blood glucose (FBG), and waist circumference. Extended pedigrees were constructed to calculate inbreeding coefficients. 23 Single Nucleotide Polymorphisms (SNPs) incurring higher risk of T2D were genotyped. All subjects were interviewed to complete food frequency and physical activity questionnaires to provide information on environmental variation between participants.

Results:

Significant inverse association was detected between inbreeding coefficients and age at diagnosis accounting for environmental covariates (β: -0.572 P-value: 0.012). In 42 families, we were able to recruit 2 healthy siblings from each. Pearson's correlation coefficient of FBG between siblings was 0.317 (P= 0.04). Correlations between siblings' FBG increased with increased range of consanguinity suggesting a stronger genetic influence leading to lower variation of FBG between siblings. The effect of consanguinity on variation of FBG was further assessed by fitting a regression line and controlling for difference in age, calorific intake, and level of physical activity (β: -0.118 P-value: 0.024).

No significant associations were detected between number of loci identical for risk alleles and age at diagnosis, BMI, FBG or waist circumference. An association of marginal significant was detected between age at diagnosis and total number of risk alleles when accounting for parental history of diabetes and inbreeding coefficients (β:-0.399 P: 0.052).

Conclusion: Study's findings suggest consanguinity might increase risk of T2D by earlier development of the disease, and by strengthening possible genetic effect on FBG.

Acknowledgment

This project is a fruit of hard labour and cooperation of supervisors, family, friends, and hundreds of participants whom I had the honour of meeting them. I must extend a special appreciation to my parents and my wife for their immense support and inspiration. By the time of writing these words, my father was lying next to me on his white bed awaiting placement of a stent in one of his Coronary Arteries to alleviate one of the complications of his chronic battle with type 2 diabetes.

Much gratitude is given to Nicola Crawford, from Sheffield Clinical Genetic Services. Crawford gave several valuable suggestions regarding ethics application, interviewing procedure and pedigrees constructions. She also suggested visiting the National Genetic Education and Development Centre website to gain more knowledge about pedigrees construction. Additionally, due to her insight regarding pedigrees construction, we were able to use her suggestions and apply them successfully during data collection.

The cooperation of the staff at the Directory of Health in Jazan was much appreciated. The Directory provided advice regarding the selection of PHCs. It was of most assistance to receive an official letter from the Directory designated to all administrators of each PHC. The cooperation was continued when we provided PHCs' administrators with the official letter. Once the investigator was introduced to the PHCs staffs by their administrators, significant assistance was made by nurses and doctors who aided in identification and approaching processes.

List of Publications

List of accepted abstracts based on the thesis:

1. Genetic Epidemiology of Type 2 Diabetes [poster presentation]. World Diabetes Congress. Dubai, UAE. 4-8 December 2011.
2. Can Inbreeding Increase the Risk of Type 2 Diabetes? [poster presentation]. 1st American Diabetes Association Middle East Congress. Dubai, UAE. 4-6 December 2012.
3. Association between Consanguinity and Fasting Blood Glucose in a Saudi Arabian Population [poster presentation]. 57th Annual Scientific Meeting Brighton and Sussex Medical School, University of Sussex. 11 –13 September 2013
4. Genetic and Familial Factors Associated with Younger Age at Diagnosis of Type 2 Diabetes. [poster presentation]. World Diabetes Congress. Melbourne, Australia. 2-6 December 2013.

Table of Contents

List of Tables

List of Figures

Glossary

Allele: one of a number of alternative forms of the same gene occupying a given position in a chromosome.

Additive effect: when the combined effects of alleles on the phenotype at different loci are equal to the sum of their individual effects.

Consanguinity: the property of being descended from a common ancestor.

Consanguineous marriage: marriage between individuals who share a common ancestor.

DNA: Deoxyribonucleic Acid a long linear polymer found in the nucleus of a cell, formed from nucleotides and shaped like a double helix; responsible for the transmission of genetic information.

Dominant: production of the same phenotype whether its alleles are identical or dissimilar.

Dominance: a deviation from additive allelic effects on phenotype within the same locus.

Gene: a segment of DNA.

Genotype: the specific pair of alleles at a single locus in a diploid organism that determines a specific trait of an individual, such as ‘Aa’ or ‘aa’.

Gluconeogenesis: the process of glucose synthesis.

Hardy-Weinberg equilibrium: a state in which genotype frequencies and proportions remain constant from generation to generation.

Heritability: proportion of population phenotypic variation attributable to variation in genetic components.

Heterozygote: having different copies of alleles at a locus.

Homeostasis: the ability of a system or living organism to adjust its internal environment to maintain a stable equilibrium.

Homozygote: having an identical copy of alleles at a locus.

Inbreeding: mating of related individuals.

Khat: Catha edulis or Khat is a plant with stimulating amphetamine-like characteristics.

Locus: the specific location of a gene or DNA sequence on a chromosome.

Metabolism: the complete set of chemical reactions that occur in living cells.

Monogenic: a condition in which inheritance of mutation in a single gene is responsible for development of a phenotype.

Mutation: changes in the DNA sequence of a cell's genome which may be caused by internal and external factors.

Nucleotide: the basic building unit of nucleic acids such as DNA and RNA. Nucleotides can be Adenine (A), Guanine (G), Thymine (T), or Cytosine (C).

Phenotype: the visible or measurable (i.e. expressed) characteristics of an organism.

Polygenic**:** a condition in which many more than one gene contribute to development of a phenotype.

Primers**:** synthetic sequence of nucleotides serve as the starting point of DNA synthesis during the PCR process.

Recessive: production of a phenotype when its alleles are identical.

Selection: certain traits or alleles of a species which may be favoured in certain environments.

Abbreviations

ATP: Adenosine Triphosphate.

BMI: Body Mass Index.

CI: Confidence Intervals.

CNF : Canadian Nutrient File.

DNA: Deoxyribonucleic Acid.

EPIC: European Prospective Investigation into Cancer Study

FBG: Fasting Blood Glucose.

HbA(1c): Glycated haemoglobin.

HOMA-B: Homeostasis Model Assessment B.

HTA: Human Tissue Authority.

IGT: Impaired Glucose Tolerance.

IC: Inbreeding Coefficient.

KASP: Kompetitive Alelle Specific PCR

OR: Odds Ratio.

PHC: Primary Health Care.

PCR: Polymerase Chain Reaction.

RR: Relative Risk .

SD: Standard Deviation.

SE: Standard Error.

SNPs: Single Nucleotide Polymorphisms.

T2D: Type 2 Diabetes.

USDA: National Nutrient Database for Standard Reference.

WC: Waist Circumference.

WHO: World Health Organization.

1Chapter One: Introduction

1.1 Background

Saudi Arabia has a high prevalence of type 2 diabetes. Several epidemiological investigations have suggested multiple factors contributing to the high level of the disease such as obesity, overweight, and reduced level of physical activity. However, type 2 diabetes is a complex disease and not solely dependent on environmental factors. Genetic factors have been reported by several studies indicating that inheriting certain type 2 diabetes risk alleles could increase the risk of developing the disease.

Several genetic variants have been suggested to increase risk of type 2 diabetes with different levels of effect. Most of detected variants are presumed to act additively. However, several epidemiological investigations were able to find an association between consanguinity and risk of developing type 2 diabetes. This might suggest a possible recessive effect increasing the risk of developing type 2 diabetes.

Saudi Arabia has a high prevalence of consanguinity compared to western societies. Almost half of marriages in Saudi Arabia are arranged between related individuals. Higher consanguinity levels are likely to increase overall homozygosity of the Saudi Arabian population. Increased homozygosity might increase the power to detect recessive effect whether on monogenic illnesses or even more complex diseases such as type 2 diabetes.

1.2 Research Problem

Despite the global increase of type 2 diabetes prevalence, there are several epidemiological investigations suggesting that certain populations could be at higher risk of developing the disease compared to others. This is clearly apparent in societies where individuals from different ethnicities, but sharing similar environment, have different levels of risk of developing type 2 diabetes. This might indicate that certain populations might be genetically more susceptible to type 2 diabetes compared to others. Genetic susceptibility could be driven by several factors influencing genetic structure of populations such as inbreeding.

1.3 Research aims and objectives

The current thesis is interested in investigating the mechanism by which consanguinity might increase risk of type 2 diabetes. Several Genome Wide Association Studies (GWAS) revealed multiple alleles incurring higher risk of the disease. However, these studies were mainly conducted in Western and East Asian and populations with low level of consanguinity.

This study hypothesized that increased level of consanguinity in the Saudi Arabian population has lead to increased overall homozygosity. In recessive effects, inheriting identical copies of the risk alleles increases the risk of developing a disease while those with

heterozygote genotypes are at lower risk of developing the disease. The study argues that with increased overall homozygosity in the Saudi Arabian population, there could be a chance to observe a cumulative recessive effect due to inheriting multiple loci identical for type 2 diabetes risk alleles.

The main objectives of this study are:

- To investigate the association between levels of consanguinity and number of detected type 2 diabetes risk alleles.
- To investigate possible cumulative recessive effect of measured alleles on earlier age at diagnosis of the disease.
- To assess cumulative recessive effect of measured alleles on continuous traits related to type 2 diabetes such as BMI, FBG and waist circumference.

1.4 Significance of this study

There is a limited research related to investigating non-additive genetic effect on development of type 2 diabetes. Although several studies have suggested synergism between consanguinity and type 2 diabetes, this investigation is the first to explore mechanisms linking between consanguinity and type 2 diabetes. Unlike other studies assessing this association, we utilized extended pedigrees to measure level of consanguinity. Additionally, this study is the first to measure SNPs related to type 2 diabetes in Saudi Arabian population. Giving lack of genotyping infrastructures in Saudi Arabia, this study provides a comprehensive guide on, solutions, logistics and methods of genotyping SNPs in a Saudi Arabian subjects which could be utilized for similar investigations.

1.5 Thesis structure

This thesis is composed of seven chapters and appendices. The current chapter provides an overview of the study and how main research questions were formulated. The second chapter is a literature review which covered several points to illustrates nature of type 2 diabetes, type 2 diabetes in Saudi Arabia, assessment of genetic risk factors of type 2 diabetes, population genetics of type 2 diabetes and genetic structure of Saudi Arabian population. The third chapter is related to generation of research hypotheses and main questions while the fourth chapter is a detailed illustrate of methods used to answer research questions. The fifth chapter is composed of descriptive and analytic sections where the main research questions were answered. The sixth chapter comprise a discussion of the study's main findings, strengths and weaknesses of the current research and gained experiences. The final chapter is a summary, recommendations and conclusions of this investigation.

2 Chapter Two: Background

2.1 Type 2 diabetes

2.1.1 Epidemiology of type 2 diabetes

Type 2 diabetes is one of the most common chronic conditions. There are about 347 million patients in the globe (1) where 90% of them are type 2 diabetes patients (2). This number is projected to increase in 2030 to 438 million patients (3). In 2004, 4.8 million people died due to the complications of the disease (2).

Formerly known as Non-Insulin Dependent Diabetes Mellitus (NIDDM), the disease is more prevalent in certain localities of the world. According to the International Diabetes Federation (IDF), Diabetes has the highest prevalence in Middle East and North African region (Prevalence 10.9%) followed by the North American and Caribbean regions (10.2%). Other parts of Africa have a much lower prevalence of diabetes (4.3%) (4). This variation in prevalence is largely explained by the variation in genetic and environmental backgrounds.

Monogenic and immune-mediated forms of diabetes are more likely to occur in younger ages (5). Type 2 diabetes is considered a disease of ageing where older populations are at more risk of getting the disease. One study conducted in the US reported that the total rates of diabetes increased with age, from 2.0% at age 20-44 years to 17.7% at age 65-74 years (6). However, although type 2 diabetes is still a disease of ageing, the number of children becoming diagnosed with the disease is progressively increasing (5).

Type 2 diabetes is equally prevalent in males and females. Nevertheless, men are more likely to acquire the disease in middle age than women (7). One study reported that females are more likely to experience symptoms related to type 2 diabetes than males. However, the complications of type 2 diabetes are very similar in both males and females, except for cataracts, which have been reported as being higher in females (odds ratio (OR) and 95% confidence intervals (CI) 1.47 (1.04, 2.08))(8).

The risk of developing type 2 diabetes stems from the presence of predisposing genetic influences and certain environmental conditions. The genetic risk is clearly apparent from the familial pattern of the disease (9, 10) and from concordance rates in twins (11, 12). A parental history of diabetes increases the risk of type 2 diabetes in offspring. The odds of having type 2 diabetes when there is paternal diabetes are 3.5 (95% CI: 2.3-5.23) and 3.4 (95% CI:2.3-4.9) in the case of maternal diabetes. The risk of getting the disease increases when both parents are affected by type 2 diabetes. The odds of having the disease when both parents are affected by type 2 diabetes increases to 6.1 (95% CI: 2.9-13.0) (13).

Despite the obvious genetic influence on type 2 development, environmental factors are similarly significant contributors (12). The presence of genetic influence of the disease might actually be augmented or prevented by other environmental factors. The westernized life style and decreased physical activity are major environmental contributors to the development of type 2 diabetes. These environmental changes can mean that obesity is more likely to occur, which creates a higher risk of the disease occurring (10).

There is an obvious association between obesity and the risk of developing type 2 diabetes. A higher BMI is an important risk factor (14, 15). A longitudinal cohort study reported an increase in the risk of type 2 diabetes in men who experience a substantial weight gain (>10%) compared to those who had a stable weight (Relative Risk [RR] 1.61 [95% CI 1.01-2.56]) after adjustment for age, and initial BMI. It has been suggested that the increased duration of high BMI can increase the risk of type 2 diabetes (16). Central obesity in particular may pose a higher risk for developing type 2 diabetes (10) (15, 17). Central obesity plays a major rule in dyslipidemia and in increasing resistance to insulin (18-20). Insulin resistance can in turn eventually lead to the development of type 2 diabetes (5).

It has also been suggested that a low birth weight may increase the risk of type 2 diabetes in later life. A systematic review was conducted to assess the association between type 2 diabetes and birth weight. It was found that birth weight is inversely associated with type 2 diabetes. The pooled odds ratio for type 2 diabetes was 0.75 (95% CI, 0.70-0.81) per kilogram, after adjustment for age and sex. Adjustment for socioeconomic status did not appear to change the association (21).

In addition to the previously mentioned risk factors, essential hypertension has been suggested as a risk factor for type 2 diabetes. The relative risk of metabolic syndromes (including type 2 diabetes) increases to 1.55 (95% CI 1.21; 1.99) in individuals with prehypertension when compared with normotensive subjects (22). This implies that hypertension may have an effect on insulin sensitivity and thus increase the risk of the disease (23). Ethnicity is another risk factor for type 2 diabetes as non-Caucasians are at higher risk of developing the disease (24). Psychological and socioeconomic conditions have also been suggested as associated factors (25, 26).

Several clinical scenarios can be experienced before the development of type 2 diabetes. An elevated blood sugar level caused by the inability of the body to metabolize sugar can form a prediabetes stage. This stage includes Impaired Glucose Tolerance (IGT) and Impaired Fasting Glucose (IFG). Individuals who exhibit this stage are at higher risk of developing of type 2 diabetes. A similar example is gestational diabetes, a condition where some women develop high levels of blood glucose during pregnancy which is associated with increase the risk of type 2 diabetes after pregnancy (27).

2.1.2 Pathophysiology of type 2 diabetes

The WHO defines type 2 diabetes as being a chronic illness where the body is unable to use the insulin it produces effectively (2). Insulin is a hormone produced by the pancreas. The pancreas is an endocrine and exocrine organ located in the upper left quadrant of the abdomen. The endocrine part of the pancreas is called the Islets of Langerhans and it contains both alpha and beta cells. Alpha cells are responsible for the production of glucagon and beta cells are responsible for insulin production (28). While insulin secretion is stimulated by high level of blood sugar, known as hyperglycemia, glucagon is secreted when there is a low level of blood sugar, known as hypoglycemia.

Hyperglycemia usually occurs after ingestion of food, especially food rich in carbohydrates (28). An enzyme called glucokinase is responsible for sensing the high level of glucose in the blood and sending signals to the pancreas to produce insulin (29). A high level of blood amino acids after absorption of protein-rich meal will also stimulate the secretion of insulin. Other digestive hormones, particularly glucose-dependent insulinotropic peptides, and parasympathetic nervous stimuli after eating food can also contribute to the induction of insulin secretion (30).

Glucose is used by body cells for energy production. Insulin is combined with membrane receptors and increases the permeability of cell membrane to facilitate the movement of glucose into the cell. Insulin reduces blood glucose, whether by increasing the utilization of glucose for energy production or increasing the storage rate of glucose as glycogen in the liver and skeletal muscles. Insulin also induces the utilization of glucose to produce amino acids, essential for protein production, and the production of fatty acids, which are later used in adipose tissue to form triglycerides (30). In addition to aiding in building protein and triglycerides, it also inhibits their degradation.

In type 2 diabetes patients, insulin is not being used effectively. Ineffective use can be due to resistance against insulin actions in peripheral tissues (5, 31, 32), or because of the deficiency of the insulin receptors of cell membranes (28). Increased insulin resistance and elevated levels of blood sugar will ultimately force the pancreas to produce more insulin as a compensatory mechanism, which may result in hyperinsulinemia (33). Hyperinsulinemia might be further augmented by the inability of the liver to uptake and degrade circulating insulin (5). Type 2 diabetes patients may suffer from defective responses to blood sugar (5) which can be caused by a lack of in glucokinase (29). A reduction in quantitative and qualitative aspects of insulin can eventually lead to an increase in the rate of hepatic gluconeogenesis and thus to more hyperglycemia (5, 31).

Abnormalities in insulin function or secretion can lead to several pathological consequences. Cells may not be able to utilize the available blood glucose. A lack of insulin can increase the

degradation of triglycerides, and this may result in increased levels of blood fatty acids, which are an alternative source of energy. Protein degradation may also increase, leading to the wasting of muscles (30).

2.1.3 Clinical presentation

There are several clinical presentations of type 2 diabetes. In certain cases, it can be asymptomatic and patients are only diagnosed by routine blood glucose measurements. Younger cases are more likely to experience a faster development of diabetes and faster appearance of its symptoms, perhaps over days or a few weeks. In older cases, type 2 diabetes can be subtle and might take months before it can be clinically presented (34).

When a normal level of blood glucose exists, the kidneys prevent the passage of glucose into the urine. In patients with diabetes, the elevated level of blood glucose will render the kidneys unable to maintain the urine in a glucose-free condition. The excretion of glucose through urine will increase the volume of urine and the frequency of voiding, creating a condition known as polyuria. Due to the high output of body fluids which pass through the urinary system, the body will eventually suffer dehydration, and increased sensations of thirst and dryness in the mouth, clinically known as polydipsia (35). To compensate for the loss of fluids, patients with diabetes drink higher amounts of water. Due to frequent fluid loss and dehydration, some patients may experience constipation (34).

Hyperglycemia and increased urinary output will cause further disturbances. The increased level of sugar in the urine will facilitate the occurrence of Candidiasis, leading to pruritus vulvae and balanitis (35). A high loss of body fluids may cause disturbances in the chemical environment of the body. This alteration of the chemical balance might lead to nervous disturbance, especially in the peripheral nerves, leading to tingling sensations in the hands and feet. A chemical imbalance may lead to weakening of the immune system, giving rise to a higher incidence of infections (34).

Since the body is unable to utilize blood glucose, storage of lipid may be utilized as an alternative energy source. This may lead to marked weight loss, even if the person has a good appetite. High levels of glucose will also have an effect on vision ability. Elevated levels of blood glucose may cause a disturbance to the focusing ability of the lens, rendering the vision hazy or blurred (34).

Some cases with diabetes might present by other symptoms caused by the long-term complications of diabetes. Chronic hyperglycemia eventually causes vascular changes where capillaries get thicker. The process of oxygen and nutrient exchange will be affected by changes in vascular nature, leading to less or diminished supplies reaching other parts of the body (28).

Furthermore, hyperglycemia increases the chance of atherosclerotic changes, leading to a decreased blood supply to several organs.

A delay in the diagnosis of type 2 diabetes may lead to macrovascular and microvascular complications being present at the time of diagnosis. Some cases can be complicated by coronary heart disease (36), cerebrovascular accidents, or peripheral vascular disease (5). Some studies suggest that increased levels of glycated haemoglobin HbA(1c) and fasting blood glucose are associated with increased risks of developing coronary heart disease (5, 37). In addition, vascular changes can also affect the kidneys, leading to diabetic nephropathy and renal failure (5).

A common complication of diabetes is retinopathy. Retinal blood vessels can become occluded due to the occurrence of vascular changes, leading to weakness of vision or blindness. After 20 years of diabetes, 60% of patients with diabetes are more likely to develop varying degrees of retinopathy (38). The risk of diabetic retinopathy is associated with longer duration of diabetes (38) and elevated levels of glycated haemoglobin (39, 40).

Macrovascular and microvascular complications are considered the leading cause of morbidity and mortality in patients with diabetes (38). This may have a significant impact on life expectancy, leading to an approximate reduction of 10 years (5). Nevertheless, patients with diabetes have different tendencies towards developing these complications. Some patients with diabetes do not develop any complications at all. This variation in the incidence of complications suggests that an elevated blood sugar level is not sufficient to cause these complications and in fact other factors have to be maintained to cause the illness (38).

2.1.4 Diagnosis of type 2 diabetes

To confirm the diagnosis of type 2 diabetes, blood glucose level need to be investigated. WHO has established diagnostic criteria for diabetes. FPG $\geq$ 7mmol/l (126mg/dl) or 2-hours blood glucose $\geq$ 11.1mmol/l (200mg/dl) are considered diagnostic criteria (41). An FPG test should be preceded by no caloric intake occurring for 8 hours before performing the test and should be repeated to further confirm the diagnosis (5).

Glycated haemoglobin (HbA1c) is used to measure the effectiveness of diabetes therapy. Recently, it has being recommended as a diagnostic test for diabetes. A HbA1c of 6.5% is recommended as a cut-off point for diagnosing diabetes. Despite this, a level below 6.5% does not rule out diabetes and blood glucose can be investigated through alternative tests (42). Although HbA1c is recommended for diagnosing type 2 diabetes, it might not be as feasible in certain parts of the world as other laboratory measures of glucose (43).

2.1.5 Management of type 2 diabetes

Management of type 2 diabetes is a comprehensive process whereby undertaking one measure is unlikely to provide meticulous control of the disease. This process is not solely dependent on pharmacological agents. Lifestyle changes can be a significant step in providing a better prognosis for the disease. Regular exercise can be a key intervention which can reduce body weight and central obesity and consequently reduce insulin resistance (44). Having low-glycemic index food is essential for better glycemic control (45). Health education is a vital tool in drawing the attention of patients with diabetes toward the importance of these lifestyle changes.

Depending on the progress of diabetes, pharmacological agents vary between oral glycemic control agents or those for the administration of insulin. Oral glycemic control drugs act in various different ways. There are insulin sensitizers which aim to increase the action of insulin, whether it appears in the liver as biguanides or peripherally as thiazolidinediones. Other drugs as such as sulfonylureas, repaglidine and nateglinide increase the secretion of insulin. Alpha-glucosidase inhibitors reduce the absorption of carbohydrate in the intestine and thus reduce the blood glucose level (5). Other pharmacological agents such as Orlistat and Sibutramine (35) may assist in weight reduction. The failure of previously mentioned management strategies will indicate the use of insulin.

In addition to conventional lifestyle changes and pharmacological agents, other surgical options have been introduced to control type 2 diabetes. Bariatric surgeries have different procedures depending on surgical intervention applied to the stomach. These procedures have different forms such as applying adjustable gastric bands, gastric bypass which involves the diversion of food flow from the stomach or sleeve gastrectomy which involves removal of parts of the stomach. All these procedure have a shared effect on reducing the level of BMI, blood glucose and induce the remission of type 2 diabetes in varying magnitudes(46).

One study compared the level of BMI, fasting blood glucose and haemoglobin A1c between two groups composed of morbidly obese type 2 diabetes patients. The first group included patients who underwent sleeve gastrectomy and the second group had a conventional medical therapy. There was a significant drop over an 18 month period from the beginning of applying interventions in both groups. In the first group means of BMI, fasting blood glucose and haemoglobin A1c dropped from 41.3 (SD: 6.0), 166.6 (SD: 68.1) mg/dL, and 7.9% (SD: 2.1%), respectively at baseline to 28.3 (SD: 5.4), 96.2 (SD: 29.4) mg/dL, and 6.0% (SD: 1.5%), respectively after 18 months. The second group did not exhibit any significant changes comparing between levels at baseline and after 18 months (47). Additionally, similar studies provided similar supporting evidence providing a stronger control of diabetes compared to medical therapy (48, 49).

Despite the significant effect of Bariatric surgery on enhancing diabetes control, it has several drawbacks. Introducing patients to surgical procedures might expose those patients to operative complications such as risk of infections. Additionally, risk of malnutrition might occur where certain nutrients will no longer be absorbed from the stomach which might mandate taking additional supplements (50).

2.1.6 Prevention of type 2 diabetes

Identifying individuals who are at higher risk of diabetes is an important element of the primary prevention of type 2 diabetes. Despite the genetic determinants, risk of type 2 diabetes can be reduced by undertaking a healthy lifestyle and balanced diet. A longitudinal cohort study investigated the effect of dietary pattern on increasing the risk of type 2 diabetes among Finnish subjects during a 23-year follow-up period. The study found that subjects who were engaged in the consumption of high calorific food items (such as potatoes and butter) were at higher risk of developing the disease compared with the group who reported higher consumption of fruits and vegetables [1.49 (95% CI: 1.11 - 2.00] (51).

One systematic review studied published randomised controlled trials which investigated the effect on the population of increased exercise and a healthier diet with regard to increased the risk of developing type 2 diabetes . This revealed that combined physical activity and balanced diet had a favourable effect on reducing the risk of type 2 diabetes (RR: 0.63, 95% CI: 0.49 to 0.79) (52). Another study implied that exercise might have a role in increasing insulin action (53).

The multi-factorial nature of type 2 diabetes indicates the importance of recognizing lifestyle risk factors in prevention of the disease in addition to the genetic risk. The ratios of developing the disease vary between 3.4 and 6.1 based on family history of type 2 diabetes (13). Additionally, reported odds ratios of SNPs incurring higher risk of type 2 diabetes varied between 1.07 and 1.7. However, despite the obvious genetic risk, we cannot neglect the effect of lifestyle on causation of type 2 diabetes. This influence is apparent by observing increased risk of type 2 diabetes due to dieting patterns rich with calorific value [RR 1.49] (51) or reduced risk due to healthier lifestyle and exercise [RR: 0.63] (52).

Several factors increased the difficultly of comparing the magnitude of risk of genetic factors with lifestyle factors. Firstly, genetic risk of type 2 diabetes is usually measured using retrospective methods where lifestyle factors could be measured prospectively. Therefore, different estimates of risk such as relative risk and odds ratios were reported. Usually, odds ratios are calculated for genetic risk factors where relative risks were calculated for lifestyle factors. Finally, in comparison with investigating the genetic effects, being able to control lifestyle factors made it possible to apply interventional methods to investigate how a change in lifestyle might increase or decrease risk of the disease

To conduct a comparison between lifestyle factors and genetic risk, population attributed risk was utilized. Several studies have reported population attributed risk of SNPs with strong effect on type 2 diabetes. Risk alleles of variants near TCF7L2, CDKN2A/B, KCNJ11and PPARG were reported to have a population attributed risk varying between 2% and 16% (54-58). Given the modest effects of these SNPs, population attributed risks were consequently small. However, a study indicated that combined population attributed risk of variants near KCNJ11and PPARG reached 28% (58).

The modest effects of detected SNPs indicate that detected genes are not strong enough to cause development of type 2 diabetes. A study which investigated population attributed risk of lifestyle factors revealed that low physical activity and consumption of processed red meat have a population attributable risk of 13% and 27% in men and 29% and 23% among women (59). Additionally, population attributable risk of high BMI is 49% among men and 50% among women (59). Given the reported population attributed risks of lifestyle factors compared to detected SNPs, it could be argued that combined lifestyle factors including level of physical activity and dieting behavior are likely to have a stronger influence on the disease compared to the detected genetic associations. However, it is possible that the combined effect of multiple SNPs is likely to have a higher population attributed risk that could be similar to the effect of combined lifestyle factors.

It is difficult to state which is more important in increasing the risk of type 2 diabetes: lifestyle or genetic risk factors. However, we might suggest that environmental factors could be augmenting the causation of the disease if the genetic risk is present. This notion might partially explain why individuals with non-healthy lifestyles do not develop diabetes or at least have a delayed development of the disease. Furthermore, if the genetic risk of type 2 diabetes is present in an individual, the risk of the disease could be reduced or the onset of the disease could be delayed by engaging in a healthier lifestyle. In addition, recognizing the effect of ageing on increasing the risk of the disease indicates the importance of applying preventive measures in ageing populations. In conclusion, behavioural and genetic risk factors of the disease are significant components of the disease and should be considered in any prevention programmes.

Secondary prevention of type 2 diabetes aims to screen individuals at risk from the disease. Undiagnosed patients are at a higher risk of developing the chronic complications of type 2 diabetes. Therefore, it is of great importance to establish cost-effective screening programmes for the disease and particularly for those with IGT who have an increased risk and who can potentially reduce the risks though lifestyle changes or drug treatment. Managing elevated blood sugar levels as early as possible can result in reductions of developing microvascular and macrovascular illnesses (60).

2.2 Type 2 diabetes in Saudi Arabia

2.2.1 Introduction

Type 2 diabetes is unquestionably a public health concern in Saudi Arabia. The prevalence of diabetes in Saudi Arabia is one of the highest in the globe (61). In 2000, the WHO reported that 890,000 individuals in Saudi Arabia were affected with diabetes and by 2030, this number is projected to increase to 2,523,000 (62). This profound increase in diabetic cases is likely to be related to several environmental and genetic factors. These factors appear to make the Saudi population more susceptible to the disease than other populations.

2.2.2 Prevalence of type 2 diabetes in Saudi Arabia

Several epidemiological studies have been conducted in Saudi Arabia to estimate the prevalence of diabetes in Saudi Arabia (63-69). An early review has explored this subject. It indicated an increase in the prevalence of diabetes from 2.5% in 1982 to 23.7% in 2004 (70). A more recent study, which was published in 2011, reported that the prevalence of diabetes in Saudi Arabia was 34.1% in males and 27.6% in females. It also reported that the mean age for the onset of diabetes in males and females was 57.5 (SD 13.1) and 53.4 (SD 13.1) years, respectively (69).

According to the International Diabetes Federation, Saudi Arabia is among the top 10 countries for the prevalence of Type 2 diabetes across the globe. The prevalence of diabetes in Saudi Arabia is almost four-fold the prevalence in the UK and twice the prevalence in the USA. The prevalence of diabetes in Saudi Arabia came second to the prevalence in Kuwait which has the highest prevalence of diabetes in the Middle East (1).

The prevalence of diabetes has been reported to have taken a different form in the Saudi population. Diabetes normally tends to be equally distributed between males and females. However, two studies conducted in the Saudi populations showed a gap between the prevalence of diabetes between men and women (68, 69). Seemingly, the prevalence of diabetes is higher in male subjects, but this suggestion remains inconclusive, as other studies suggest opposite results (64, 67). Furthermore, one study reported a higher prevalence of diabetes among the urban population (67). This might be explained by the differences in lifestyle and physical activity between the rural and urban populations.

2.2.3 Attitudes, knowledge and behaviour toward type 2 diabetes in Saudi Arabia

In addition to the increased prevalence of diabetes in Saudi Arabia, the situation is worsened by limited levels of knowledge about the disease and the preventive measures mentioned above. A study conducted in the eastern province of Saudi Arabia reported that less than a half of the primary care centres attendees knew about the risk of diabetes and how to prevent it (71). Another study conducted among intermediate and secondary school students suggested that most pupils do recognise the role of obesity in development of illnesses, in general but not as a risk factor for development of diabetes (72).

In addition to low public awareness about type 2 diabetes risk factors, health care providers in Saudi Arabia do not seem to be properly advocating the measures necessary for prevention of the disease. One study was conducted to measure the effort applied by primary care physicians in Riyadh City to educate their patients in the importance of physical activity. Only 24% of physicians were satisfactorily engaged in such educational activity and only 45% of the physicians were properly exercising themselves (73). These reports suggest a lack of proper public educational effort to raise awareness of diabetes.

2.2.4 Overweight and obesity in Saudi Arabia

Increased prevalence of overweight and obesity in the Saudi population gives an important explanation for the increase in type 2 diabetes. Table 2.1 shows several epidemiological studies estimating the prevalence of overweight and obesity in the Saudi population. Classification of weight is based on BMI, which is calculated by dividing weight in kilograms by the square of height in metres. Individuals with BMI below 18.5 are considered underweight, individuals with BMI between 18.5 and 25 are considered normal. Those with BMI above 25 are considered overweight and those above 30 are presumed obese (74).

The prevalence of overweight individuals ranges between 13.8% and 30.7%, while for obesity it ranges between 14.2% and 26.4%. Another two studies included diabetes and hypertension patients and revealed a noticeably higher prevalence of overweight and obesity among them than in the normal population (75, 76).

Overweight and obesity in the Saudi population seem to be influenced by several factors. Studies conducted of younger people suggest a smaller prevalence of obesity and overweight compared to older subjects. Furthermore, studies conducted in diabetes and hypertension patients yielded higher percentages with elevated BMIs. The influence of gender on the prevalence of overweight and obesity varies across different studies.

Table 2.1: Prevalence of overweight and obesity in the Saudi population:

(Reference) publication year	Sample Size	Age	Overweight		obese		Study population
			Males	Females	Males	Females	
(77) 2003	894	12-20	13.8%	-	20.5%	-	Males only
(78) 2007	19598	0-75	30.7%	28.4%	23.6%	14.2%	Mixed
(79) 2008	1072	8-12	-	-	-	14.9%	Females only
(80) 2010	7056	2-18	18%	20%	26.4%	19%	Mixed
(81) 2010	239	13-18	21.35%	23.41%	-	-	Mixed
(82) 2011	1869	11-19	11.5%	15.5%	11.8%	13.9%	Mixed

Several studies reported different reasons behind the increased prevalence of overweight and obesity. One study reported a high intake of carbohydrates and fat among adolescents in Jeddah city in Saudi Arabia (81). Another study reported family history as a risk factor for high BMI in Saudi adolescents (77). Moreover, physical inactivity is another contributor to the increased prevalence of overweight and obesity in Saudi individuals (77, 83-86). Saudi men are more likely to engage in physical activities than women (72).

Amin et al compared patterns of leisure physical activity between males and females where six out of 12 types of sports and leisure activities were not even practised by females(87). Another study including adult subjects confirmed this association indicating that females were more likely to exhibit a sedentary lifestyle compared to males (p value <0.001) (88). Additionally, a similar trend was observed in Saudi adolescent subjects (89). This could be a result of the cultural background in Saudi Arabia, which limits the physical activity of females. Furthermore, a lack of resources can be another barrier leading to both men and women engaging in less physical activities (83, 90).

A cross-cultural study compared levels of physical activity in male and female adolescents aged between 15 and 17 years. Three groups were selected from Birmingham and Coventry in the UK and a third one from the Al-Ahsa region in Saudi Arabia. This study showed that the proportion of adolescents with low levels of physical activity in Saudi Arabia is higher than those reported in the UK. Additionally, the proportion of individuals engaged in physical activities in the UK was higher compared to Saudi adolescents (91). This study signifies the effect of cultural attitude toward reduced physical activity observed in Saudi Arabia.

In addition to the environmental factors causing overweight and obesity, there might be other non-environmental factors that render the Saudi population more susceptible to increased BMI. One study reported a higher speed of intestinal glucose absorption in Saudi subjects in comparison with other African individuals (92). Faster absorption of glucose will eventually lead to higher levels of blood sugar and it is more likely that some of this glucose will be stored as fat and that this will increase the risk of obesity and consequently of type 2 diabetes.

2.2.5 Complications of type 2 diabetes in Saudi Arabia

An increase in diabetes will eventually lead to an increased incidence of the complications related to diabetes. One of the commonest complications is diabetic retinopathy. Its prevalence among type 2 diabetes patients reached 25.3% (93). Diabetic retinopathy was reported to be higher in patients with diabetes residing in urban areas. Its risks were also reported to be higher with a longer duration of diabetes and poor control of the disease (94).

In addition to diabetic retinopathy, other complications have been reported as affecting patients with diabetes in the Saudi population. According to the Saudi Stroke Data Bank, diabetes is the second commonest risk factor for cerebrovascular accidents in Saudi Arabia (95). Another study conducted in the Eastern province in Saudi Arabia reported that 37% of stroke cases in the province were affected by diabetes (96). Diabetic nephropathy has been reported to be the most common cause of end stage renal failure in the Al-Madinah province in Saudi Arabia (97). The list of illnesses associated with diabetes in Saudis also includes coronary heart disease (98), peripheral arterial disease (99), neuropathy (100) and erectile dysfunction (101).

Complications of diabetes are more likely to increase when there is poor management and control of the disease. The majority of patients with diabetes in Saudi Arabia do not achieve good glycemic control. Only 27% to 33.8% of patients with diabetes manage to reach a normal glycemic level (102, 103). A study reported that higher age, smoking, and lack of physical activities were significantly associated with poor control of diabetes (102).

2.2.6 Conclusion

- Type 2 diabetes is a major health concern in Saudi Arabia.
- Its prevalence is progressively increasing. This increase has been augmented by a lack of appropriate public knowledge about the risk of the disease and how to prevent it.
- The increase in prevalence of overweight and obesity is an important explanation of the prevalence of elevated type 2 diabetes.
- Increase of type 2 diabetes is associated with the increased prevalence of several complications such as diabetic retinopathy, cerebrovascular accidents and diabetic nephropathy.

2.3 Genetics of type 2 diabetes

2.3.1 Introduction

The genetic impact on illnesses and phenotypes has a wide range of variations. On certain occasions, a change of a single nucleotide can result in a disease such as sickle cell anaemia. Alternatively, genetic determinants may not be enough to cause the disease and environmental components are deemed essential. Type 2 diabetes is an example of a disease where the genetic determinants are not solely responsible for the development of the disease and the augmenting environmental causes are important.

2.3.2 Familial risk

A simple way of attributing the genetic impact to a trait or an illness is to observe its recurrence in families. Higher recurrence of a disease or a trait within family members could be evidence of a common genetic effect. The presence of this effect renders family members more vulnerable to a particular phenotype than other members of the population. Furthermore, observing the pattern of the trait among family members and the fashion of its transmission might illustrate the mode of inheritance (104).

Type 2 diabetes is a disease which tends to appear within members of the same family. Having a member of the family with type 2 diabetes increases the risk of developing the disease (105). The risk of type 2 diabetes is higher among first degree relatives (106). Although this kind of aggregation signifies an inherited risk of the disease, we cannot neglect the importance of shared environment in causing the disease.

Another way of recognizing the genetic impact of a phenotype is by observing the concordance of traits in twins. Monozygotic twins have identical DNA and any difference in their phenotype is assumed to be caused by a difference in their environments. Dizygotic twins only share half of their genes, so any difference between their phenotypes is due to genetic and environmental differences. In the case of type 2 diabetes, concordance rates have been reported to be higher in monozygotic twins than dizygotic twins (107-109). Concordance rates have also been reported to increase with ageing among monozygotic twins, which signifies the importance of ageing in causing type 2 diabetes (108).

2.3.3 Heritability estimates of type 2 diabetes

2.3.3.1 Estimation of heritability

Differences in genetic and environmental factors can help in creating a wide range of phenotype variations. On a population level, the proportion of trait variance accounted for by genetic variance is called heritability. Family and twin studies have been utilised to estimate the heritability of different phenotypes and illnesses. Higher heritability estimates indicate a higher influence of genotypic variations in causing the variation of a particular phenotype. Lower heritability estimates are more likely to signify the role of environmental variations in causing trait variations.

Heritability has been further divided to broad and narrow sense heritability. Broad sense heritability measures phenotype variations that are attributed to genotype variation. Genotype variation has been further divided to variation caused by additive variation, dominance variation or interactive variation. The proportion of phenotype variation that is attributed by additive variation of genotypes is defined as narrow sense heritability.

Trait resemblance between parents and their offspring indicates the narrow sense heritability that is contributed by passing alleles from parents to their offspring. The assumption remains valid if we assume a similar environment shared between parents and their offspring. The regression coefficient of association between mid-parent value of a particular trait and their offspring represents the narrow sense heritability of the trait. However, estimating heritability from full siblings is complicated by the probability of inheriting identical genotypes rather than just identical alleles. The later notion suggests an additional effect of dominance variation on the overall phenotype variation (110).

Monozygotic twins mostly share identical alleles and identical genotypes. High trait correlation in monozygotic twins might indicate strong genetic influence. However, correlations in monozygotic twins might be also influenced by the shared environment. To solve this issue, correlation in monozygotic twins is compared to correlation in dizygotic twins who share half of their alleles and one quarter of their genotypes in addition to the assumption of sharing a similar environment.

Heritability in twin study is estimated by doubling the difference of correlation between monozygotic twins and dizygotic twins (110). For continuous traits, correlations are calculated to estimate heritability. However, for binary traits, such as having a disease or not, twin concordance rate is usually utilised to estimate heritability.

Several heritability estimates have been provided for type 2 diabetes. The heritability of type 2 diabetes ranges between 26% and 73% (12, 111-113). The lower heritability estimates suggest the importance of environmental components. Even when there is genetic propensity to develop type 2 diabetes, environmental factors have to be present for the disease to develop.

2.3.3.2 Heritability of type 2 diabetes risk factors

The risk of developing type 2 diabetes is linked to several continuous traits. These traits include BMI, waist circumference, and blood glucose. Genetic components have been reported to affect the variation of these traits. Consequently, genetic components affecting the variation of these traits are likely to contribute to the variation of incidence pattern of type 2 diabetes in the population.

Heritability of BMI has been investigated by various family and twin studies which have been carried out in different populations. A review by Maes et al reported a 20% to 90% heritability of BMI. They also state that the levels of heritability reported in twin studies are usually higher than those in family studies. That might be due to the higher levels of environmental variation seen among families rather than among twins (114). Another systematic review studied the heritability of BMI among children by reviewing twin and adoption studies. Although it reported similar levels of heritability of BMI, it emphasized the importance of common environmental factors during mid-childhood (115).

A variety of factors seem to have an influence on the heritability of the BMI. Ageing has been reported to increase the heritability of BMI (115-117). Another study reported a relatively higher heritability of BMI among twins with parental type 2 diabetes (118). It has also been suggested that the heritability of BMI has been modified by environmental factors. Physical activity has been reported to decrease the heritability of BMI, indicating an interaction between genetic and environmental factors (119-121). Furthermore, it has been proposed that BMI may be genetically correlated with dieting behaviour (122).

Most family and twin studies have been concerned with estimating heritability of BMI have been conducted in white European populations. Lack or rarity of family and twin studies related to other populations has limited our ability to study the variations in BMI heritability between populations. However, the small number of family and twin studies conducted on the Asian population have revealed similar BMI heritability estimates to those reported in Caucasians (123-127). One study reported a similar heritability estimate of BMI among African individuals living in Jamaica, USA, and Nigeria (128). It seems that having a similar genetic background might have a stronger influence on BMI heritability, despite the fact that the individuals concerned are living in different environments.

The genetic components of waist circumference variation seem to be similar to those of BMI. Table 2.2 shows heritability estimates for waist circumference. In a similar way to what has been seen in heritability of BMI, twin studies reported higher heritability of waist circumference than family studies. Twin studies reported a 56% to 80% heritability of waist circumference while family studies reported a heritability of between 28% and 59%.

Variations in age, gender and ethnicity do not show an obvious effect on waist circumference heritability. However, a study conducted on subjects from the South of India has revealed the lowest estimate of waist circumference heritability. Furthermore, two family studies conducted in the USA on a mixture of black and white individuals revealed two different estimates (129, 130).

Table 2.2: Heritability of Waist Circumference reported in twin and family studies.

Reference	Year	Type	Age	Sample Size	Population	Heritability	Comments
(131)	2000	Family	35-65	953	Amish (USA)	55%	Included diabetics
(129)	2005	Family	38-86	554	White (90%), Black(10%) (USA)	55%	Included diabetics
(130)	2007	Family	15-53	818	White and Black (USA)	44%	Healthy subjects
(132)	2008	Family	33-62	2506	White (Netherland)	40%	
(133)	2009	Family	19-57	524	Indian (South India)	28%	Included diabetics
(111)	2009	Family	34-70	400	Black (Mauritius)	59%	Included diabetics
(127)	2010	Family	15-66	425	Asians(China)	51%	Healthy subjects
(134)	2009	Family	24-74	293	White (Italy)	38%	Included diabetics
(135)	2005	Twin	50-74	302	White (Finland)	68%	Healthy Subjects
(136)	2008	Twin	18-34	768	White (Belgium)	80% (Men) 78%(Women)	Healthy subjects
(137)	2009	Twin	7-11	186	White (UK)	74%	Healthy subjects
(138)	2009	Twin	30-53	1942	Asians (Korea)	59%	Healthy subjects
(119)	2009	Twin	22-27	4343	White (Finland)	56% (Men) 71%(Women)	Healthy subjects
(124)	2010	Twin	13-20	1613	Asians (China)	77%(Men) 78%(Women)	Healthy subjects

Environmental influences on FBG might be stronger than genetic ones. Table 2.3 shows the FBG heritability reported in family and twin studies. This ranges from between 10% and 75% in both twin and family studies. Unlike other traits related to type 2 diabetes, the heritability of FBG appears relatively similar in both study types. There is no clear evidence of any difference in heritability of FBG between ethnicities, genders, and health conditions.

Table 2.3: Heritability of FBG reported in twin and family studies.

Reference	Year	Type	Age	Sample Size	Population	Heritability	Comments
(139)	1990	Family	0-70	2113	Black and White (North America)	33%	Healthy subjects
(131)	2000	Family	35-65	953	Amish (USA)	10%	Included diabetics
(140)	2003	Family	29-57	1379	Black and Hispanic (USA)	28%	Healthy subjects
(141)	2004	Family	41-55	811	White (UK)	72%	Healthy subjects
(142)	2006	Family	48-56	852	White (Netherland)	42%	Healthy subjects
(130)	2007	Family	15-53	818	White and Black (USA)	34%	Healthy subjects
(133)	2009	Family	19-57	524	Indian (South India)	24%	Included diabetics
(111)	2009	Family	34-70	400	Black (Mauritius)	47%	Included diabetics
(143)	2010	Family	12-64	425	White (France)	20%	Healthy subjects
(127)	2010	Family	15-66	425	Asians(China)	47%	Healthy subjects
(144)	1999	Twin	37-51	418	White (Netherland)	50%(Women)	Healthy subjects
(145)	1999	Twin	39-65	302	White (USA)	75%	Healthy subjects
(146)	2003	Twin	18-67	607	White (Denmark)	38% (Men) 12%(Women)	Healthy subjects
(135)	2005	Twin	50-74	302	White (Finland)	45%	Healthy subjects
(147)	2008	Twin	20-45	185	White (Netherland)	66%	Healthy subjects
(136)	2008	Twin	18-34	768	White (Belgium)	59% (Men) 70%(Women)	Healthy subjects

2.3.4 Inheritance and Detecting Genetic Variants

Segregation analysis helps in identifying the mode of inheritance of traits or illnesses. Without requiring measured genotypes, observing the recurrence and aggregation of a particular trait among family members might indicate how the genes are being transmitted (104). If the assumption is that a single major gene affects a certain trait, its inheritance could be recessive or dominant, with different degrees of penetration.

Segregation analysis studies concerned with type 2 diabetes have revealed contradicting results. One study conducted on Pima Indians revealed a major gene effect with different degrees of expression based on environmental circumstances (148). Another two studies have suggested the possibility of dominant inheritance of type 2 diabetes (149, 150). However, the polygenic mode of inheritance has been suggested as the best fit for type 2 diabetes inheritance (149-152).

Several approaches have been utilized to identify genes related to type 2 diabetes. A candidate gene approach is based on previous understanding of the metabolic pathway associated with a protein of interest. An effort is then made to locate the specific gene responsible for the coding of the protein (153, 154). Other approaches do not require prior knowledge of metabolic pathways. The linkage analysis approach utilizes measured phenotypes and investigates the transmission of specific DNA segments between family members. On a population level, association studies are used to identify loci related to a phenotype by comparing cases with unrelated controls (104).

So far, about 42 loci have been shown to be significantly associated with type 2 diabetes (155). Another more recent review revealed that this number has increased to about 60 loci detected in European and Asian populations (156). The large number of genetic variants proves the multiple genes influence on type 2 diabetes. However, these loci only explain about 10% of type 2 diabetes heritability (157).

2.3.5 Candidate Gene approach

Candidate genes studies have been successful in detecting loci associated with type 2 diabetes. Several candidate genes were selected based on prior knowledge about their roles in glucose metabolism, insulin action, and other relevant pathways, such as metabolic rates and energy expenditure (158). More than 200 candidate genes have been investigated (158, 159). However, most of these studies have provided weak associations which lacked reproducibility (160). Nevertheless, few candidate genes were significantly associated with type 2 diabetes or were replicated in other associated studies.

Table 2.4 shows loci associated with the risk of type 2 diabetes found by candidate gene studies, which were confirmed by large-scale genome-wide association studies. These loci are

KCNJ11 (158, 159, 161, 162), PPARG (159), and WFS1 (163). KCNJ11 gene codes can be used for a K+ inwardly-rectifying channel which is associated with insulin secretion in beta cells (160). The PPARG (Peroxisome proliferator-activated receptor Gama) gene encodes a nuclear receptor which is a target for thiazolidinedione compounds used for sensitizing insulin (160). A WFS1 (Wolfram syndrome 1) gene may cause Wolfram Syndrome. Diabetes mellitus is one of the manifestations of this syndrome (164).

Table 2.4: Genetic variants associated with type 2 diabetes reported by candidate genes studies and replicated by genome wide association studies.

Locus	Chr	Marker	Allele (Effect/Other)	Effect: OR (95% CI)
KCNJ11	11	rs5219	T/C	1.15 (1.09–1.21)(165)
PPARG	3	rs1801282	C/G	1.15 (1.10-1.21)(166)
WFS1	4	rs1801214	T/C	1.13 (1.07-1.18)(167)

2.3.6 Linkage Analysis approach

Linkage analysis has been successful in identifying rare genetic loci with a strong effect on illnesses. However, its role in identifying multiple genetic loci which have a modest effect on phenotypes has been limited. Therefore, its role in identifying genetic variants for complex illnesses has been less fruitful (157). Nevertheless, one locus was found by linkage analysis to be associated with an increased risk of type 2 diabetes.

A TCF7L2 (transcription factor 7-like 2) has been reported by a linkage analysis study to be associated with increasing the risk of type 2 diabetes. The study was conducted on Mexican American families and the locus was suggested to be on chromosome 10 (168). The locus was also suggested to be on chromosome 10 by another linkage analysis conducted on the Icelandic population (169). TCF7L2 was later to be replicated by association studies in European, Asian, and African ethnicities (170) and was found to have the strongest effect on the type 2 diabetes risk (160). A meta-analysis reported higher odds of type 2 diabetes when inheriting the risk allele of TCFL2 [OR 1.46 (95% CI: 1.42-1.51)] (170).

2.3.7 Genome-wide Association

Genome-wide association studies (GWAS) are able to scan the whole genome and recruit larger numbers of individuals. Having a larger number of scanned SNPs and the ability to compare unrelated cases with controls has increased the power to detect variants with modest effect on phenotypes. Several novel loci and several previously discovered ones have been reported by different GWAS and replication studies to be associated with type 2 diabetes.

Table 2.5 illustrates loci reported by genome-wide association studies and replicated by other association studies. Although there are other novel loci not reported in this table, these loci in particular were replicated among several European and Asian populations. One study including Chinese subjects reported higher odds of the disease when inheriting loci near HHEX and CDKAL1 (171). However, there was no significant reported difference in the odds ratio among these populations. They are almost similar across populations and their confidence intervals are overlapping. This might indicate the similar role of these loci in terms of influencing the risk of type 2 diabetes across these populations.

Apart from the marker reported in SLC30A8 gene, all other markers listed in table 2.5 are located in non-coding areas. SLC30A8 codes for Zinc transporters in beta cells (172). CDKAL1 (CDK5 regulatory subunit associated protein 1-like) has also been suggested to increase the risk of type 2 diabetes by affecting insulin secretion (173). HHEX (hematopoietically expressed homeobox) gene might be involved in pancreatic cell development (174). CDKN2B (cyclin-dependent kinase inhibitor 2B) gene and IGF2BP2 gene (coding for insulin-like growth factor 2 mRNA binding protein) have no known functional correlation to type 2 diabetes. However, the later has been suggested to affect beta-cell function (175). MTNR1B is a melatonin receptor expressed in pancreatic cells. Melatonin might influence the level of insulin through the circadian rhythm (176).

Table 2.5: Most replicated novel loci related to type 2 diabetes reported by GWASs:

Gene	Marker	Chr.	Allele (Effect/ Other)	Effect: OR (95% CI)	Population
SLC30A8	rs13266634	8	C/T	1.07 [1.00–1.16]	Finnish and Swedish (165)
				1.18 [1.09–1.29]	Finnish (177)
				1.19 [1.06–1.33]	Chinese (178)
				1.19 [1.08–1.31]	Icelandic (178)
				1.51 [1.25–1.81]	US White (178)
HHEX	rs1111875	10	C/T	1.14 [1.06–1.22]	Finnish and Swedish (165)
				1.43 [1.18-1.72]	Korean (179)
				1.10 [1.01–1.19]	Finnish (177)
				1.64 [1.25–2.15]	Chinese (171)
CDKN2B	rs10811661	9	T/C	1.20 [1.12–1.28]	Finnish and Swedish (165)
				1.47 [1.23-1.75]	Korean (179)
				1.20 [1.07–1.36]	Finnish (177)
				1.31 [1.12–1.54]	Chinese (171)
				1.19 [1.11-1.28]	White British (180)
	rs564398	9	C/T	1.13 [1.08–1.19]	White British (180)
IGF2BP2	rs4402960	3	T/G	1.17 [1.11–1.23]	Finnish and Swedish (165)
				1.18 [1.08–1.28]	Finnish (177)
				1.11 [1.05-1.16]	White British (180)
CDKAL1	rs7754840	6	C/G	1.49 [1.27–1.75]	Chinese (171)
				1.44 [1.23-1.68]	Arab Lebanese (181)
				1.77 [1.50-2.10]	Korean (179)
				1.08 [1.03–1.14]	Finnish and Swedish (165)
				1.12 [1.03–1.22]	Finnish (177)
	rs7756992	6	G/C	1.23 [1.10–1.37]	Icelandic (178)
				1.21 [1.10–1.33]	Danish (178)
				1.25 [1.11–1.40]	Chinese (178)
				1.38 [1.17–1.62]	Chinese (171)
				1.35 [0.75- 1.09]	Arab Lebanese (181)
				1.23 [1.15-1.32]	Chinese (182)
	rs10946398	6	A/C	1.47 [1.25–1.73]	Chinese (171)
				1.16 [1.10–1.22]	White British (180)
	rs9465871	6	C/T	1.41 [1.21–1.66]	Chinese (171)
KCNQ1	rs2237895	11	C/T	1.34 [1.26-1.43]	Japanese (183)
				1.25 [1.12-1.39]	Chinese (183)
				1.27 [1.08-1.49]	Korean (183)

				1.32 [1.20-1.45]	Japanese (184)
				1.14 [1.03-1.26]	Singaporean (184)
				1.24 [1.16-1.32]	Danish (184)
	rs231362	11	G/A	1.08 [1.06-1.10]	European (167)
				1.24 [1.08-1.43]	Punjabi (185)
	rs2237892	11	C/T	1.31 [1.10-1.56]	Korean (179)
	rs2237897	11	C/T	1.41 [1.29–1.55]	Japanese (184)
				1.22 [1.09–1.35]	Singaporean (184)
				1.36 [1.16–1.60]	Danish (184)
MTNR1B	rs10830963	11	G/C	1.09 [1.06-1.12]	European (186)
				1.16 [1.03-1.31]	Chinese (187)
	rs2166706	11	A/G	1.21 [1.06-1.38]	Indian Asian (188)

Several studies have reported the association between adiponectin, an adipocyte-derived protein, and a lower risk of Type 2 diabetes. A systematic review revealed that the relative risk of diabetes decreased with every increase in the level of adiponectin (RR: 0.72 [95% CI: 0.67-0.78]). This effect was observed in Europeans, East Asians, Asian Indians, African Americans and Native Americans (189). A genome-wide association study of European ancestry found several loci near gene ADIPOQ (which encodes adiponectine) associated with the level of adiponectine. Alleles near that gene which reduce the level of adiponectine, were associated with an increased risk of type 2 diabetes (190). However, a study conducted in Saudi Arabian subjects failed to replicate the effect of variants near the ADIPOQ gene on the risk of Type 2 diabetes (191).

2.3.8 Shared variants with monogenic forms of diabetes

Genes which are influencing monogenic forms of diabetes might be involved in the causation of type 2 diabetes. GWASs were able to detect variants related to type 2 diabetes near genes which were previously known to cause neonatal diabetes (ND) and maturity-onset diabetes of the young (MODY). Although these forms of diabetes were caused by a single gene, variants detected near these genes are not independently strong enough to cause type 2 diabetes (192) but they might give a clarification on how these variants participate in the physiology of glucose metabolism and insulin action.

Table 2.6 list genes causing neonatal diabetes and maturity-onset diabetes of the young that are associated with increased risk of type 2 diabetes. GCK (glucokinase) gene is responsible for causing neonatal and maturity-onset diabetes of the young type 2 (MODY2). Patients with neonatal diabetes were homozygous for mutation in this gene while patients with MODY2 were heterozygote (193).GCK was later on found to increase the risk of type 2 diabetes (186).

KCNJ11 has been reported to cause neonatal diabetes. However, this gene in particular has different sorts of mutations and different modes of inheritance(194). HNF1α and HNF1β (hepatic necrosis factors 1 α and β) mutations cause other forms of MODY. These genes are structurally related and are both nuclear transcription factors (195).

Table 2.6: Genes causing neonatal and MODY and related to type 2 diabetes:

Gene	Disease	Chr	Inheritance	T2D Risk Marker	Allele (Effect/ Other)	Effect: OR (95% CI)
GCK	ND and MODY 2	7	Autosomal recessive (196)	rs4607517	A/G	1.07 (1.05–1.10) (186)
KCNJ11	ND	11	Autosomal dominant or recessive (194)	rs5219	T/C	1.15 (1.09–1.21) (165)
HNF1α	MODY 3	12	Autosomal dominant (197)	rs7957197	T/A	1.07 (1.05-1.10) (167)
HNF1β	MODY 5	17	Autosomal dominant (195)	rs757210	A/G	1.12 (1.07-1.18) (167)

2.3.9 Genetic variants affecting traits related to type 2 diabetes

A large number of genetic variants have been reported to be associated with continuous traits related to type 2 diabetes. Some of these variants do not have a direct impact on type 2 diabetes risk but may act on intermediate phenotypes such as elevating blood sugar, increasing risk of obesity, or decreasing insulin secretion and sensitivity. However, several variants have been reported to increase the risk of type 2 diabetes and affect continuous traits related to the disease. Table 2.7 illustrates the genetic variants that affect more than one trait related to type 2 diabetes.

Variants reported in table 2.7 were detected in healthy subjects. All these variants seem to be associated with an increased level of FBG. Moreover, they independently increase the risk of type 2 diabetes except for three variants, G6PC2 (glucose-6-phosphatase, catalytic, 2), FADS1 (fatty acid desaturase 1), and GLIS3 (GLIS family zinc finger 3). These variants do not have a direct impact on type 2 diabetes susceptibility but might increase the risk by increasing the level of blood glucose.

Variants near GCKR (glucokinase regulator), ADCY5 (adenylate cyclase 5), and DGKB-TMEM195 (diacylglycerol kinase, beta 90kDa - transmembrane protein 195) have weaker associations with type 2 diabetes risk. However, variants with moderate and weak association might be acting additively and increasing the risk of the disease. A study reported a higher risk type 2 diabetes when inheriting several risk alleles related to GCK, GCKR, G6PC2, AND MTNR1B. The odds ratio of type 2 diabetes increased to 2.05 (1.50-2.80) when inheriting 6-8 risk alleles near these genes (198).

Homeostasis Model Assessment B (HOMA-B) is a method of estimating beta cell function. Most of the variants reported in table 2.7 are associated with decreased beta cell function. In contrast, GCKR has been suggested to increase beta cell function and insulin secretion. In addition, variants near IGF1 (insulin-like growth factor 1) gene has been associated with increased fasting insulin level (0.14 mmol/l increase per risk allele SE: 0.57) (186).

The majority of variants reported in table 2.7 are associated with increased risk of higher level of HbA1(c). Nonetheless, Variants near TCF7L2, and SLC30A8 are suggested to decrease the level of HbA1(c). Another meta-analysis reported a higher HbA1(c) when inheriting risk alleles of TCF7L2 and SLC30A8 but did not reach genome-wide significance (186).

Table 2.7: Genetic variants associated with traits related to type 2 diabetes:

Gene	Marker	Allele (Effect/ other)	FPG (mmol/l) B (SE or 95%CI)	HOMA-B B(SE)	HbA1(c) % B(SE)
TCF7L2	rs7903146	T/C	0.023 (0.004) (186)	–0.020 (0.004)(186)	−0.054 (0.014) (199)
G6PC2	rs560887	C/T	0.075 (0.003) (186) 0.06(0.04, 0.09) (198)	–0.042 (0.004) (186)	0.032 (0.004) (186)
MTNR1B	rs10830963	G/C	0.067 (0.003) (186) 0.056 (200) 0.08 (198)	–0.034 (0.004) (186)	0.024 (0.004) (186)
	rs1387153	T/C	0.058 (0.038-0.078) (200)		0.028 (0.004) (201)
GCK	rs4607517	A/G	0.062 (0.004) (186)	–0.025 (0.005(186)	0.041 (0.005) (186)
	rs1799884	T/C	0.075 (0.049-0.101) (200)		0.038 (0.004) (201)
SLC30A8	rs13266634	C/T	0.027 (0.004) (186)	–0.016 (0.004) (186)	-0.0189 (202)
GCKR	rs780094	C/T	0.029 (0.003) (186)	0.014 (0.003) (186)	
ADCY5	rs11708067	A/G	0.027 (0.003) (186)	–0.023 (0.004) (186)	
DGKB-TMEM195	rs2191349	T/G	0.030 (0.003) (186)	–0.017 (0.003) (186)	
FADS1	rs174550	T/C	0.017 (0.003) (186)	–0.020 (0.003) (186)	
GLIS3	rs7034200	A/C	0.018 (0.003) (186)	–0.020 (0.004) (186)	

Variants near FTO (fat mass and obesity associated) genes are presumed to increase the risk of type 2 diabetes. The odds of having type 2 diabetes increases by 1.15 (95% CI: 1.09-1.22) when inheriting the risk allele of rs8050136 near the FTO gene (166). However, the effect of the FTO gene on type 2 diabetes is more likely to be mediated through its effect on BMI (203). In addition to the effect of FTO on BMI and type 2 diabetes, another study reported an increased risk of higher paediatric BMI when inheriting the risk allele of variant rs 7923837 near the HHEX gene (B: 0.06523 increase in BMI, SE: 0.02865 (204)), which as has already been suggested, may increase the risk of type 2 diabetes.

The influence of BMI on type 2 diabetes susceptibility has more dimensions of complexity. Variant rs8050136 near FTO gene has been revealed to impose more risks of type 2 diabetes in obese individuals (RR: 1.49 (95% CI: 1.34-1.66)) while variant rs7901695 near TCF7L2 has a weaker impact in obese individuals (RR: 1.07 (95% CI: 0.97-1.19)). In contrast, the rs7901695 risk allele has been suggested to have more influence on type 2 diabetes susceptibility in non-obese subjects (RR 1.53 (95% CI: 1.37-1.71)) (205). Two studies have reported the similar stronger influence of rs7903146 near TCF7L2 on type 2 diabetes risks in non-obese individuals in the Tunisian (206) and French (207) populations. Additionally, a variant (rs8090011) near the LAMA1 gene was revealed to be significantly associated with a higher increase in the risk of developing type 2 diabetes in lean individuals compared to obese ones (OR 1.13 [95% CI: 1.09-1.18] in lean individuals compared to OR 1.03 [95% CI: 1.00-1.06] (208).

2.3.10 Conclusion

- Familial recurrence and heritability indicate a genetic influence on type 2 diabetes risk.
- The genetic risk was confirmed by several approaches to detecting genetic variants increasing the risk of type 2 diabetes.
- Linkage analysis revealed the gene TCLF7 as a risk factor for type 2 diabetes, and this was later confirmed to have had the strongest genetic effect yet detected.
- GWASs were the most successful in detecting multiple risk alleles with smaller effects.
- All the variants detected only explain a small part of type 2 diabetes heritability.
- Discovering several genetic variants related to type 2 diabetes indicates a polygenic influence.
- These variants appears to act additively to increase the risk of type 2 diabetes.
- Inheriting more risk alleles increases the risk of developing type 2 diabetes.
- There are other alleles which might act as surrogates. These variants were not shown to influence type 2 diabetes directly but to influence other traits related to type 2 diabetes, such as blood glucose and insulin levels.

2.4 Interpretation of detected type 2 diabetes risk alleles

2.4.1 Introduction

GWASs assume no biological pathways when choosing single nucleotide polymorphisms. However, it is rather important to understand how these variants confer the risk of type 2 diabetes biochemically. Figure 2.1 shows how these genes might affect biochemical mechanisms and eventually lead to an impaired glucose metabolism and reduced insulin secretion and action. It has been suggested that these genes affect the development and functioning of beta cells, and the ability of insulin to activate glucose metabolism in other cells.

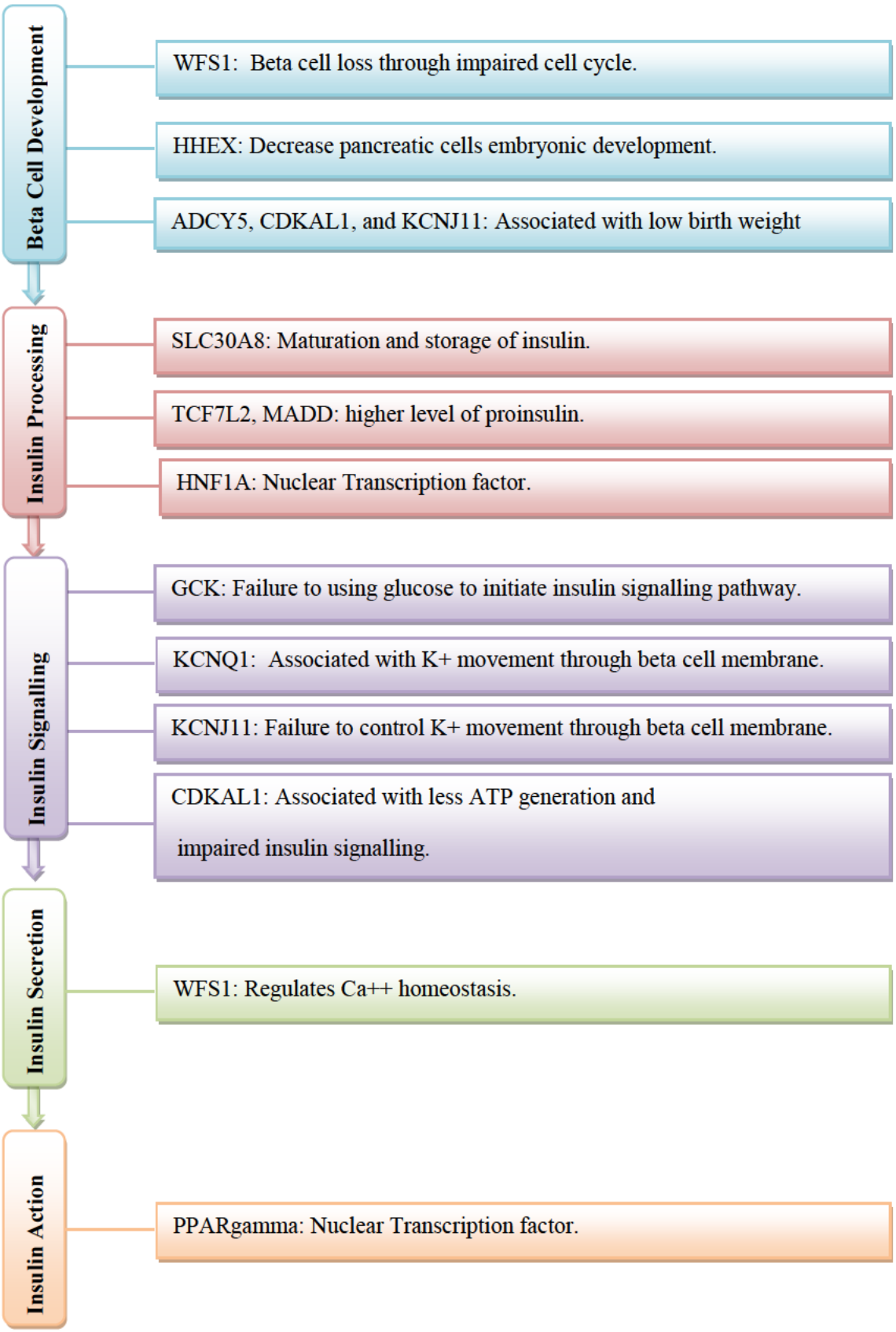


Figure 2.1: Variants influencing type 2 diabetes and related biochemical pathways.

2.4.2 Production and function of insulin

Figure 2.2 shows the process of conversion of insulin to an active form. The production of insulin proceeds through three steps. Firstly, insulin is produced in the endoplasmic reticulum of beta cells in the form of preproinsulin. At this stage, it is composed A, B, C chains and a signal sequence. The signal sequence's function lies in guiding the protein to secretory granules in beta cells cytoplasm. Once the preproinsulin reaches the secretory granules, it loses its signal sequence and becomes proinsulin (209). Elevated blood sugar triggers the cleavage of the C chain of the proinsulin, converting it to active insulin and making it ready for secretion through the blood stream (209-211). Zinc has also been reported to participate in the maturation of insulin in these secretory granules (212).

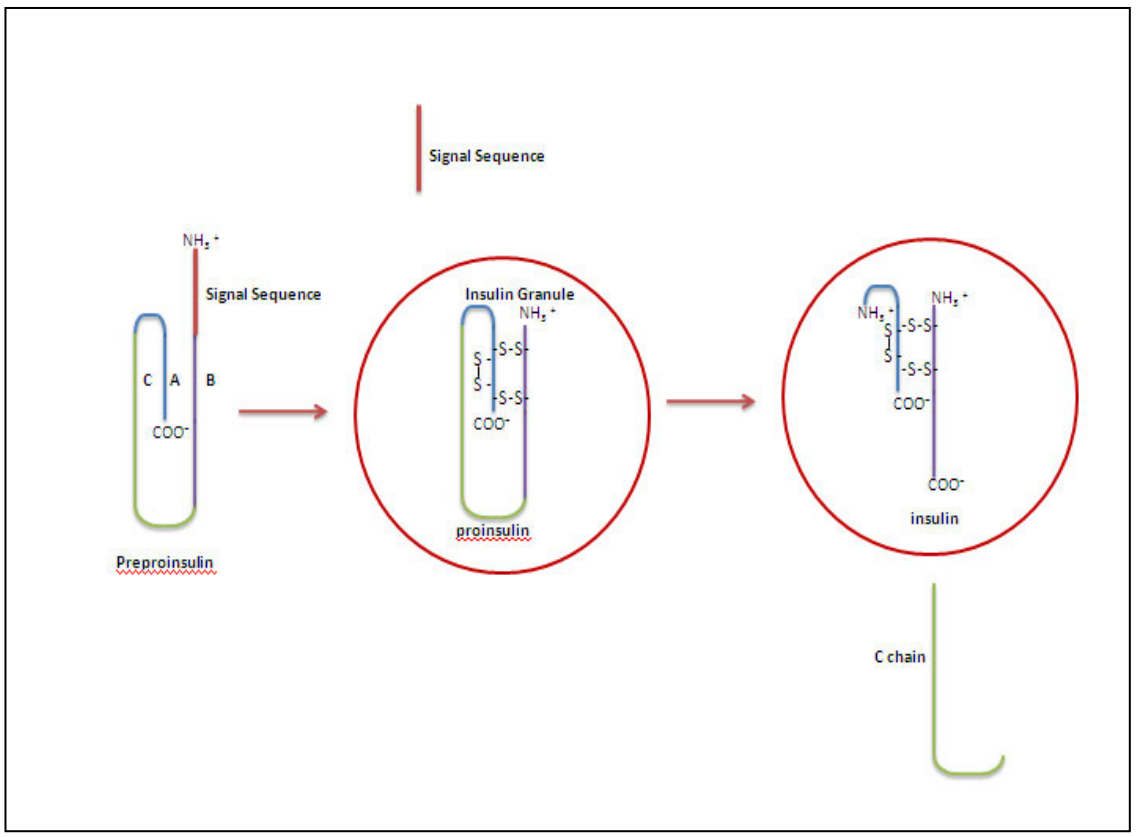


Figure 2.2: Steps in the conversion of insulin to an active form.

The secretion of insulin from beta cells is initiated by the entrance of glucose to the cells. Figure 2.3 shows the process of insulin secretion. GLUT2 are glucose transporters which are predominantly located in a beta cell (213). Similar to other cells in the body, degradation of glucose (glycolysis) is started by Glucokinase (209, 214). The end result of glycolysis is Pyruvate and Adenosine Triphsophate (ATP). ATP forces ATP-gated K+ channel to close leading to depolarization of cell membrane. This depolarization leads to activation of voltage-gated Ca++ channel leading influx of Ca++ to beta cell. In addition to the Ca++ released from the endoplasmic reticulum, the increased level of Ca++ activates the release of active insulin from the granules to blood stream (209).

Figure 2.4 depicts how insulin acts on target cells such as muscle cells. Insulin does not enter cells but attaches itself to an insulin receptor (215). An insulin receptor is composed of 2 alpha subunits and 2 beta subunits (214, 215). Once insulin is attached to the receptor, the beta subunits of the receptor become closer to each other and exhibit autophosphorylation (214). The activated beta subunits activate insulin receptor substrate-1, which in turn initiates a cascade of events ending with the transcription and translation of around 100 genes related to insulin action (215).

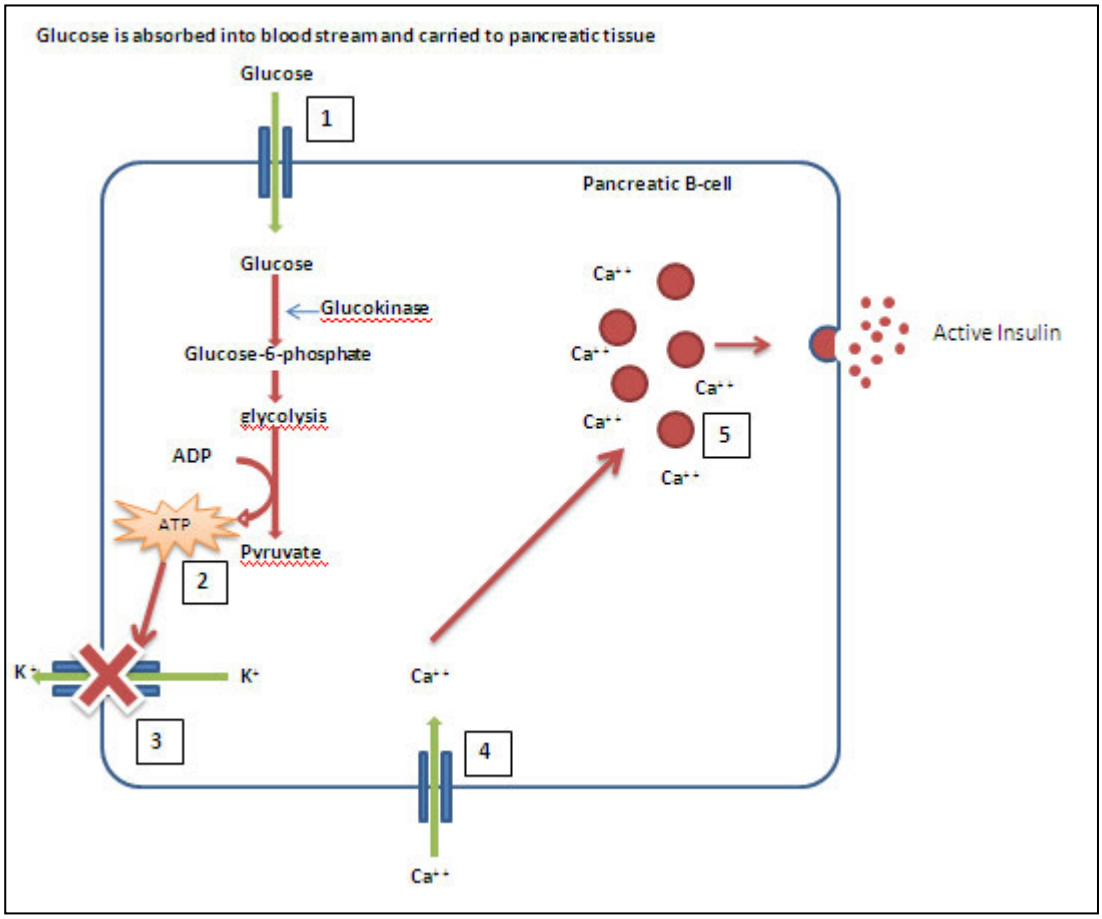


Figure 2.3: Steps of insulin secretion.

One of the main actions of insulin is increasing the influx of glucose into the cells' cytoplasm. Glucose is transported to muscle and fat cells by GLUT 4, which is different to the receptors present in pancreatic cells (213). However, GLUT4 is present in small quantities in the cell membranes of fat and muscle cells and GLUT4 transporters are usually stored inside vesicles in the cytoplasm of these cells. Having the insulin attached to its receptors in these cells will eventually initiate the exocytosis of the GLUT4-containing vesicles on the cell membranes. The increased amount of GLUT4 in the cell membranes increases the transportation and utilization of glucose. Once insulin secretion is reduced, these transporters return and are stored in their vesicles (214, 216).

2.4.3 Genes affecting Beta cells

Certain genes have been reported to decrease the development of beta cells or increase their cell death. It has been noticed that the HHEX gene influences embryonic pancreas development. Mutations in this gene are associated with a failure of pancreatic specification, while the liver is not affected (174). Patients with Wolfram's syndrome were found to have their beta cells selectively absent (164). Furthermore, WFS1 gene deficiency leads to impaired beta cell cycle and increased apoptosis (217). Variants detected near these genes related to type 2 diabetes might indicate some degree of beta cell death or impaired function although not strong enough to solely cause type 2 diabetes.

An association between insulin resistance and fetal development has been suggested. Impaired sensing of glucose and reduced insulin secretion might be associated with reduced birth weight. Genetic causes of this association are seen in monogenic forms of diabetes affecting glucose sensing and insulin secretion and causing low birth weight (218). Several studies have investigated the association between type 2 diabetes risk variants and low birth weight. Variants near ADCY5 (219, 220), KCNJ11 (221), HHEX and CDKAL1 (220, 222) have been revealed to be associated with increased risk of type 2 diabetes and a higher risk of low birth weight (table 2.8). This genetic association between birth weight and type 2 diabetes could be explained by the influence of birth weight on beta cell function in the later period of life (219).

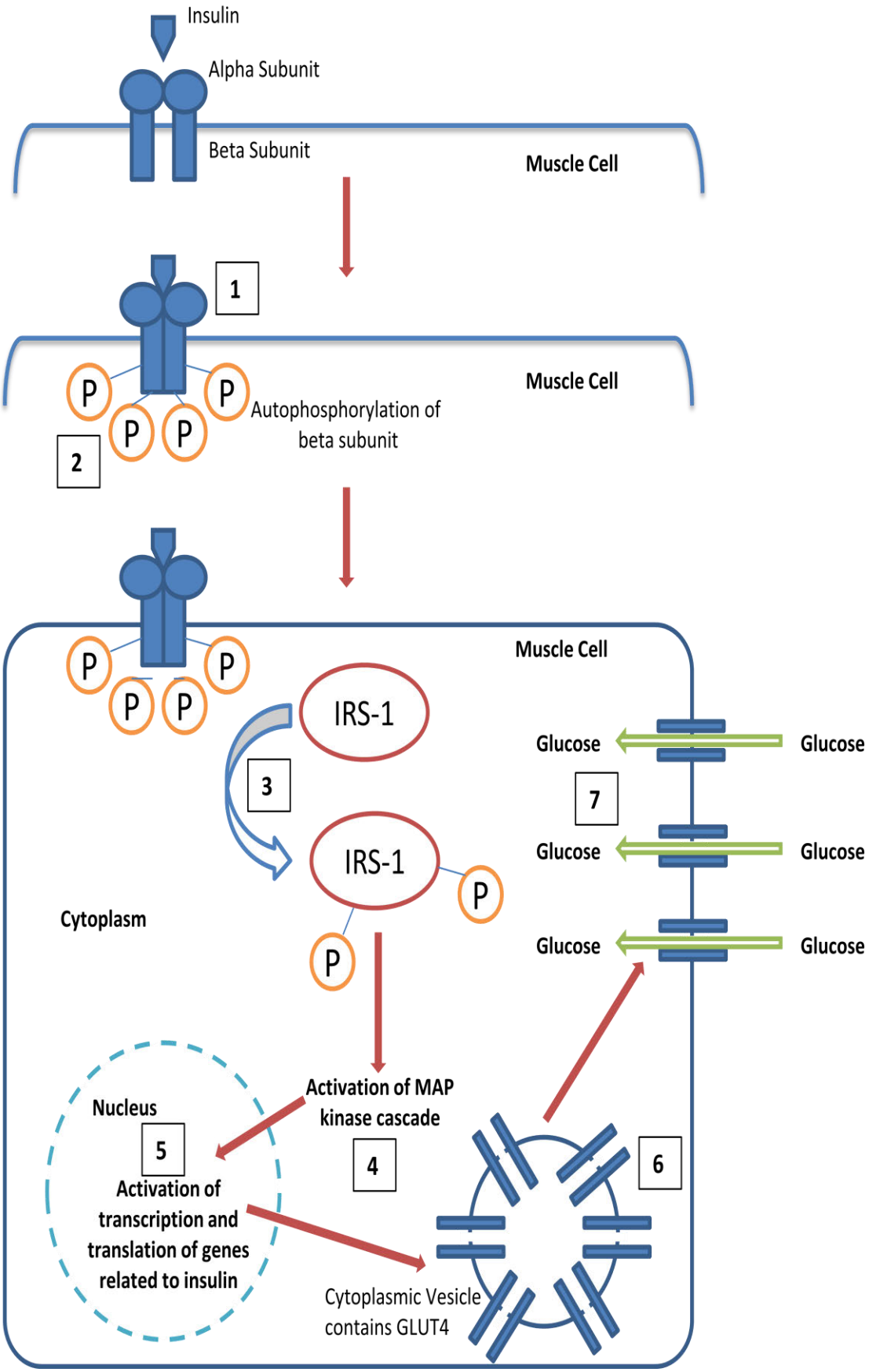


Figure 2.4: Insulin signalling and glucose entrance in muscle cell. IRS-1: Insulin Receptor Substrate-1.

Table 2.8: Type 2 diabetes variants associated with the risk of reducing birth weight:

Locus	Chr	Marker	Allele (Effect/Other)	Effect:
KCNJ11	11	rs5219	T/C	OR and 95% CI : 1.90 (1.04-3.50) (221)*
ADCY5	3	rs11708067	A/G	Reduction per risk allele (g):-33 (95% CI:-10,-55) (220)
CDKAL1	6	rs7756992	G/C	Reduction per risk allele (g):-22 (95% CI:-1,-43) (220)
HHEX	10	rs1111875	C/T	Reduction per risk allele (g):-16 (95% CI:-08,-24) (220)

*The odds of having low birth weight for gestational age when inheriting the risk allele.

2.4.4 Genes affecting insulin processing

SLC30A8 gene codes for ZnT-8 which is a zinc transporter are expressed in pancreatic beta cells (172). These transporters are located in the membrane of secretory granules containing insulin. The increased transportation of zinc into these granules has been suggested to aid in the maturation and storage of insulin in these granules (223). A SLC30A8 gene knockout in mice resulted in the reduction of plasma insulin (224). An association study indicated that type 2 related variant rs13266634 near SLC30A8 gene was associated with increased proinsulin level (Beta: 0.031, SE 0.0067) (225). Having a higher level of proinsulin when inheriting the risk allele of this variant might indicate the significance of zinc transporters in the conversion of proinsulin to insulin which would consequently aid in the pathogenesis of type 2 diabetes.

In addition to SLC30A8, other variants have been suggested to increase the level of proinsulin. Variant rs7903146 near TCF7L2 gene has been revealed to have a higher effect on the proinsulin level than SLC30A8 (Beta: 0.045, SE 0.0066) (225). However, the mechanism in which TCF7L2 increases the level of proinsulin is not evidently apparent. Other variants were reported to increase proinsulin level but not to increase the type 2 diabetes risk. Variants rs7944584 near MADD (MAP-kinase activating death domain) were noted to have a strongest effect on the proinsulin level (Beta: 0.10 SE: 0.0059) (225) but no risk of developing type 2 diabetes (186). On a nuclear level, transcription factor HNF1 alpha has been shown to act on beta cells through regulating the expression of gene coding for insulin and GLUT2, thus affecting insulin production (226).

2.4.5 Genes affecting insulin signalling

We have seen that sensing glucose elevation and its effect on insulin release goes through several steps. Any impairment in any of these steps will eventually lead to impaired insulin signalling and secretion. Hence, glucokinase plays a key role in sensing glucose elevation and the initiation of insulin signalling, and any impairment in its activity will have an impact on subsequent events. In addition to its major role in other monogenic forms of diabetes, variants near the GCK gene have been shown to increase the risk of type 2 diabetes. This risk is more likely to be caused by mutations in the GCK gene, leading to an inability of the enzyme to phosphorylate glucose and initiate insulin signalling. Furthermore, GCKR increases the risk of type 2 diabetes through its regulatory effect on glucokinase (227).

An ATP-sensitive potassium channel can be activated by the energy produced during glycolysis. Abnormality in ATP production will end in failure to control the K+ movement through a beta cell membrane and will eventually reduce insulin secretion. CDKAL1 knockout in mice has been shown to decrease the amount of glucose-induced ATP generated in beta cells (173). However, another study suggested that the variant rs7754840 near CDKAL1 is more likely to affect the genetic expression of another gene called SOX4 (sex determining region Y-box 4) which is located near the CDKAL1 gene. This implies that the SOX4 gene is more likely to inflict the risk of type 2 diabetes than CDKAL1 gene (228).

ATP-sensitive potassium channel is composed of two subunits, namely, the sulphonylurea receptor (SUR) and the K+ inwardly rectifier channel (KIR) (229). These subunits are coded by ABCC8 and KCNJ11 genes, respectively, and are located 4.5Kb of each other within chromosome 11 (158). Abnormality in ATP-sensitive potassium channel is associated with impairment of beta cell membrane depolarization, less Ca++ influx and subsequently decreased insulin secretion. Patients who suffer mutations in the KCNJ11 gene are shown to have an increase in insulin secretion when taking Sulfonylurea treatment and have more effective glycemic control than when taking insulin injections (230). Due to the polygenic nature of type 2 diabetes, Sulfonylurea has a different effect among people with type 2 diabetes. One study suggested that patients with a risk allele of variant rs12255372 near TCF7L2 are more likely to encounter sulfonylurea failure to control blood glucose (1.95 95% CI 1.23-3.06) (231).

KCNQ1 has been suggested to increase the risk of type 2 diabetes by limiting insulin secretion. KCNQ1 gene codes for voltage-gated potassium channels which control the movement of potassium. Functional loss of this gene has been linked to cardiac arrhythmias due to impairment in cardiac current (232). A voltage-gated potassium channel has been reported to increase the influx of K+ in beta cells (233). The increased influx of K+ will decrease the Ca++ influx to beta cells and will subsequently reduce insulin secretion.

2.4.6 Genes affecting insulin secretion

Since calcium is important in the excretion of stored insulin from beta cells, a reduction in the quantity of calcium will lead to decreased excretion of insulin. WFS1 gene codes for a transmembrane protein are located in the endoplasmic reticulum of the beta cells in addition of other cells. This gene has been revealed to be associated with Ca++ homeostasis and the increased filling of endoplasmic reticulum of Ca++ needed for insulin secretion (234). In addition to its effect on Ca++ homeostasis, WFS1 gene deficiency has been recently revealed to influence insulin-containing granule acidification. It has been suggested that the acidification of these granules in mice may be important in the insulin maturation process and lack of this gene resulted in increased proinsulin level (235).

2.4.7 Gene affecting insulin action at target cells

Nuclear transcription factors have been revealed to modulate the action of insulin in target cells. It has been suggested that PPAR gamma can increase the risk of type 2 diabetes by impairing the transcription of genes related to insulin and glucose homeostasis (236). Another study proposed that PPAR gamma inhibits the transcription of GLUT4 and thus decreases the transportation and utilization of glucose in target cells. Rosiglitazone, an anti-diabetic oral agent, was revealed to increase insulin sensitivity by detaching PPARgamma from the promoter area coding for GLUT4 (237).

2.4.8 Conclusion

- Correlating detected type 2 diabetes risk alleles with impairments in insulin and glucose biochemical pathways indicates how complex this disease is.
- Discovered biochemical pathways indicates the polygenic nature type 2 diabetes.
- A better understanding of roles of these genes might help in better diagnostic and therapeutic applications, especially when addressing impairment in multiple biochemical steps.

2.5 Variation in type 2 diabetes across populations

2.5.1 Introduction

Variation in the incidence of diabetes across ethnicities is well documented. White European populations are at a lower risk of developing diabetes than other ethnicities. In the USA, the incidence of diabetes was 19 times higher among the Pima Indians than among the white population in Rochester, Minnesota (238). Similarly, the incidence of diabetes among Indian Asian Immigrants living in the USA was reported to be higher than whites, Hispanics and Africans (239). In the UK, incidence of type 2 diabetes was reported to be up to 6 times higher in South Asian descent compared to white Europeans. The risk of type 2 diabetes was 3 times greater for those of African and African-Caribbean descent living in the UK compared with whites (240).

Pima Indians and Asian Indians are experiencing a relatively similar environment to White European individuals living in the UK or the USA. Despite the similarity in exposure to the western lifestyle in these populations, people from American Indian and Asian Indian ethnicities are at higher risk of developing type 2 diabetes. Having a variation in incidence of type 2 diabetes despite having a relatively similar environment might indicate a significant variation in genotypes.

2.5.2 Variation in allelic frequencies

Having a higher incidence of a certain illness in certain populations might indicate a higher aggregation of genetic risk factors. The effect of genes on influencing the incidence of illnesses depends on the effect size and on the frequency of these genes. In the case of type 2 diabetes, variants related to the disease have weak or moderate effects. However, it appears that certain risk variants are experiencing a significant frequency variation between several ethnicities.

Figure 2.5 depicts risk allele variation among cases of type 2 diabetes in different ethnicities. It is apparent from the figure that populations who are ethnically related have relatively similar allelic frequencies. For example, the risk alleles frequencies of variants rs1111875, rs10811661 and rs2237895 are almost similar among Asian populations. On the other hand, risk alleles frequencies of variants rs1111875, and rs10811661 are higher among European than Asian populations. Similarly, the risk alleles frequencies of variants rs7756992 and rs13266634 are higher among West African than Asians and Europeans. However, other variants near TCF7L2 and MTNR1B did not show wide allelic frequency variation across ethnicities (170, 188).

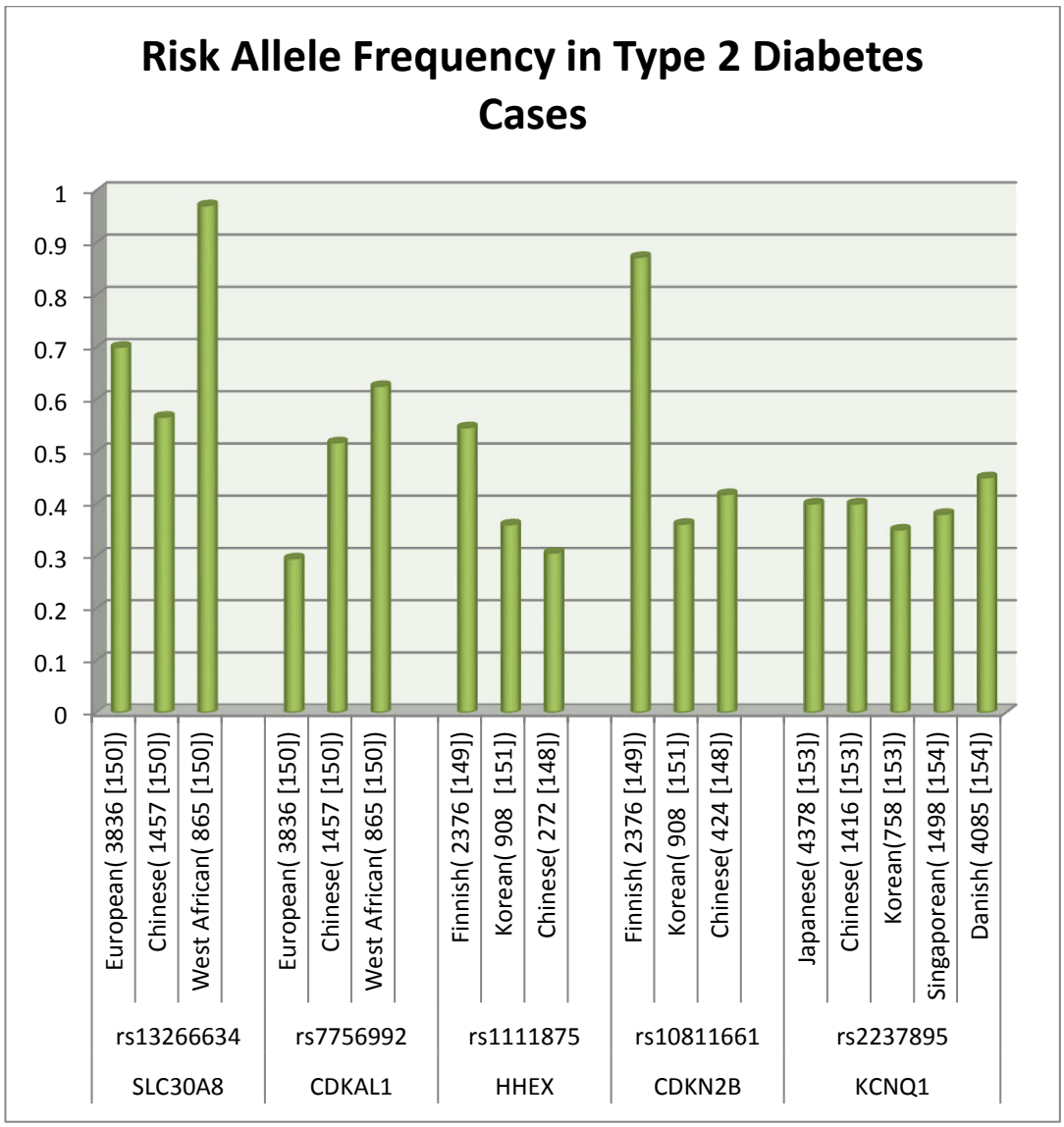


Figure 2.5: Variation of certain risk alleles related to type 2 diabetes across populations from different ethnicities.(Number of type 2 diabetes cases [study reference])

2.5.3 Forces of variation

The Hardy-Weinberg law states that allelic frequencies are constant when genes are passed from one generation to another, as long as mating is done randomly in a large population with no selection, no migration or mutations (110). However, these conditions are hard to maintain and consequently populations do exhibit deviation in their allelic frequencies. Hence, different populations are exposed to different degrees of these forces, whereas allelic frequencies are expected to vary across populations.

It has been suggested that natural selection is one of the forces that may inflict the risk of type 2 diabetes on certain populations. Natural selection influences allelic frequencies based on the ability of individuals with certain genotype to survive in certain environments. When there is a genotype which has been selectively favoured, its frequency increases. If a person is carrying an unfavoured genotype, that person is more likely to have reduced viability and fertility and consequently may be less able to pass the unfavoured genotypes to the next generation (110).

The Thrifty Genotype Hypothesis has been postulated to explain the variation in the incidence of diabetes among ethnicities. This hypothesis claims that certain genotypes were advantageous during periods of starvation and low food availability in the past. Natural selection might have lead to an increased frequency of favoured genotypes, which later became risky genotypes in periods when food was abundant (241).

Although this hypothesis might sound biologically plausible, there is no consistent evidence to support it. However, certain studies have suggested a positive selection for certain type 2 diabetes risk alleles. Variants near the gene TCF7L2 were reported to show evidence of natural selection (242, 243). Another study suggested that a variant near HHEX might have also been subjected to natural selection in East Asians (244). Nonetheless, other type 2 diabetes related loci did not show consistent evidence of selection (243).

In addition to natural selection, non-random mating is another important force that influences the genetic composition of populations. When individuals are more likely to mate due to their physical similarity or because they have a common ancestor, the mating is no longer random and this will eventually affect the allelic frequencies of the subsequent generations. In the case of inbreeding, individuals who are related share similar genes which have been passed down from their common ancestor. The mating of these related individuals will eventually increase the prevalence of homozygosity and reduce heterozygote genotypes. The likelihood of inbreeding will also increase when there is a small isolated population and eventually individuals will be more likely to become genotypically similar (110).

Inbreeding may lead to aggregation of type 2 diabetes risk alleles. Two epidemiological studies have suggested a synergism between type 2 diabetes and consanguineous marriages (245) (246). The increased degree of consanguinity increases the inbreeding coefficient and consequently increases the risk of inheriting a common allele shared by a common ancestor. On certain occasions, the shared allele could be a type 2 diabetes risk allele. However, no study has attempted to investigate the effect of inbreeding on measured type 2 diabetes risk alleles aggregation.

2.5.4 Conclusion

- There is an observed variation in type 2 diabetes prevalence between populations.
- Two major forces have been proposed to increase the risk of type 2 diabetes
- Natural selection may increase the risk by increasing the aggregation of risk alleles which were previously considered advantageous in low food availability periods.
- Inbreeding might increase the risk of type 2 diabetes by increasing the possibility of inheriting homozygote type 2 diabetes risk alleles in inbred individuals.

2.6 Population genetics of Saudi Arabia

2.6.1 Introduction

Saudi Arabia has unique ecological and cultural conditions that make its population different to other populations. Due to these unique conditions, natural selection, migration, and inbreeding are very likely to be the major forces that have created the genetic admixture of Saudi Arabian population. Acknowledging these forces may help us correlate them to illnesses affecting the Saudis, such as type 2 diabetes.

2.6.2 Natural selection

The Kingdom of Saudi Arabia is a relatively new country and was established about eight decades ago. It is dominated by deserts and upon its establishment, it suffered a major lack of resources and was one of the poorest countries in the world. Having this environmental condition might indicates that natural selection could be an important force favouring individuals who were able to survive in this environment.

Over the last 3-4 decades, Saudi Arabia has witnessed an economic surge caused by discovery of the world's largest reserve of oil in eastern parts of the country. This change was associated with an increased availability of food and more exposure of Saudis to western life-style. Subsequently, this change in life-style was associated with the increased prevalence of overweight and obesity which could be partially explained by the thrifty genotype hypothesis.

2.6.3 Migration and genetic flow

Several studies have investigated the diversity of genetic structures in Saudi Arabia. The country is located in a migratory pathway between Europe, Africa and Asia. This location has aided in creating a genetically diverse population. Most of the genetic flow came from the West Asians area. In addition, a significant gene flow was found to be from African Areas (247-249). One study reported that Saudi Arabia might have received a minor genetic flow from central Asia, Indonesia, and even Australia (248).

2.6.4 Inbreeding

Another well studied force affecting the genetic structure on Saudi Arabian populations is inbreeding. The percentage of consanguineous marriages ranged between 54% and 57% and most of these marriages were between first degree cousins (250-252). The prevalence of consanguineous marriages was also found to be higher in rural areas (252).

The prevalence of consanguinity depends on several traditional and social factors. Families in Saudi Arabia are adherent to cultural behaviour and marriage between first degree cousins is almost a norm. It has been argued that marriage between relatives is more likely to be easily mediated, hence mating individuals are brought up together in the same environment. Furthermore, such arrangements may ease spousal adjustment after marriage and enhance the stability of marriages. However, other economic reasons have been suggested, since marriage between relatives is a method of keeping fortunes and properties within the same family or tribe (250).

2.6.5 Conclusion

- Natural selection, migration, and inbreeding are three major forces that may have defined the Saudi Arabia population genetic structure.
- Inbreeding is evidently the most apparent and quantified force.
- Unlike other European or American populations, marriage between first degree cousins is a common practice in Saudi Arabia.
- Approximately, half of marriages in Saudi Arabia occur between related individuals.

3 Chapter Three: Aims and Objectives

3.1 Inbreeding and susceptibility to diseases

If parents share a common ancestor, they have a bigger chance of having similar inherited alleles. Once they pass their similar alleles to their offspring, their children will be at more risk of inheriting 2 identical copies of the same alleles which have been passed from the common ancestor. If the allele inherited from a common ancestor is a risk allele, having 2 identical copies will increase the risk of acquiring monogenic illnesses which can be inherited recessively. In this case, 2 copies of the risk allele are necessary for disease development.

The effect of inbreeding is rather apparent in illnesses where genes have a strong effect in developing diseases, especially at an early age. There are some monogenic forms of diabetes such as mutations in GCK (196) and KCNJ11 (194). These genes have been found to recessively cause neonatal diabetes. One study reported a higher risk of congenital heart disease among consanguineous families in Saudi Arabia (OR 2.12 (1.27-3.57) (251). Another study suggested that neonatal mortality is higher among inbred families, probably because of congenital malformations (253)

In addition to its effect on monogenic illnesses, it has been noticed that inbreeding can increase the risk of low birth weight. Several studies conducted in Arabic (254), Indian (255), Pakistani (256) and Norwegian (257) populations have revealed a significant association between consanguinity and decline in birth weight. After controlling for medical and sociodemographic covariates, consanguinity was shown to reduce birth weight by 1.8% (beta = -0.018, 95% confidence interval: -0.027, -0.008) (254).

3.2 Can inbreeding increase the risk of type 2 diabetes?

A study conducted in Qatar reported higher odds of developing type 2 diabetes among the consanguineous marriages of first degree cousins (OR=1.59, 95% CI=1.11-2.29). Furthermore, the odds of developing type 2 diabetes among consanguineous marriage of second degree cousins was smaller than in the case of marriage between first degree cousins. However, the later odds ratio was not statistically significant, probably because the number of marriages between second degree relatives was less in comparison to mating between first degree cousins (245).

Another study conducted in Saudi Arabia revealed that a positive family history of type 2 diabetes is more likely to be reported among families with consanguinity than families without consanguinity (OR: 13.4, 95% CI=8.64-23.98) (246). However, this study only reported the odds of having familial history of type 2 diabetes and did not actually provide the effect of

having a history of consanguinity on type 2 diabetes disease status. Although these studies appear to show a significant association between consanguinity and susceptibility to type 2 diabetes, another similar study conducted in the European population showed contradicting results.

A study by Rudan et al studied the effect of inbreeding on several late-onset complex illnesses including type 2 diabetes in an European population (258). This study revealed a significant effect of inbreeding on several complex illnesses excluding type 2 diabetes. Although this study utilized a good measurement of inbreeding using a pedigree-based inbreeding coefficient, only 480 individuals were studied to assess the effect of inbreeding on 10 different complex illnesses. The paper does not report the number of type 2 diabetes patients included. However, if this study had a small sample size of type 2 diabetes patients it would have had insufficient statistical power to provide evidence of a lack of association between inbreeding and type 2 diabetes.

Based on the available evidence, a synergism between type 2 diabetes and inbreeding can be suspected. Recognizing how inbreeding increases the risk of type 2 diabetes is an important step in strengthening the effect evidence. However, there is no evidence explaining how inbreeding might lead to an increased risk of type 2 diabetes. The following section will propose several hypotheses explaining the association.

3.3 How could inbreeding increase the risk of type 2 diabetes?

Due to the possibly polygenic nature of type 2 diabetes, inheriting more risk alleles has been seen to increase risk (198). In each locus, one might inherit a single or two copies of risk alleles. Those who inherit two copies of risk alleles at each locus could be at higher risk of developing type 2 diabetes than those who inherit a single risk allele at a locus. Since inbreeding increases the chance of inheriting identical alleles at loci, it can also increase the risk of inheriting two identical copies of particular risk alleles at each locus and thus increasing the risk of type 2 diabetes. This is augmented by the chance of inheriting identical copies of risk alleles at several loci to inflict a possible cumulative recessive risk of the disease.

Another approach to explain the association between inbreeding and type 2 diabetes risk is simply to look at the effect of inbreeding on type 2 diabetes risk factors. Individuals who have suffered a low weight at birth are at a higher risk of developing type 2 diabetes at a later age. However, inbreeding is a risk factor for having a low birth weight. This association between inbreeding and low birth weight can partially explain how inbreeding increases the risk of type 2 diabetes. We can hypothesize that inbreeding increases the risk of low birth weight and that low birth weight in turn increases the risk of having type 2 diabetes at a later age.

Increased BMI, blood glucose, and central obesity are known to be risk factors for type 2 diabetes. Part of the variation of these continuous traits can be explained by the existence of variations in genotypes, which signify an important genetic influence on these phenotypes. Since these traits incur genetic risk factors, we can hypothesize that inbreeding can help in aggregating risk alleles, leading to the elevation of these risk factors, and subsequently increasing the risk of type 2 diabetes.

3.4 Aim and Research Questions

The aim of this study is to examine the possible mechanisms by which inbreeding might increase the risk of type 2 diabetes in a Saudi population.

First question: Does an inbreeding coefficient predict the extent of aggregation of type 2 diabetes risk alleles?

Second question: Does the aggregation of type 2 diabetes risk alleles predict the age at diagnosis of the disease?

Third question: Does the aggregation of type 2 diabetes risk alleles predict the extent of type 2 diabetes risk factors (BMI, waist circumference, fasting blood glucose)?

4 Chapter Four: Methodology

4.1 Introduction

Answering the proposed research questions required different study designs, different sorts of collected data, different methodologies of collection and different analytic methods. Table 4.1 summarises the research methods used in this study with reference to each research question. The process of approaching and recruiting study subjects was similar in each question.

4.2 Study designs

As illustrated in Table 4.1, different designs were utilised to answer the research questions. The study utilised family-based study designs. It included a comparison between type 2 diabetes patients with different degrees of inbreeding and ages of onset of the disease. The study also made comparisons between unaffected adult subjects in order to examine continuous traits related to type 2 diabetes.

Table 4.1: Research method summary:

Research question	Population	Design	Variables	Data collection	Analysis
1st question: Does inbreeding coefficient predict the extent of aggregation of type 2 diabetes risk alleles?	Patients and unaffected siblings.	Comparison between subjects with different degrees of inbreeding.	Inbreeding coefficient and number of loci identical for type 2 diabetes risk alleles.	Pedigrees for inbreeding coefficient. Genotyping for type 2 diabetes risk alleles	Correlation between inbreeding coefficient and number of loci identical for type 2 diabetes risk alleles.
2nd question: Does the aggregation of type 2 diabetes risk alleles predict the age at diagnosis of the disease?	Type 2 diabetes patients	Comparison between type 2 diabetes patients with different age at diagnosis of the disease.	Age diagnosis of type 2 diabetes, number of loci identical for type 2 diabetes risk alleles, food frequency and physical activity	Questionnaires for age at diagnosis, food frequency and physical activity. Genotyping for type 2 diabetes risk alleles	Correlation between age at diagnosis of type 2 diabetes and number of loci identical for type 2 diabetes risk alleles. Data collected using questionnaires were used to adjust for possible confounding caused by environmental variation between patients.
3rd question: Does aggregation of type 2 diabetes risk alleles predict the extent of type 2 diabetes risk factors (BMI, waist circumference, FBG)?	Adult unaffected relatives of recruited type 2 diabetes patients.	Comparison between unaffected adults who will have the same inbreeding coefficients of their affected relatives.	Number of loci identical for type 2 diabetes risk alleles, BMI, waist circumference, FBG, food frequency and physical activity	Genotyping for type 2 diabetes risk alleles. Physical measurement for continuous variables. Questionnaires for food frequency and physical activity	Correlation between number of loci identical for type 2 diabetes risk alleles and BMI, waist circumference and FBG. Data collected using questionnaires were used to explain phenotype variation caused by environmental variation between individuals.

4.3 Ethical Approval

4.3.1 Introduction

Application for ethical approval to collect data was conducted in two steps. The first step was to secure ethical approval from the Directory of Health in Jazan, Saudi Arabia where data collection was performed. The second part was to secure ethical approval from the School of Health and Related Research (ScHARR) Ethics Committee. The process of securing ethical approvals is detailed in the following sections.

4.3.2 Directory of Health in Jazan ethical approval

Ethical approval from Directory of Health in Jazan was secured on 17th of July 2011. The approval was granted by the Training Department, Continuous Medical Investigation and Clinical Supervision Department. To secure the approval from the department, a research proposal in both Arabic and English was submitted to the department accompanied by an official letter from the primary supervisor explaining the purpose of the research. This letter of approval is illustrated in appendix 1.

4.3.3 ScHARR Ethical Approval

4.3.3.1 Ethical concerns raised

One of the concerns raised by Scharr's Ethic Committee was related to the process of identifying and approaching participants. We assured the committee that identification and approaching participants will be strictly performed by healthcare staff and not by study investigators. Additionally, the committee questioned our method of providing participants with information sheets and consent forms. The committee suggested using the postal system to send relevant forms and await the participant's reply to confirm willingness to participate. However, we explained that the postal services in Saudi Arabia are not reliable especially in rural areas and therefore the using postal service was not an appropriate option for contacting potential participants.

The ethic committee provided several other ethical concerns. One of these concerns involved using birth registries to aid in estimating inbreeding coefficients instead of using extended pedigrees. The committee questioned our ability to retrieve sufficient information to generate informative genealogies. However, we provided explanations related to the pragmatic use of extended pedigrees in the target population.

Using constructed pedigrees for estimating the inbreeding degree was the best choice according to the available facilities in Saudi Arabia. Using birth registry records would not have provided sufficient data to have a good estimation of inbreeding. The Saudi Arabian Ministry of Health was established approximately six decades ago and the Directory of Health in Jazan was established two decades ago. Therefore, the availability of informative genealogies in the target population and birth registry records would not be sufficient to estimate the degree of inbreeding.

The committee questioned our ability to retrieve information related to the genealogies of participating families. The committee thought that it might be difficult or impossible to construct extended pedigrees expanding for several generations. However, we assured the committee that by using our interviewing technique, we would be able to retrieve valuable information for our research from constructed pedigrees. The technique is explained in section 4.7.4.2**.**

The committee thought that asking study's participants about the history of consanguinity is likely to be stressful for participants. Asking about consanguinity history and mating between cousins is very unlikely to cause discomfort or psychological harm to participants from the targeted population. Almost 50% of marriages in Saudi Arabia are between cousins and this practice is highly respected for familial, traditional, and economic reasons. Mating between cousins is a lawful action in Saudi Arabia. Therefore, a stigma caused by marriage between cousins does not exist in Saudi Arabia.

The committee expressed concerns about other aspects of data collection. The committee requested the investigator to competently inform participants about possible risks and inconveniences that might be encountered during data collection. For example, inconveniences caused by requesting healthy participants to fast for eight hours before glucose measurement. Additionally, we were advised to provide a clear and easy-to-understand explanation of a blood glucose measurement and what nuisance it might cause to participants such as the pain caused by finger pricking.

In relation to the level of FBG in healthy participants, the committee requested a clear description of events and procedures taken when a healthy participant appears to have a high level of FBG. We clarified this point to the committee and stated that this test is not sufficient to clinically detect the status of the disease. Additionally, we stressed that we would be requesting participants to seek medical advice from their family doctors and not to take clinical advice from the study investigator.

One of the most important concerns shared by the committee was related to collecting buccal swabs for the purpose of DNA analysis. The committee recommended seeking approval from bodies related to licensing activities related to importing and analysing DNA samples in the UK. Ethical approval was only secured when we provided evidence from the relevant authorities stating that our protocol is adherent to the regulations laid down by these authorities.

Other concerns related to the collection of DNA samples involved issues related to the safety of samples. The committee was assured that collected swabs were very unlikely to carry any risk of transmitting infections whether to the investigator or other individuals handling the samples. Additionally, the contamination of samples after collection was prevented by securely saving samples in designated isolating tubes.

The committee requested a clear description of the methods used to store DNA samples. The committee was informed that the principal investigator was responsible for securing the collected buccal swabs. Since there were no specific laboratories in the Jazan region where the investigator could securely store the samples and since the samples did not require laboratory facilities to store them, the principle investigator provided a secure cabinet to store the samples in his office. No freezing facilities were required hence samples could be stored at normal room temperature.

After explaining the storage procedures, we had to ensure that the process of storing samples in the UK is in good order. We explained to the committee that related authorities were contacted where the nature of the research had been clarified. In order to be able to import the samples to the UK, we had to ensure that those samples would not be stored in the UK for a period longer than a week and would not be stored once genetic analysis has been conducted.

Apart from the ethical issues related to collecting, storing and transporting samples, the committee stressed the notion of making approached participants entirely aware of the purpose of collecting their DNA samples. We had to assure the participants that those samples were merely collected to investigate selected risk factors related to type 2 diabetes only. We had to emphasise the thought of not utilising this sample for the investigation of any other illnesses. Additionally, we had to assure that those samples would never be used to investigate the certainty of sibship between siblings of each family.

Security of identifiable data was a concern of the committee. The committee recommended not to use paper-based methods when dealing with data that could lead to the identification of participants. Therefore, we utilised an electronic database which can be encrypted to store information related to addresses and contact details of approached participants. We also used the same database to indicate those who we were not recruited to prevent re-approaching. Finally, we had to ensure that all other data was anonymised and identifiable data was only accessed by the principal investigator. We also stated that all identification data would be deleted at the end of data collection.

ScHARR's ethical approval was granted on 18th of June 2012. The letter of approval is included in Appendix 2. All requirements are detailed in the following sections:

4.3.3.2 Human Tissue Authority Licence

The committee postulated the need for a licence from the Human Tissue Authority (HTA) to enable the researcher to import collected DNA samples to the UK. After contacting the HTA and explaining the principles of the research, the authority informed the investigator that there was no need to obtain a licence, hence the collection of DNA samples was only needed for research purposes only, given that samples were anonymised and consent was taken from participants to use their samples in the research. The HTA response was forwarded as a part of the ethics application and accepted as an official form exempting the study from the licence.

4.3.3.3 University Insurance

The committee advised the applicant to seek University Insurance as a part of the ethics application. The committee explained that the insurance is needed since the investigation involves the collection of DNA samples from participants. The insurance was granted on 19th of April 2012. University insurance is included in Appendix 2.

4.3.3.4 Construction of Participants' Information Sheets

Two information sheets were made. The first one was for type 2 diabetes patients and the second was for healthy participants. Several adjustments were made to the information sheets to make them easier to read and understand for a lay person. The information sheets explained to the potential participants the importance of their participation in the research and that they have the absolute right to reject the participation and withdraw at any stage. The sheets also included detailed descriptions of methodology and what participants are expected to do once they agree to participate. The sheets explained to the potential participants that the results of the research will be anonymised and will not disclose any personal information in case of publishing the results. All information sheets were translated to Arabic. Participants' information sheets are included in Appendix 3.

4.3.3.5 Construction of consent form

A consent form was designed which participants were requested to sign before the commencement of data collection. The consent included items related to the declaration of understanding information provided in the information sheets, understanding their right to withdraw at any stage, and declaring their willingness to give DNA samples. The consent form was translated to Arabic. The consent form is included in Appendix 4.

4.3.3.6 Construction of Interview Script

The script explained the process of interviewing participants. The interviews were conducted in two main stages. The first stage included pedigrees construction. The applicant provided a flow chart to explain the process of asking participants to construct pedigrees. The second part of the interview included interviewing participants to fill food frequency and physical activity questionnaires.

4.4 Study Settings and Recruitment

Recruitment of subjects took place in the Jazan region. It is located in the south west of Saudi Arabia and is composed of several urban and rural areas. Certain PHCs were targeted, based on their having specialised diabetes clinics. The selection of these centres was discussed with the Directorate of Health Care in Jazan upon the commencement of the data collection.

On 14th of July 2012, a visit was made to Directory of Health in the Jazan region. The visit was made to discuss the selection of the most appropriate PHCs. During the visit, we were advised to visit selected PHCs within three sectors, namely Jazan, Sabia and Abuarish. These sectors cover several large PHCs and cover several urban and rural areas. The Directory of Health Care in Jazan brought our attention to the fact that there are no PHCs designated for rural areas and most PHCs are located in urban areas. However, large centres are likely to cover individuals who come from both urban and rural areas nearby.

We received a letter from the Directory of Health in Jazan stating the research mission. The letter was directed to all PHCs covered in the three sectors. This letter was used to illustrate the procedure of the research to the staff of targeted PHCs. In addition to the letter, the principal investigator explained in detail the procedure of the research to the staff working at the diabetes clinics. Illustrations to staff included how to identify the target population and specific statements used to approach patients.

Dr Mousa Altherwee, director of clinical training in the directory, provided a list of the largest and earliest-established PHCs in the three sectors. Recommended PHCs were targeted and the first recruitment of participants was made on 16th of July 2012. Thirteen PHCs were visited during data collection. Map 1 illustrates the distribution of PHCs visited during data collection.

Map 4.1: Distribution of visited PHCs during data collection:

Figure 4.1 shows the research procedure flow chart. Identification of potential participants was conducted by PHC staff where identification of patients was based on diagnosis made by treating doctors. Treating doctors or nurses approached patients during follow-up sessions providing information about the research. The principal investigator was available at the PHCs when health professionals were informing their patients. Those who appeared to be interested were informed by their doctors that they could have a brief meeting with the principal investigator immediately after the follow-up session.

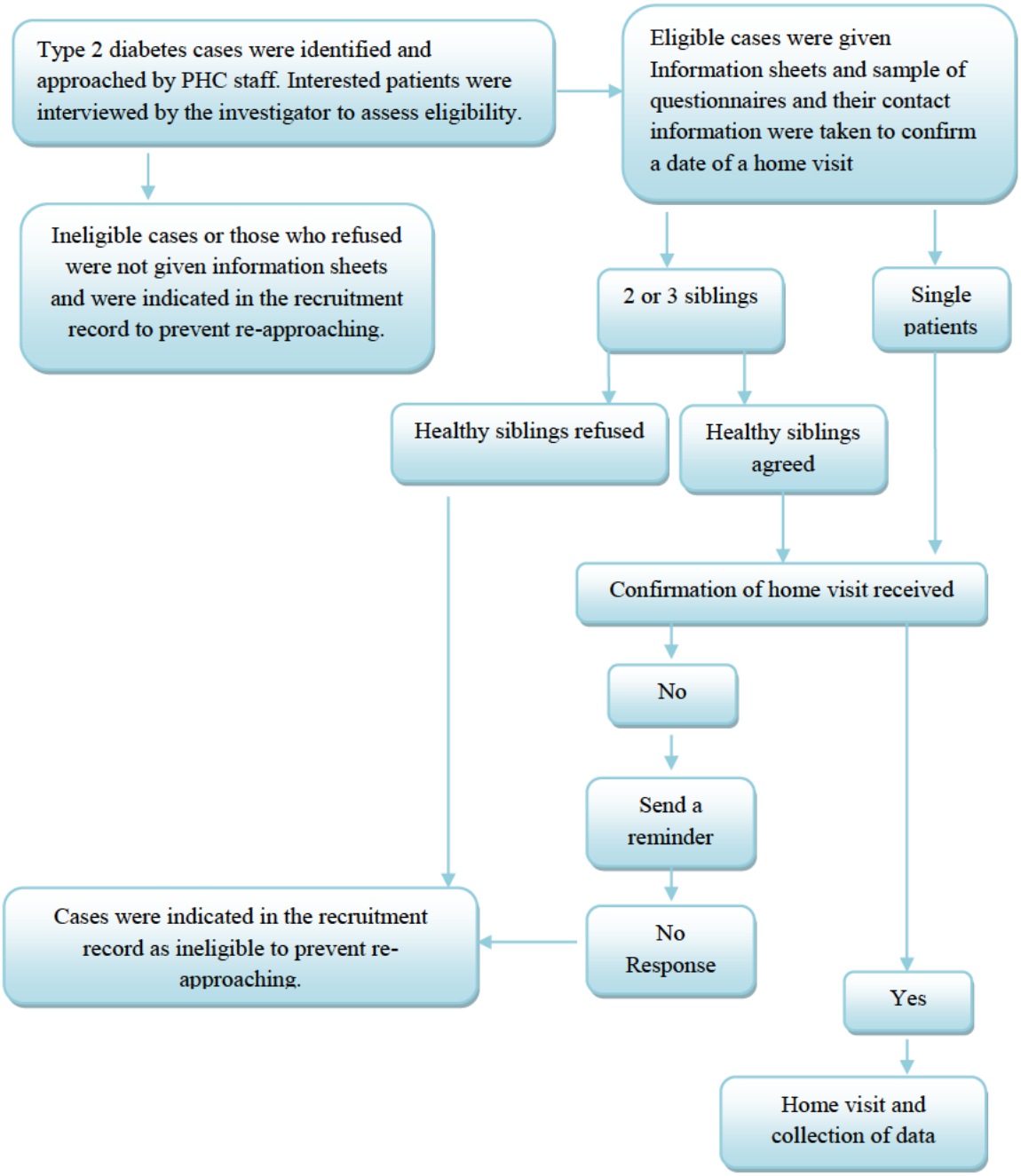


Figure 4.1: Research procedure flow chart.

Once patients agreed to meet the principal investigator, they were asked by the principal investigator to assess the availability of healthy siblings. The second step involved the identification of healthy participants by asking patients to enquire if their healthy adult siblings would be willing to participate. Patients who were not eligible were not recruited. Those who were not eligible during the identification and approaching stage were indicated in the recruitment record to prevent re-approaching of patients.

Approaching the study's participants was performed in four different stages, where each stage involved recruiting 100 participants. An electronic recruitment record was established by the principal investigator to facilitate the process of approaching and recruiting the study's participants. This record involved information about all selected potential participants during the identification stage. The record was in Microsoft Excel 2007 format where identifiable information such as names, address, and contact details were saved. The Excel file was secured by encrypting the file and the principal investigator was the only person who had access to this file. Table 4.2 is the recruitment record used to record the details of approached individuals. Those who refused to participate, or agreed but not eligible were indicated in the other remarks option.

Table 4.2: Recruitment record:

S. No	Date	PHC	Name of patient	Given ID	Names of Healthy Participants	Contact Number	Address and agreed collection date	Other Remarks

During the identification stage of patients, the principal investigator provided eligible patients with information sheets and samples of questionnaires. Healthy siblings received information sheets and questionnaires from their affected siblings. Information sheets included the contact details of the principal investigator. The principal investigator informed the patients that they and their siblings had the freedom to call the principal investigator whenever they wished to confirm their participation. The principal investigator also informed the approached case that they would be reminded by phone about their potential participation after one week of receiving the study information sheets. The principal investigator requested the approached case to provide a contact number, preferably a mobile number, to facilitate sending a reminder.

Potential participants had the freedom to decide the time needed before giving consent. The period before consenting varied between approached individuals. However, for the purpose of completing the recruitment on a timely manner, four weeks was presumed as a waiting period between providing documents to approached individuals and giving notice of approval to recruit. In the majority of cases, approached participants tended to give a decision about participation within two or three days.

Cases or siblings who did not call the investigator to confirm participation were reminded by calls or by sending an SMS text after one week of handing out the sheets. Then we waited for the response for another three weeks. If no response was received from approached participants, or if they rejected participation, they were presumed uninterested and this was noted in the recruitment record. Those who agreed to participate were asked to select a date and address to facilitate a home visit and recruitment. Healthy siblings who contacted the principal investigator and agreed to participate were requested to provide a contact number to aid in reminding them about the fasting needed for the glucose test. Recruitment time and place were stated in the recruitment record.

For the information to be useful at least one case and one healthy sibling must agree to participate and in case any of them refused or were unable to participate, neither of them was recruited. The exception to this condition was made where the patient had at least one sibling affected with type 2 diabetes and there were no healthy siblings available to recruit. In situations where cases or their healthy siblings disagreed or could not participate in the research, their contact information was deleted from the recruitment record. The recruitment record indicated those who refused or could not participate. However, the names of index cases were not removed from the list to prevent re-recruitment and they were permanently removed once the recruitment of participants had ended at each stage.

Due to the cultural background of Saudi Arabian populations, families are large compared to western populations and siblings are more likely to live in the same neighbourhood. Adult siblings, who may have their own nuclear families, are very likely to be neighbours. This common traditional behaviour in Saudi Arabia facilitated the process of approaching potential study participants.

Participating cases and healthy siblings were asked if it was possible to recruit them in one home visit. Otherwise, recruitments were planned according to the convenience of cases and their siblings. In addition, healthy siblings were called or sent a reminding SMS about fasting for eight hours before the measurement of blood glucose a day before the planned home visit. However, asking healthy participants to fast before the data collection appointment seemed to

reduce the participation rate. Therefore, we decided to give the participants the convenience of choosing the most appropriate time to have their FBG measured. In most of the situations, healthy participants tended to choose having their blood glucose measured early in the morning before their breakfast.

On the arranged data collection date, cases and their siblings were requested to sign a consent form. They were informed that they had the choice of not participating even if they had signed the consent form. They were also informed that it was up to them to answer or give some data but not other. Once the consent form was signed, they were ready for data collection. If for any reason, healthy siblings were not able to fast before the measurement of blood glucose, they were asked if it was possible to rearrange the measurement of blood glucose at their convenience.

Recruitment of participants was conducted in four stages. Selected cases and siblings at each PHCs were approached and recruited in each stage. Once the target of number of participants per stage had been achieved, investigators were ready to start the identification, approaching, and recruitment process in the following stage in a similar manner to the previous stage. Recruitment records were permanently deleted by the end of each stage.

Subjects and unaffected family members were interviewed to fill a questionnaire covering information about demographics, family tree, food frequency and physical activity. After the interviews, unaffected individuals had their physical measurements taken. Height and weight were measured to calculate their BMI. We also measured their waist circumference. Additionally, we measured their FBG. The final step was the collection of buccal swabs from all recruited subjects.

4.5 Study Population

Patients were identified based on diagnoses made by general practitioners. The investigator informed PHC staff about excluding type 1 diabetes cases. In situations where an approached case would be suspected to be a type 1 diabetes case, the investigator asked a few questions to clarify this issue, such as asking about age at diagnosis, course of treatment, initial use of insulin and magnitude of blood glucose upon diagnosis. Patients were excluded if they were confirmed type 1 diabetes cases or if answers to any of these questions suggested that the patient approached was likely to be a type 1 diabetes case.

The initial intention was to recruit cases with available healthy siblings. However, a proportion of patients did not have healthy siblings but multiple siblings affected with type 2 diabetes. A criterion for inclusion was changed to include patients with multiple affected siblings with no available healthy siblings. This change of inclusion criterion was made to avoid selection bias, since patients with multiple affected siblings are very likely to have a strong genetic risk of the disease.

Patients were not recruited based on their parental consanguinity. Nonetheless, the frequency of consanguinity was expected to be 50% in the targeted population. Approximately half of the recruited subjects had a history of parental consanguinity. The subjects of this study were restricted to individuals over the age of 18.

The third part of this study examined the effect of inbreeding on continuous traits. Recruiting type 2 diabetes cases to assess these traits might not be appropriate. Cases are likely to be taking medication and experiencing lifestyle changes to manage their disease and these factors will consequently have an effect on their BMI, FBG, and waist circumference. To avoid these confounding factors, we targeted healthy siblings of the recruited cases to answer questions related to continuous traits.

In the beginning of the study, we were aiming to recruit participants from both genders. However, we restricted the study to male participants. Female participants were extremely difficult to recruit due to two reasons.

Upon visiting the first PHC, we realised that it was quite difficult to recruit female participants. We could not work with female staff at PHCs hence PHCs in Saudi Arabia are divided in two sections, for males and females. Male staff are not allowed to enter or work in areas designated for female patients. Therefore, it was not possible to identify or approach female participants. However, we must clarify that this policy is not usually applied in hospitals where working with female staff is more acceptable than PHCs.

The second reason adding difficulties to the recruitment of females was due to the cultural background in Saudi Arabia. Apparently, recruiting male siblings was easier as male siblings, in

most occasions, were living together. However, the situation for females was reversed. Female siblings were less likely to live in the same neighbourhood. Females in Saudi Arabia are very likely to be married younger and once they get married, they move to other regions depending on the original location of their husbands. This is a standard social behaviour in the country. Therefore, female siblings were very unlikely to live together in the same neighbourhood and thus recruitment of female participants was found to be impractical.

Giving the circumstances, recruiting females would have required a larger budget to fulfil two issues. Firstly, we would have required a female investigator to aid in identifying and approaching female patients and participants. Secondly, since married female siblings were unlikely to live in the same neighbourhood, having to commute to collect data would have incurred more cost and time required. For these reasons, female participants were not recruited.

4.6 Random Sampling

Several issues related to this study limited our ability to recruit eligible individuals on a random basis. Firstly, we needed to find Type 2 diabetes patients with available siblings who could be recruited. Secondly, the recruitment pattern varied according to the cooperation of PHC staff. In certain PHCs, the extent of staff cooperation in identifying and approaching patients affected the recruitment rate. Therefore, the mode of sampling was mainly opportunistic sampling based on the number of patients attending PHCs, cooperation of staff and the number of eligible patients.

4.7 Data Collection Tools

4.7.1 Introduction

The study utilised different collection tools at different settings. At PHCs, a recruitment record was utilised to collect data about index cases and information needed to arrange home visits. Upon family visits, pedigrees construction was performed, interviews were made to fill questionnaires, and physical measurements for body weight, height, waist circumference, and FBG were conducted. Finally, buccal swabs were collected for DNA extraction and genotyping.

4.7.2 Initial collection at PHCs

Once index cases agreed to participate in the research, details in the recruitment record were filled. A reference number was utilised to identify families and link siblings to each other once data had been anonymised. For example, families were coded by letters and each individual was given a number. So the first recruited family was coded AA and the first recruited member was numbered 01. So the study reference number of the first recruited case was AA 01.

Single patients were coded in a similar manner. The letter ‘S’ was designated to indicate that the patient recruited had no available healthy siblings to recruit and a number was given based on the order of the patient. For example, the third single patient was given the code S 03.

4.7.3 Data Collection Forms

Two forms were mainly used to record collected data. Both forms were anonymised and did not include any identification data except for participants' codes. The first form is illustrated in Box 4.1 and is a checklist. This checklist was used to ensure the comprehensiveness of collected data. It was also used to indicate measurements which were not possible to be collected at the time forms were being filled. Using that form enabled us to indicate whether the item of interest was successfully collected, the participant rejected collection, or collection of item was rescheduled. Rescheduled time and date of items were stated accordingly. The same form was used for single patients or participants who belonged to the same family.

Box 4.1: Data Collection Checklist:

Family Code:……. Commencement Date:…….

1) Patient Code: ……

Data	Collected	Refused	Rescheduled	Rescheduled Date and Time
Pedigree				
FFQ				
PAQ				
Buccal Swab				

2) Healthy Participant 1 Code: ……

Data	Collected	Refused	Rescheduled	Rescheduled Date and Time	
Blood Sugar					
FFQ					
PAQ					
Buccal Swab					
Weight					
Height					
WC					
BW*					PHC Name: Record Number:

3) Healthy Participant 2 Code: ……

Data	Collected	Refused	Rescheduled	Rescheduled Date and Time	
Blood Sugar					
FFQ					
PAQ					
Buccal Swab					
Weight					
Height					
WC					
BW *					PHC Name: Record Number:

*For BW, state name of registered PHC and record number .

FFQ: Food Frequency Questionnaire, PAQ: Physical Activity Questionnaire, WC: waist circumference, BW: Birth Weight.

The second form is illustrated in Box 4.2. This form was used to collect the data of single patients and participants who belong to the same family. In the patients' section we recorded the age of diagnosis of type 2 diabetes, and data related to food frequency and physical activity. For healthy participants, we added rows for physical measurement such as weight, height, waist circumference, fasting blood glucose.

Box 4.2: Data Collection Form:

Data Collection Sheet

Date :…… Patient's Code:……

Data	Value
Age at diagnosis with diabetes	
Daily Calorific Intake	
Number of Weekly Sedentary hours	
Number of hours spent performing moderate activities	
Number of hours spent performing vigorous activities	

Date :…… Healthy Participant 1 Code:……

Data	Value
Daily Calorific Intake	
Number of Weekly Sedentary hours	
Number of hours spent performing moderate activities	
Number of hours spent performing vigorous activities	
Fasting blood glucose	
Weight	
Height	
Waist Circumference	

Date :…… Healthy Participant 2 Code:……

Data	Value
Daily Calorific Intake	
Number of Weekly Sedentary hours	
Number of hours spent performing moderate activities	
Number of hours spent performing vigorous activities	
Fasting blood glucose	
Weight	
Height	
Waist Circumference	

4.7.4 Pedigrees Construction

4.7.4.1 Introduction

Constructing pedigrees of recruited families enabled us to calculate inbreeding coefficients. It aided in illustrating the mode of mating between parents in each family pedigree. Participants codes were indicated in the constructed pedigrees as illustrated in Figure 4.2. The pedigree drawing tool is depicted in Figure 4.3 and it was freely provided by the National Genetic Education and Development Centre – NHS.

4.7.4.2 Purpose of using pedigrees to calculate inbreeding coefficients

Although using pedigrees is a simple yet effective method to calculate inbreeding coefficients, the limitedness of extending pedigrees to cover several generations would result in the inability to provide the exact estimate of inbreeding coefficients. Lately, another method has been introduced to measure the inbreeding coefficients from genomic data. However, this method requires large number of markers to be measured in order to enable a precise estimate of coefficients. For example, one study selected 1000 markers expanded on several regions in the DNA to measure inbreeding coefficients (259).

Given the large cost of estimating inbreeding coefficients from genomic data and due to the limitedness of our study, we had to resort to using genealogical data to estimate inbreeding coefficients. Additionally, the relation between inbreeding coefficients calculated from genomic data and genealogical data has been reported to be correlated. A study using both methods to calculate inbreeding coefficients in a sheep population reported a reasonable agreement between the two methods ($R2$ of 0.5353, β: 0.6092) (260).

4.7.4.3 Construction Process

It was preferable to arrange a home visit where both the case and healthy siblings were available to improve the pedigree construction process. If interviewing both the case and healthy sibling was not possible on the same occasion, interviews were conducted separately. If one member of the family (whether the case or the healthy sibling) was interviewed and was able to construct a pedigree where all required information was confidently given, there was no need to ask the other member questions related to pedigree construction as they will have the same inbreeding coefficient. In certain occasions, participants unconfident about their family tree were able to ask older family members who were able to give missing information.

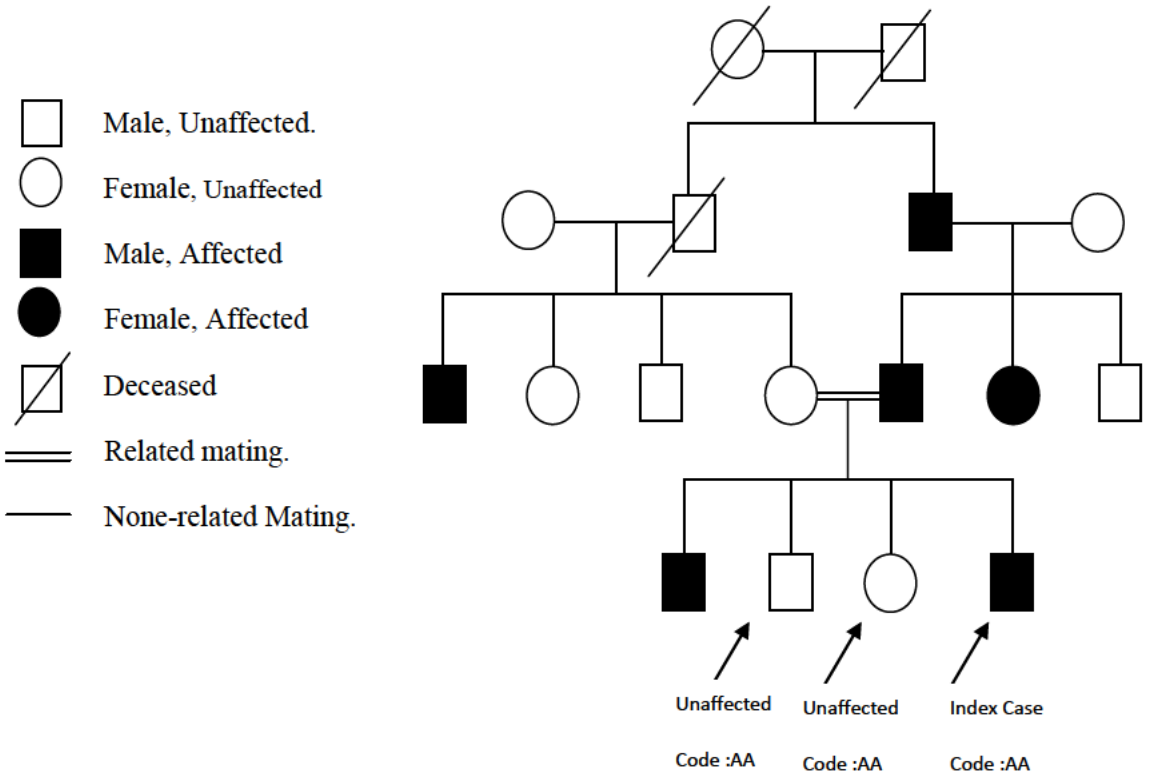


Figure 4.2: Example of a pedigree and explanation of symbols.

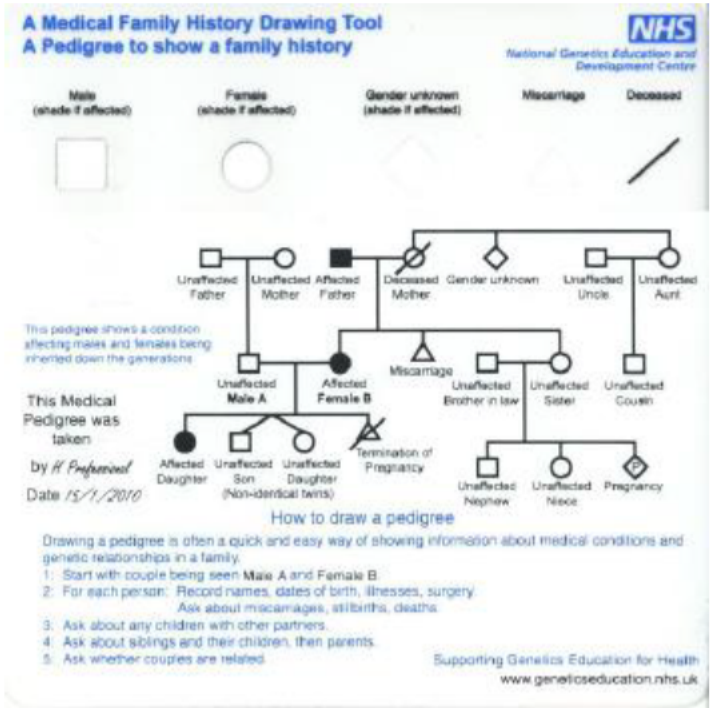


Figure 4.3: Pedigree drawing tool.

Pedigree constructions were performed in two steps. In the first step, the researcher asked the participants about the names of siblings, parents, grandparents or great grandparents and stated their positions in the pedigree accordingly. This step made it easier for participants to remember their family history. The second step included replacing all written names with anonymised symbols and codes. Pedigrees which contained names were given to participants to keep.

Pedigrees were drawn on A4 paper. Anonymised pedigrees did not contain any names or any identifiable data. It only contained codes linking participants to other collected data. The investigator drew family trees using questions depicted in Figure 4.4. These questions were ordered where the investigator drew symbols according to answered questions starting from the bottom of the page.

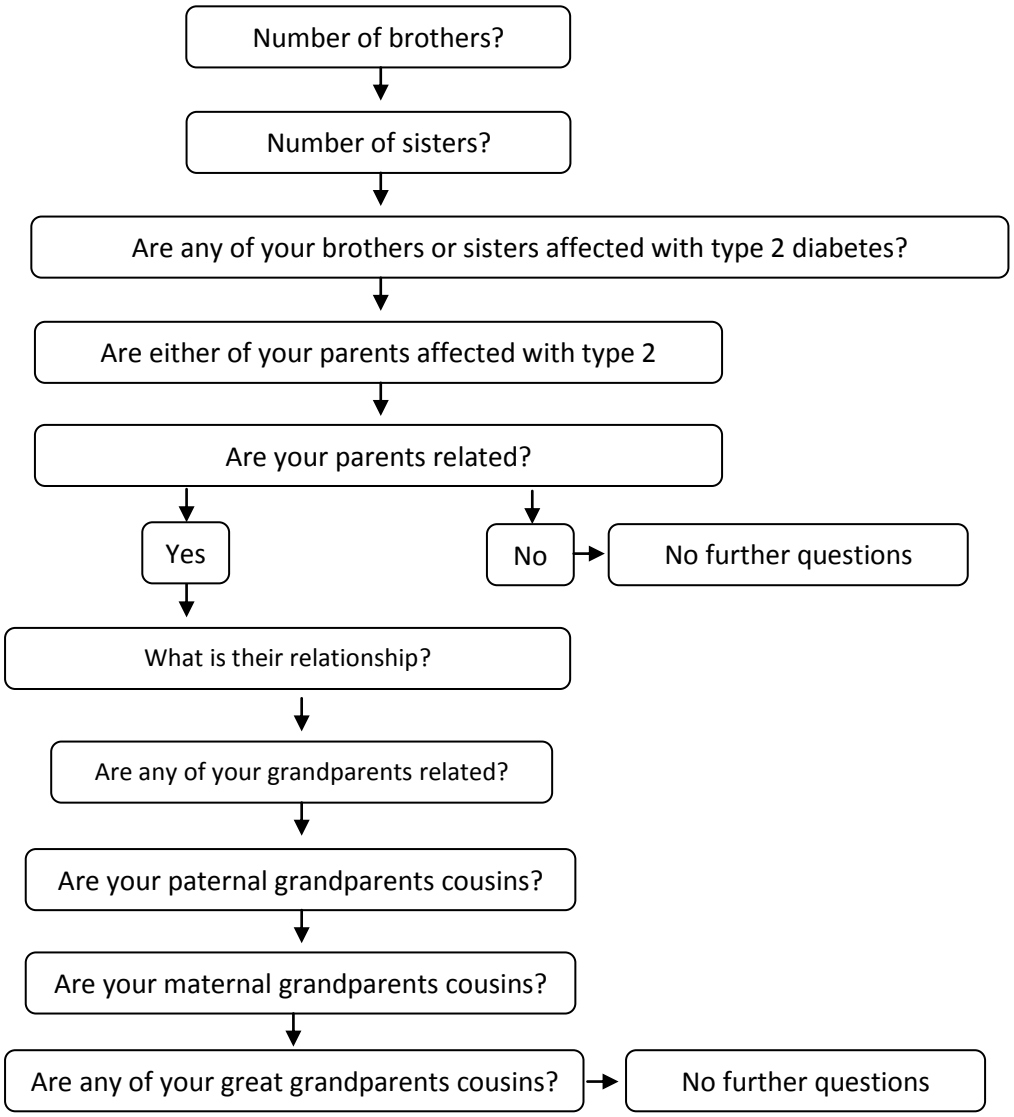


Figure 4.4: Pedigrees construction guidelines.

4.7.4.4 Calculating inbreeding coefficient

Inbreeding coefficients calculations were based on drawn pedigrees. An inbreeding coefficient measures the probability of having two alleles that are identical by descent from a common ancestor. One of the methods of calculating inbreeding coefficients is called the path method (261). This method relies on the number of parents traced from an individual to a shared common ancestor. Inbreeding coefficients using the path method can be calculated by using the following formula:

$$Fx = \sum \left(\frac{1}{2}\right)^{n} (1 + FA)$$

*F*x is inbreeding coefficient of individual.

n is the number of parents connecting between the individual and the common ancestor.

*F*A is the inbreeding coefficient of the common ancestor.

4.7.4.5 Introducing new concepts in pedigree construction

Several mating systems had been observed during the data collection of the first phase. Marriages were, as expected, between non-related, first cousins, second cousins, third cousins, and marriage between an individual and a cousin of his parents. However, we observed two new relatedness values that have to be considered during the calculation of inbreeding coefficients. Therefore, we developed two new concepts when drawing pedigrees to aid in measuring these relatedness degrees. Marriage can be between individuals who are from the same tribe or subtribe. This relatedness degree was described when the participant interviewed did not exactly know who the common ancestor was.

It was relatively easier for those who could not describe the relatedness between their parents as third degree cousins or less to reveal that their parents are from the same tribe or subtribe. Figure 4.5 is an example of a tribe structure observed during data collection. It can be easily observed the sort of variation within the tribe and the degree of relatedness between individuals of interest.

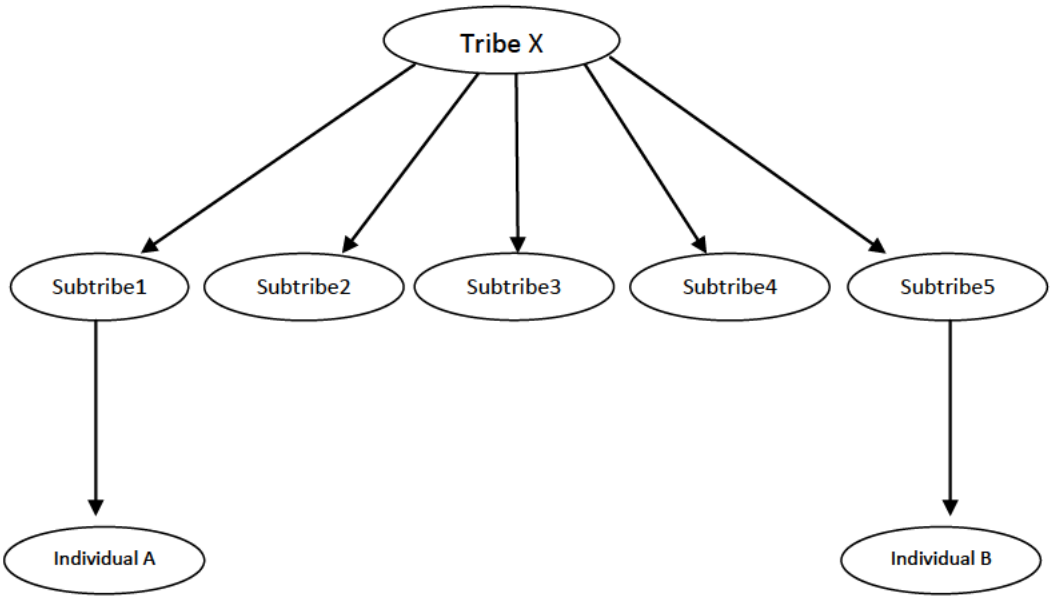


Figure 4.5: Example of mating of individuals from shared tribe.

Here we examine mating between Individual A and Individual B. Individual A belongs to a subtribe 1, and individual B belongs to a subtribe 5 and both subtribes belong to the tribe X. It is not clearly apparent what sort of inbreeding coefficient could be estimated for their offspring. However, since they belong to the same tribe and are distantly related, there should be a small degree of inbreeding which is definitely smaller than those who are siblings of first degree or second degree cousins. Similarly, parents who belong to the same subtribe are likely to have an inbreeding coefficient smaller than those who are first cousins but definitely bigger than those who belong different subtribes but share the same tribe.

Estimating the exact inbreeding coefficients of individuals whom their parents are from the same tribe or subtribe is almost impossible. Therefore, we made an assumption to generate coefficients for those who belong to the same tribe or subtribe. The estimation is explained in Figure 4.6.

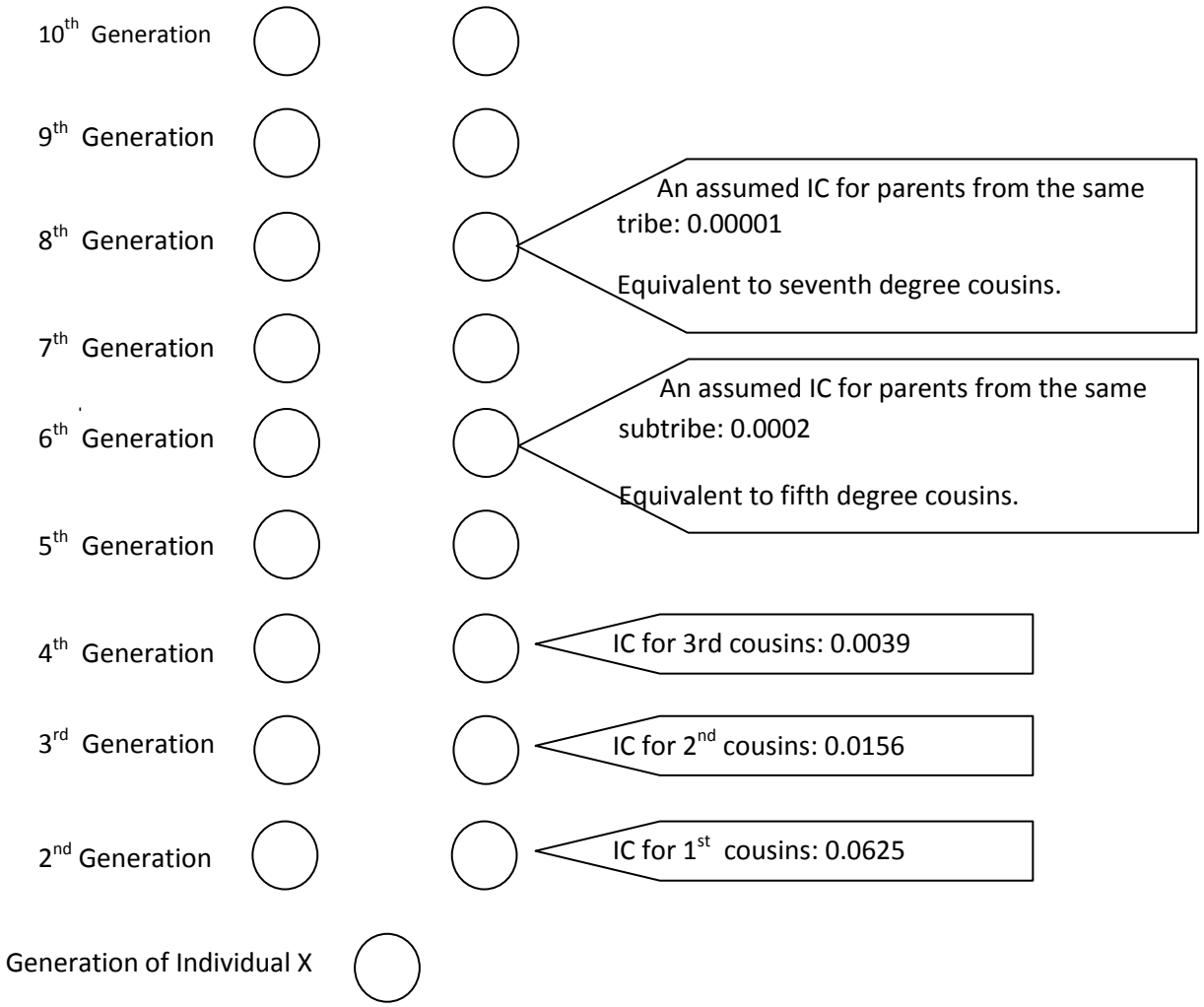


Figure 4.6: Estimation of inbreeding coefficients for mating of individuals with a shard tribe or subtribe. IC: Inbreeding Coefficient.

Figure 4.7 is a pedigree drawn to explain how inbreeding coefficients are estimated when utilising the new concepts introduced. The parents of the individual of interest are second degree cousins. In addition, there is a complicated network of relationships where several great grandparents belong to the same subtribe or tribe. We have calculated 15 pathways related to the shared tribe, one pathway related to the shared subtribe and two pathways related to a degree of second cousins. Calculated the inbreeding coefficients of parents with second degree cousins is 0.0156 but once all other distant pathways are considered the coefficient is increased to 0.0169. This estimation although not precisely correct, takes into consideration other pathways that are difficult to ascertain such as a common ancestor in a subtribe or a tribe.

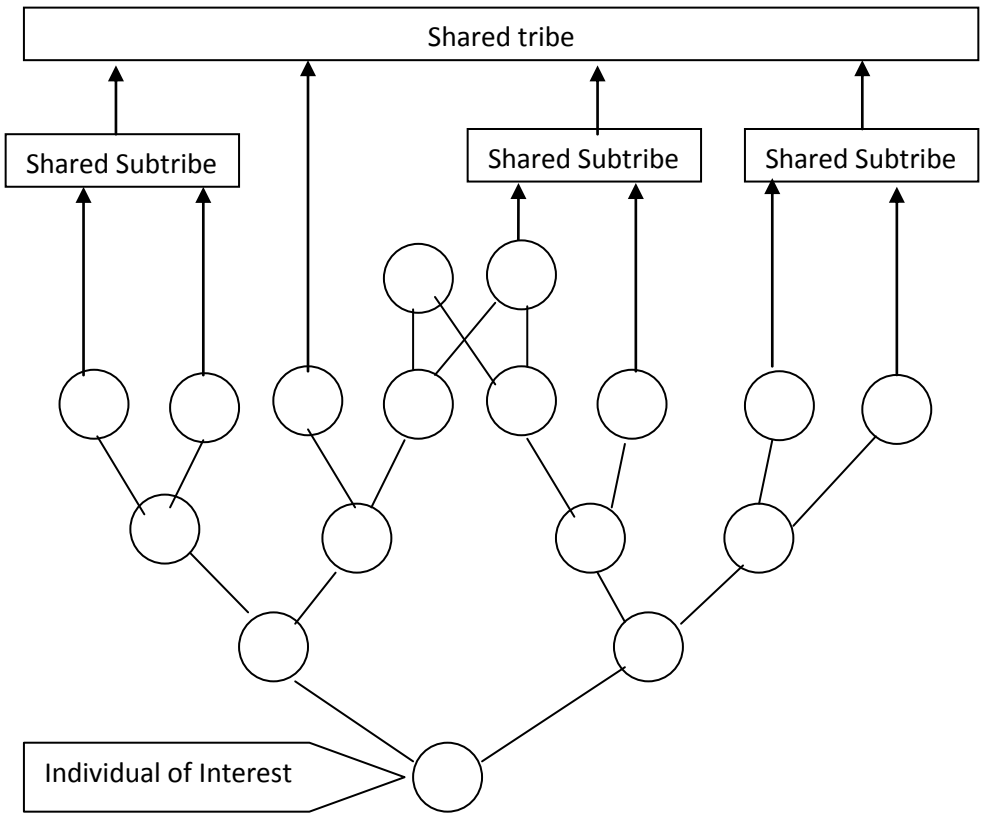


Figure 4.7: Estimating of inbreeding coefficients for a person utilizing the new concepts.

4.8 Questionnaires

4.8.1 The Purpose of Using Questionnaires

Environmental variation explains a large proportion of type 2 diabetes variations. Therefore, measuring the environmental variation between individuals is an important step in explaining the variation in the incidences of type 2 diabetes. Measuring daily calorific intake and level of physical activity may partially explain type 2 diabetes variations. Once the measurable environmental conditions of subjects has been established, we can then have a better understanding of the genetic effect on type 2 diabetes incidences.

In addition to measuring environmental variations between study subjects, questionnaires were utilised to collect the demographic data of the participants. The demographic section included questions about age. For type 2 diabetes cases, a question was added to ask about the age of the onset of the disease.

4.8.2 Measuring food intake

4.8.2.1 Exploration of methods used to measure food intake

4.8.2.1.1 Dietary Recall

Dietary recall is one of the methods used to measure food intake. It depends on the ability of individuals to recall food and beverages consumed in a specific period of time, usually 24 hours (262). However, performing one interview might not be enough to provide an overall of dietary intake of participants. UK Low Income Diet and Nutrition Survey (LIDNS) performed four different 24 hours dietary recalls within a 10 day period (263). Conducting several interviews using this method is needed to avoid variations of food and beverage consumption exhibited between different days of the week.

Interviews conducted to record dietary intake during the previous 24 hours usually go through several steps. Firstly, interviewers would ask participants to recall all the items taken during the previous 24 hours. Once all items consumed have been provided, further details are requested for each food item. Details include exact time of consumption, a description of food item or specific brand name, recipes and method of cooking, quantity and portion of each food item and any combination of consuming several food items in one occasion. A final chronological review of all descriptions given is conducted by the end of the interview (263).

This method has advantages such as a proper record of the quantity and portion size of food items and shorter period of recall given to estimate food intake. However, it has other disadvantages including increased cost and time needed to collect data over several occasions. Therefore, this type of measurement is usually applied to individual cases and less likely to be suitable for large mass investigations (264).

4.8.2.1.2 Food Records

In this method, participants are asked to record their food intake on a daily basis. It does not rely on memory and participants can prospectively record their intake of the day. Participants are given a booklet which contains instructions about recording the type of food, brand name, cooking method and recipes. It also includes a description about portions of consumed food and beverages, and in some situations the weight of consumed food and waste food can be measured and recorded in these booklets (265).

This method does not require trained interviewers. However, participants have to be literate and educated about using the booklets and measuring weights of consumed food. Unlike dietary recall, seven days of recording is usually conducted to avoid bias caused by recalling food intake in only one day of the week (265).

Food records have the advantage of precisely recording consumed items and avoiding recall error. Additionally, weighing consumed items is likely to give a better estimate of intake compared to the recall of portion size. Although this method is less expensive as no interviewers are required, illiterate individuals cannot participate. Additionally, the process of recording food items might vary between different individuals based on motivation levels and literacy levels and thus lead to biased estimates of food intake. Finally, participants might tend to change their eating habits while recording their intake since they might become more conscious about their food intake.

4.8.2.1.3 Food Frequency Questionnaires

In this method, participants are asked how frequently they consume several food items over a fixed time period (usually a year). Several frequencies can be reported varying from never to several times per day. Frequencies are usually fixed and similar in all items listed. A comprehensive food frequency questionnaire usually contains a large number of items commonly consumed in the target population.

This method can be self-administered or completed during interviews. These questionnaires are usually closed-ended but can contain open-ended sections to add food items not listed. Additional to the food items list, methods of cooking or the addition of other nutritional supplements can be integrated to the questionnaire. An average portion size is estimated for each food item.

Although this method represents a comprehensive coverage of frequencies of consuming variety of food items, this method intensively relies on recall over a longer period of time compared to the 24 hours recall. Additionally, the estimation of average portion sizes is difficult and may vary between different individuals. Unlike food diaries and 24 hours recall, a food frequency questionnaire designed for a specific population cannot be utilised in another population and modifications are usually needed. Finally, the cost of using this method might increase if used in populations with high illiteracy rates.

Nonetheless, this method is more suitable for large-scale epidemiological studies compared to 24 hours food recall or food diaries. Using this method in large studies can be less expensive, easier and faster to conduct. Because responses to the questionnaires are usually fixed, analysis of food intake can be easier and less time consuming compared to other methods (266).

4.8.2.1.4 Agreement between food frequency questionnaires and other methods

Several differences between administration and analysis of methods used to measure food intake mandate different analytic outcomes from each method. However, several studies have been conducted to measure the variability of outcomes of these methods. These studies revealed a reasonable agreement between the usages of food frequency questionnaires, 24-hour recall and food diaries.

One study made a comparison between three-day food diaries and semi-quantitative food frequency questionnaires. Three-day food records were conducted four times in a year to detect seasonal variations. The overall average calorific intake calculated from food diaries was similar in both methods. The Pearson correlation coefficient between the energy intake calculated in both methods was 0.42 after adjusting for sex (267). Similarly, another study assessed the association between seven day diary records and food frequency questionnaires and revealed a similar agreement (268). However, one study reported a significant effect of gender on the difference between energy intakes calculated from seven days diary records and food frequency questionnaires. While the energy intake in both methods was similar in males, the energy intake estimated from food frequency questionnaires was higher in female subjects compared to seven days food diaries (269).

A similar agreement has been observed between using food frequency questionnaires and 24-hour recall. A study investigated the association by conducting 12 different sessions to measure 24-hour recall on a monthly basis over a year. Food frequency questionnaires were answered. The Pearson's correlation coefficient between energy intakes estimated in both methods was 0.77 in males and 0.62 in females (270).

4.8.2.1.5 Selecting food intake measurement method

Selecting a convenient and robust method to measure the variation of food intake between participants is a vital issue. However, utilising food diaries and food recodes would have been costly and time consuming if applied to our study design. Due to the reasonable agreement between food frequency questionnaires and other methods, using a food frequency questionnaire was a suitable option. Additionally, the large scale of this study mandated using food frequency questionnaires for easier administration and analysis.

An effort was made to find a comprehensive and easy method to fill in food frequency questionnaires. In addition, we were aiming to find questionnaires relevant to the Saudi population. We found several studies which measured the eating habits of Saudi subjects. We contacted the authors of these studies but unfortunately our efforts were fruitless. Out of three authors, only one replied to our request and the one who replied informed us about the impossibility of using their questionnaire.

Although we could not find food frequency questionnaires related to the Saudi population, we did find several questionnaires related to other populations. We found questionnaires which had been utilised in the USA, UK and China. Comparing these questionnaires clearly indicates a difference in listed food items. This signifies the importance of expecting variations between the food items listed in these questionnaires and the kind of food available in Saudi Arabia.

By comparing found food frequency questionnaires, we found a questionnaire designed by the European Prospective Investigation into Cancer Study (EPIC) Norfolk to be the most comprehensive tool for measuring food frequency (Appendix 6). In addition, its layout makes it easy to understand and fill in. We contacted one of the EPIC team members for permission to use the questionnaire and this was granted. However, we needed to perform several adjustments to make this questionnaire suitable for Saudi subjects.

4.8.3 Measuring physical activity

In addition to finding questionnaires related to food frequency, we found several questionnaires concerned with measuring physical activity. Similarly, we could not find physical activity questionnaires conducted in a Saudi population. However, a physical activity questionnaire designed by EPIC was the best one to utilise in this study due to the same reasons found in the food frequency questionnaire (Appendix 7). Permission to use was granted, although adjustments were necessary to make this questionnaire more relevant to Saudi society.

4.8.4.Adjusting questionnaires

The food frequency questionnaire and physical activity questionnaire produced by the EPIC were mainly used to measure environmental variations in European subjects. In this study, several adjustments were proposed to make these questionnaires more relevant to Saudi subjects. A list of all changes is reported in Appendices 8 and 9.

4.8.5 Questionnaire Translation and Answering

All questionnaires were transformed from paper-based questionnaires to electronic forms. Questionnaires were designed in a similar format to the EPIC questionnaire using Visual Basic 2010. All questions were written in English during the programming and construction of the interface. Instant translation of questions from English to Arabic was made by the investigator during the data collection sessions. All participants were interviewed by the investigator and all data was entered immediately into the programme during data collection.

4.8.6 Food Frequency Analysis Programme

4.8.6.1 Development of Nutrition Database

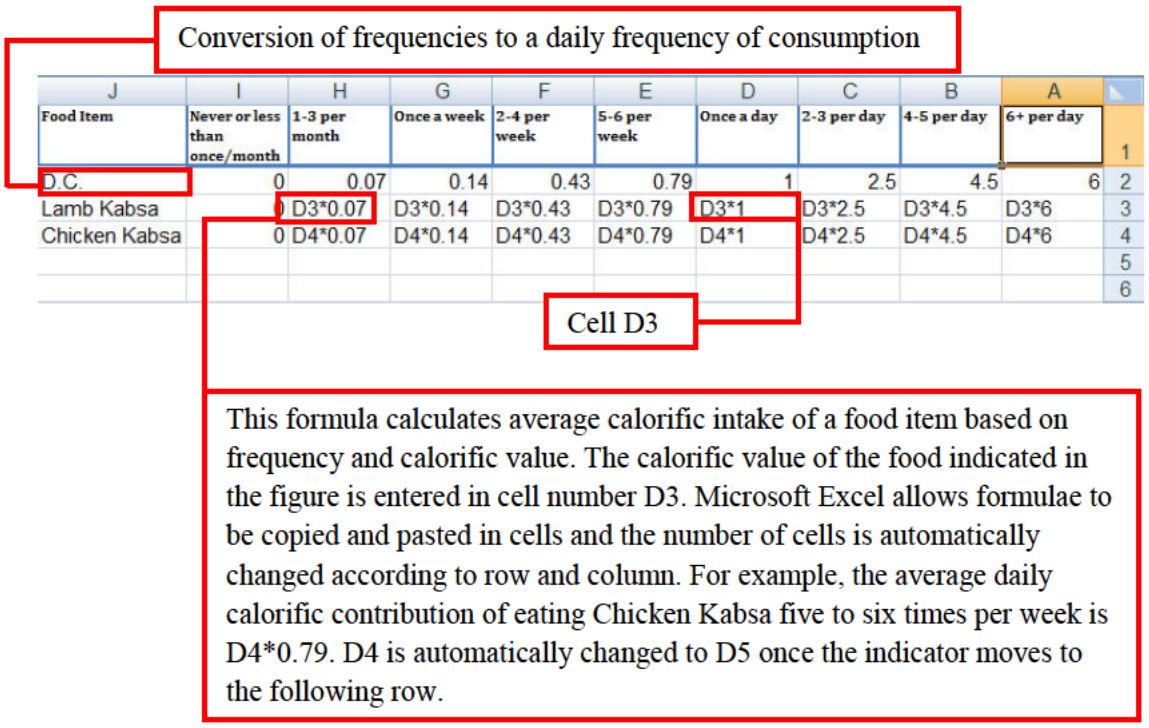
4.8.6.1.1 Purpose of constructing Nutrition Database

The main purpose of constructing this database is to provide estimates of the average calorific value of a variety of food items most frequently consumed in Saudi Arabia. No effort was made to estimate nutrients or minerals available in each food as this goes beyond the interest of this research. Reported calorific values were later fed into the programme to facilitate the calculations of average calorific intakes.

4.8.6.1.2 Format of Nutrition Database

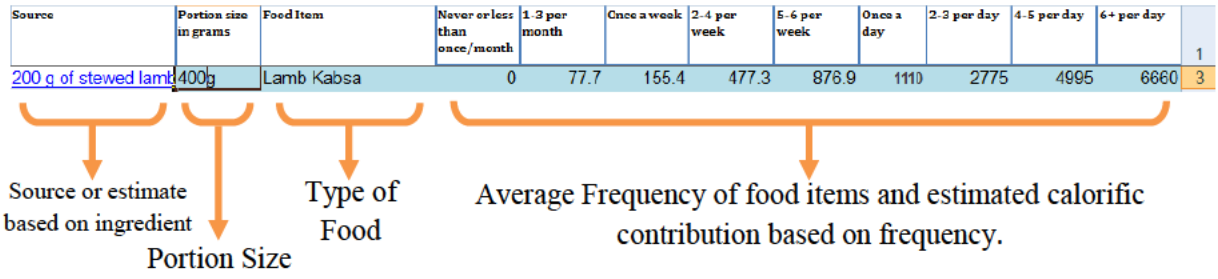
A database was created using Microsoft Excel 2007. Using an Excel database is a convenient choice as it allows easier data entry utilising a spreadsheet. Excel database also facilitates performing calculations which can easily be copied and pasted on multiple rows. Figure 4.8 gives an example of how the calculation of average calorific intake is estimated based on the frequency of food consumption. Figure 4.9 is a copy of the first line of the database. The database contained columns for type of foods, portion size, calorific values based on average frequency, source of calorific estimate.

The daily contribution of each food item was estimated in a similar manner to the way the EPIC food frequency questionnaire was analysed using Compositional Analysis from Frequency Estimates (CAFE) software (271). It was estimated by multiplying the average frequency of each food item by the calorific value of each food item. Those who consume a food item never or less than once in a month were given a value of zero for daily contribution. Other frequencies were converted to daily frequencies. A food item which is consumed from one-three times per month was given a daily frequency of 0.07. This daily frequency was calculated by dividing two (which is a midway point between one and three indicated in the choice) by 30. The resulting number was rounded to 0.07. A food item which is consumed once a week was given a daily contribution of 0.14. This was calculated by dividing one by seven. Remaining figures were estimated in a similar manner.



J	I	H	G	F	E	D	C	B	A	
Food Item	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day	1
D.C.	0	0.07	0.14	0.43	0.79	1	2.5	4.5	6	2
Lamb Kabsa	0	D3*0.07	D3*0.14	D3*0.43	D3*0.79	D3*1	D3*2.5	D3*4.5	D3*6	3
Chicken Kabsa	0	D4*0.07	D4*0.14	D4*0.43	D4*0.79	D4*1	D4*2.5	D4*4.5	D4*6	4
										5
										6

Figure 4.8: Calculation of average calorific intake based on frequency of consumption.



Source	Portion size in grams	Food Item	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day	1
200 g of stewed lamb	400g	Lamb Kabsa	0	77.7	155.4	477.3	876.9	1110	2775	4995	6660	3

Figure 4.9: Copy of the first two rows of the Nutrition Database.

4.8.6.1.3 Estimation of calorific values

The calculation of the average calorific value of each food item was based on the type of food and average portion size. Calorific values were retrieved from two main databases, USDA National Nutrient Database for Standard Reference (USA) and Canadian Nutrient File (CNF). Food items which were not found in these two databases were retrieved from other databases or websites. Certain local food items in Saudi Arabia were not available in any database. The calorific values of these items were estimated using recipes of each kind of food. The calorific values of ingredients were retrieved from the USDA and CNF. The source or ingredients of each food item are indicated in our nutrition database including URL links to certain websites. constructed Nutrition Database is illustrated in appendix 10.

The method of cooking was incorporated in our estimates of calorific value. Rather than asking about the frequency of consuming raw food, we asked about the frequency of consuming food items the way they are served. Using this method, we incorporated all ingredients including usage of oil or butter in preparing food items. The choice of oil or butter was based on the ingredients of each recipe. Methods of cooking were consulted from several recipe websites as indicated in Appendix 10. Certain food items were more likely to be cooked using oils while others were more likely to be prepared using butter. In addition, the methods of cooking food items were also included where different estimates were given for the same food based on whether it was grilled, fried or stewed.

Average estimates of serving size were given according to type. There is no specific database where portion size of Saudi food items could be retrieved. However, for food items available in Saudi Arabia and are similar to items available in the USDA or CNF, serving sizes suggested by these two databases were used as a portion size. For other items unique to the Saudi society, we had to give best estimates based on serving methods.

Portion size varies according to the type of food served. Portion sizes are detailed in the nutrition database. For example, Kabsa, a famous Saudi dish was assumed to be served in a 400g portion because this item is served as a lamb Kabsa or chicken Kabsa which is cooked and served with rice. However, for dishes which are likely to contain meat only, a 200g portion size was assumed. For most pastries, sandwiches, salads, and vegetables portion size varied between 100g to 170g. For fruits, an average medium fruit, slice, or serving was specified for each kind. For sweets, biscuits and cakes, the average serving size varied between 100g to 150g except for snack bars and toffees which were given smaller portion sizes according to the way they are packaged. For dairy products, 100ml portion size was assumed. Other dairy products such as yogurt and cheese were given a portion size between 50g and 100g. A spoon of mayonnaise, jam, or honey was given a portion size of 15ml. Milk added to coffee or tea was given portion size of 15ml. Most drinks were given a value of 100ml except for soft drinks which were given a portion size of a standard can.

4.8.6.2 Construction of programme interface by Visual Basic 2010

An easy-to-operate interface was designed to facilitate interviewing participants and entering their responses into the programme immediately. The programme is composed of five main components: database navigation bar, identification data section, questionnaire, database, and calculation button. These components are indicated in figure 4.10. Visual Basic 2010 was used to create this programme. The style of the interface is similar to the format of the EPIC food frequency questionnaire.

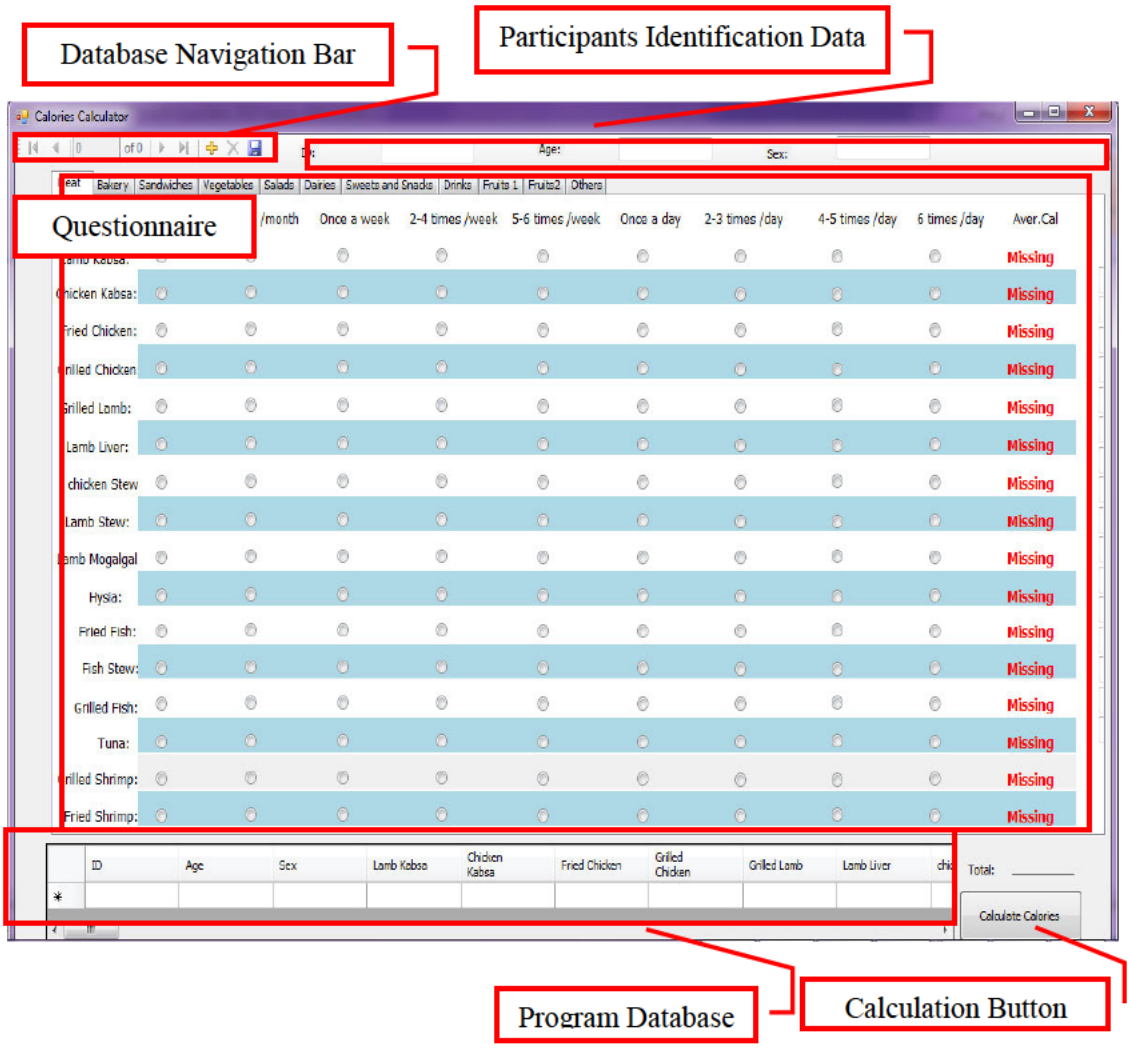


Figure 4.10: Food Frequency Analysis Programme.

The main function of the database navigation bar is to allow the creation of new fields for each interviewed participant. It can also enable navigation between fields. The navigation bar has a saving icon which enables answered questions to be saved in the programme database.

The programme has a feature of identifying participants via entering given codes, age and sex of each participant. Once this information is entered in their designated text boxes, the interviewer was able to save this data alongside their answered questions. Upon designing the programme database, the ID was selected to be a unique value and thus used codes cannot be used again. Saving identification data is performed by pressing the saving button in the navigation bar.

The questionnaire section of the programme was designed in a similar format to the EPIC food frequency questionnaire. Figure 4.11 shows a description of the first four lines of the questionnaire. The first line is composed of 11 tabs where the user is able to navigate between types of food while interviewing participants. The second line states the frequency of food consumption in a similar way to the EPIC food frequency questionnaire. The final column of the questionnaire was added to give an estimate of the calorific contribution of selected food and selected frequency. Different background colours were used to facilitate distinguishing between different food items. The interviewer is able to select the frequency of consuming each food item by selecting a corresponding radio button.

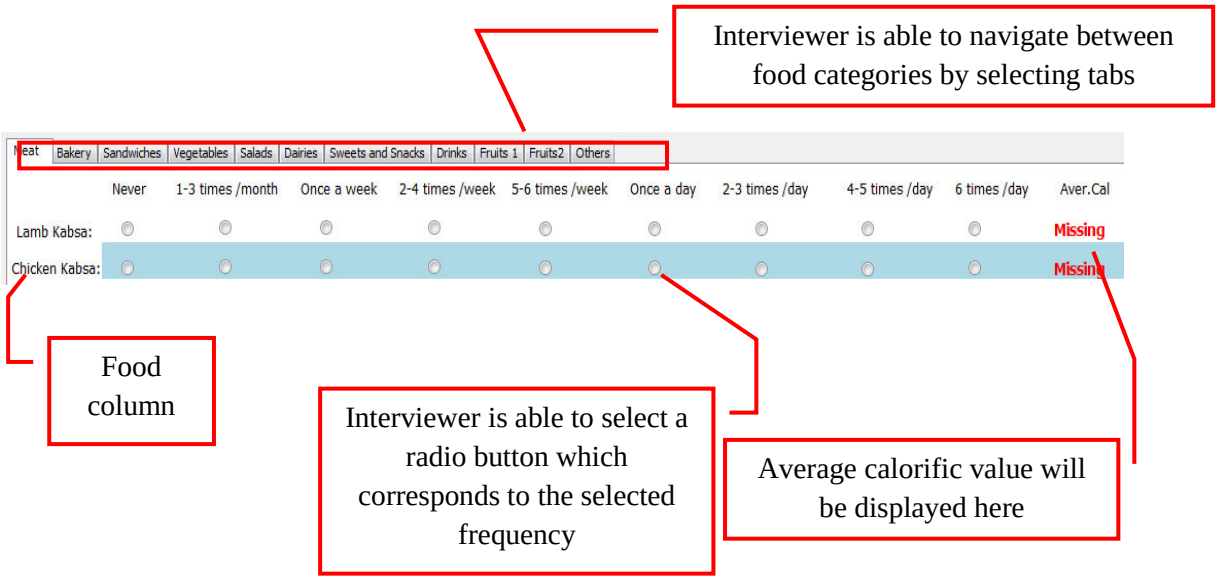


Figure 4.11: Sample of the questionnaire.

Figure 4.12 shows a sample from the program database and calculation button. Once all questions are answered and the calorific contribution of each food item has been calculated, the interviewer will be able to save all entered data into this database. The data of each participant will be saved in a designated row. Each row is linked to the anonymised identity of participants by their unique IDs, age and sex. A database scrolling bar will allow navigation between food items.

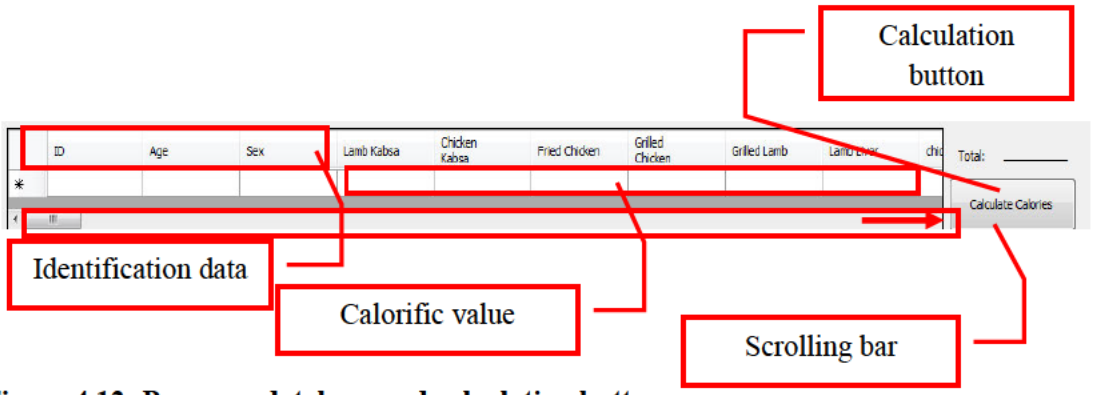


Figure 4.12: Program database and calculation button

4.8.6.3 Programming

The programme was designed to perform three main functions. The first function is to facilitate interviewing subjects about their frequency of consuming various food items. The second function is to perform calculations of average calorific intake. The third function is to save performed calculations.

Coding for calculations was performed through several steps. The first step was to assign a value for each food item. 137 different food items were given numeric values. For example, the first food item was given the value of (num 1) using the following code
"Dim num1 As Single". "Dim" symbolises an identification method used to assign the indicated value (num1 in this example) as a single to allow saving the value as a numeric value.

Figure 4.13 illustrates the second step of coding. This step is initiated by assigning a label for each food item that is displayed on the interface. For example, the label indicating the calorific value of Lamb Kabsa was named "Label11". Labels are usually coded in text format and a code was used to convert labels to numeric values. For example, label11.text was converted to a numeric value by using the code "Val(label11.text)". The converted value was linked to the designated numeric value of the food item by typing "Num1= Val(label11.text.). Using these codes allows numbers displayed in labels to be used in formulae needed to calculate the average calorific value of each food item and the total calorific intake of each participant.

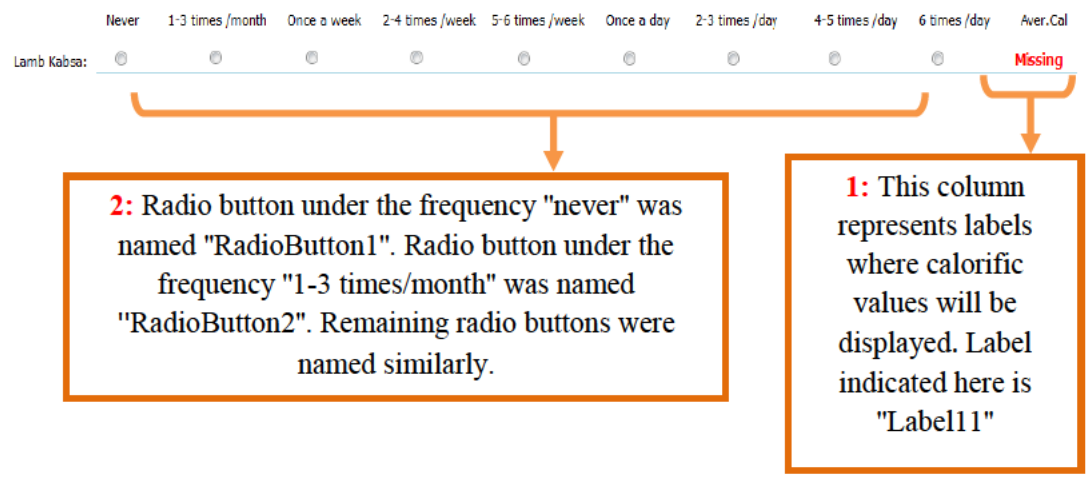


Figure 4.13: Second step of coding.

Each line (as displayed in Figure 4.13) contains nine radio buttons. Each button represents a specific frequency. Each radio button has a name and was assigned a numeric value. These values were assigned by manually feeding values calculated in the nutrition database into these radio buttons. An example of the process of naming radio buttons is explained in Figure 4.13.

Assigning numeric values to each radio button is performed using codes displayed in the following example:

```
If RadioButton631.Checked = True Then
            Label121.Text = 141 * 6
        ElseIf RadioButton632.Checked = True Then
            Label121.Text = 141 * 4.5
        ElseIf RadioButton633.Checked = True Then
            Label121.Text = 141 * 2.5
        ElseIf RadioButton634.Checked = True Then
            Label121.Text = 141
        ElseIf RadioButton635.Checked = True Then
            Label121.Text = 141 * 0.79
        ElseIf RadioButton636.Checked = True Then
            Label121.Text = 141 * 0.43
        ElseIf RadioButton637.Checked = True Then
            Label121.Text = 141 * 0.14
        ElseIf RadioButton638.Checked = True Then
            Label121.Text = 141 * 0.07
        ElseIf RadioButton639.Checked = True Then
            Label121.Text = 0
        End If
```

Using these codes allows the programme to display calorific values based on selected frequency. For example, if "RadioButton634" was checked, the programme will display the value "141" in label121. Radio button 634 corresponds to the calorific value of one fried egg consumed once per day. Other radio buttons are given values in a similar manner.

All labels created in this programme and assigned a numeric value will display the average daily calorific contribution of each food item. As all labels were converted to numeric values, it is possible to apply a formula to calculate the average daily calorific intake of each participant. The formula was created by using the following code

total.Text = num1 + num2 + num3… +num137

This code combines all numeric values of all selected food items. Calculating total daily calorific intake is the final step of calculations performed by the programme.

The final step of coding of the programme is codes used to save calculated values. The daily calorific contributions of food items are displayed in designated labels. Labels were then linked to the programme database using a code shown in the following example:

The code "Lamb_KabsaTextBox.Text" represents the assigned cell in the programme database. Pressing the saving icon will send all calorific values shown in the labels to their corresponding cells in the programme database.

4.8.6.4 How to operate the programme

Steps of using the programme are displayed in Figure 4.14. The steps are in order from first to last. The process is ended by pressing the saving icon which will save all entered data and calculated values in the database.

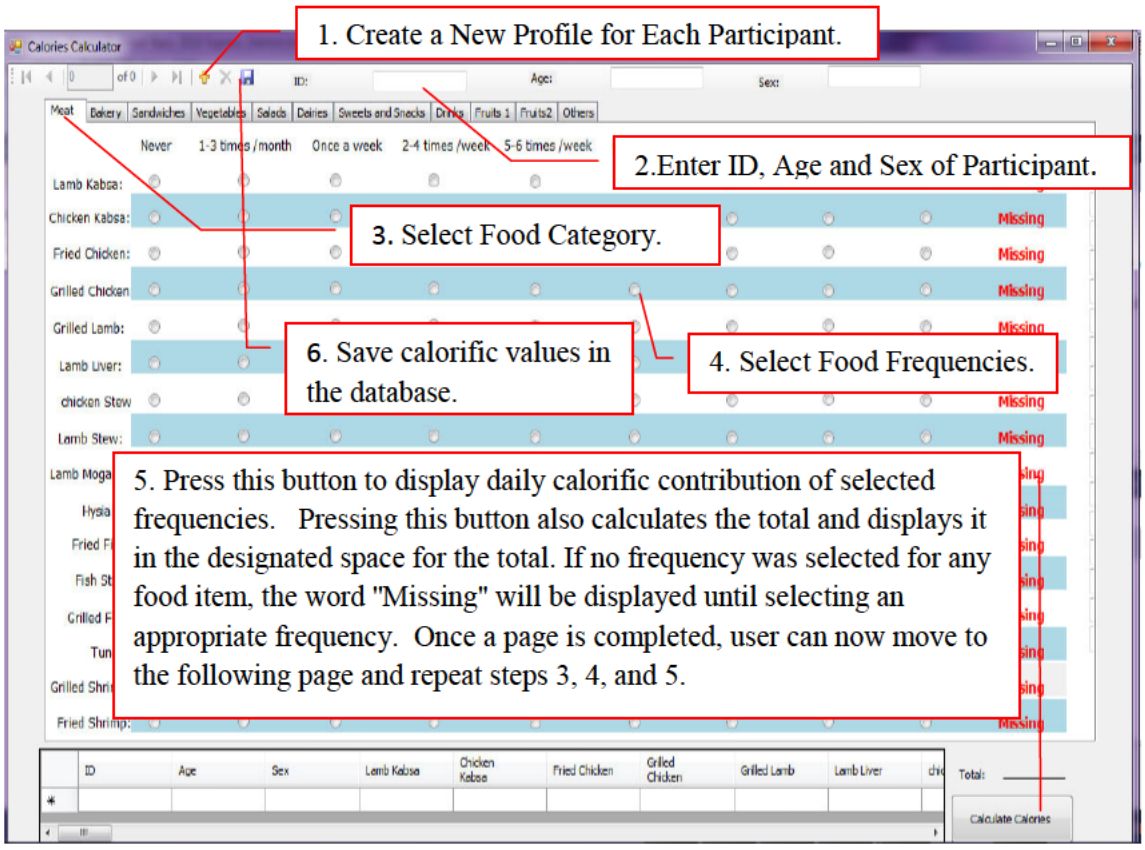


Figure 4.14: Food Frequency Programme Manual.

4.8.6.5 Adding food items not included in the programme

We added another feature to the programme to allow adding food items not coded in the programme. Figure 4.15 illustrates the format of this feature. This feature was integrated in the last page of the software and it simply allows the interviewer to directly add food items and add given frequencies. This new information is saved in the software database once the operator presses the button save. The calorific values of the newly-added items are estimated in a similar manner used to construct the nutrition database.

Figure 4.15: Feature used to add new items and frequencies.

4.8.7 Physical Activity Analysis Programme

4.8.7.1 Purpose and analysis:

The physical activity analysis program was designed to facilitate interviewing participants, feeding their responses into the programme to facilitate data entry and perform data analysis. The programme is designed to look similar to the EPIC physical activity questionnaire. However, the analytic approach intended for this questionnaire mandated changes to certain categories. Once all questions have been answered, the programme will be able to calculate the hours per week spent in sedentary activities or hours spent performing moderate or vigorous activities.

4.8.7.2 Constructing interface by Visual Basic 2010

Visual Basic 2010 was used to create an easy-to-use interface to facilitate interviews and data entry. The programme is composed of five components, database navigation bar, identification data section, questionnaire, results and calculation button, and the programme database. The format of the programme is slightly different than the EPIC questionnaire where tables have been replaced by "Combo Boxes" which is a good alternative aiding in making the questionnaire shorter and easier to handle. Figure 4.16 describes the components of physical activity analysis programme.

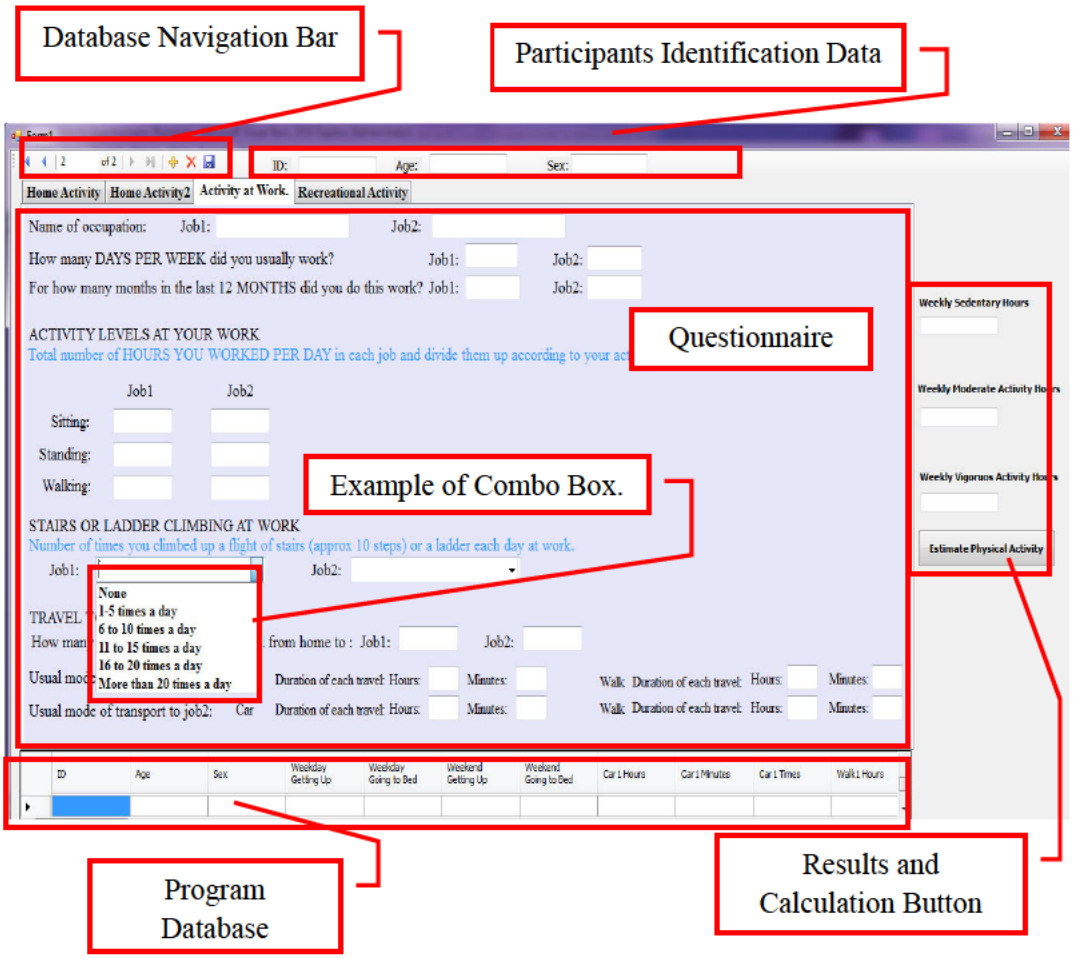


Figure 4.16: Physical Activity Analysis Programme.

The database navigation bar, participants identification data, programme database, and calculation button are similar in design and function to the ones explained in the food frequency analysis programme. However, the questionnaire part of the programme is composed of four different tabs where each tab has a category of physical activity. In addition, several data entry tools have been introduced to this questionnaire such as using combo boxes where interviewers will be able to choose from a list of choices and text boxes into which the interviewer can enter numeric values.

4.8.7.3 Programming

4.8.7.3.1 Assigning numerical values and combo boxes

Similar to assigning numerical values to variables in the food frequency analysis programme, physical activities were assigned numerical values to aid in the analysis process. For example, the number of hours slept during a week is assigned the value of "numtotalsleeping". Assigning the value was performed using the code: "Dim numtotalsleeping As Single". "Dim" symbolises an identification method used to assign the indicated value (numtotalsleeping in this example) as a single to allow saving the value as a numeric value. Using the method will allow applying formulae on entered data.

Combo boxes were used to replace lengthy columns and tables. Coding of all combo boxes in this programme was similar. However different values were utilised for certain questions. For example, the frequency of climbing stairs at home was designed using a box called "ComboBox1". A list of choices were entered in the box using the following codes:

.

```
ComboBox1.Items.Add ("None")
    ComboBox1.Items.Add ("1-5 times a day")
    ComboBox1.Items.Add ("6 to 10 times a day")
    ComboBox1.Items.Add ("11 to 15 times a day")
    ComboBox1.Items.Add ("16 to 20 times a day")
    ComboBox1.Items.Add ("More than 20 times a day")
```

Once these codes have been entered, the combo box would appear as shown in figure 4.17.

.

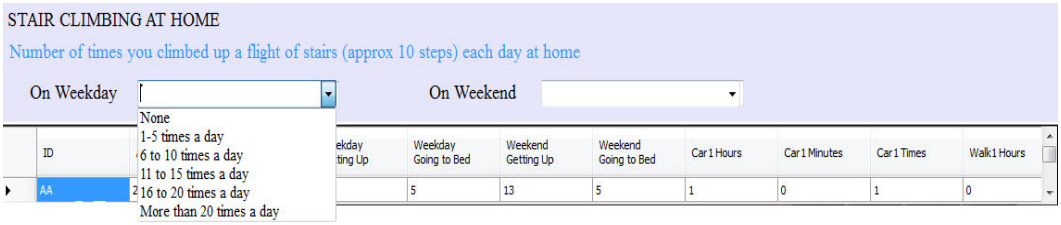


Figure 4.17: List of choices used in ComboBox1.

4.8.7.3.2 Coding Formulae

The analytic function of this programme was designed to calculate the average number of hours per week spent performing sedentary activities, moderate activities or vigorous activities. Table 4.3 summarises the process of sorting physical activities into these categories. These activities were already indicated in EPIC physical activity questionnaire. Activities which are more likely to cause an increase in heart rate and consume more calories are assumed to be vigorous activities. The total number of hours per week spent performing sedentary, moderate, or vigorous activities were calculated by summing the total of hours per week spent performing activities in each column.

Table 4.3: Sorting of physical activities:

Sedentary activities	Moderate activates	Vigorous activities
Sleeping	Walking	Climbing Stairs
Watching TV	Cooking	Walking carrying things at work
Driving	Shopping	Jogging
Sitting at work	Cleaning the house	Running
	Laundry	Football
	Caring for children	Weight lifting
	Caring for elderly	Conditioned exercise
	Standing at work	Swimming competitively
	Volleyball	Mountain climbing
	Leisure swimming	Tennis
	Fishing	
	Gardening	
	Home maintenance	
	Table tennis	

4.8.7.3.2.1 Coding Combo Boxes

A part of calculating the number of hours spent performing physical activity is to assign numerical values to choices listed in combo boxes. Fixed values were used to facilitate the analysis of choice. Table 4.2 gives an illustration of the values assigned for choices given in combo boxes.

Steps of calculating number of hours per week spent climbing stairs (AT HOME):

1. The number of 10 steps was assumed to be a standard number of steps climbed during each flight as assumed in the EPIC physical activity questionnaire. For the purpose of estimating time spent climbing stairs, one second was assumed as the time needed to climb one step. Therefore, 10 seconds would have been spent while climbing 10 steps.
2. Number of minutes (approximately) spent climbing stairs in each choice was assumed as follows:

- Choice 1: None: given a value of 0 minutes.
- Choice 2: 1-5 times a day: given a value of 1 minute spent climbing stairs every day.

3. Minutes were converted to hours per week according to each choice by using the following formula:

   Choice 2: (1 minute / 60) * 7 = 0.12 hours per week

4. As participants might give different a frequency of climbing stairs at home during weekdays and weekend, the following formulae were used to assign different values to time spent climbing stairs at home during weekends or weekdays:

   Choice 2: Stair climbing during weekdays = (5/7) * 0.12 hours per week = 0.08 hours/week

   Stair climbing during weekends = (2/7) * 0.12 hours per week = 0.04 hours/week

5. These values were later transferred to the programme during the coding phase as illustrated in Table 4.2.

Steps of calculating number of hours per week spent climbing stairs (AT WORK):

1. Number of steps and time spent climbing was calculated in a similar manner to stair climbing at home.
2. Number of minutes (approximately) spent climbing stairs in each choice was assumed in a similar manner to stairs climbing at home as follows:
   - Choice 1: None: given a value of 0 minutes.
   - Choice 2: 1-5 times a day: given a value of 1 minute spent climbing stairs every day.
3. Proportion of months spent working at job1 during last year = number of months spent working at job1/12
4. Proportion of months spent working at job2 during last year= number of months spent working at job 2/12
5. Hours spent climbing stairs per week at job1 was calculated according to each choice by using the following formula:

   ((Number of minutes spent climbing stairs every day at job1/60) * number of days per week attended at job1)) * Proportion of months spent working at job1 during last year

6. Hours spent climbing stairs per week at job2 was calculated according to each choice by using the following formula:

   ((Number of minutes spent climbing stairs every day at job2 /60) * number of days per week attended at job2)) * Proportion of months spent working at job2 during last year

Giving values for other combo boxes is relatively easier as the number of hours was stated in the choices or stated in text boxes linked to it. For example, a list of choices regarding activities around the house were given a medium number of hours listed in the choice such as a choice which states "from 1 to 3 hours" was given the value of 2 hours. Further illustrations are given in Table 4.4. As for recreational activities, the values of choices were converted to the weekly frequency of each activity based on the choice given. The frequency is later multiplied by the number of hours spent performing each activities. Further illustrations are given in Table 4.4.

Table 4.4: illustration of values given in combo boxes:

Activity name	List of Choices	Assigned Values	Codes used to assign values
Stair climbing at home (weekday) Coding Name: ComboBox1	None	0 Hours / week	Case 0 Stair1_WeekdayTextBox.Text = 0
	1-5 times a day	0.08 Hours / week	Case 1 Stair1_WeekdayTextBox.Text = 0.08
	6 to 10 times a day	0.16 Hours / week	Case 2 Stair1_WeekdayTextBox.Text = 0.16
	11 to 15 times a day	0.2 Hours / week	Case 3 Stair1_WeekdayTextBox.Text = 0.2
	16 to 20 times a day	0.35 Hours / week	Case 4 Stair1_WeekdayTextBox.Text = 0.35
	More than 20 times a day	0.42 Hours / week	Case 5 Stair1_WeekdayTextBox.Text = 0.42
Stair Climbing at Home (weekend)	Coding Name: ComboBox2 Values were assigned in a similar manner to the above example with changes of values.		
Stair Climbing at Work	Coding Name: ComboBox9 and 10 for Job1 and Job2 Values were assigned in a similar manner in the above example except for the part of calculating number of days where work was attended for job1 or job2.		
Activities in and Around the House Example of coding is given for caring for elderly activity. Remaining activities are coded similarly	None	0	Case 0 Adult_CaringTextBox.Text = 0
	Less than 1 hour a week	0.5 hour/ week	Case 1 Adult_CaringTextBox.Text = 0.5
	1 to 3 hours a week	2 hours/week	Case 2 Adult_CaringTextBox.Text = 2
	3 to 6 hours a week	4.5 hours/week	Case 3 Adult_CaringTextBox.Text = 4.5
	1 hour per day	7 hours/week	Case 4 Adult_CaringTextBox.Text = 7
	2 hours per day	14 hours/week	Case 5 Adult_CaringTextBox.Text = 12.5
	More than 15 hours a week	16 hours/week	Case 6 Adult_CaringTextBox.Text = 16
Recreational Activities Volley ball is given as an example of recreational activity. Remaining activities are coded similarly	None	0	Case 0 Volleyball_TextBox.Text = 0
	Less than once a month	0.2 times /week	Case 1Volleyball_TextBox.Text = 0.2
	Once a month	0.25 times /week	Case 2

			Volleyball_TextBox.Text = 0.25
	2 to 3 times a month	0.5 times /week	Case 3 Volleyball_TextBox.Text = 0.5
	Once a week	1 time /week	Case 4 Volleyball_TextBox.Text = 1
	2 to 3 times a week	2.5 times /week	Case 5 Volleyball_TextBox.Text = 2.5
	4 to 5 times a week	4.5 times /week	Case 6 Volleyball_TextBox.Text = 4.5
	6 times a week or more	6 times /week	Case 7 Volleyball_TextBox.Text = 6
Activity name	List of Choices	Assigned Values	Codes used to assign values

4.8.7.3.2.2 Calculation of number of hours per week spent performing sedentary activities

Several steps were used to calculate the number of hours per week spent in sedentary activities. Firstly, the number of hours spent sleeping was calculated by the following formula:

Total number of hours slept per week = daily number of hours slept during a weekday * 5 + daily number of hours slept during weekend * 2

Secondly, the number of hours spent watching TV or playing video games per week was calculated using the following formula:

Total number of hours spent watching TV or Playing video games per week = Number of hours spent watching TV or playing video games per day in weekdays * 5 + Number of hours spent watching TV or playing video games per day in weekends * 2

Thirdly, the number of hours spent driving the car per week (apart from going to work) was calculated by the following steps:

1. Number of hours spent driving during every day = (number of minutes spent driving / 60) + number of hours spent driving
2. Number of hours spent driving car per week (apart from going to work) = number of hours spent driving every day * 7

Fourthly, the number of hours per week spent driving to a job was calculated using the following steps:

1. Duration of travel by car to Job1= (number of minutes spent driving to job1 /60) + number of hours spent driving to job1
2. Duration of travel by car to Job2= (number of minutes spent driving to job1 /60) + number of hours spent driving to job2
3. Number of hours spent driving to Job1 per week = (duration of travel by car to job 1 * number of travels per week) * proportion of months spent doing job1
4. Number of hours spent driving to Job2 per week = (duration of travel by car to job 2 * number of travels per week) * proportion of months spent doing job2
5. Total number of hours spent driving to work per week = number of hours spent driving to job1 per week + number of hours pent driving to job2 per week

Fifthly, the number of hours per week spent sitting at work was calculated using the following steps:

1. Proportion of months spent working at job1 during last year = number of months spent working at job1/12

2. Proportion of months spent working at job2 during last year= number of months spent working at job 2/12
3. Number of hours spent sitting at job1 per week= (number of hours spent sitting every day at job1* number of days attended at job1 per week) * proportion of months spent working at job1 during last year.
4. Number of hours spent sitting at job2 per week= (number of hours spent sitting every day at job2* number of days attended at job2 per week) * proportion of months spent working at job2 during last year.
5. Total number of hours per week spent setting at work during last year = Number of hours spent sitting at job1 per week + Number of hours spent sitting at job2 per week

Finally, the total number of hours per week spent performing sedentary activities = number of hours per week spent sleeping + number of hours per week spent driving (apart from going to work) + number of hours spent travelling to work by car + number of hours spent sitting at work.

4.8.7.3.2.3 Calculation of number of hours per week spent performing moderate activities

The calculations in this section are similar to the calculations performed to calculate the hours spent in sedentary status. The following steps are a detailed explanation of the process:

Firstly, the number of hours spent walking per week (apart from going to work) was calculated by the following steps:

1. Number of hours spent walking during every day = (number of minutes spent walking / 60) + number of hours spent walking
2. Number of hours spent walking per week (apart from going to work) = number of hours spent walking in every day * 7

Secondly, activities performed in or around the house such as activities listed in Table 4.3 under the moderate activities section are calculated by assigning a fixed value to each choice in the combo boxes. Table 4.4 shows the values assigned to choices in activities in and around the house section. All of these activities are assumed to be moderate activities and the values are assigned as hours spent every week performing each activity.

Thirdly, the number of hours per week spent standing at work was calculated using the following steps:

1. Proportion of months spent working at job1 during last year = number of months spent working at job1/12
2. Proportion of months spent working at job2 during last year= number of months spent working at job 2/12
3. Number of hours spent standing at job1 per week= (number of hours spent standing every day at job1* number of days attended at job1 per week) * proportion of months spent working at job1 during last year.
4. Number of hours spent standing at job2 per week= (number of hours spent standing every day at job2* number of days attended at job2 per week) * proportion of months spent working at job2 during last year.
5. Total number of hours per week spent standing at work during last year = Number of hours spent standing at job1 per week + Number of hours spent standing at job2 per week

Fourthly, the number of hours per week spent walking to work was calculated using the following steps:

1. Proportion of months spent working at job1 during last year = number of months spent working at job1/12
2. Proportion of months spent working at job2 during last year= number of months spent working at job 2/12
3. Duration of walking to Job1= (number of minutes spent walking to job1 /60) + number of hours spent walking to job1
4. Duration of walking to Job2= (number of minutes spent walking to job2 /60) + number of hours spent walking to job2
5. Number of hours spent walking to Job1 per week = (duration of walking to job 1 * number of travels per week) * proportion of months spent doing job1
6. Number of hours spent walking to Job2 per week = (duration of walking to job 2 * number of travels per week) * proportion of months spent doing job2
7. Total number of hours spent walking to work per week = number of hours spent walking to job1 per week + number of hours pent walking to job2 per week

Fifthly, the calculation of hours per week spent performing moderate recreational activities. These activities are listed in Table 4.3 as moderate activities. Its calculation is based on the value assigned for each list in the combo boxes as illustrated in Table 4.4 under the section of recreational activities. It is also based on entered number of hours and minutes spent performing each activity as illustrated in Figure 4.18.

Example of calculating number of hours spent performing moderate recreational activities:

Number of hours spent fishing per week = number of times spent fishing * average time per episode

Example: the number of hours spent fishing per week if a person performs the activity once per month= weekly frequency (from table 4.4 = 0.25 times weekly) * Spent an hour and half in each episode = 0.37 hours per week.

Finally, the total number of hours per week spent performing moderate activities is the sum of hours per week calculated performing all moderate activities.

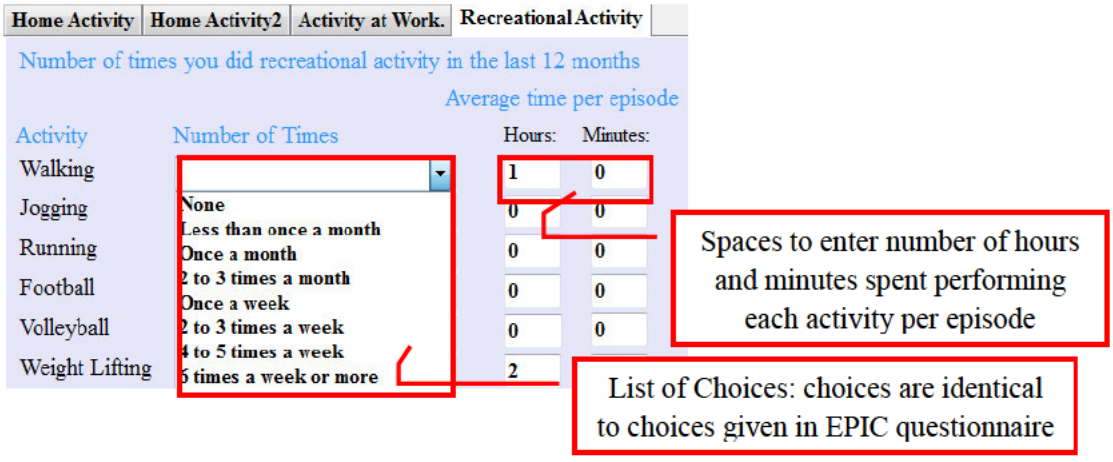


Figure 4.18: Data entry for recreational activities.

4.8.7.3.2.4 Calculation of number of hours per week spent performing vigorous activities

It involves all the activities listed in Table 4.3 under the vigorous activities section. To calculate the number of hours per week spent performing these activities, the programme included coding of formulae as follows:

Firstly, the number of hours spent climbing stairs at home:

The calculation of this section is already explained in section 8.7.3.2.1. Number of hours spent climbing stairs at home will be the total of number of hours per week spent climbing stairs during weekdays and number of hours spent climbing stairs during the weekend.

Secondly, the number of hours per week spent walking carrying things at work is calculated as follows:

1. Proportion of months spent working at job1 during last year = number of months spent working at job1/12
2. Proportion of months spent working at job2 during last year= number of months spent working at job 2/12
3. Number of hours spent walking at job1 per week= (number of hours spent walking every day at job1* number of days attended at job1 per week) * proportion of months spent working at job1 during last year.
4. Number of hours spent walking at job2 per week= (number of hours spent walking every day at job2* number of days attended at job2 per week) * proportion of months spent working at job2 during last year.
5. Total number of hours per week spent walking at work during last year = Number of hours spent walking at job1 per week + Number of hours spent walking at job2 per week

Thirdly, the number of hours per week spent climbing stairs at work is calculated as explained in section 8.7.3.2.1. The total number of hours per week spent climbing stairs at work is the sum of hours per week spent performing this activity at job1 and job2.

Fourthly, the number of hours per week spent performing vigorous recreational activities are calculated in a similar manner to the moderate recreational activities.

Finally, the total number of hours per week spent performing vigorous activities is the sum of all hours per week spent performing activities listed in Table 4.3 under the vigorous activities section.

4.8.7.4 How to operate the programme

The steps for using this programme are similar to the steps explained in the food frequency analysis programme. The steps for using the programme are displayed in Figure 4.19. Steps are ordered from first to last. The process is ended by pressing the saving icon which will save all entered data and calculated values in the database. The number of hours per week spent performing different types of activities will be separately displayed in the results section.

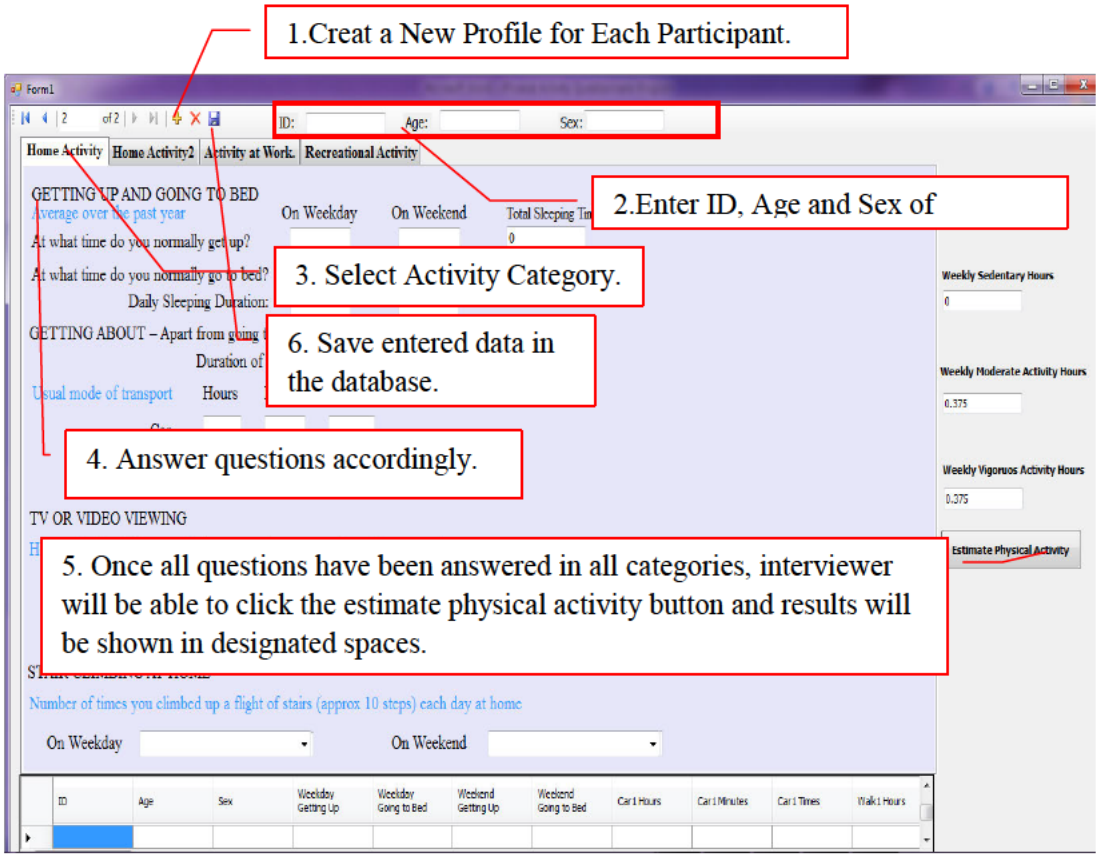


Figure 4.19: Steps of using physical activity analysis program.

4.9 Physical measurements

4.9.1 Body Weight

The body weight of healthy participants was measured using the Beurer PS05 scale. All participants were measured using the same device. At the time of measurement, participants were wearing light clothing, and standing barefoot. Participants tended to look downward to inspect the readings of the scale. However, this action appeared to influence the reading due to the scale's sensitivity to minor movements. Therefore, participants were advised to remain stable until the reading was complete and shown on the scale's screen. Body weight was recorded in kilograms.

We also noticed that using the scale on non-hard or non-flat surfaces provided faulty readings. Those surfaces included – but were not limited to – carpets, sand, and uneven pavements. We tended to measure body weight on hard and flat surfaces to provide accurate measurements. To ensure the competent function of the scale, we also insured to always carry additional batteries to avoid the inability to perform measurements due to low batteries.

4.9.2 Height

Height was measured using the Seca 213 Portable Stadiometer as illustrated in Figure 4.20. As all measurements were taken during homes visits, it was pragmatic to use this stadiometer as it can easily be dismantled and assembled. Once it is dismantled, it can be easily fitted to the case of the stadiometer and easily transported as depicted in picture A. In addition, the scale weighs 2.4kg which enabled easy transportation. Once assembled, it will appear as shown in picture B.

Participants were advised to be barefoot while having their height measured. They were also advised to remove any unnecessary clothing worn on their heads. Once participants were standing on the base of the stadiometer, the indicator shown in picture C was lowered until the lower surface of the indicator was touching the head of the participants. Measurements of height were taken according to the location of the two red arrows on the height scale. Height was taken in centimetres. After converting height to metres, we were able to calculate BMI based on weight and height.

Figure 4.20: Measuring height using Seca 213.

4.9.3 Waist Circumference

Waist circumference was measured using a measuring tape shown in Figure 4.21. Measurements were made according to the regulations of the Joint British Societies' guidelines with minor exceptions. According to the Joint British Societies' guidelines on the prevention of cardiovascular disease in clinical practice, abdominal circumference needs to be measured after clothes are removed, while standing, and the abdomen should be relaxed, feet should be about 25-30 cm apart, and arms held loosely at the side. The tape shall be placed horizontally at the midpoint between the iliac crest (the upper edge of the hip bone) and the lower edge of the lower rib. After asking the subject to gently breathe in and out, measurements can then be taken (272). In this study, the measurements were not performed in clinical settings therefore, measurements were made while participants were wearing light clothing.

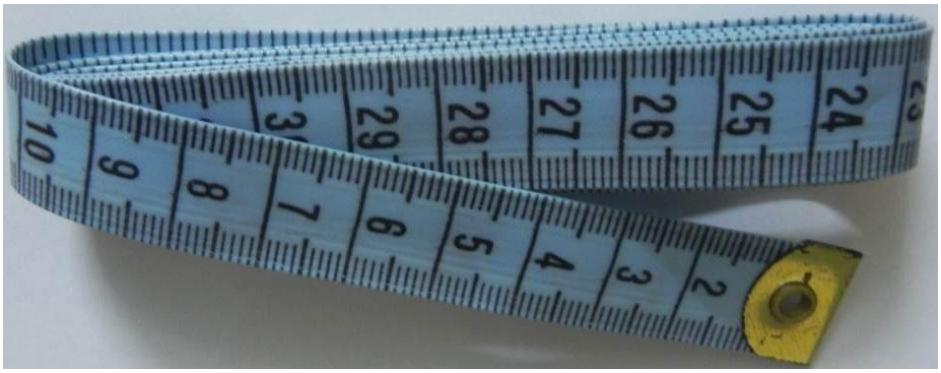


Figure 4.21: Measurement tape used to measure waist circumference.

4.9.4 Fasting Blood Glucose

For several reasons, FBG was measured using a Glucometer. Firstly, this test is suitable to be performed in non-clinical settings. Secondly, it has been suggested as being more suitable for screening purposes than other measures such as Hb1Ac (273). Finally, this test is cheaper and more convenient than other laboratory tests. In addition to having a continuous measure of FBG, using this test also aided in assuring that unaffected cases did not have potentially undiagnosed type 2 diabetes. If an apparently normal subject appeared to have above the normal level of FBG (>126 mg/dl) they were not excluded from the study. Such subjects were advised to seek medical advice from their family physicians.

Measurements of blood glucose were made using Optium Xceed glucometer Figure 4.22 (A). Before performing a measurement of blood glucose, we ensured that participants were fasting for the previous eight hours. Otherwise, measurements had to be arranged. We used lancets (Figure 4.22 B) to acquire a capillary blood drop. Before applying the lancets, the site of applying the lancets was disinfected using an alcohol swab (Figure 4.22 C). Usually, participants tended to have their thumbs pricked. We only used new capped and sterile lancets as shown in Figure 4.22 (B).

Lancets were loaded in the lancet applicator (Figure 4.22 D). In order to insert the lancet in the applicator, the plastic cover of the applicator (Figure 4.22 E) had to be removed. Once the lancet was inserted in the applicator, we were able to remove the cap of the lancet and cover the applicator. The investigator had to make a judgment regarding the strength of the prick.

.

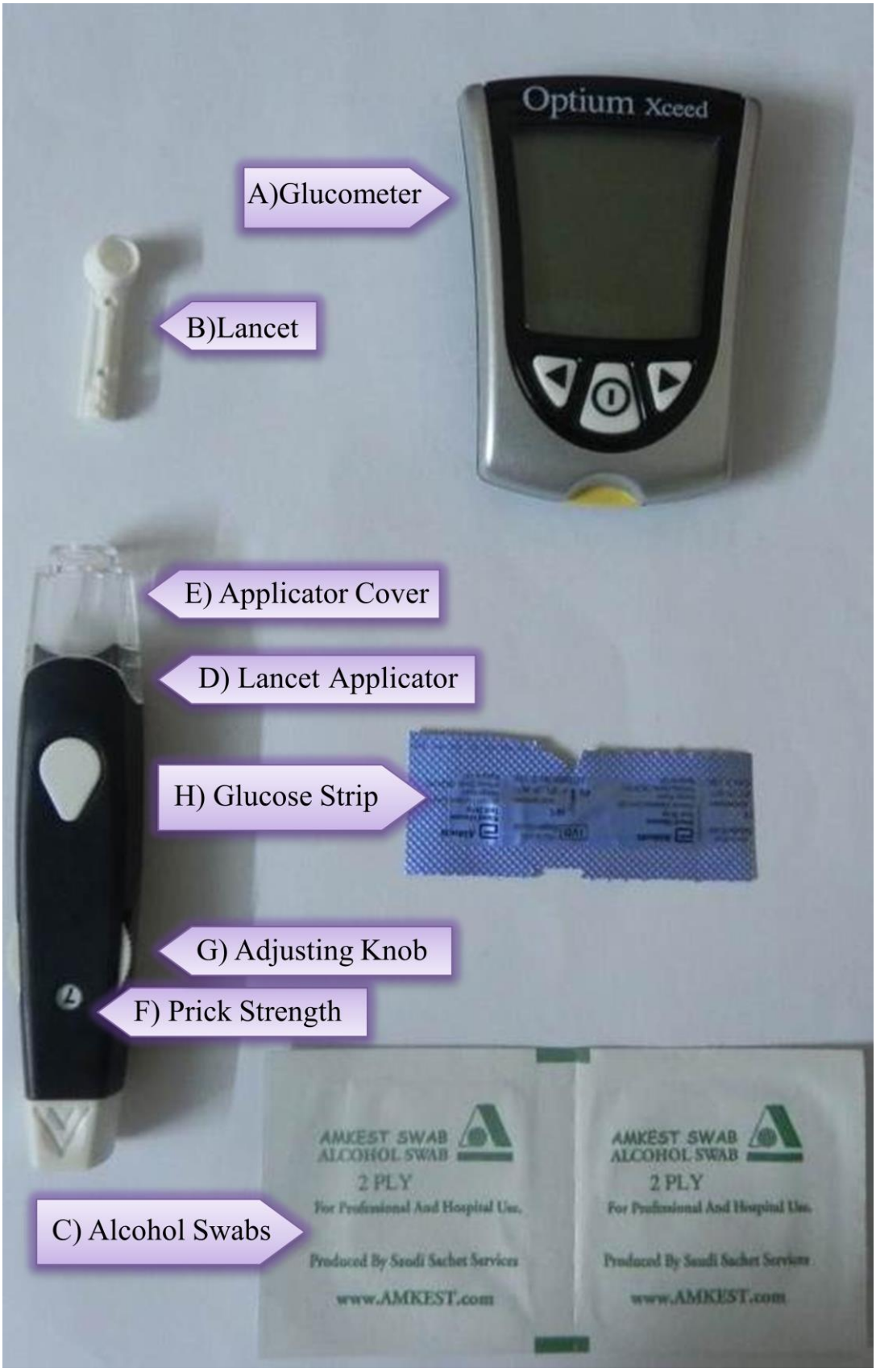
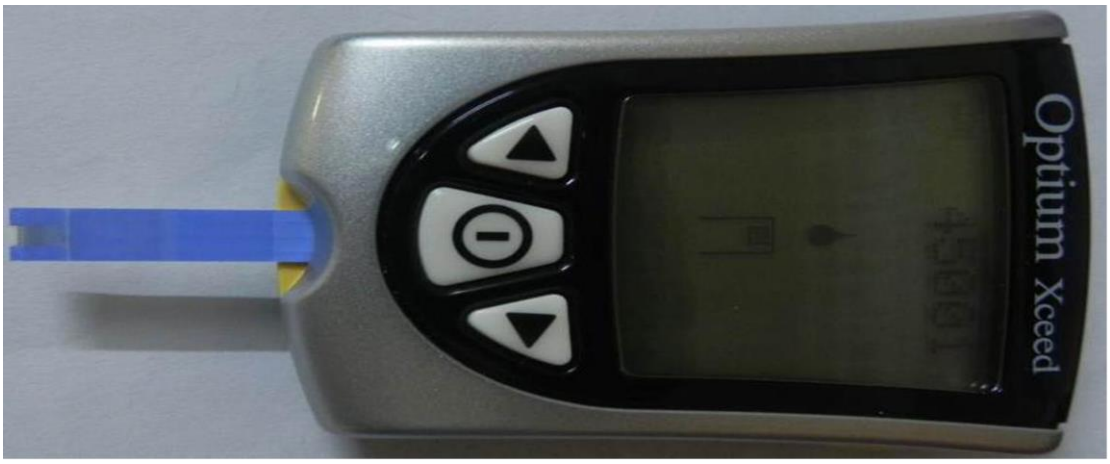


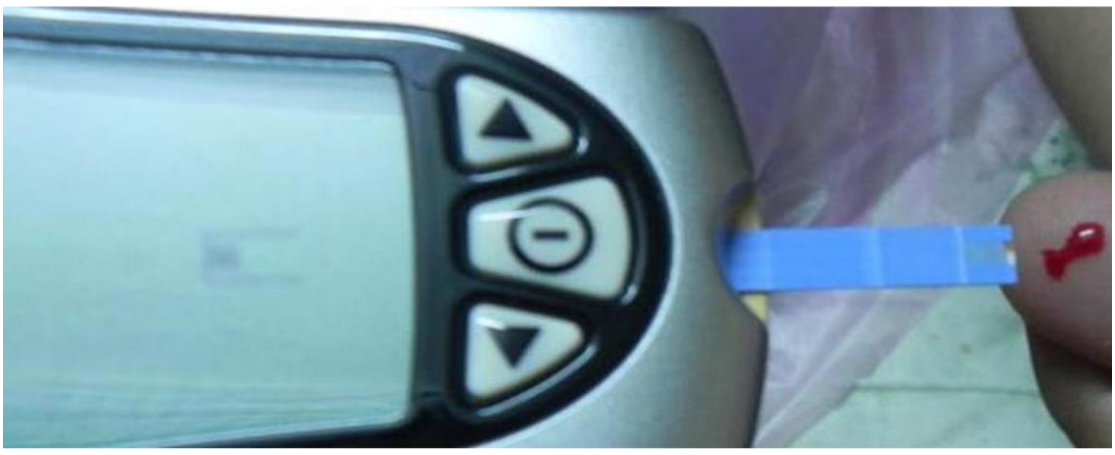
Figure 4.22: Tools of measuring blood glucose.

The strength of the prick is indicated in numbers as shown in Figure 4.22 (F). For individuals who have thicker thumb skin, we had to set the strength indicator at 8 to successfully acquire a blood drop. Otherwise, acquiring the drop would not have been possible. For individuals with thinner thumb skin, a smaller unit was used. The most often used unit was 7. Changing the strength unit was made using adjustment knob indicated in Figure 4.22 (G).

Glucose strips were used to apply blood drops to the glucometer. Glucose strips are indicated in Figure 4.22 (H). A new strip was used for each participant. Inserting strips into the glucometer is shown in Figure 4.23 (A). Once the strip was inserted, the glucometer started immediately and was ready to have the measurement taken. Once a blood drop was acquired, the glucose measurement was taken as shown in Figure 4.23 (B). Finally, the blood glucose level would appear on the glucometer screen in mg/dl. Afterward, participants were given new and sterile cotton balls to apply at the site of the finger prick.



A) Applying Glucose Strip.



B) Blood Glucose Measurement.

Figure 4.23: Applying Glucose Strip to Glucometer and blood glucose measurement.

To ensure that all used lancets were disposed of appropriately, we needed to use a medical waste bin. However, we could not find a medical waste bin in the Jazan region available for sale. Therefore, we had to use a normal bin and had it pierced to allow the insertion of strips without opening the bin. The bin used is shown in Figure 4.24 (A). To ensure our safety from being pricked by used lancets, we removed the lancet using the method shown in Figure 4.24 (B) and disposed of lancets as shown in Figure 4.25 (C). The bin was saved in a secure location and was disposed of appropriately by the end of the data collection.

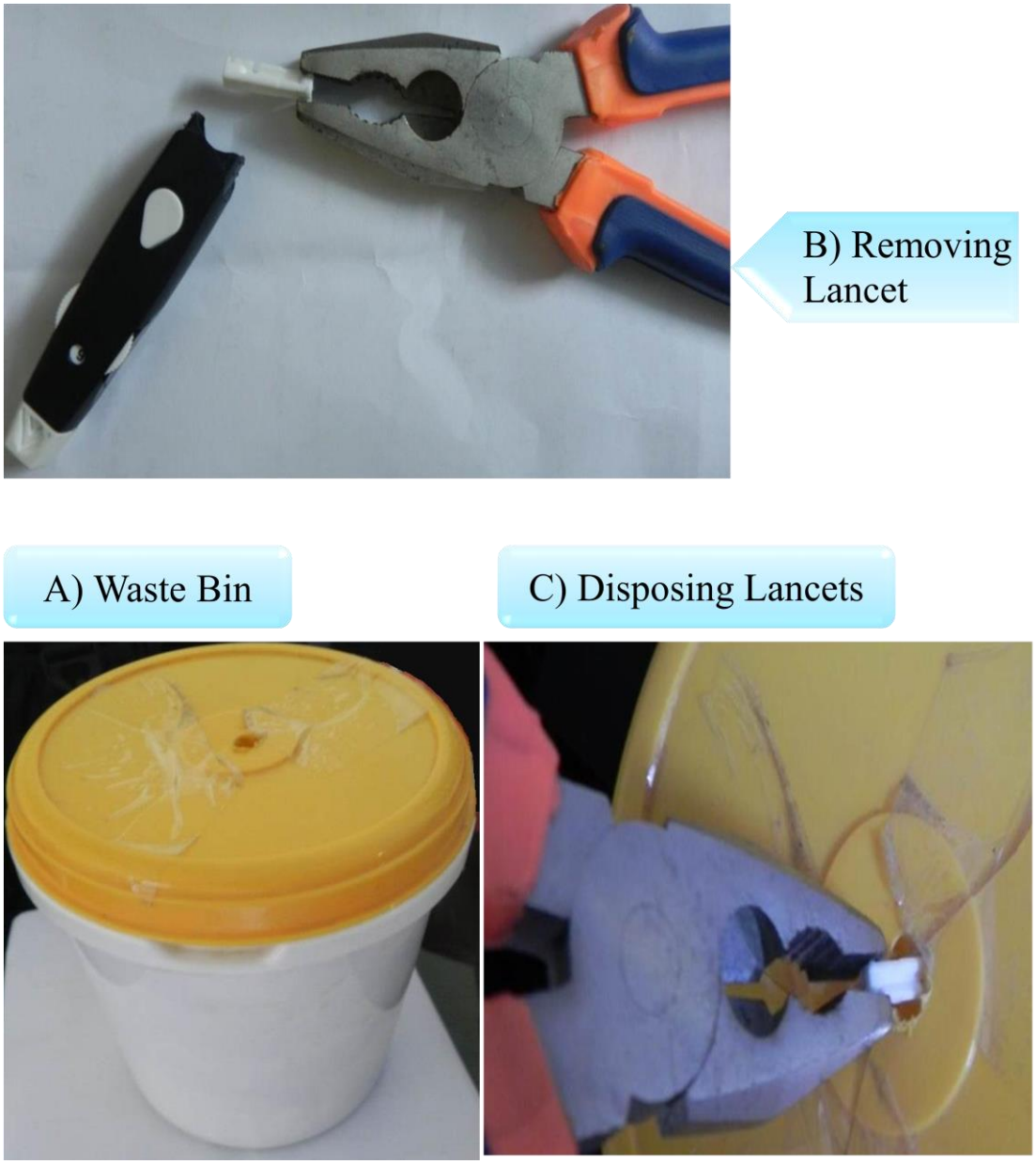


Figure 4.24: Waste Bin and disposal of Lancets.

4.10 DNA Samples and Genetic Analysis

4.10.1 Collection Tool

We used buccal swabs as an option of collecting DNA samples from participants. Unlike the collection of blood samples, it is a non-invasive procedure, painless, and does not introduce patients to the risk of infections caused by phlebotomy. In addition, it can be performed by non-professionals in non-clinical settings. Using buccal swabs appeared to increase the participation rate due to its convenience. Buccal swabs (SK-1S DNA Swab Pack) were purchased from Isohelix (a division of Cell Projects, UK). Each swab includes swab, tube and a cap as illustrated in Figure 4.25.

Figure 4.25: Swab used to collect DNA samples.

According to the Isohelix, swabs had to be taken one hour after eating or drinking and the mouth had to be rinsed immediately with water prior to taking the sample. We were advised by Isohelix to take clean samples to avoid any possibility of contamination with any materials with anti-PCR (Polymerase Chain Reaction) qualities. Anti-PCR materials can be food or chemical materials which might interfere with DNA amplification. Participants were advised to hold the end of the swab and enter the swab in the mouth toward the inner side of the cheek. They were later instructed to move the swab in a vertical movement and applying mild pressure toward the cheek for at least one minute.

Extreme measures were taken to avoid contamination of the swabs by the participants or by the investigator. If for any reason, a buccal swab was not appropriately placed in the tube in a secure condition, another swab was taken instantly with disposal of the unfit one. By the end of collecting the swab, we were able to break the stick to allow storing of the swab in the tube. Place of breaking the stick is indicated by a red arrow in Figure 4.25.

In order to store the swabs for a long period we used preservative materials. Preservative capsules (SGC Dri-Capsules) were provided by Isohelix which allowed storing the samples at

normal room temperature. Samples were stored in a secured cabinet which was only accessed by the principal investigator.

While new capsules were yellow, a used one tended to be greenish. According to the guidelines of using the capsules, dri-capsule which are greenish in colour before using are not suitable for use. However, all unused capsules were yellow in colour once they were evacuated from their containers. The method of storing the swab in the tube and applying dri-capsules is illustrated in Figure 4.26. A comparison between new and old dri-capsules is also indicated in Figure 4.26.

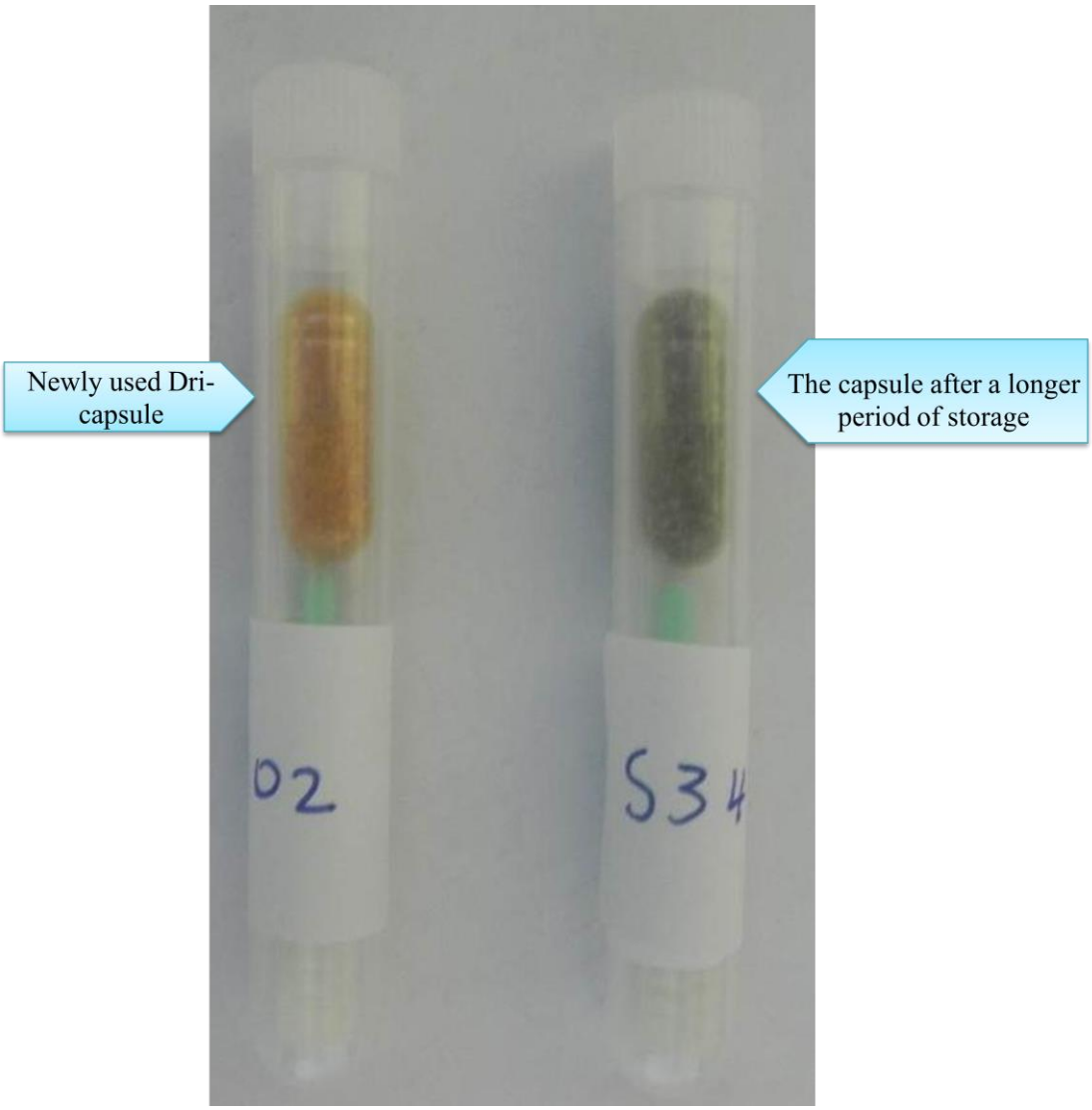


Figure 4.26: storing buccal swabs using Dri-capsules.

4.10.2 Labelling Samples

Buccal swabs were anonymously labelled with codes given for each participant. Codes were written on white stickers. Stickers were applied to the storage tubes as indicated in Figure 4.26. We used additional adhesive material to ensure the strong attachment of the stickers.

4.10.3 Importing Samples

Importing samples to the UK was conducted in several steps. Firstly, we had to ensure that samples were given necessary clarifications and permissions to be imported. In addition to contacting Human Tissue Authority, we contacted The Department of Health in the UK. The Department of Health replied by advising to seek approval from H&M Revenue and Customs.

The H&M Revenue and Customs directed our attention to fill needed forms which are illustrated in Appendix 11. We illustrated the academic nature of the research and therefore, the sample was exempt from duty or taxes. Further clarifications of the shipment were detailed in an official letter (Appendix 12).

Once all necessary permissions were taken, samples were shipped using Fedex. As samples could be stored at normal temperature, no further regulations were given to the courier regarding the condition of transportation except not exposing the shipment to heat. All forms and shipment letters were attached to the shipment box.

We were able to track the location of the samples using the tracking systems provided by the courier. The shipment was delayed in Manchester and a call had to be made by the principal investigator to give more clarification about the samples and the purpose of importing the sample to the UK. Once the call was made and clarifications were given, the samples were released from UK Customs. The route of shipping the samples is illustrated in Figure 4. 27.

All shipment travel activity is displayed in local time for the location

Date/Time	Activity	Location	Details
Jan 11, 2013 9:56 AM	**Delivered**	SHEFFIELD GB	
Jan 11, 2013 9:02 AM	On FedEx vehicle for delivery	SHEFFIELD GB	
Jan 11, 2013 8:53 AM	At local FedEx facility	SHEFFIELD GB	
Jan 11, 2013 5:27 AM	International shipment release - Import	MANCHESTER GB	
Jan 10, 2013 10:51 AM	International shipment release - Import	MANCHESTER GB	
Jan 10, 2013 8:17 AM	At local FedEx facility	MANCHESTER GB	
Jan 10, 2013 8:08 AM	Clearance delay - Import	MANCHESTER GB	Description provided is insufficient to classify commodity.
Jan 10, 2013 7:08 AM	In transit	MANCHESTER GB	
Jan 10, 2013 7:08 AM	In transit	MANCHESTER GB	Package available for clearance
Jan 10, 2013 4:31 AM	At destination sort facility	BIRMINGHAM GB	
Jan 10, 2013 2:42 AM	Departed FedEx location	ROISSY CHARLES DE GAULLE CEDEX FR	
Jan 10, 2013 1:30 AM	Arrived at FedEx location	ROISSY CHARLES DE GAULLE CEDEX FR	
Jan 10, 2013 1:30 AM	In transit	ROISSY CHARLES DE GAULLE CEDEX FR	
Jan 9, 2013 8:25 PM	Departed FedEx location	ROISSY CHARLES DE GAULLE CEDEX FR	
Jan 9, 2013 7:38 PM	In transit	ROISSY CHARLES DE GAULLE CEDEX FR	
Jan 9, 2013 6:18 PM	Arrived at FedEx location	ROISSY CHARLES DE GAULLE CEDEX FR	
Jan 9, 2013 2:05 PM	In transit	DUBAI AE	
Jan 9, 2013 2:05 PM	In transit	DUBAI AE	
Jan 9, 2013 1:21 PM	In transit	DUBAI AE	
Jan 8, 2013 9:13 PM	In transit	RIYADH SA	
Jan 8, 2013 9:12 PM	Left FedEx origin facility	RIYADH SA	

Figure 4.27: Shipment of samples. Description of delay of samples release is indicated in the red rectangle.

4.10.4 SNPs Selection

GWAS and replication studies are used to discover and confirm the association between risk alleles and the disease of interest. Currently, no one has performed an association or replication study in Saudi Arabia concerning type 2 diabetes. Therefore, no one knows if the type 2 diabetes risk alleles which are detected in European and Asian populations are replicable in the Saudi Arabian population.

In this study, we have made an effort to avoid genotyping risk SNPs which might not be replicable in the Saudi population. The study genotyped selected SNPs based on several criteria. Firstly, we selected SNPs which were replicated in more than one population. Secondly, we genotyped SNPs which have been shown to affect more than one continuous trait related to type 2 diabetes. Another criteria was based on selecting SNPs which were shared with reduction of birth weight. Finally, the study genotyped SNPs which were found to be shared with other forms of monogenic diabetes. A list of genotyped SNPs is illustrated in table 4.5.

Table 4.5: list of genotyped SNPs and reasons for selection of each SNP:

No.	Marker	Locus	Chr	Allele (Effect/Other)	Reason for Selection
1	rs7903146	TCF7L2	10	T/C	Most replicated among ethnicities. Affects FBG, Insulin, and HbA1(c) levels
2	rs13266634	SLC30A8	8	C/T	Replicated in European and Asian populations. Affects FBG, Insulin, and HbA1(c) levels.
3	rs1111875	HHEX	10	C/T	Replicated in European and Asian populations. Associated with birth weight reduction.
4	rs10811661	CDKN2B	9	T/C	Replicated in European and Asian populations.
5	rs7754840	CDKAL1	6	C/G	Replicated in European Asian, and Arab Lebanese populations.
6	rs7756992	CDKAL1	6	G/C	Replicated in European Asian and Arab Lebanese populations. Associated with birth weight reduction.
7	rs10946398	CDKAL1	6	A/C	Replicated in European and Asian populations.
8	rs2237895	KCNQ1	11	C/T	Replicated in European and Asian populations.
9	rs231362	KCNQ1	11	G/A	Replicated in European and Asian populations.
10	rs2237897	KCNQ1	11	C/T	Replicated in European and Asian populations.
11	rs560887	G6PC2	2	C/T	Affect FBG, insulin, and HbA1(c) levels.
12	rs10830963	MTNR1B	11	G/C	Affect FBG, insulin, and HbA1(c) levels.
13	rs1387153	MTNR1B	11	T/C	Affects FBG and HbA1(c)
14	rs4607517	GCK	7	A/G	Affect FBG, insulin, and

					HbA1(c) levels. Shared with new natal diabetes and MODY.
15	rs1799884	GCK	7	T/C	Affects FBG and HbA1(c)
16	rs780094	GCKR	2	C/T	Affects FBG and insulin level.
17	rs11708067	ADCY5	3	A/G	Affects FBG and insulin level. Associated with birth weight reduction.
18	rs2191349	DGKB-TMEM195	7	T/G	Affects FBG and insulin level.
19	rs174550	FADS1	11	T/C	Affects FBG and insulin level.
20	rs7034200	GLIS3	9	A/C	Affects FBG and insulin level.
21	rs5219	KCNJ11	11	T/C	Shared with neonatal diabetes. Associated with birth weight reduction.
22	rs7957197	HNF1a	12	T/A	Shared with MODY 3
23	Rs757210	HNF1B	17	A/G	Shared with MODY 5

4.10.5 DNA Extraction and Genotyping

The process of DNA extraction and genotyping was performed by K-biosciences, UK. To gain knowledge about the steps of performing DNA extraction and genotyping, we visited K-biosciences facilities in Hoddesdon, UK, to observe the process on-site and visited their website for more detailed information (www.lgcgenomics.com). Several steps were taken to ensure the extraction of DNA, preparing DNA for genotyping and reporting allelic variability in selected SNPs.

The process of extracting and genotyping DNA was performed using the following steps. Buccal swabs were stored in an incubator oven where DNA was extracted from buccal cells using a silica-based method. DNA attaches to the silica allowing other materials to be eliminated and therefore only maintaining participants' DNA. This method is human-DNA specific and DNA from other sources would be eliminated.

In order to perform genotyping of selected SNPs, K-biosciences requested at least 50 bases on either sides of the SNP of interest. These sequences were used to establish assays needed in the genotyping process. KASP (**K**ompetitive **A**lelle **S**pecific **P**CR) genotyping assays were designed to aid in the amplification of DNA and detecting variants of interest.

KASP genotyping assay contains several elements. Firstly, primers were built based on the sequence we provided for each SNP. Two forward primers were made with two different alleles at the 3' end and one reverse primer is added to form the KASP Primer Mix which is the first element of the KASP assay. The second element of the KASP assay is the DNA sample provided by the participants. Thirdly, a master mix is added which contains specific sequence-binding fluorescent dyes. Each sequence at the 5' ends of the two forward primers contains specific sequence only matching a specific sequence-binding fluorescent dye.

Once DNA was extracted in a liquid format, the RepliKator was used to transfer liquid format material containing DNA to a 96 well-plates. Four plates were used to cover the whole samples of this study. A Plate Mate system was used to assign identification labels to each sample in the plate and linking it to the coding system we initially used during data collection. Barcodes were also used to identify each plate and were integrated in the K-bioscience laboratory workflow management system which is called Kraken.

Once assays were ready to be used, a machine called Meridian was used to transfer assays to the DNA samples already dispensed in the plates. This machine is automated so there is no human contact needed to transfer the samples. The machine aspires the needed assays and dispenses them into the plates with no risk of cross contamination.

After dispensing assays into the plates, plates were sealed using Fusion 3. This machine uses non-heat laser technique to seal plates. Once the plates were sealed, the plates were ready to be used in the Hydro Cycler.

The Hydro Cycler is a water bath thermal cycler. It uses water to aid in conducting heat to the PCR samples. Several thermal cycles were conducted leading to the replication of DNA through several rounds. In the first round, heat results in the separation of two strands of DNA which is called denaturation. Once the strands are separated, the matching forward primer attaches to the DNA sample and elongates by adding nucleotides.

If we assume that the variation at the target SNP is either T or G, then the two forward primers assembled in the KASP assay will end in either A or C. If the allele in the SNP of interest is T, then the premier ending with A would attach to the DNA sequence where the primer ending with C would not attach. Once the matching primer is attached (in a process called Annealing), the polymerase reaction would start elongating the primer starting from the 5' end to generate a new sequence containing the matching primer.

In the second cycle of PCR, the newly formed double strand of the original DNA sample and the matching primer sequence would denaturant. Then, the reverse primer attaches to the matching primer sequence and start creating a reverse copy (a complement). The final round of PCR includes denaturation of the double strand composed of the matching primer sequence and its complement strand. This is followed by the attachment of a specific fluorescent dye binding sequence to the complementary sequence. Once a new double strand is made containing a strand complement sequence of the matching primer sequence and one sequence containing fluorescent dye, amplification of this double strand leads to creation of multiple copies.

The final step of genotyping will be the detection of allele at the SNP of interest. Because the final round of PCR results in the amplification of DNA strands containing specific fluorescent dye, the type of alleles detected will be determined based on the colour of the fluorescent dye observed. The strength of the signal fluorescent dye will depend on how many of the final strands of DNA had been replicated.

The final report of allelic distribution of each SNP will be depicted as shown in Figure 4.28. Different colours are used to illustrate the variability of genotypes of each SNPs. Homozygote genotypes are depicted in red or blue and heterozygote genotypes are shown in green. Pink clusters indicated samples which were not successfully genotyped for the SNP of interest. Black clusters serve as control and are called NTCs (No Template Controls). Clustering of signals around NTCs points will indicate the invalidity of the genotyped sample.

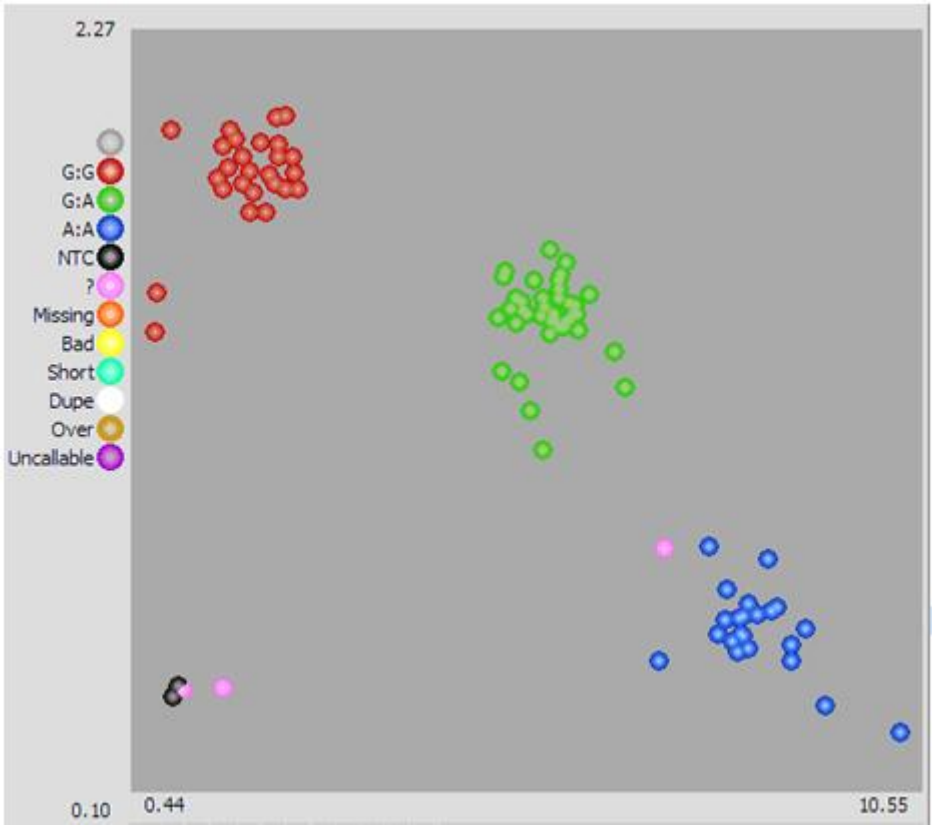


Figure 4.28: Signals used to depict detected genotypes of a particular SNP.

The final report was made in an Excel format and an additional SNP viewer provided by K-biosciences. Staff at K-bioscience explained that cross contamination with human DNA is a major risk factor influencing the successful yield of DNA. This risk of reducing the yield is higher when the contamination of the participants' DNA sample with other human DNA compared to contamination of a DNA sample with a non-human DNA. Furthermore, the duration spent using the buccal swab from each participant is another factor influencing the yield of DNA. Participants who spend a shorter period using the buccal swabs to retrieve buccal cells are less likely to yield a significant amount of DNA compared to those who spend longer periods.

4.11 Data Entry

Data entry was performed in several steps. Firstly, during home visits, we were able to save collected data using a food frequency analysis programme and physical activity analysis programme. At that stage, data was saved in the programmes' database. Data was then transformed from the programmes' database to a Microsoft Office Excel file. This measure allowed us to have several secured copies of collected data. Copying data to a memory stick was performed on a weekly basis to provide spare copies in case of any loss of the original data.

By the end of each of the four stages of data collection, all data was transformed from Excel files to SPSS. The final destination of all collected data is the designated SPSS database. All calculated inbreeding coefficients were fed to the SPSS database. All steps of data entry were performed by the principal investigator.

4.12 Study Variables and analysis

Study variables and analysis are explained in Table 4.1. All of the variables explained in the table are continuous. Correlation analysis was utilised to assess the association of study variables. Regression analysis was utilized to account for multiple independent factors. SPSS was utilised to perform the analysis.

4.13 Data Collection Team

Identification and approaching patients was the responsibility of PHCs staff. Recruitment of all participants was made by the principle investigator. All home visits and interviews and collection of anthropometric measures and biospecimens were performed by the principle investigator.

4.14 Piloting

We needed to assess the ability of the food frequency analysis and physical activity analysis programmes to perform their tasks. We made interviews with 10 individuals to assess the programmes. Every individual was interviewed using both programmes. Several faults and adjustments were made to the software. Other adjustments involved the process of interviewing subjects.

Regarding the food frequency analysis programme, the questionnaire performed well with most participants. However, a few additional food items were suggested by each participant. The action taken was to add spaces to the programme to allow entering open ended answers and frequency of consumption.

During the pilot stage of the food frequency analysis programme, we noticed that one of the items was not functioning properly. Selecting radio buttons of that item did not reveal corresponding calorific values. This problem was solved by rewriting commands of calculating the calorific value of that item.

Another issue was observed relating to reporting frequencies. Some individuals tend to assume that they have consumed certain food items on monthly basis which in reality was not true. This assumption of consuming variety of food at least once a month resulted in exaggerated total calorific values. To resolve this issue, interviewees were clearly asked if a month would pass without consuming the item of interest. If the participant answers yes, then the frequency would be "never". This action resulted in giving reasonable values of calorific intake once applied in the actual collection of data.

Similarly, several changes were made to the physical activity analysis programme. The EPIC physical activity questionnaire included a section about sleeping time. During the pilot stage, it was observed that asking participants about approximate duration spent sleeping on a daily basis was easier than asking participants about the time they go to bed at night and the time they wake up in the morning as written in the EPIC questionnaire. That issue was complicated further if they have two episodes of sleeping in one day which was not indicated in the original EPIC questionnaire. The adjustment appeared to be beneficial in considering those who have more than one episode of sleep every day.

Asking about total duration spent driving or walking on a daily basis appeared to be more appropriate than asking about the duration of a single journey and then multiplying it by the number of journeys. Some individuals might have several journeys per day with different durations thus making the current format of the questionnaire inappropriate. Therefore, we changed this question to only ask about the total duration of hours or minutes spent driving or walking on daily basis without asking about the number of journeys.

A distinction has to be made regarding considering walking as a vigorous activity during work. If walking during work included carrying heavy items, the activity was considered a vigorous activity. Otherwise, walking was assumed to be a moderate activity.

4.15 Sample Size

To calculate the sample size needed to answer the questions, we consulted sample size tables (274) and GPower 3.1 software for sample size calculation. Table 4.6 shows the sample size needed for each research question and the correlation coefficients could be detected. These correlation coefficients are calculated assuming a two-sided type 1 error rate of 5% and 80% power to detect a true correlation. The total sample size was estimated to be 400. Approximately 150 subjects were planned to be type 2 diabetes cases and 250 were planned to be healthy subjects.

Table 4.6: Sample size calculations:

Question	Population	Correlation Coefficient	Sample Size
1st question	All sample	0.12	400
2nd question	T2D patients	0.2	150
3rd question	Adult unaffected relatives of recruited T2D patients.	0.16	250

Although we were aiming to recruit 400 subjects, several factors interfered with our ability to achieve that objective. 362 participants were recruited which represents approximately 90% of the target sample size. 179 were type 2 diabetes patients and 183 were healthy participants. The recruitment pattern with the reasons behind the reduced number of recruited individuals are illustrated in Figure 4.29. Three cessations were due to national or religious holidays. During these seasons, participants were more likely to be engaged in religious and recreational activities which interfered with the recruitment process.

The study was originally aiming to recruit 250 participants with available birth weight record to investigate the association between inbreeding, low birth weight and risk of diabetes. However, since we were not able to retrieve the birth weight of participants, the number of participants needed to answer this question was no longer needed. In addition, several patients only had affected siblings. Families which had no healthy siblings were likely to have a strong genetic risk incurring the development of type 2 diabetes among its members. We thought that by excluding these families, we were neglecting an important part of the population who were

likely to provide a very good insight regarding genetic and familial causes of type 2 diabetes in the target population.

Based on the previous remarks, we made several changes to sample size. We reduced the number of healthy participants to from 250 to 200 which would be enough as presumed earlier to answer the remaining questions. We also increased the number of patients from 150 to 200. The excess 50 patients was devoted to recruit single patients from families with no healthy siblings and multiple type 2 diabetes patient siblings.

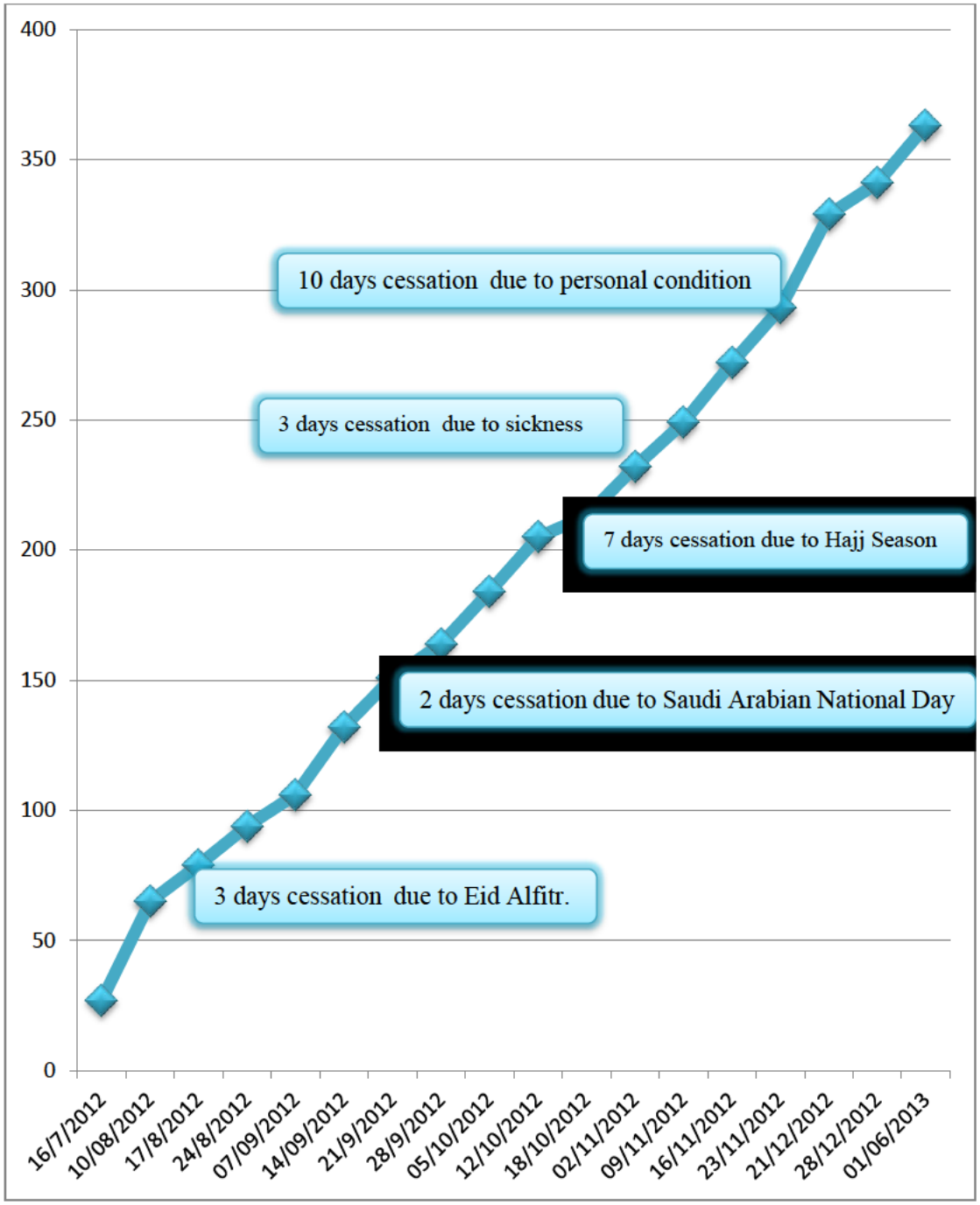


Figure 4.29: Recruitment pattern with cessations explanations.

4.15 Budget

Measuring several variables increased the cost of this study. Most of the study's fund was dedicated to the DNA extraction and genotyping process. In order to get the best available rates, we contacted several companies. The best rates were given by K-bioscience. Additionally, significant cost was also paid to cover shipping collection tools to Saudi Arabia and shipping the collected samples to the UK. Further details are illustrated in Table 4.7. The Saudi Cultural Bureau in the UK funded of budget of this study. Travelling costs to Saudi Arabia and back to the UK were also covered by Saudi Cultural Bureau.

Table 4.7: Estimation of the study cost:

Item	Cost
SK-1S 100 x 1 swab with 5ml tube	£352.8
SGC-50 50 Isohelix Dri-Capsules	£688.8
DNA Extraction Service from Buccal Swaps for 400 samples	£2000
Genotyping 23 SNPs for 400 samples	£3253
Glucose Meter (Optium Xceed)	£46
Glucose strips (4 Packs Optium Xceed strips)	£86
Relijject Alcohol Cleansing Wipes	£9
One Touch UltraSoft Lancets	£24
Small Sharps Bin	£4
Cotton Balls	£5
Portable Scale	£20
Portable Stadiometer (Seca 213)	£70
Printing forms. (Approximate cost for 400 subjects)	£20
Transportation cost (eg. Car Fuel) (approximate cost)	£120
Total	£6689.6

5 Chapter Five: Results

This chapter provides a comprehensive descriptive and inferential analysis of variables collected in this study. In the first section we provide summary statistics relating to the recruitment process. Following the descriptive statistics, inferential statistics are presented in two parts. The first part is restricted to answering the study's research questions and in the second part several exploratory investigations are performed.

5.1 Descriptive Statistics

5.1.1 Introduction:

Several descriptive statistical measures were applied in this study. We calculated means and medians of continuous variables. In addition to using means and medians to assess the normal distribution of continuous variables, we used histograms with normal curves to give a better judgment of distributions. In addition to using standard deviations to assess the variation around means, we also used maximum and minimum values to have an overview of the spread of data.

In addition to giving an overview of the recruitment process results, descriptive analysis was performed for the following variables:

- Ages of all participants and summary of the ages of healthy participants and patients.
- Age at diagnosis of recruited patients.
- Variables related to familial structures such as the number of siblings, the number of affected parents and the siblings in each family.
- Inbreeding coefficients and description of degree of relatedness.
- Variables related to food frequency and physical activity questionnaires.
- Anthropometric variables: BMI, FBG and waist circumference.

The final section of this chapter is dedicated to illustrating descriptive aspects related to genetic analysis. We give a description of genotyping success. Genotypes and alleles frequencies are described in tables and visualized in histograms. Frequencies are reported for patients and healthy subjects.

5.1.2 Recruitment summary

Table 5.1 gives a description of the total number of participants recruited in this study. The total number of participants was 362; of these, 179 participants were type 2 diabetes patients. 74% of type 2 diabetes patients who were approached agreed to participate. Only three healthy participants did not give a buccal swab because they were chewing Khat during the visit, and, due to issues relating to communication, we could not arrange to revisit them. Approximately 84% of recruited healthy participants had their FBG measured. The remaining 16% refused to give glucose measurement due to reasons such as a fear of being diagnosed with type 2 diabetes, a fear of needles and a lack of participants' proper organization caused by Khat chewing.

Table 5.1: Recruitment summary:

Total number of participants	362
Number of Healthy participants	183
Number of families with three participants	51*
Number of families with two participants	80
Number of Single patients	48**
Number of approached patients	240
Number of patients who were not possible to recruit	61***
Number of participants who refused to give buccal swabs	3
Number of participants who refused to give FBG test	29

*One family included 4 participants: a monozygotic twins where one twin was type 2 diabetes patient and the other was healthy and additional two healthy siblings.

**Patients whom all or most siblings were affected with type 2 diabetes and no healthy siblings were available for recruitment.

**Reasons interfered with recruitment of patients:

33 patients refused to participate, 24 patients had no available siblings or none of their siblings were affected with type 2 diabetes. Three patients agreed to participate but their healthy sibling refused to participate. One patient was not Saudi.

Table 5.2 illustrates number of participants recruited in each PHC. 69 participants were recruited from a single PHC. This is the largest number of participants recruited from a single PHC. Although there are other PHCs similar to AlZabiah PHC in size, recruitment in ALZabiah centre was high due to cooperation of the PHC staff and due to large number of registered type 2 diabetes patients. Additionally, this centre covered several urban and rural areas.

The smallest number of participants recruited from a single PHC was 9. This PHC was smaller than other PHCs and number of registered type 2 diabetes patients was smaller. Sabia and ALMabouj PHCs were two of the largest centres in addition to ALZabiah. However recruitments percentage from those two centres were 6% from Sabia PHC and 3.5% from ALMabouj PHC due to factors related to cooperation of the staff and participants.

Table 5.2 Number of recruited participants per PHC:

PHC	Number of participants	Percentage
AlBaher	29	8%
AlFeqra	24	7%
AlGhara	31	9%
AlAreesh	9	2.5%
AlZabiah	69	18%
Sabia	21	6%
Sambah	25	7%
ALBedie	27	7.5%
ALRayian	13	3.5%
Mehliah	22	6%
AlMabouj	13	3.5%
AlQa'ariah	38	10.5%
AlHusaini	41	11%
Total	362	100%

5.1.3 Age distribution

Table 2 shows the ages of patients and healthy participants. The mean age of patients was 45 years compared to 36 years in the healthy subjects group. The youngest patient was 23 years old and the oldest one was 61 years old. The oldest healthy participant was 71 years old. There were 40 recruited families where patients were younger than the healthy participants. In two families, we recruited two pairs of twins where one was an identical pair and the other was non-identical. In both twin pairs, one twin was a patient and the other one was healthy. Figure 1 depicts the distribution of ages in both groups.

Table 5.3: Age of patients and healthy participants:

Age	Patients (n=179)	Healthy Subjects (n=183)
Mean	45.84	36.40
Median	47.00	35.00
Std. Deviation	8.901	10.837
Minimum	23	18
Maximum	61	71

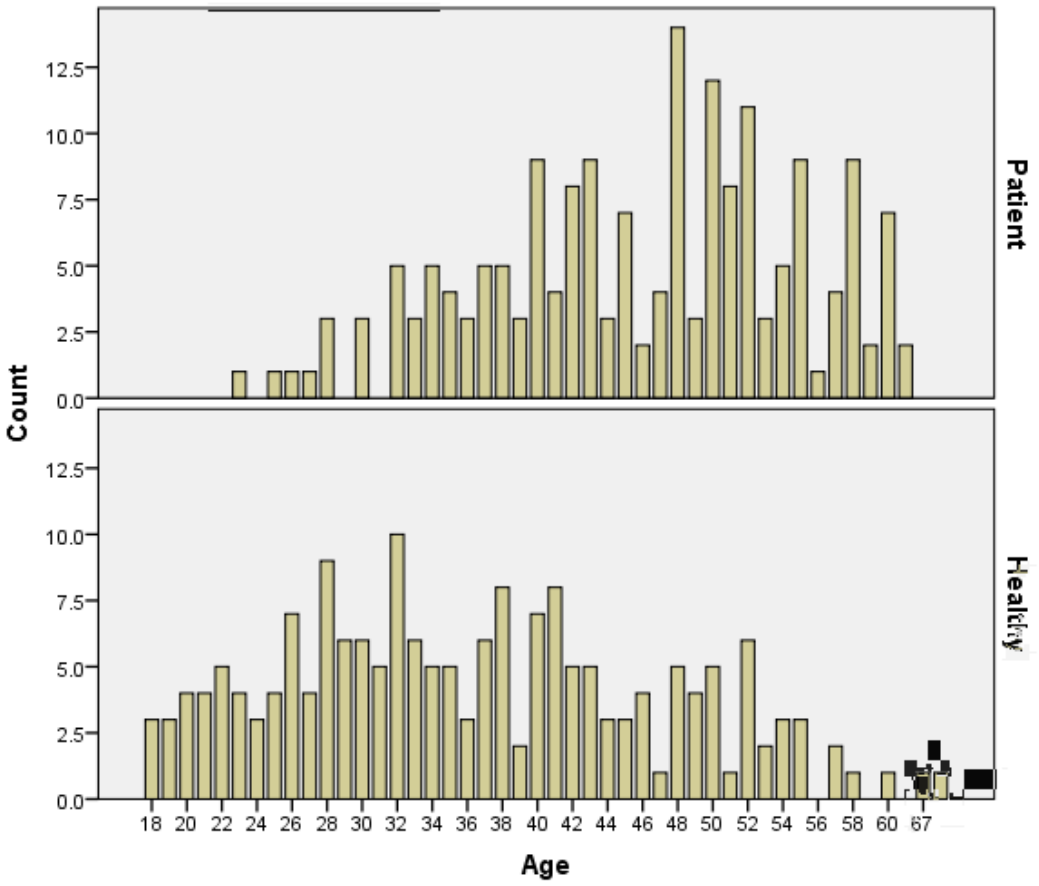


Figure 5.1: Comparison of age distribution in cases and healthy subjects.

5.1.4 Age at Diagnosis with Type 2 Diabetes

Summary statistics of age at diagnosis are described in table 5.4. The mean age at diagnosis in the sample was 38.05 years (SD: 8.9 years). Age at diagnosis varied between 20 and 57 years. Figure 5.2 depicts an approximately normal distribution of age at diagnosis.

Table 5.4: Statistics summary of age at diagnosis with type 2 diabetes:

Age at Diagnosis	
Mean	38.05
Median	38.00
Std. Deviation	8.883
Minimum	20
Maximum	57

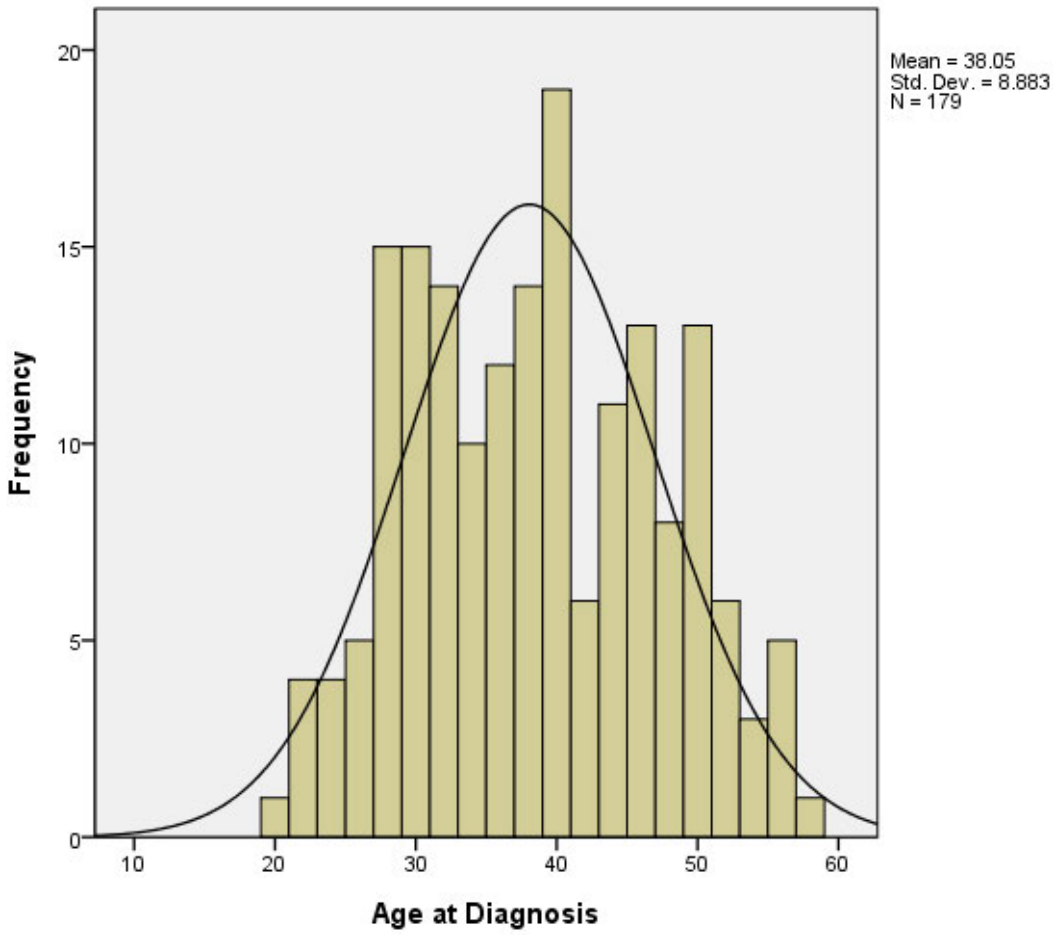


Figure 5.2: Distribution of age at diagnosis with type 2 diabetes.

5.1.5 Familial patterns

Table 5.5 gives an overview of the number of siblings per family in the target population. As expected, the number of siblings per family is large in comparison to western societies. The mean number of siblings per family in the target population is 7.03 (SD= 3.04 siblings). Families with only one offspring were not recruited. Six families only had two siblings while approximately 23% of families reported having 10 siblings or more. Figure 5.3 illustrates the distribution of the number of siblings per family. The distribution appears to be approximately normal in the collected sample.

Table 5.5: Number of siblings per family:

Number of Siblings per family	Frequency	Percentage
2	6	3.4%
3	13	7.4%
4	16	9.1%
5	18	10.3%
6	29	16.6%
7	17	9.7%
8	15	8.6%
9	21	12.0%
10	14	8.0%
11	15	8.6%
12	7	4.0%
13	1	0.6%
14	2	1.1%
18	1	0.6%
Total	175	100%

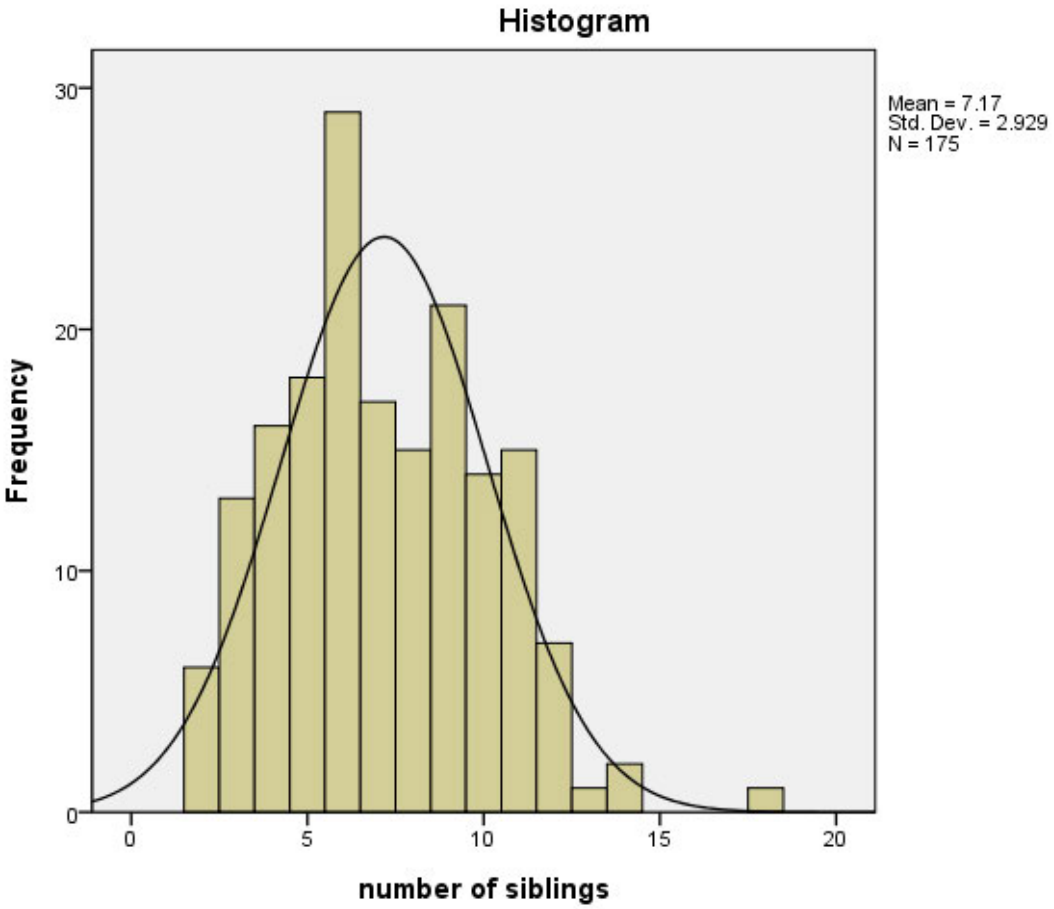


Figure 5.3: Distribution of number of siblings per family.

In addition to the number of siblings per family, the constructed pedigrees reported several other variables. The total number of male siblings reported in the constructed pedigrees was 745 while the number of female siblings reported was 514. In the constructed pedigrees, 316 male siblings were affected with type 2 diabetes which represented 42% of reported male siblings. On the other hand, only 14% of reported females in pedigrees had type 2 diabetes. However, the low percentage of reported female siblings with type 2 diabetes might have been due to the fact that only the male siblings of each family were interviewed.

Table 5.6 illustrates the number of affected male and female siblings reported in constructed pedigrees. The highest number of siblings affected in the same family was 7 patients. 59% of the constructed pedigrees were revealed as having 2 or more siblings affected with type 2 diabetes.

Table 5.6: Number of siblings affected with type 2 diabetes in each family:

Number of siblings affected in each family	Frequency	Percent
1	74	41.3%
2	52	29.1%
3	25	14.0%
4	11	6.1%
5	11	6.1%
6	4	2.2%
7	2	1.1%
Total	179	100%

Constructed pedigrees reported different patterns which were related to a parental history of type 2 diabetes. The number of families reported a paternal history of type 2 diabetes was 80 (45%). 71 families reported a maternal history of type 2 diabetes (40%). Only 34 families reported having both a maternal and a paternal history of type 2 diabetes (19% of constructed pedigrees). 57 families reported having at least one grandparent with a history of type 2 diabetes.

5.1.6 Inbreeding Coefficients

Table 5.7 describes calculated inbreeding coefficients of recruited participants. 63% of constructed pedigrees revealed a history of consanguinity; in 37% of constructed pedigrees, participants reported that their parents were from different tribes. Constructed pedigrees revealed different and complex mating systems between individuals from the same tribe or the same subtribes, and parents who were third, second and first cousins. On several occasions, different combinations of mating patterns were observed in the same family.

Table 5.7: Distribution of inbreeding coefficients:

Inbreeding coefficient	Frequency	Percentage
0%	66	36.9%
0.001%	8	4.5%
0.002%	12	6.7%
0.004%	9	5%
0.024% - 0.044%.	8	4.5%
0.084% - 0.196%	4	2.25%
0.39% - 1.5%.	14	7.8%
1.56% - 6.24%.	24	13.5%
6.25% - 12.5%	32	18%

Using our method of calculating inbreeding coefficients revealed 57 distinct inbreeding coefficients. The calculated inbreeding coefficient of 8 (4.5%) families was 0.001%. In these families, the only relationship found between the parents was that one pair of grandparents either from the maternal or paternal side was from different subtribes but within the same tribe. Another 12 (6.7%) families had inbreeding coefficient of 0.002% where three grandparents shared the same tribe but did not share a subtribe. Similarly, 9 (5%) families had inbreeding coefficient of 0.004% where all grandparents shared the same tribe.

8 (4.5%) families had calculated inbreeding coefficients ranging between 0.024% and 0.044%. In these families, grandparents belonged to a shared subtribe in addition to belonging to the same tribe. Families which were given inbreeding coefficients of 0.044% had three grandparents who shared the same subtribe and all grandparents shared the same tribe. Four (2.25%) families had inbreeding coefficients varying between 0.084% and 0.196% where all grandparents belonged to the same subtribe.

14 families had a history of mating between third degree cousins (7.8% of families). Calculated inbreeding coefficients for these families varied between 0.39% and 1.5%. Variation between inbreeding coefficients of these families was augmented by a variation in the mating of grandparents form shared tribes or subtribes. Similarly, 24 (13.5% of families) inbreeding coefficients were calculated for families where parents were second cousins and coefficients varied between 1.56% and 6.24%. 32 families reported a history of mating between first degree cousins (18% of recruited families). Inbreeding coefficients of families of first degree consanguinity varied between 6.25% and 12.5% where the latter figure represents inbreeding coefficient of mating between parents who were double first cousins.

Studies which have measured the prevalence of consanguinity in Saudi Arabia have provided different results. The range of consanguinity varied between 52% and 67% (250-252, 275), which is similar to our estimated rate of consanguinity. However, these studies reported that the frequency of marriages between first cousins was high compared to other marriages between distantly related subjects. In our study, only 18% of recruited families had a mating history of parents who were first degree cousins. This largely explains why the inbreeding coefficients calculated by those studies (0.02 to 0.03) were high compared to the average inbreeding coefficient of our sample (0.016). However, we must consider that their methods of categorizing mating as between first cousins, second cousins and distant cousins is quite different from our pedigree-based method.

5.1.7 Questionnaires analysis

5.1.7.1 Food Frequency Questionnaire

Summary statistics of average calorific intake are illustrated in table 5.8. The mean calorific intake in this sample was 3816 (SD: 1151.49) calories per day. According to the British Nutrition Foundation, the recommended average daily calorific intake for men is about 2500 calories (276). Therefore, there is a tendency in our sample to exceed the recommended daily calorific intake. The distribution of average calorific intake per day in this sample appears to be approximately normally distributed (figure 5.4). However, there were several outliers with an average daily consumption exceeding 7000 calories per day.

Table 5.8: Average Calorific Intake of patients and healthy subjects (n=362):

Average Calorific Intake	
Mean	3816.7557
Median	3639.2180
Std. Deviation	1151.49314
Variance	1325936.451
Range	6921.62
Minimum	1706.42
Maximum	8628.04

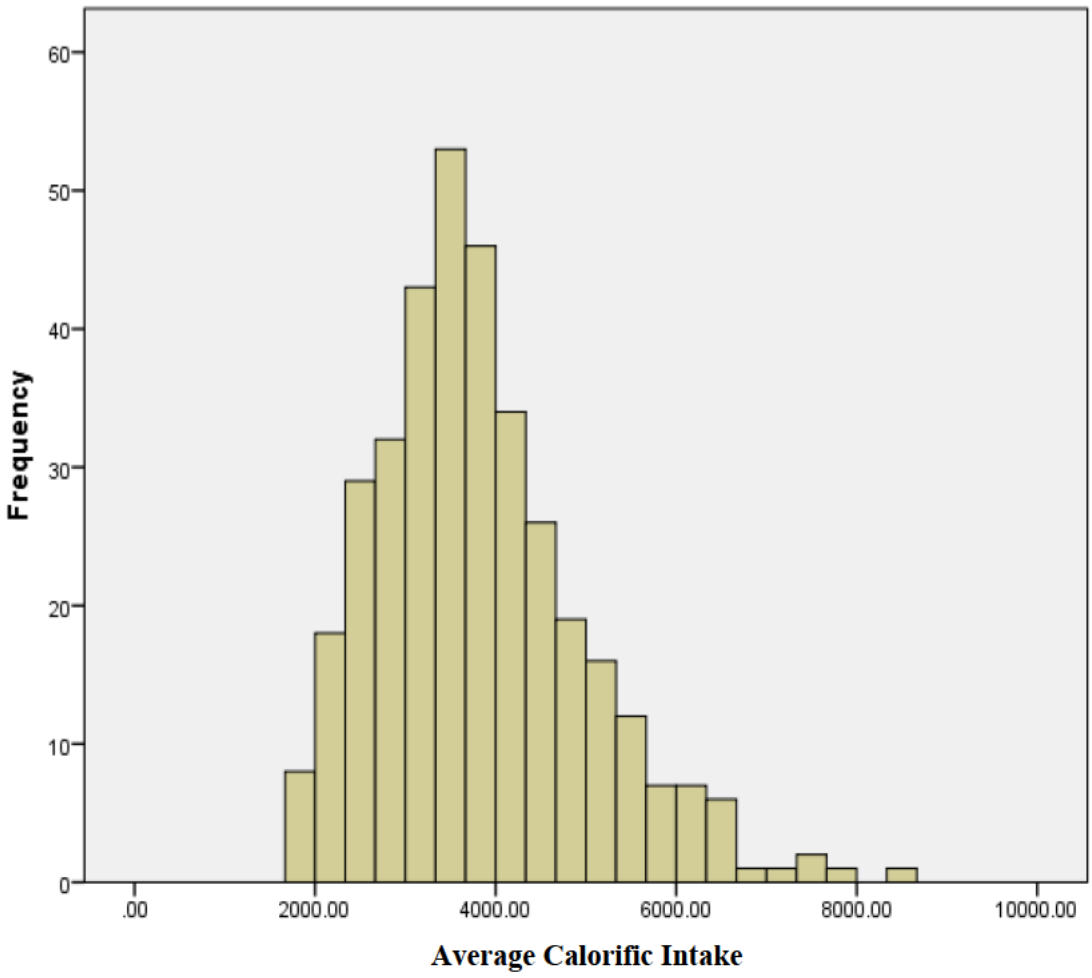


Figure 5.4: Distribution of Average Calorific Intake of the Sample. (n=362).

A comparison between the average calorific intake of patients and healthy subjects is depicted in table 5.9. The mean intake in patients is apparently lower than those of healthy subjects. Mean average daily calorific intake of healthy subjects was on average 275 calories more than patients. Furthermore, the distribution of both patient and healthy samples appear to be in normal distribution as depicted in figure 5.5.

Table 5.9: Comparison between Average calorific intake of patients and healthy siblings:

Average Calorific Intake	Patients (n=179)	Healthy Subjects (n=183)
Mean	3677.7079	3952.7642
Median	3508.2430	3805.6720
Std. Deviation	1128.54	1160.50
Minimum	1706.42	1895.59
Maximum	8628.04	7958.94
Sum	658309.71	723355.85

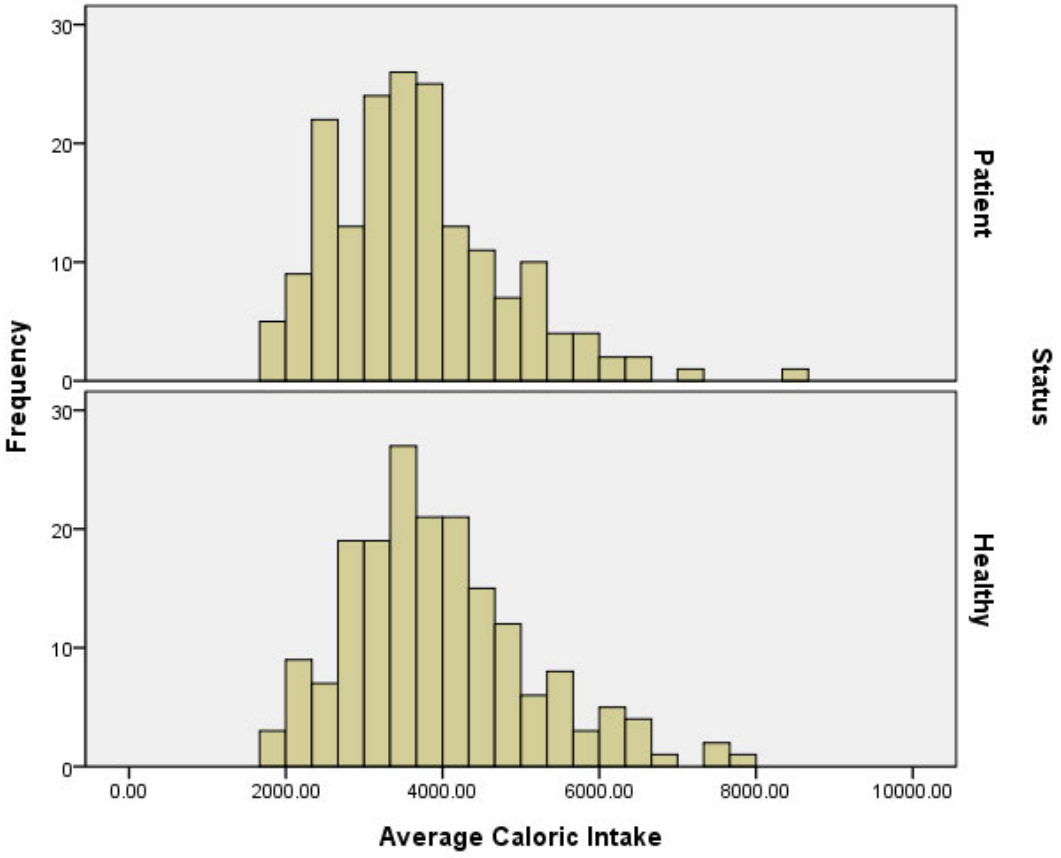


Figure 5.5: Comparison of distribution of average Daily Calorific Intake between patients and healthy subjects.

A study was conducted in Saudi Arabia to measure the effect of the diet consumed by adolescents on their nutritional status **(81)**. This study utilized 24-hours diet recall and 17-items food frequency questionnaires. This study calculated the average daily caloric intake in addition to other nutrients. However, the daily caloric intake estimated by this study is smaller than our estimate (2025 Kcal compared to our study 3816 Kcal). This large difference could be caused by the age difference between the populations in each study. However, no similar studies were found where the average caloric intake was calculated in adult Saudi subjects.

Table 5.10 illustrates a comparison between different studies which have calculated an average daily caloric intake using food frequency questionnaires. All reported values were related to male subjects recruited in those studies to make it comparable to our sample which did not include females. The average calorific intake in our study is relatively high compared to other studies. Although there is a chance of an overestimation of calorific intake in our study, there could be several factors contributing to the higher calorific intake in our sample.

As we illustrated earlier, the prevalence of obesity and type 2 diabetes in the Saudi Arabian populations is high compared to other populations. A contribution to this increased prevalence could be the high calorific intake as illustrated in our sample. However, the food list length in our study was relatively high compared to the remaining studies, which could have contributed to the recording of higher average caloric intakes.

Our sample included a large proportion of young participants, and, as shown in our analysis, younger subjects are likely to have a high calorific intake compared to older adults. Finally, there could be racial variations leading to different consumption of different food types with different calorific values as in our sample there was a tendency for subjects of our study to report frequent intake of high calorie food items.

Table 5.10: Description of studies used food frequency questionnaires to calculate average daily calorific intake in male subjects:

Study Reference	**Population**	**Sample size**	**Food list**	**age**	**calorific intake (SD)**
(267)	Korean	33	103	42-55	1953.3 (473.5)
(268)	UK	457	127	39–61	2292 (716)
(269)	Australian	37	129	25–75	2269 (559)
(277)	Greek	5633	150	25-65+	2740 (870)
Current study	Saudi	362	157	18-71	3816 (1151)

5.1.7.2 Contributors to Calorific Values

Certain food items appear to be popular in the selected sample. These food items contributed about 40% of the total calorific value of the whole sample. These items are listed in table 5.11. Differences between patients and healthy siblings are also illustrated in table 5.11. However, there is similarity of most favoured food items between the two groups.

In both groups, the most popular food items were Marsa, Chicken Kabsa and bread. These items were consumed several times on a weekly basis in the recruited sample. Marsa has a high calorific value because it contains wheat, banana, butter and honey in its ingredients. The high calorific value of the chicken Kabsa is principally contributed by chicken, rice and cooking oil. In addition to the items listed in table 5.11, other items appear to be important contributors but to a lesser extent. These items include, but are not limited to, lamb stew, fried fish, fried chicken, dates, biscuits, fries, fried egg and boiled egg sandwiches.

Table 5.11: Top Contributors to Calorific Values:

Patients		**Healthy Subjects**	
Food Item	Contribution Percentage	Food Item	Contribution Percentage
Marsa	10%	Marsa	10%
Chicken Kabsa	9%	Chicken Kabsa	10%
Bread	8%	Bread	7%
Lamb Kabsa	5%	Rice	4%
Grilled Chicken	4%	Grilled Chicken	4%
Rice	4%	Lamb Kabsa	3%
Fish Stew	3%	Cakes	3%
Total	43%	Total	41%

5.1.7.3 Physical Activity Questionnaire Analysis

The physical activity questionnaire was analysed by calculating the number of hours spent performing sedentary, moderate or vigorous activities. Table 5.12 shows summary statistics of physical activities reported in the total sample (n=362). An average of 88 hours (SD: 24 hours) per week was spent performing sedentary activities. The distributions of hours spent performing different forms of activities per week are depicted in 5.6.

The distribution of number of hours spent performing sedentary activities appears to be approximately normal. However, the distribution of the number of hours spent performing moderate and vigorous activities per week are positively skewed. The medians of moderate and vigorous activities indicate that, on average, about 11 hours (SD: 11 hours) are spent performing moderate activities per week and almost four minutes (SD: 2.6 hours) are spent performing vigorous activity per week.

Physical activity levels in our sample were markedly deviated toward sedentary activities. Higher levels of moderate and vigorous activities were reported more in the healthy group compared to patients.

Table 5.12: Summary Statistics of physical activities:

Physical Activity Category	Weekly Sedentary Hours	Weekly Moderate Activity Hours	Weekly Vigorous Activity Hours
Mean	87.89	13.15	0.88
Median	85.10	10.75	0.06
Std. Deviation	23.74	10.60	2.64
Minimum	41	0	0
Maximum	165	69.3	32.08
Sum	31819.16	4762.50	318.68

5.1.7.4 Comparison of Physical Activity between Patients and Healthy Siblings

Referring to table 5.13, we can observe that the mean number of hours spent performing sedentary activities per week was 5.5 hours more in healthy subjects in comparison to patients. However, the medians of the number of hours spent on moderate or vigorous activities per week were higher in the healthy group. Figure 8 compares the distribution of physical activity between patients and healthy subjects.

Table 5.13: Comparison between level of activity between patients and healthy subjects:

	Sedentary activities*		**Moderate Activities***		**Vigorous Activities***	
	Patients (n=179)	**Healthy (n=183)**	**Patients (n=179)**	**Healthy (n=183)**	**Patients (n=179)**	**Healthy (n=183)**
Mean	85.9	89.84	11.44	14.8	0.33	1.4
Median	81.25	89.31	9.50	12.71	0.055	0.12
Std. Deviation	25.32	21.98	9.42	11.42	1.15	3.46
Variance	641.32	483.24	88.78	130.63	1.34	12.01
Minimum	41.26	41.00	0	0.50	0	0
Maximum	165.00	147.50	57.00	69.30	12.30	32.08
Sum	15377	16442	2049	2713	60.45	258.238

*Number of hours per week.

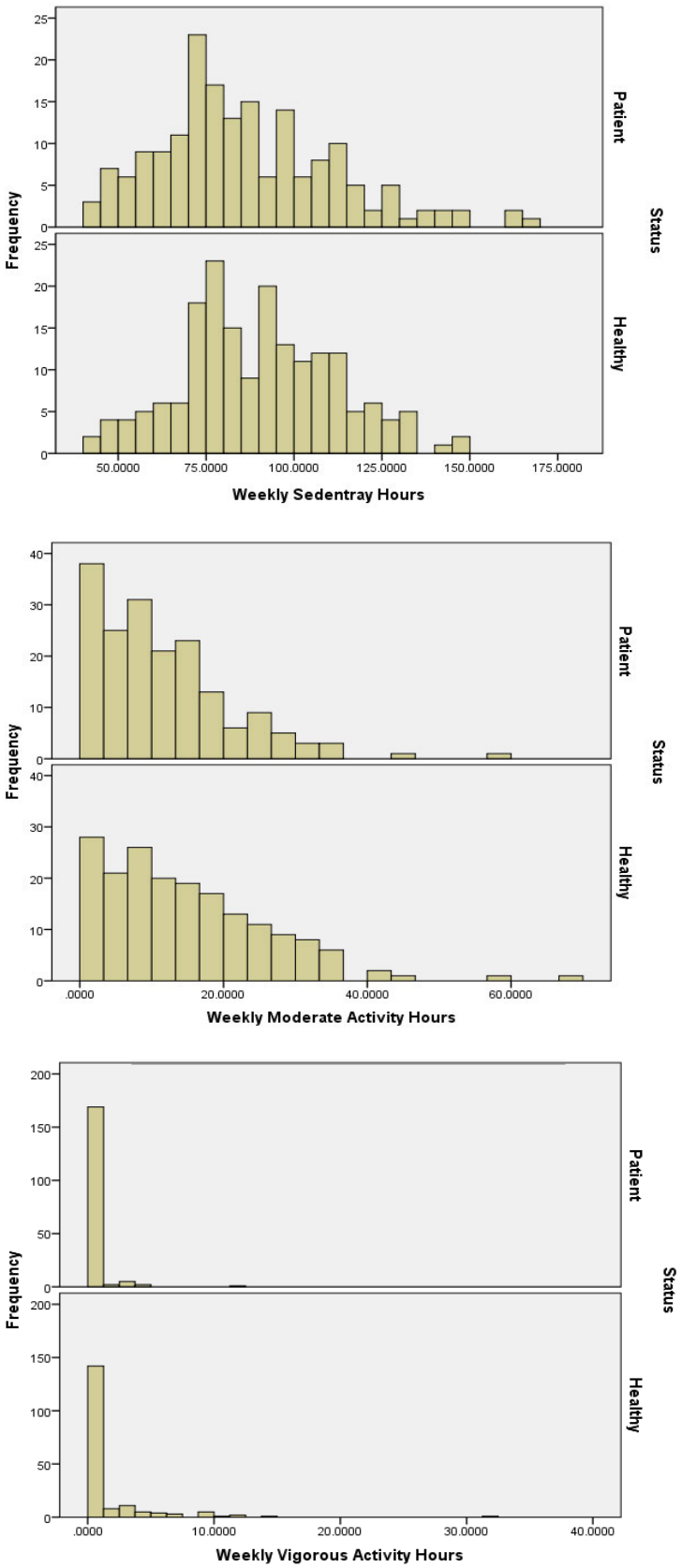


Figure 5.6: distribution of physical activity in patient and healthy groups.

5.1.7.5 Physical activity behaviour

The number of hours per week spent performing sedentary activities varied between 41 and 165 hours per week. Sedentary activities included sleeping, watching TV or other devices, driving and sitting during work. There was a higher tendency to drive as a means of transportation compared to walking among recruited participants. Only 7% (n= 26) of the participants revealed walking to their work as a means of transportation. The vast majority of participants used cars as the usual mode of transportation.

Certain physical activities appeared to be more frequently reported in our sample. Table 5.14 illustrates the frequency of performing these activities in patient and healthy groups. The frequency of performing these moderate activities was almost similar between the two groups except caring for children which was higher amongst the patient group and gardening activities which appear to be reported more in the healthy group.

Table 5.14: Most frequently reported moderate activities:

Activity	**Number of patients (Percentage) (n=179)**	**Number of healthy subjects (Percentage) (n=183)**
Shopping	163 (91%)	161 (88.5%)
Walking	79 (44%)	69 (38%)
Caring for Children	68 (38%)	63 (25%)
Gardening	23 (13%)	45 (25%)
Home maintenance	8 (4.5%)	5 (3%)
Fishing	6 (3.3%)	9 (5%)

Table 5.15 illustrates most reported vigorous activities. There was a clear tendency for healthy subjects to report performing vigorous activities in comparison to patients. The most common sport reported by participants was football and it was reported more amongst healthy subjects (33%).

Table 5.15: Most Frequently Reported Vigorous Activities:

Activity	**Number of patients (Percentage) (n=179)**	**Number of healthy subjects (Percentage) (n=183)**
Football	10 (7%)	33 (18.5%)
Jogging	3 (1.6%)	17 (9%)
Conditional Exercise	4 (2%)	7 (4%)
Weight Lifting	3 (1.6%)	3 (1.6%)

Table 5.16 gives an illustration of the occupations reported by participants. Most of recruited participants were teachers or soldiers. 27% of patients were retired at the time of recruitment. Most of these occupations are more likely to exhibit sedentary or moderate level of activity. Only a minority of jobs were reported which might have included vigorous physical activities such as mechanics or occupations which include carrying heavy machinery.

Table 5.16: Description of most reported occupations:

Patients (n=179)		**Healthy Subjects (n=183)**	
Occupation	Frequency (Percentage)	Occupation	Frequency (Percentage)
Teachers	49 (27%)	Soldiers	48 (26%)
Retired	48 (27%)	Teachers	42 (23%)
Soldiers	39 (22%)	Students	24 (13%)
Clerical Jobs	24 (13%)	Clerical Jobs	18 (10%)
Others	13 (7%)	Others	22 (12%)
Unemployed	6 (4%)	Retired	15 (8%)
		Unemployed	14 (8%)

5.1.8 Anthropometric Variables

5.1.8.1 Fasting Blood Glucose

Descriptive statistics of FBG are stated in table 5.17. We were able to collect 154 values of FBG which represents 84% of the total sample of healthy subjects. The mean FBG was 95 mg/dl (SD: 23.89 mg/dl). The distribution of FBG appears approximately to be normal as shown in figure 5.7. About 90% of FBG values were within the normal range (60 – 126 mg/dl). Four readings were below the normal range of FBG and 12 values were above the normal range. The minimum value reported was 41 mg/dl and the maximum one reported was 248 mg/dl.

Participants who were not diagnosed with type 2 diabetes when they were recruited were presumed to be healthy. However, since type 2 diabetes can be an asymptomatic illness, detecting several healthy participants with high levels of FBG was expected.Twenty subjects had FBG levels varying between 6.1 mmol/l (110 mg/dL) and 6.9 mmol/l (125 mg/dL) and were very likely to be diagnosed as having impaired fasting glucose once consulting their health professionals. Twelve subjects had FBG above 7mmol/l (126 mg/dl) upon blood glucose measurement. Those subjects are very likely to be diagnosed with type 2 diabetes once they consult their health professionals.

Subjects suspected with impaired fasting blood glucose and FBG above 126 mg/dl were classified as healthy subjects in this study, given that they were not adhering to any type 2 diabetes management plans or change of lifestyle. Engaging in lifestyle change or taking medications is likely to influence BMI, FBG and waist circumference. Therefore, those participants remained in the control arm of the study to investigate aspects related to these continuous traits, since they were blind to their possible disease status and were not engaged in any change in lifestyle or taking medication.

Table 5.17: Description of Fasting Blood Glucose:

Number of cases	Valid	154
	Missing	29
Mean		95.4784
Median		93.00
Mode		84.00
Std. Deviation		23.89
Minimum		41.00
Maximum		248.00

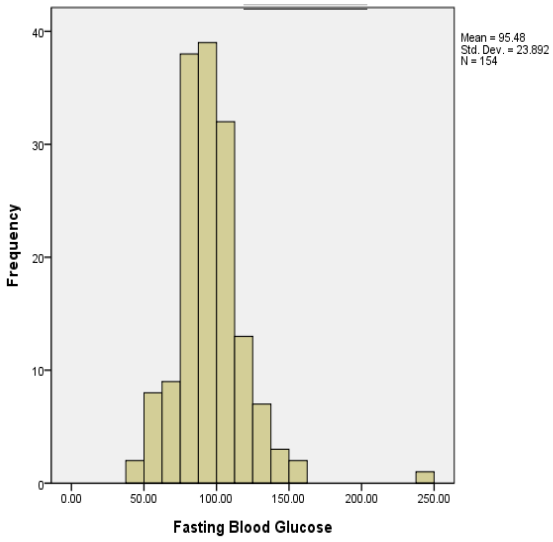


Figure 5.7: Distribution of Fasting Blood Glucose

5.1.8.2 Body Mass Index

Descriptive statistics of BMI are illustrated in table 5.18. The mean BMI was 28.15 (SD: 6.8). The distribution of BMIs is depicted in figure 5.8 which seems to be a normal distribution with a few extreme values slightly skewing it.

Table 5.18: Descriptive Statistics of Body Mass Index:

BMI (n=183)	
Mean	28.15
Median	27.40
Std. Deviation	6.858
Minimum	15.10
Maximum	67.00

By observing table 5.19 we can detect a prevalence of overweight and obese individuals in our sample. 4 participants were reported to have below normal BMI and only 30% of the healthy subjects had normal BMI. The percentage of overweight and obese individuals in this sample was about 68% and 8 participants were morbidly obese.

Table 5.19: Interpretation of BMIs:

BMI Category	**Frequency (Percentage) (n=183)**
Underweight (<18)	4 (2.2%)
Normal Weight (18 –25)	55 (30.1%)
Overweight (25 – 30)	69 (37.7%)
Obese (30 - 40)	47 (25.6%)
Morbidly Obese (> 40)	8 (4.4)

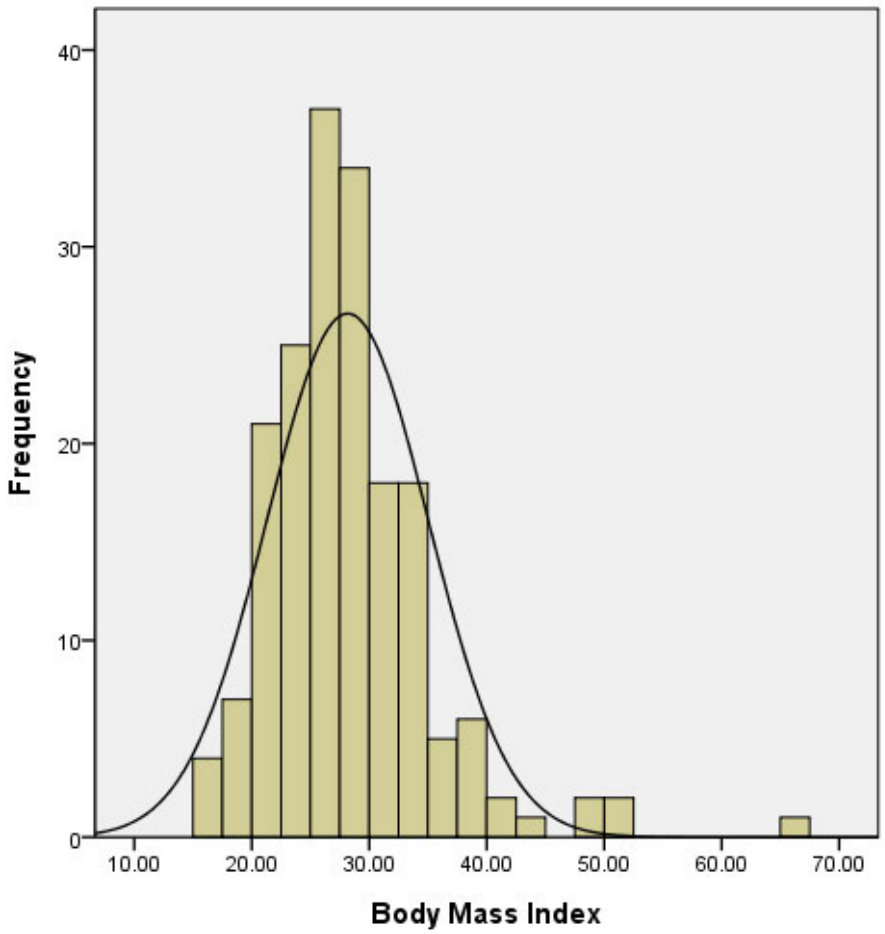


Figure 5.8: Distribution of BMI in healthy.

Most previously published studies concerned with the measurement of the prevalence of overweight and obese healthy subjects in Saudi Arabia reported overall rates among children. Only one study reported the rates of overweight and obese subjects in a population which included adults. This study was published in 2007 and reported that 30.7% of male subjects were overweight and 23.6% of male participants were obese **(78)**. The rates provided by this study are smaller than rates provided in our study.

In the current study, the proportion of healthy participants who were overweight was 37.7%. Additionally, 30% of healthy participants were obese. Upon comparing the prevalence of obesity in our sample to the rates reported by other studies we can observe a tendency of our sample to have had higher body weights. This could be largely explained by the fact that all of our healthy participants had a family history of diabetes and consequently, had the genetic and environmental risk of developing obesity and perhaps type 2 diabetes.

5.1.8.3 Waist Circumference

Table 5.20 describes summary statistics of the waist circumference of healthy participants. Mean waist circumference was 96.22 cm (SD: 14.96 cm). Distribution of waist circumferences is depicted in figure 5.9 and appears to be approximately normal.

Table 5.20: Summary statistics of waist circumferences:

Waist Circumference (n=182)	
Mean	96.22
Median	95.00
Std. Deviation	14.96
Minimum	60.00
Maximum	151.00

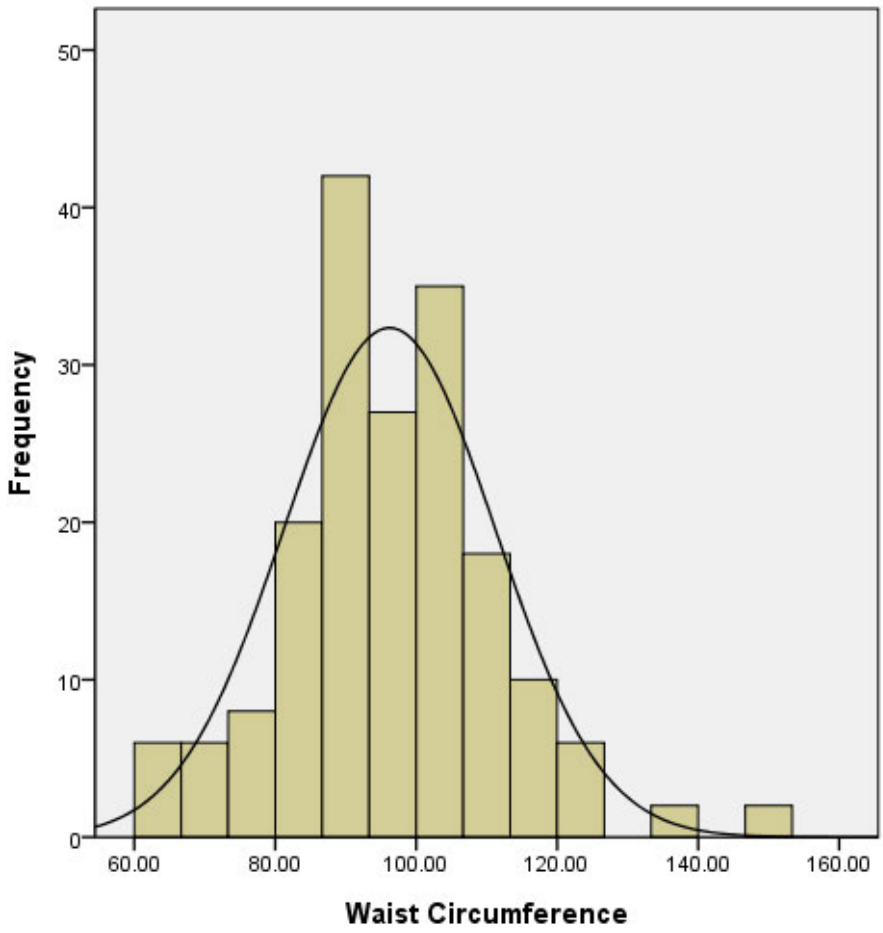


Figure 5.9: Distribution of waist circumferences.

5.1.9 Descriptive analysis of genetic data

5.1.9.1 Genotyping Success

Out of 362 participants, 3 participants refused to give buccal swabs during data collection. Additionally, 5 collected samples were not reported by the genotyping agent and were likely to have been lost during DNA extraction and analysis. Out of the remaining 354 samples, undetected signals of each SNP varied between 15 and 30. The proportion of successfully genotyped SNPs varied between 91% and 95%. Further details are illustrated in table 5.21.

Although we made an effort to avoid taking swabs of subjects who were actively chewing Khat or other materials, we collected three samples which were visibly contaminated with Khat, chewing tobacco and food. However samples of these participants were successfully genotyped. Additionally, other subjects for whom buccal swab collection was rescheduled due to Khat chewing during home visits yielded valid samples. However, nothing peculiar was detected upon reviewing the data of subjects who did not provide good DNA samples and there was no report of Khat chewing during their data collection sessions. We could conclude that Khat chewing is unlikely to have anti-PCR qualities and it is unlikely to interfere with DNA extraction and genotyping.

Upon reviewing SNPs signal density charts provided by the genotyping company, we observed variations between call rates of each used plate. Four plates were used during genotyping. Figure 5.10 illustrates an example of SNPs signal density.

It can be observed in figure 5.10 how the genotype of each subjects was decided. Cases with calls clustering toward one axis of the chart were determined to be homozygote genotypes. Calls which clustered in the middle were presumed to be heterozygote genotypes. However, if certain calls were not clustering toward any group, they were then called undetermined and are coloured in pink in figure 5.10. Additionally, DNA samples which had a bad quality are depicted in yellow as shown in wells diagram in the bottom of figure 5.10.

Table 5.21: Detected Signals of each SNP:

No.	SNP	Detected	Undetected	Call Rate
1	rs10946398	334	20	94%
2	rs10811661	338	16	95%
3	rs1111875	332	22	93%
4	rs13266634	330	24	93%
5	rs7754840	339	15	95%
6	rs560887	335	19	94%
7	rs2191349	328	26	92%
8	rs10830963	328	26	92%
9	rs7034200	329	25	92%
10	rs174550	324	30	91%
11	rs11708067	333	21	94%
12	rs4607517	330	24	93%
13	rs7903146	334	20	94%
14	rs1799884	339	15	95%
15	rs5219	335	19	94%
16	rs7756992	336	18	94%
17	rs780094	333	21	94%
18	rs2237897	336	18	94%
19	rs2237895	332	22	93%
20	rs1387153	330	24	93%
21	rs231362	327	27	92%
22	rs7957197	334	20	94%
23	**rs**757210	327	27	92%

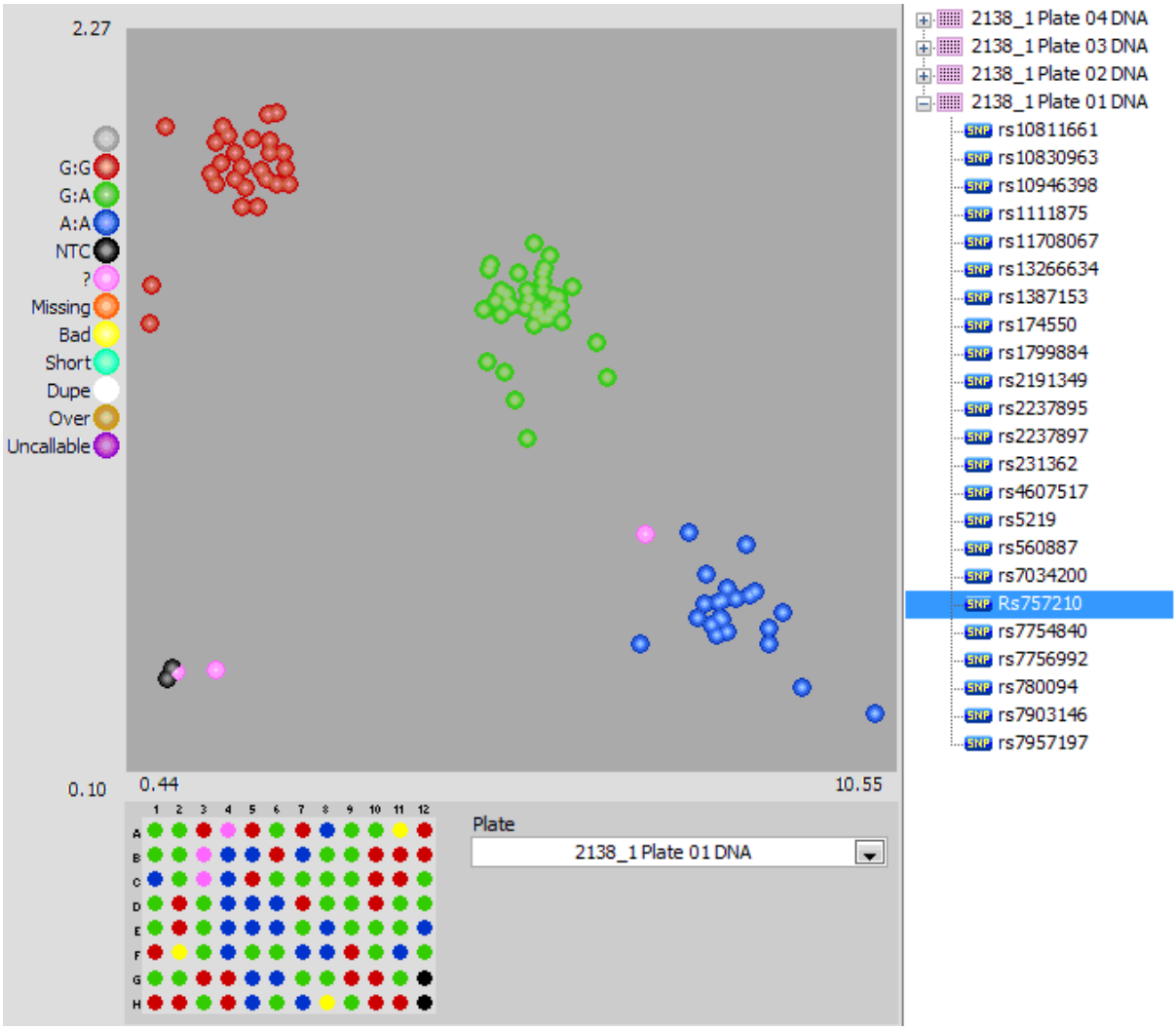


Figure 5.10: SNP rs757210 signal density detected in plate 1.

In table 5.22, we illustrate the number of undetected genotypes of each SNP in each plate. In this table we only included the number of calls which were not given specific genotypes but not DNA samples which were described as bad and failed genotyping in all SNPs. By observing the table, we can notice a difference in the number of undetermined calls according to used plate.

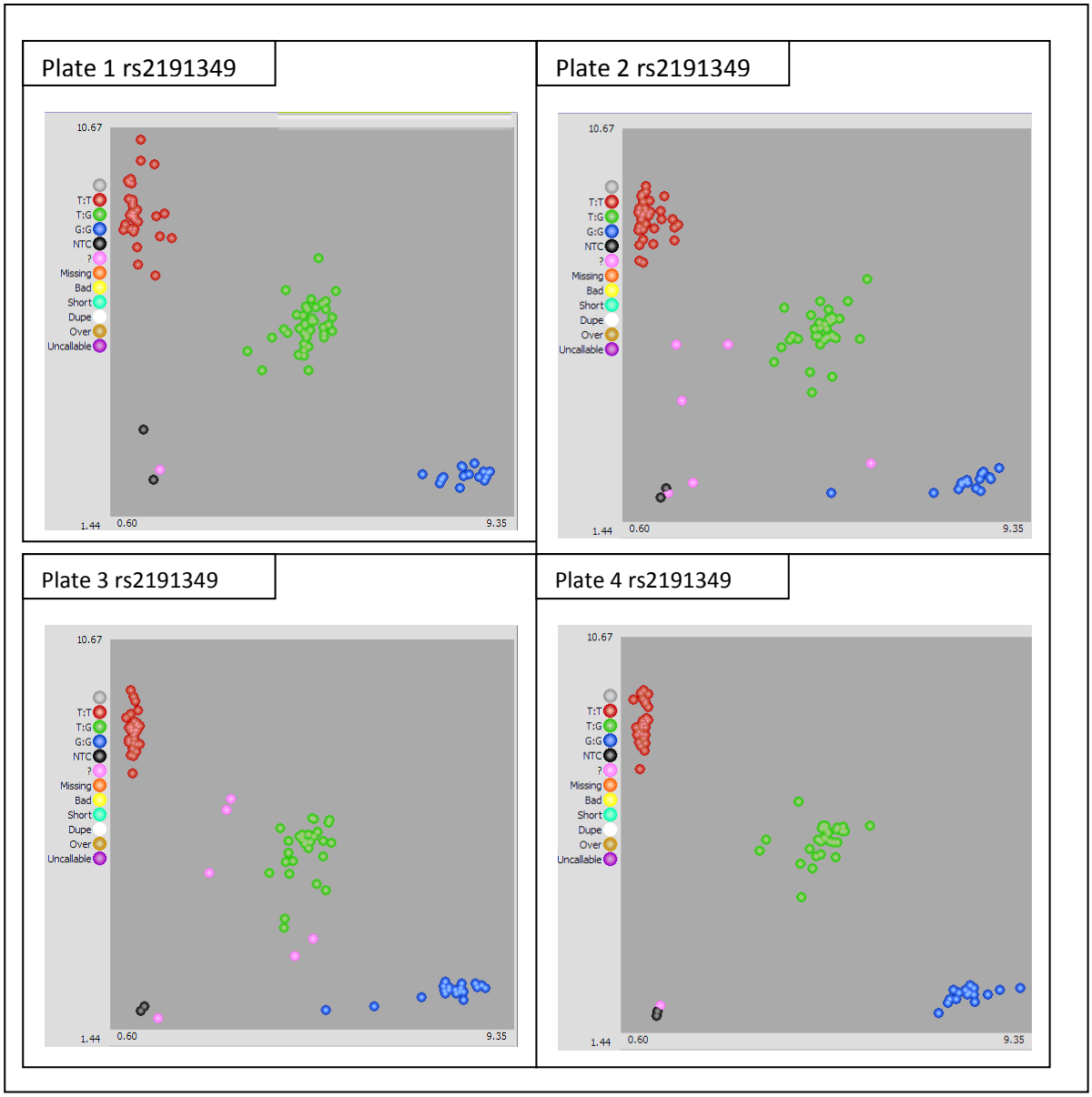
Table 5.22: Frequency of undetected calls of each SNP according to the plate used (number of samples):

No.	SNP	Plate 1 (96)	Plate 2 (96)	Plate 3 (96)	Plate 4 (81)
1	rs10946398	1	3	3	1
2	rs10811661	0	2	1	0
3	rs1111875	1	2	2	3
4	rs13266634	0	5	5	3
5	rs7754840	0	0	1	0
6	rs560887	2	2	3	0
7	rs2191349	1	6	6	1
8	rs10830963	1	4	4	3
9	rs7034200	4	5	2	2
10	rs174550	1	6	6	2
11	rs11708067	1	3	3	1
12	rs4607517	4	3	4	0
13	rs7903146	2	3	1	1
14	rs1799884	1	1	2	0
15	rs5219	1	1	2	2
16	rs7756992	0	2	1	1
17	rs780094	0	3	3	3
18	rs2237897	0	1	2	3
19	rs2237895	1	4	2	1
20	rs1387153	2	2	5	1
21	rs231362	1	6	4	2
22	rs7957197	1	4	1	1
23	**r**s757210	1	5	5	3
	Total (Proportion)	26 (1.2%)	73 (3.3%)	68 (3.1%)	34 (1.8%)

According to the genotyping process, PCR can vary according to the condition which was applied to each plate. The performance of PCR and the number of calls detected can therefore vary according to the plates. We observed that the rate of undetected genotypes was higher in plates two and three compared to plates 1 and 2. Box 2 is an additional example of the variation of call rate according to the plate used.

Box 5.1 illustrates signal density graph of SNP rs2191349 according to the plate used. We can notice that for the same SNP only 1 case was not detected in plates 1 and 4 compared to 6 undetected cases in plates 2 and 3. This might indicate that the condition in which each plate was exposed to during the reaction might have contributed to the overall quality of the genotyping.

Box 5.1: Missing Calls (Pink) of rs2191349 according to plate used.



Our analysis mandates using all SNPs in estimating the association between detected alleles and study variables. The level of successfully genotyped SNPs of each participant varied greatly. Apart from missing swabs and samples with a bad DNA quality, the number of failures of reporting SNPs of each participant varied between failure of genotyping one SNP to 15 SNPs per sample. Because we are counting the number of identical and total risk alleles of each SNP, we needed to ensure that all samples were comparable to each other. For example, there might have been a chance of measurement bias if we compared a subject where all SNPs were successfully genotyped to a subject for whom 10 (out of 23) SNPs failed to be genotyped.

To resolve this issue, we needed to exclude all subjects where there was a failure to report 3 or more SNPs. There is no specific reason behind choosing these inclusion criteria. However, we wanted to take the least risk of introducing measurement bias and at the same time not losing large number of participants who had a minimal number of SNPs that failed genotyping. 27 subjects had a failure of genotyping 3 or more SNPs. Figure 5.11 illustrates the distribution of cases with different magnitudes of genotyping failure. After removing bad quality samples, the number of valid samples for the analysis was reduced from 354 to 316 (157 were patients and 159 were healthy participants).

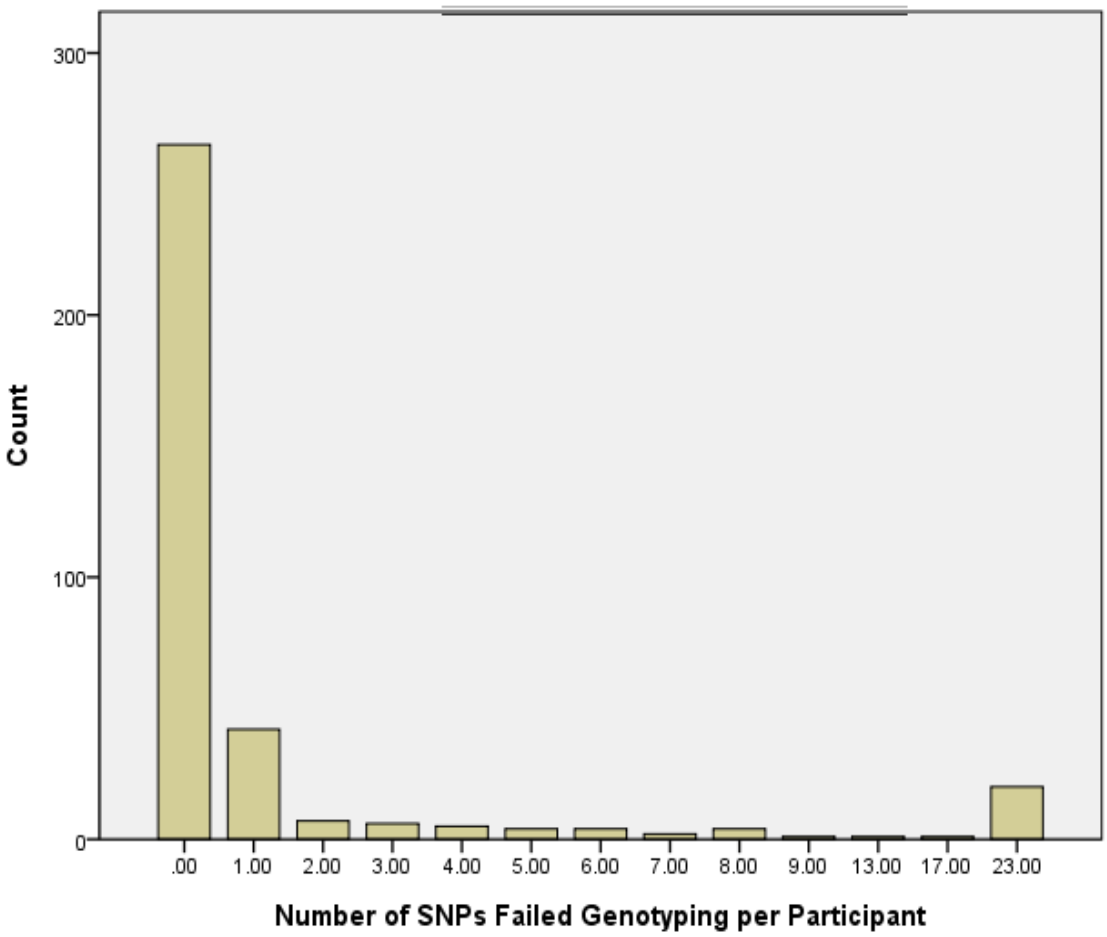


Figure 5.11: Distribution of number of non-genotyped SNPs according to number of participants.

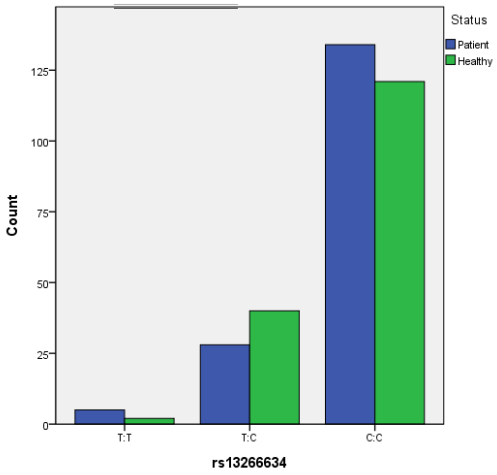
5.1.9.2 Alleles and Genotypes Frequencies

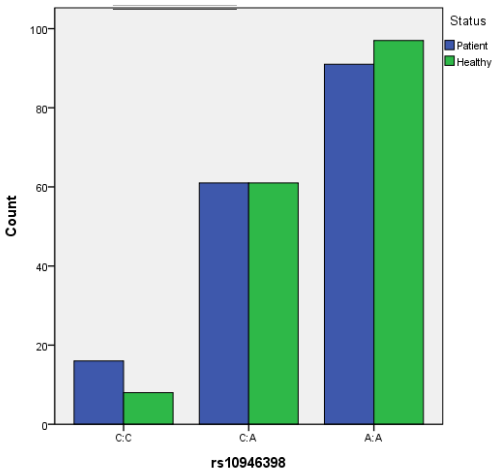
Table 5.23 illustrates the overall frequencies of genotypes and alleles in the total sample. The risk status of alleles and genotypes were depicted according to results of related GWAS studies in other populations. Box 5.2 is a visualization of genotype frequencies in patients and healthy subjects.

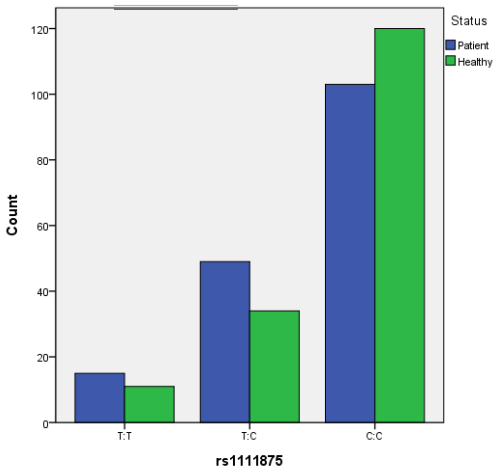
Table 5.23: Alleles and genotypes frequencies Number (Percentage):

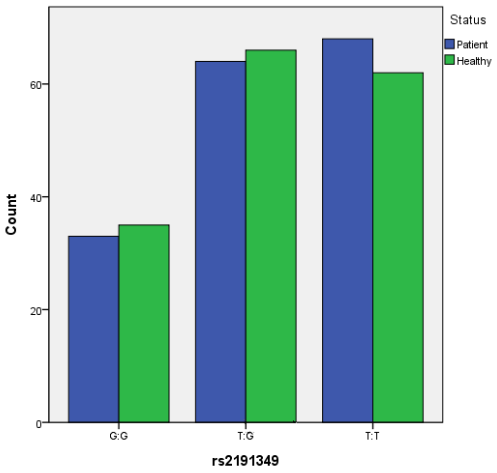
No	SNP	Homozygote genotypes of risk alleles	Heterozygote	Homozygote genotype of other alleles	Risk alleles	Other alleles
1	**rs10946398**	(A:A) **188** (56.3%)	(C:A)**122** (36.5%)	(C:C) **24** (7.2%)	**498** (74.5%)	**170** (25.4%)
2	**rs10811661**	(T:T) **223** (66%)	(T:C) **102** (30.2%)	(C:C) **13** (3.8%)	**548** (81%)	**128** (18.9%)
3	**rs1111875**	(C:C) **223** (67.2%)	(T:C) **83** (25%)	(T:T) **26** (7.8%)	**529** (79.6%)	**135** (20.3%)
4	**rs13266634**	(C:C) **255** (77.3%)	(T:C) **68** (20.6%)	(T:T) **7** (2.1%)	**578** (87.5%)	**82** (12.4%)
5	**rs7754840**	(C:C) **24** (7.1%)	(G:C) **125** (36.9%)	(G:G) **190** (56%)	**173** (25.5%)	**505** (74.4%)
6	**rs560887**	(C:C) **199** (59.4%)	(T:C) **116** (34.6%)	(T:T) **20** (6%)	**514** (76.7%)	**156** (23.2%)
7	**rs2191349**	(T:T) **130** (39.6 %)	(T:G) **130** (39.6%)	(G:G) **68** (20.7%)	**390** (59.4%)	**266** (40.5%)
8	**rs10830963**	(G:G) **20** (6.1%)	(G:C) **102** (31.1%)	(C:C) **206** (62.8%)	**142** (21.6%)	**514** (78.3%)
9	**rs7034200**	(A:A) **132** (40.1%)	(C:A) **152** (46.2%)	(C:C) **45** (13.7%)	**416** (63.2%)	**242** (36.7%)
10	**rs174550**	(T:T) **226** (69.8%)	(T:C) **91** (28.1%)	(C:C) **7** (2.2%)	**543** (83.7%)	**105** (16.2%)
11	**rs11708067**	(A:A) **257** (77.2%)	(A:G) **74** (22.2%)	(G:G) **2** (2 .6%)	**588** (88.2%)	**78** (11.7%)
12	**rs4607517**	(A:A) **29** (8.8%)	(A:G) **102** (30.9%)	(G:G) **199** (60.3%)	**106** (17.5%)	**500** (82.5%)
13	**rs7903146**	(T:T) **50** (15%)	(T:C) **148** (44.3%)	(C:C) **136** (40.7%)	**215** (33.8%)	**420** (66.1%)
14	**rs1799884**	(T:T) **34** (10%)	(T:C) **115** (33.9%)	(C:C) **190** (56%)	**183** (26.9%)	**495** (73%)
15	**rs5219**	(T:T) **23** (6.9%)	(T:C) **94** (28.1%)	(C:C) **218** (65.1%)	**140** (20.8%)	**530** (79.1%)
16	**rs7756992**	(G:G) **22** (6.5%)	(A:G) **135** (40.2%)	(A:A) **179** (53.3%)	**179** (26.6%)	**493** (73.3%)
17	**rs780094**	(C:C) **143** (42.9%)	(T:C) **131** (39.3%)	(T:T) **59** (17.7%)	**417** (62.6%)	**249** (37.3%)
18	**rs2237897**	(C:C) **328** (97.6%)	(T:C) **7** (2.1%)	(T:T) **1** (0.3%)	**663** (98.6%)	**9** (1.3%)
19	**rs2237895**	(C:C) **90** (27.1%)	(T:C) **151** (45.5%)	(T:T) **91** (27.4%)	**331** (49.8%)	**333** (50.1%)
20	**rs1387153**	(T:T) **31** (9.4%)	(T:C) **136** (41.2%)	(C:C) **163** (49.4%)	**198** (30%)	**462** (70%)
21	**rs231362**	(G:G) **116** (35.5%)	(A:G) **153** (46.8%)	(A:A) **58** (17.7%)	**385** (58.8%)	**269** (41.1%)
22	**rs7957197**	(T:T) **253** (75.7%)	(T:A) **71** (21.3%)	(A:A) **10** (3%)	**577** (86.3%)	**91** (13.6%)
23	**rs757210**	(A:A) **91** (27.8%)	(A:G) **140** (42.8%)	(G:G) **96** (29.4%)	**322** (49.2%)	**332** (50.7%)

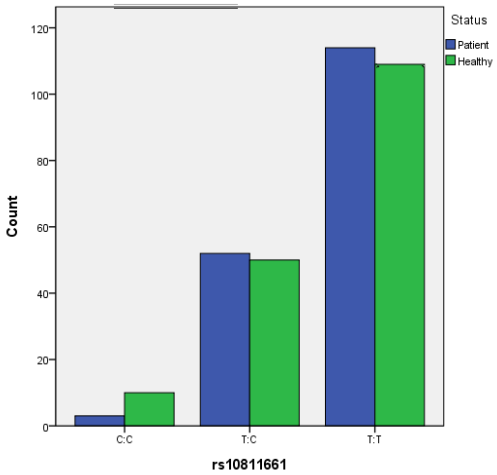
Box 5.2: Visualization of genotypes variations in patient and healthy subjects:

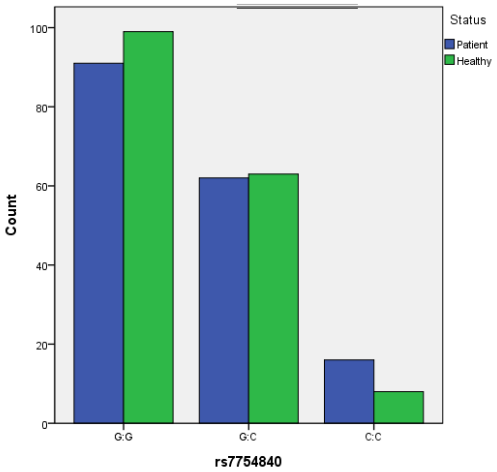




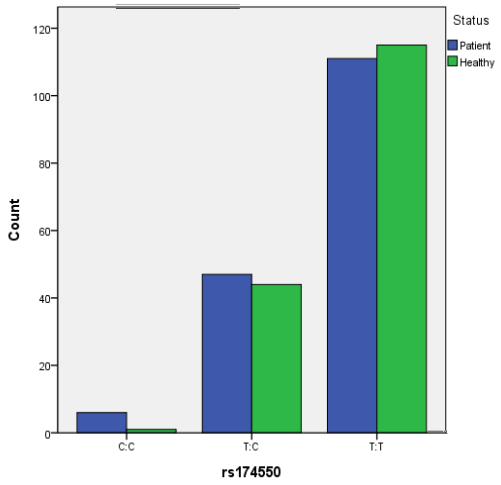


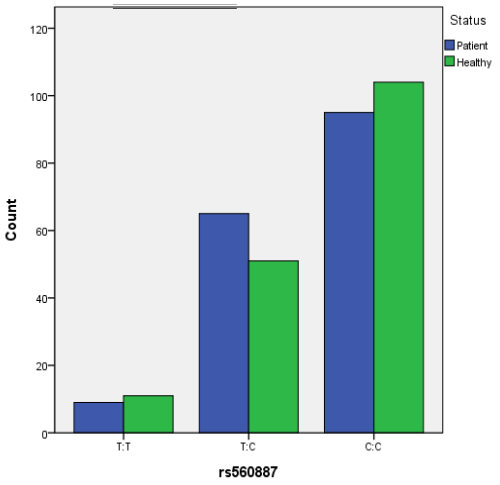


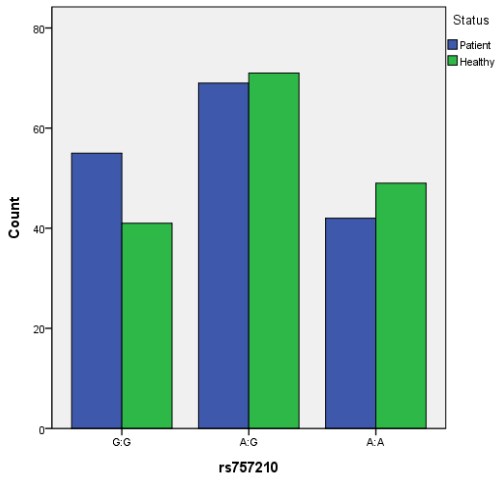


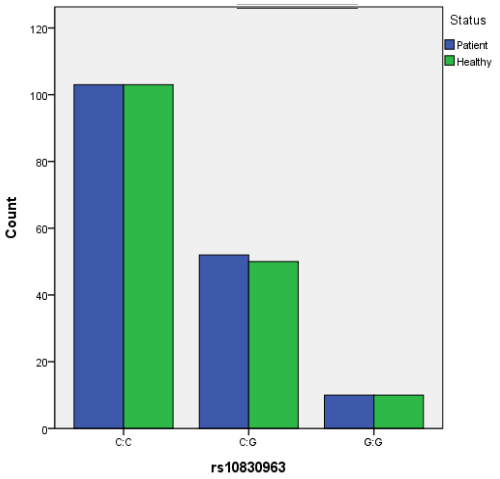


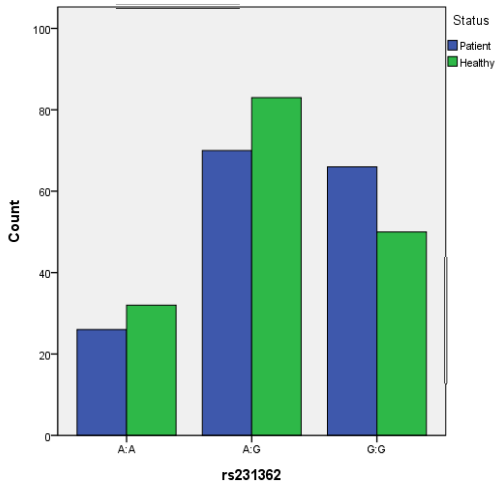
Box 5.2: Continue:

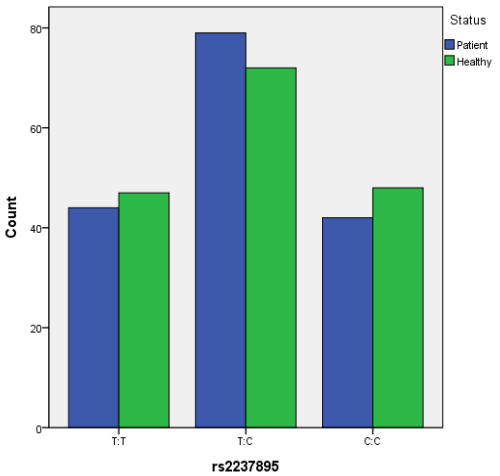




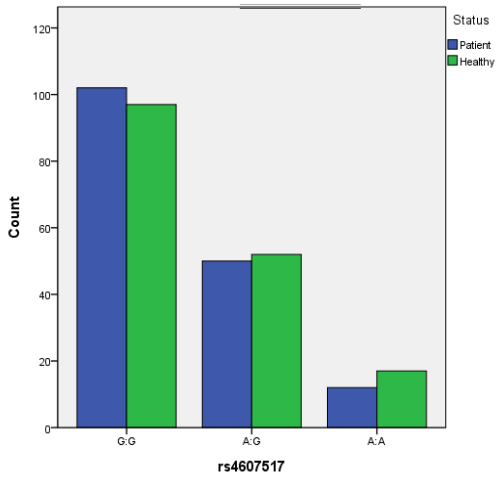


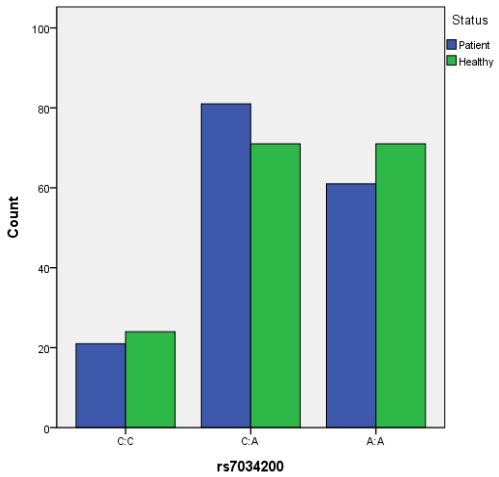


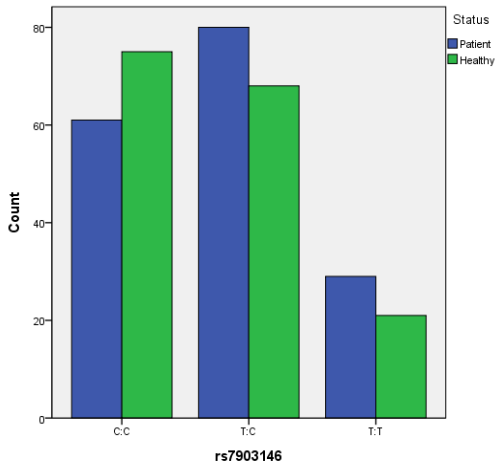


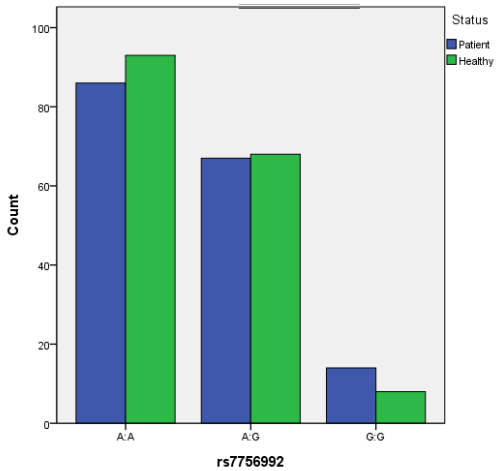


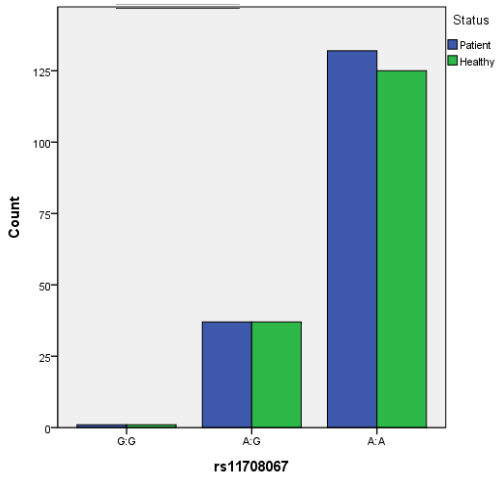
Box 5.2: Continue:

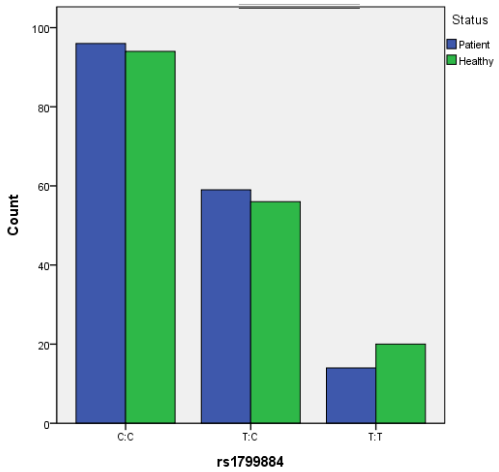




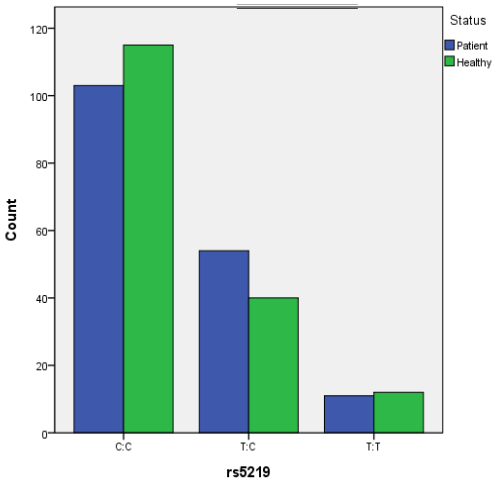


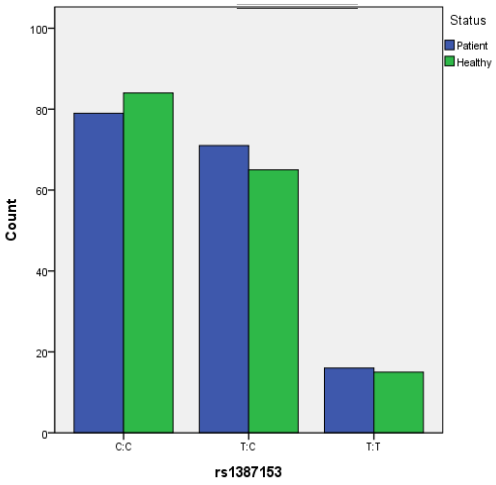


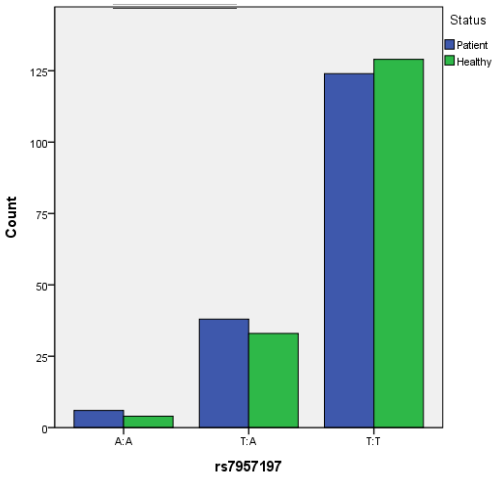


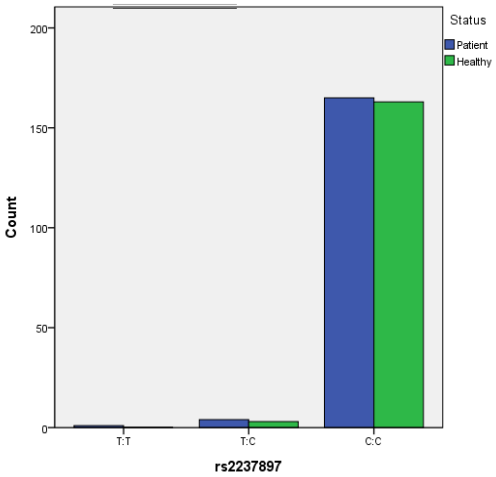


Box 5.2: Continue:









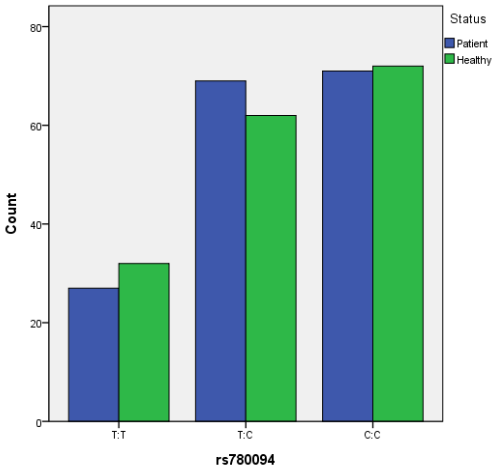


Table 5.24 illustrates summary statistics of several variables relating to the reported number of homozygote genotypes and number of risk alleles of the total sample. On comparing patients and healthy subjects, the number of inherited risk alleles and the levels of homozygosity are almost identical in both groups.

5.1.9.3 Distribution of number of homozygote genotypes and detected risk alleles

Figure 5.12 illustrate distribution of homozygote genotypes in all subjects. This distribution appears to be negatively skewed. This skewedness indicates that there is a tendency of deviation towards carrying homozygote genotypes in the sample. This is consistent with that expected in inbred populations. However, by observing figure 5.13, we can notice normal distribution of homozygote genotypes for risk alleles. The last observation might indicate that inbreeding increases level of homozygote genotypes regardless to the risk status of the allele.

Table 5.24: Genotypes summary statistics:

	Total number of all homozygote genotypes			Total Number of homozygote genotypes for risk alleles			Number of all risk alleles		
Status	Total	Patients	Healthy	Total	Patients	Healthy	Total	Patient	Healthy
No.	316	157	159	316	157	159	316	157	159
Mean	14.99	14.80	15.18	9.06	9.15	9.15	25.86	25.89	25.84
Median	15.00	15.00	15.00	9.00	9.00	9.00	26.00	26.00	26.00
Mode	16	16	16	9	9	9	24	27	26
SD	2.818	2.964	2.662	1.956	1.993	1.920	3.199	3.165	3.241
Minimum	5	5	7	1	1	2	17	18	17
Maximum	21	21	21	14	14	14	34	34	34

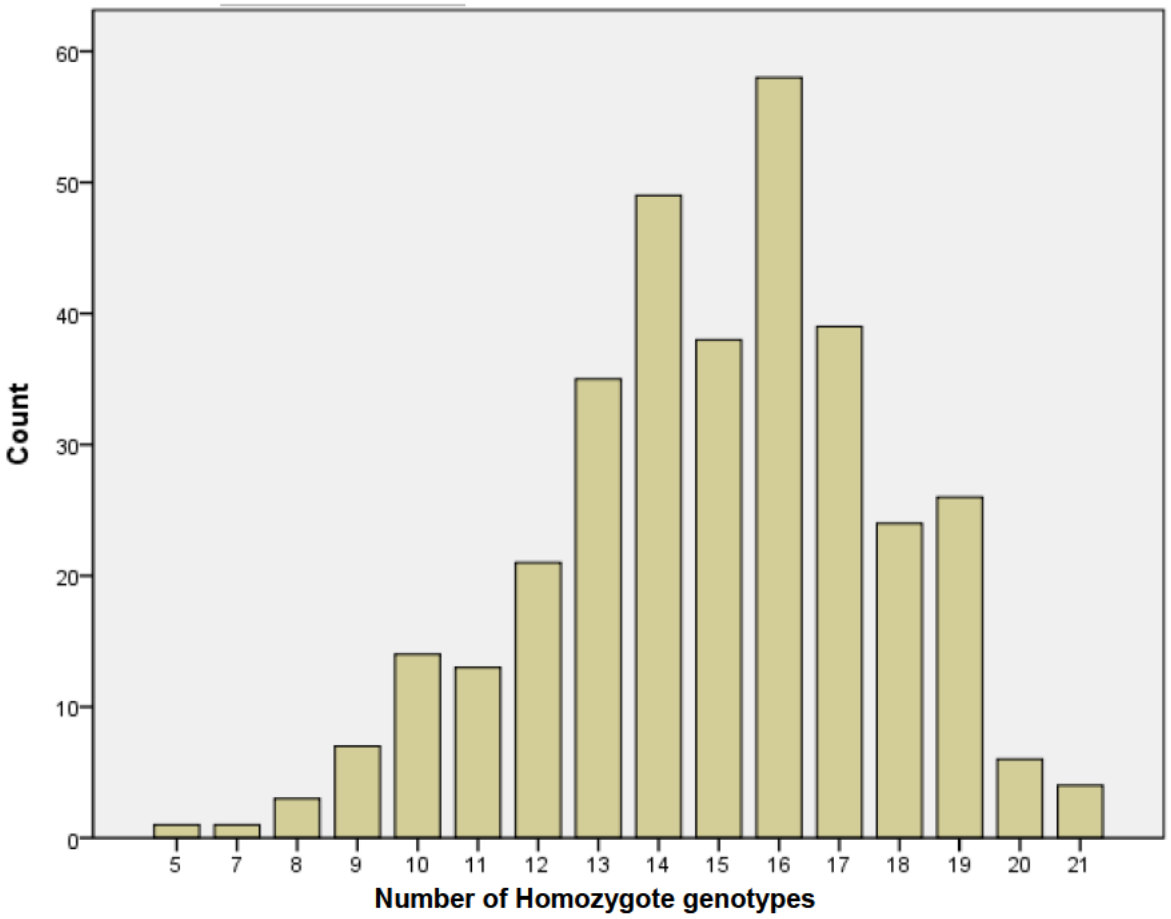


Figure 5.12: Distribution of number of homozygote genotypes of the whole sample.

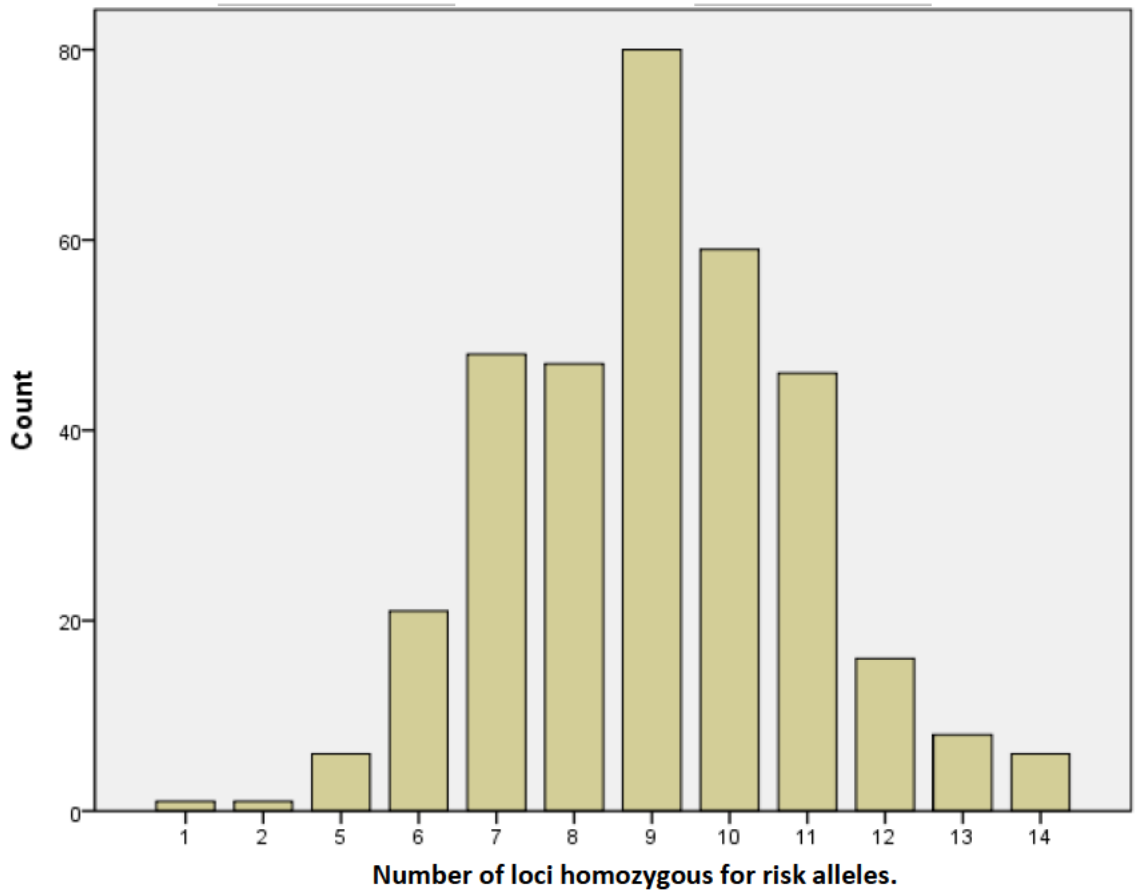


Figure 5.13: Number of loci homozygous for risk alleles per participant.

In figure 5.14, we compared the distribution of the number of homozygote genotypes between patients and healthy subjects. There is similar deviation toward an increased number of homozygote genotypes in patients and healthy subjects. Additionally, there is no consistent difference in homozygote genotypes number between patients and healthy subjects. In figure 5.15, we compared the distribution of homozygote genotypes of risk alleles of patients and healthy subjects.

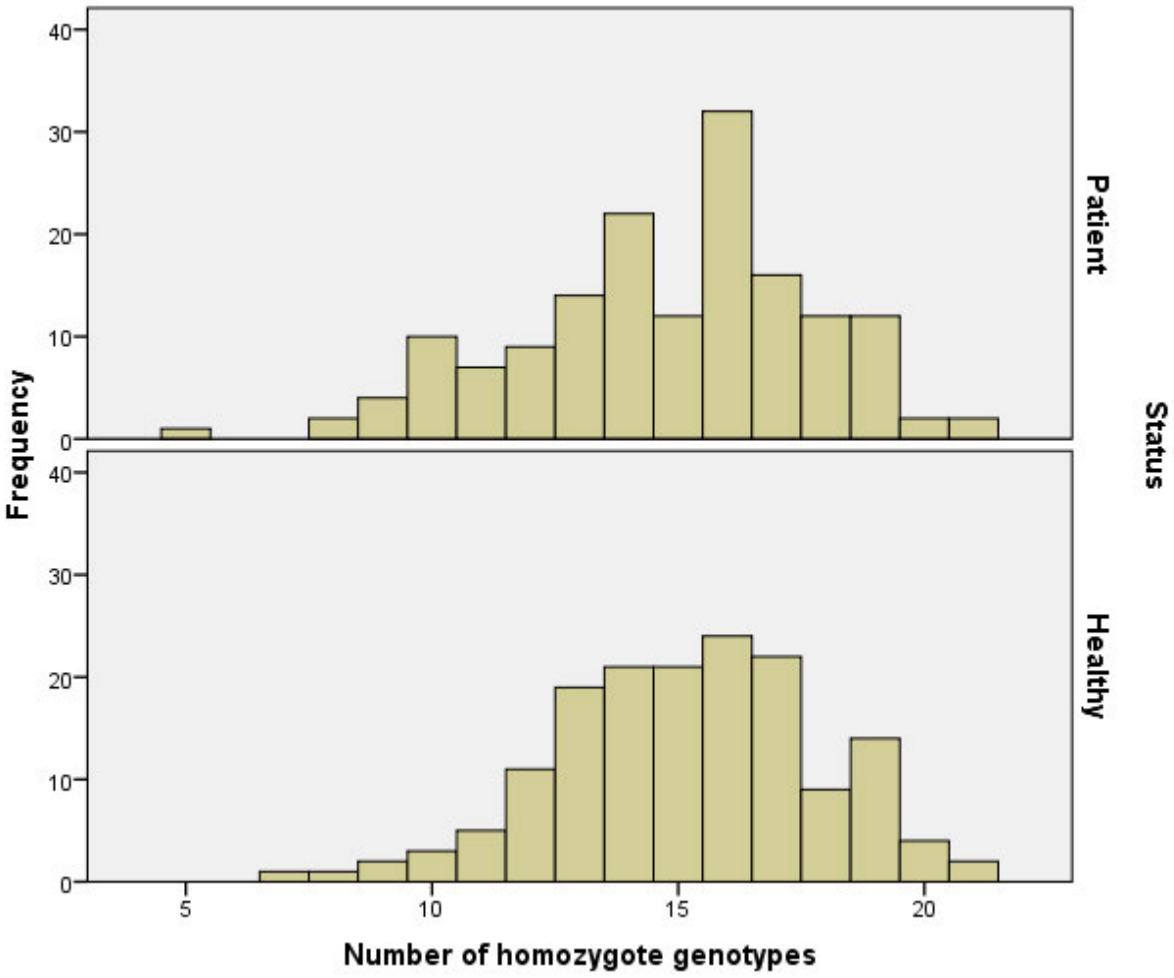


Figure 5.14: comparison of distribution of number of homozygote genotypes in patient and healthy subjects.

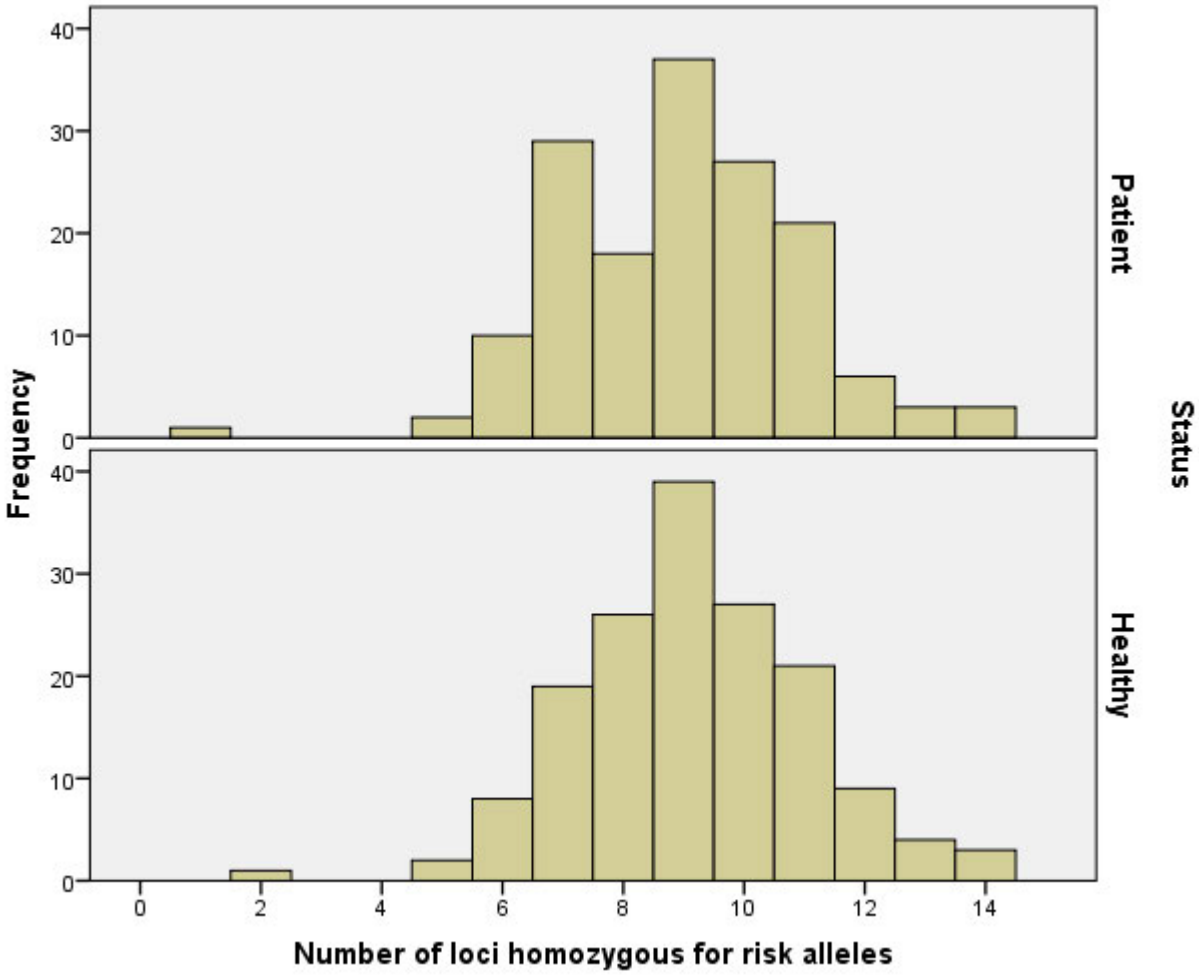


Figure 5.15: Comparison of number of loci homozygous for risk alleles in patients and healthy subjects.

Figure 5.16 illustrates approximately a normal distribution of number of detected type 2 diabetes risk alleles in the total sample. In figure 5.17, we compared the distribution of detected risk alleles of patient and healthy subjects. Similar to what has been observed in figure 5.15, illustrating distribution of homozygote genotypes of risk alleles, both distributions of detected risk alleles are approximately normal. Additionally, in a similar pattern, the distribution of detected risk alleles appears to be more consistent with normal in the healthy subjects' group compared to patients' group.

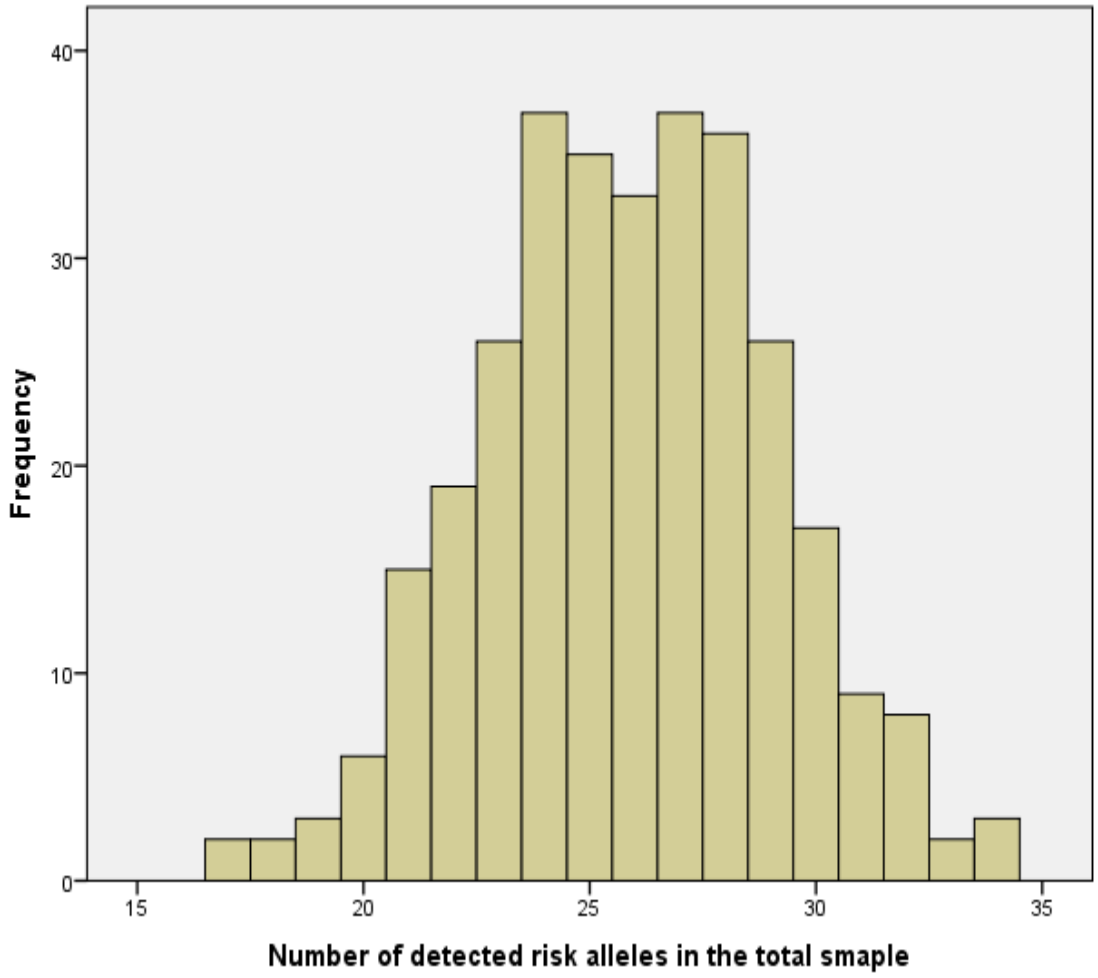


Figure 5.16: number of detected risk alleles in the sample.

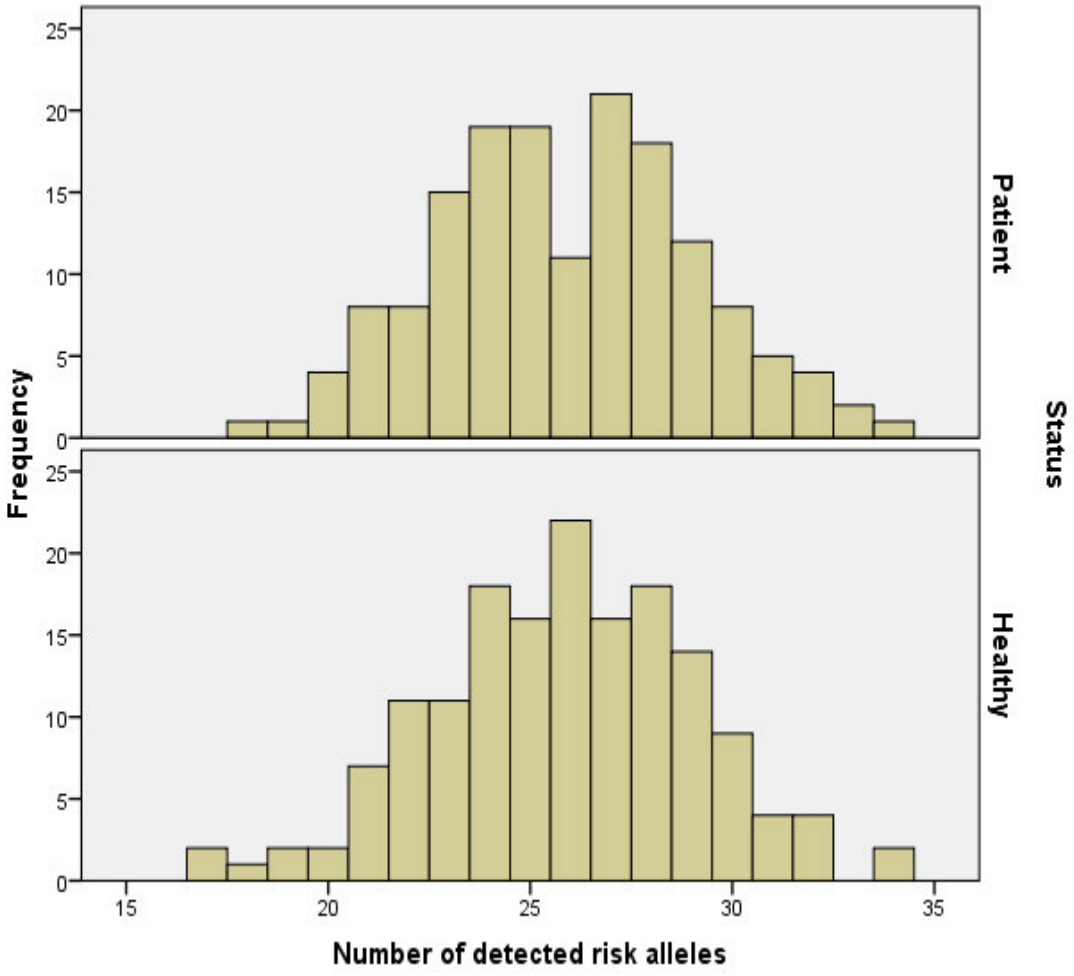


Figure 5.17: number of detected risk alleles in patient and healthy subjects.

5.1.10 Conclusion

- The total number of participants was 362, of whom 179 participants were type 2 diabetes patients.
- Several participants were reluctant to have their FBG measured but did not hesitate to give buccal swabs.
- Patients were relatively older compared to healthy participants.
- As expected, the number of siblings in Saudi families was large compared to that in western societies. The mean number of siblings per family in this sample was 7.03 (SD= 3.04).
- The prevalence of a history of consanguinity was 63% in our sample which is similar to the proportion of consanguineous marriages reported in studies conducted in Saudi Arabia. However, the number of marriages between first degree cousins was relatively lower in our sample compared to similar studies.
- More than half of participants were exceeding recommended daily calorific values.
- Healthy participants reported having a higher daily calorific intake compared to patients.
- Healthy participants reported having a higher tendency to engage in sedentary activities compared to patients. In contrast, healthy subjects reported higher engagement with moderate and vigorous activity compared to patients.
- The mean FBG was 95 mg/dl (SD: 23.89 mg/dl). Four readings were below the normal level of fasting blood glucose and 12 values were above the normal.
- Mean BMI of healthy participants in this sample was 28.15 (SD: 6.8). There was a tendency of increased prevalence of overweight and obese individuals in our sample compared to a similar study conducted in Saudi Arabia.
- The overall genotyping success was about 90%.
- Khat chewing is less likely to interfere with genotyping quality. Genotyping success variation was observed between plates used during the genotyping process.
- Overall genotype frequencies indicated a higher tendency to observe homozygosity in this sample.

5.2 Inferential Statistics

5.2.1 Introduction

This chapter is composed of two main sections. The first section is dedicated to answering the study's research questions. The second section is a secondary analysis emphasizing the possible effect of consanguinity on traits related to type 2 diabetes. Questions stated in the secondary analysis were only established after completion of data collection.

We used Pearson's correlation to test for the association between continuous variables. If distribution of variables was skewed, we utilized Spearman's test. In both tests, significant a level was chosen at 0.05 and all tests were two-tailed. In addition to using a correlation test, we used scatter plots to visualize significant associations. Simple linear regression was utilized to measure effect levels and multiple linear regression was utilized to assess the effect of multiple independent variables. T-test was utilized to assess difference of means of selected variables.

5.2.2 Primary Analysis

5.2.2.1 First question

Does an inbreeding coefficient predict the extent of aggregation of type 2 diabetes risk alleles?

Aim: In this question, we were aiming to investigate the possible association between the degree of consanguinity and the number of loci homozygous for risk alleles. We hypothesized that consanguinity is likely to increase aggregation of risk alleles by increasing the number of loci homozygous for risk alleles.

Analytic Method: Spearman's rank correlation coefficient.

Variables:

- Number of loci homozygote for risk alleles.
- Inbreeding coefficients.

Subjects: 316 patients and healthy participants.

Results:

Spearman's correlation coefficient indicates statistically non-significant association between degree of consanguinity and number of loci homozygote for risk alleles (0.01 p-value 0.856).

Interpretation:

Calculated inbreeding coefficients might not be powered enough to predict an increase in homozygosity of measured SNPs. However, the direction is consistent with the hypothesis of increased number of loci identical for risk alleles with increased level of consanguinity. When assessing the association between the number of all risk alleles and inbreeding coefficients, the association was statistically not significant (Spearman's coefficient: -0.045, p-value: 0.427). The association between heterozygote genotypes and inbreeding coefficients was not statistically significant (Spearman's coefficient: -0.087, p-value: 0.122). However, the inverse association with heterozygote loci is consistent with the notion of inbreeding reducing levels of heterozygote genotypes. Finally, measuring the association between aggregation of risk alleles and inbreeding coefficients when separating patients and healthy subjects did not result in any statistically significant associations.

5.2.2.2 Second question

Does the aggregation of type 2 diabetes risk alleles predict the age of onset of the disease?

Aim: In this question, we were aiming to investigate the possible association between age at diagnosis and allelic aggregation induced by loci homozygous for risk alleles. The study utilized age at diagnosis as a proxy for age of onset. If there were a cumulative recessive effect, we would expect to see individuals with higher number of loci homozygous for risk alleles with earlier age at diagnosis compared to those with lower numbers of loci homozygous for risk alleles.

Analytic Method: Simple linear regression to assess the association between age at diagnosis and number of loci homozygous for risk alleles. Multiple linear regression was later utilized to account for environmental factors.

Variables:

- Dependent variable: Age at diagnosis.
- Independent variables:
  - Number of loci homozygous for risk alleles.
  - Average daily calorific intake.
  - Number of weekly hours spent performing sedentary activities.
  - Number of weekly hours spent performing moderate activities.
  - Number of weekly hours spent performing vigorous activities.

Subjects: 157 type 2 diabetes patients.

Results:

The univariate analysis reveals statistically non-significant association between number of loci homozygous for risk alleles and age at diagnosis (β: -0.106, p-value: 0.762). This association remained statistically non-significant using the multivariate analysis.

Interpretation:

Measured SNPs did not provide any statistical evidence of cumulative recessive effect on age at diagnosis. However, the negative association might suggest a recessive effect that could be detected with a larger sample or larger number of SNPs. Additionally, we examined the association between the number of detected risk alleles (contributed from homozygote loci for risk alleles and heterozygote loci). However, no significant associations were detected between overall number of risk alleles and other traits. Only one association between age at diagnosis and total number of risk alleles became marginally significant after accounting for parental

history of diabetes and inbreeding coefficients (β: -0.399, p-value: 0.052) (Table 5.25 and Figure 5.18). We hypothesized that parental history of diabetes and consanguinity history might have captured additional genetic effect, which was not effectively captured by measured SNPs. In this final model, we categorized patients according to parental history of diabetes and only the maternal history of type 2 diabetes was statistically associated with reduced age at diagnosis. Other categories of parental history were not statistically associated with any effect on age at diagnosis. The final model suggests that increased number of detected risk alleles, higher level of consanguinity and maternal history of diabetes are likely to be associated with reduced age at diagnosis.

Table 5.25: Association between age at diagnosis and number of all risk alleles:

Model	Unstandardized Coefficients		t	Sig.
	B	Std. Error		
(Constant)	51.709	5.400	9.576	.000
Number of risk alleles	-.399	.204	-1.958	.052
Mother Affected	-5.307	1.319	-4.023	.000
Inbreeding coefficients (%)	-.533	.235	-2.266	.025

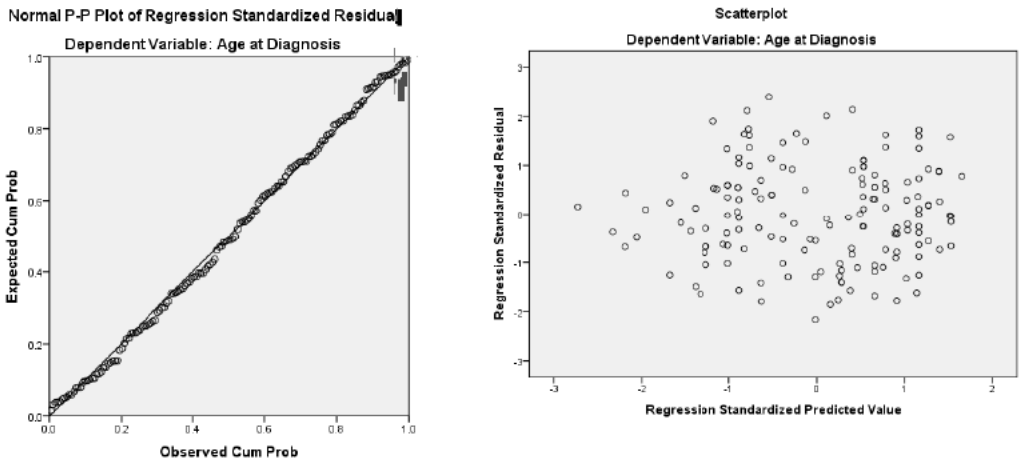


Figure 5.18: Normal probability plots and residuals of association between age at diagnosis and number of risk alleles accounting for parental history and consanguinity.

5.2.2.3 Third question

Does aggregation of type 2 diabetes risk alleles predict the extent of type 2 diabetes risk factors (BMI, waist circumference, FBG)?

Aim: This question assesses the association between continuous risk factors and allelic aggregation induced by the number of loci homozygous for risk alleles. We hypothesized that individuals with higher number of loci identical for risk alleles could have been influenced by a cumulative recessive effect leading to higher BMI, waist circumference and FBG.

Analytic Method: Simple linear regression to assess the association between risk factors and number of loci homozygous for risk alleles. Multiple linear regression was later utilized to account for age and environmental variables.

Variables:

- Dependent variables: BMI, waist circumference, and FBG.
- Independent variables:
  - Number of loci homozygous for risk alleles.
  - Age.
  - Average daily calorific intake.
  - Number of weekly hours spent performing sedentary activities.
  - Number of weekly hours spent performing moderate activities.
  - Number of weekly hours spent performing vigorous activities.

Subjects: 159 healthy participants for BMI and waist circumference and 133 participants for FBG.

Results:

The univariate analysis reveals statistically non-significant association between number of loci homozygous for risk alleles and continuous risk factors (BMI: β: 0.221, p-value: 0.438. waist circumference: β: -0.278, p-value: 0.658. FBG: β: 0.448, p-value: 0.652). This association remained statistically non-significant using the multivariable analysis.

Interpretation:

Similar to what was observed in the first two research questions, no statistically significant cumulative recessive effect was detected. Even when utilizing number of all risk alleles, no additive effects were detected (BMI: β: 0.207, p-value: 0.217, waist circumference: β: -0.169, p-value: 0.645), FBG: β: 0.967, p-value: 0.159). Not being able to find any recessive or additive effect on traits related to type 2 diabetes might suggest a lack of power or possibility of the contribution of unmeasured SNPs.

5.2.3 Secondary Analysis

5.2.3.1 Association between age at diagnosis and inbreeding coefficients

Aim: In this section, we were interested in testing whether higher levels of inbreeding coefficients are associated with earlier age at diagnosis.

Analytic Methods: Firstly, simple linear regression was utilized to test the association between age at diagnosis and inbreeding coefficients. Secondly, multiple linear regression was preformed to control for other environmental variables.

Variables:

- Dependent variable: Age at diagnosis.
- Independent variables:
  - Inbreeding coefficients (percentage).
  - Average daily calorific intake.
  - Number of weekly hours spent performing sedentary activities.
  - Number of weekly hours spent performing moderate activities.
  - Number of weekly hours spent performing vigorous activities.

Subjects: 179 type 2 diabetes patients.

Results:

Table 5.26: Univariate analysis of association between age at diagnosis and inbreeding coefficients:

Model	Unstandardized Coefficients		Standardized Coefficients	Sig.
	B	Std. Error	Beta	
(Constant)	39.044	.765		.000
Inbreeding coefficients (%)	-.593	.236	-.186	.013

a. Dependent Variable: Age at Diagnosis

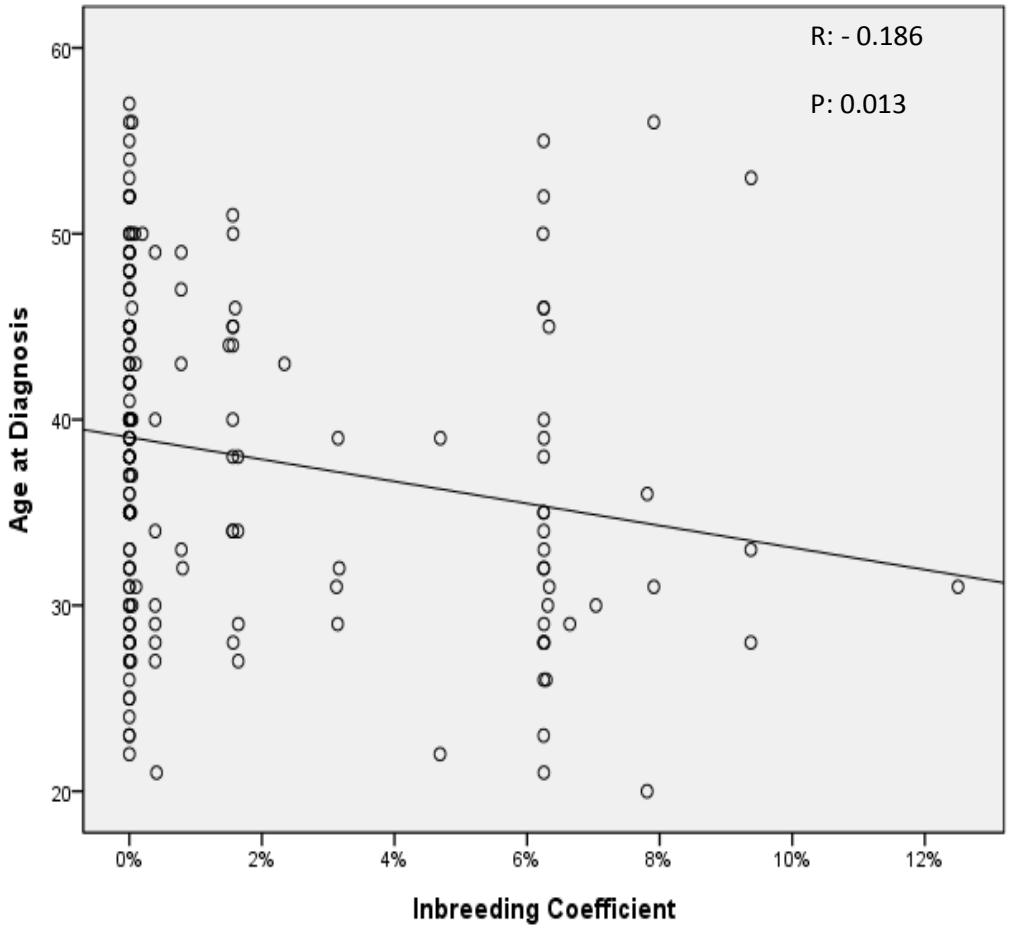


Figure 5.19: Scatter plot of inbreeding coefficients and age at diagnosis.

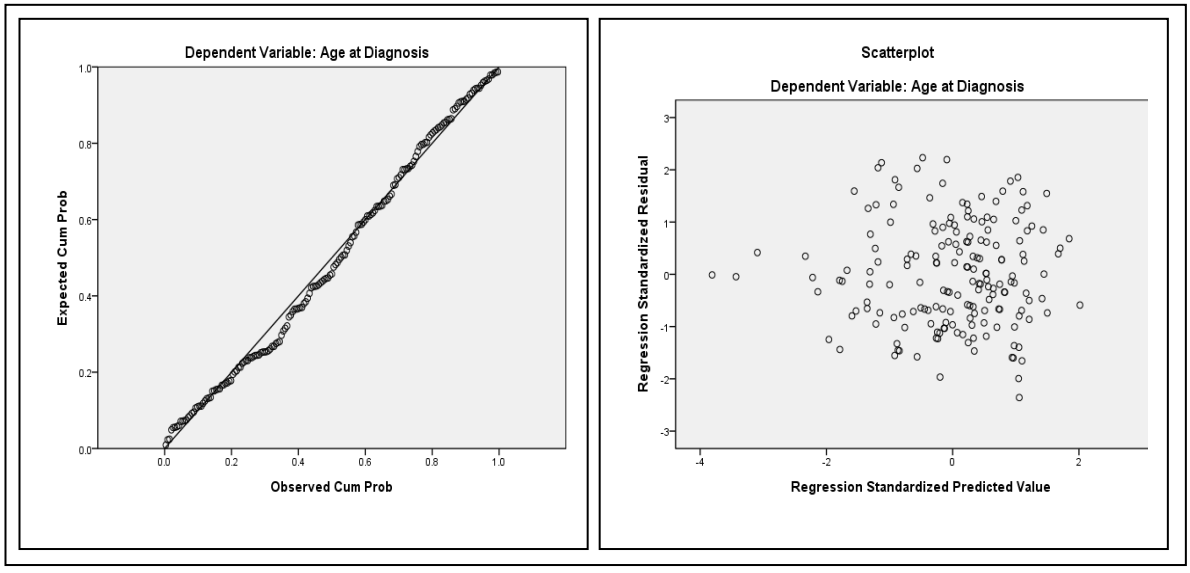


Figure 5.20: Normal probability plot and scatter plot of standardized residuals against standardized predicted values of association between age at diagnosis and inbreeding coefficients.

Table 5.27: Multivariate analysis of association between age at diagnosis and inbreeding coefficients:

Model	Unstandardized Coefficients		t	Sig.
	B	Std. Error		
(Constant)	48.555	2.405	20.193	.000
Inbreeding coefficients (%)	-.572	.225	-2.544	.012
Weekly Sedentray Hours	-.087	.025	-3.513	.001
Weekly Moderate Activity Hours	-.142	.066	-2.138	.034
Weekly Vigorous Activity Hours	-1.380	.538	-2.566	.011

a. Dependent Variable: Age at Diagnosis.

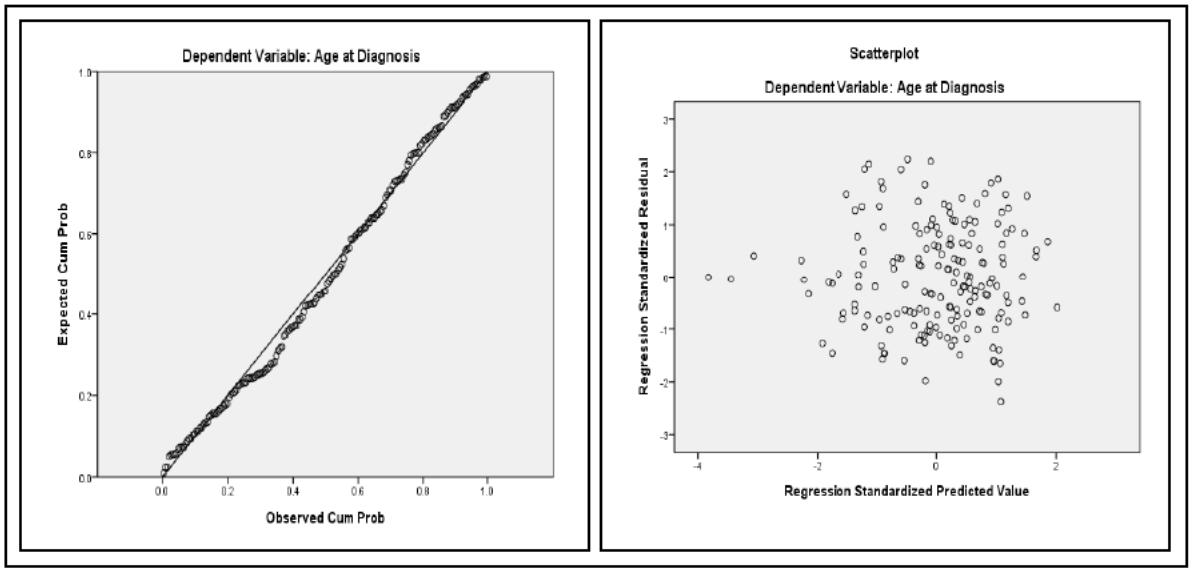


Figure 5.21: Normal probability plot and scatter plot of standardized residuals against standardized predicted values of association between age at diagnosis and inbreeding coefficients after accounting for environmental variables.

Interpretation:

The univariate analysis indicates significant inverse association between inbreeding coefficients and age at diagnosis. Although this negative association is weak, it gives us an indication that those who have higher inbreeding coefficients are likely to be associated with increased risk of developing type 2 diabetes at a younger age. Figure 5.20 shows an approximate linear distribution of residuals. Environmental variables were added to the model to test for possible effect on the association between inbreeding and age at diagnosis. However, after environmental variables were added to the model the association remained significant.

5.2.3.2 Association between age at diagnosis and inbreeding coefficients accounting for parental history of diabetes

Aim: To assess whether parental history of diabetes has an effect on the association between inbreeding coefficients and age at diagnosis.

Analytic Methods: Patients were divided into four groups based on parental history of diabetes. In each group, median of inbreeding coefficients, mean and standard deviation of age at diagnosis and Spearman's correlation coefficients were calculated.

Variables:

- Parental history of diabetes (four categories: both parents affected, neither affected, only father affected, only mother affected).
- Inbreeding coefficients.
- Age at diagnosis.

Subjects: 179 patients.

Results:

Table 5.28: Inbreeding coefficients, age of diagnosis and their correlations according to parental history of type 2 diabetes:

Parental History of Type 2 Diabetes		**Paternal History**	
		Yes	**No**
Maternal History	**Yes**	Median of inbreeding coefficients: 0.23% Mean of age at diagnosis: 36.94 (SD: 9 years). Correlation between IC and age at diagnosis: -0.502 (p-value: 0.003) N = 35	Median of inbreeding coefficients: 0. 04% Mean of age at diagnosis: 34.51 (SD: 7 years). Correlation between IC and age at diagnosis: 0.117 (p-value: 0.491) N = 37.
	No	Median of inbreeding coefficients: 0.02% Mean of age at diagnosis: 38.96 (SD: 9 years). Correlation between IC and age at diagnosis: -0.228 (p-value: 0.128) N = 46	Median of inbreeding coefficients: 0. 04% Mean of age at diagnosis: 40.26 (SD: 9 years). Correlation between IC and age at diagnosis: -0.158 (p-value: 0.223) N = 61

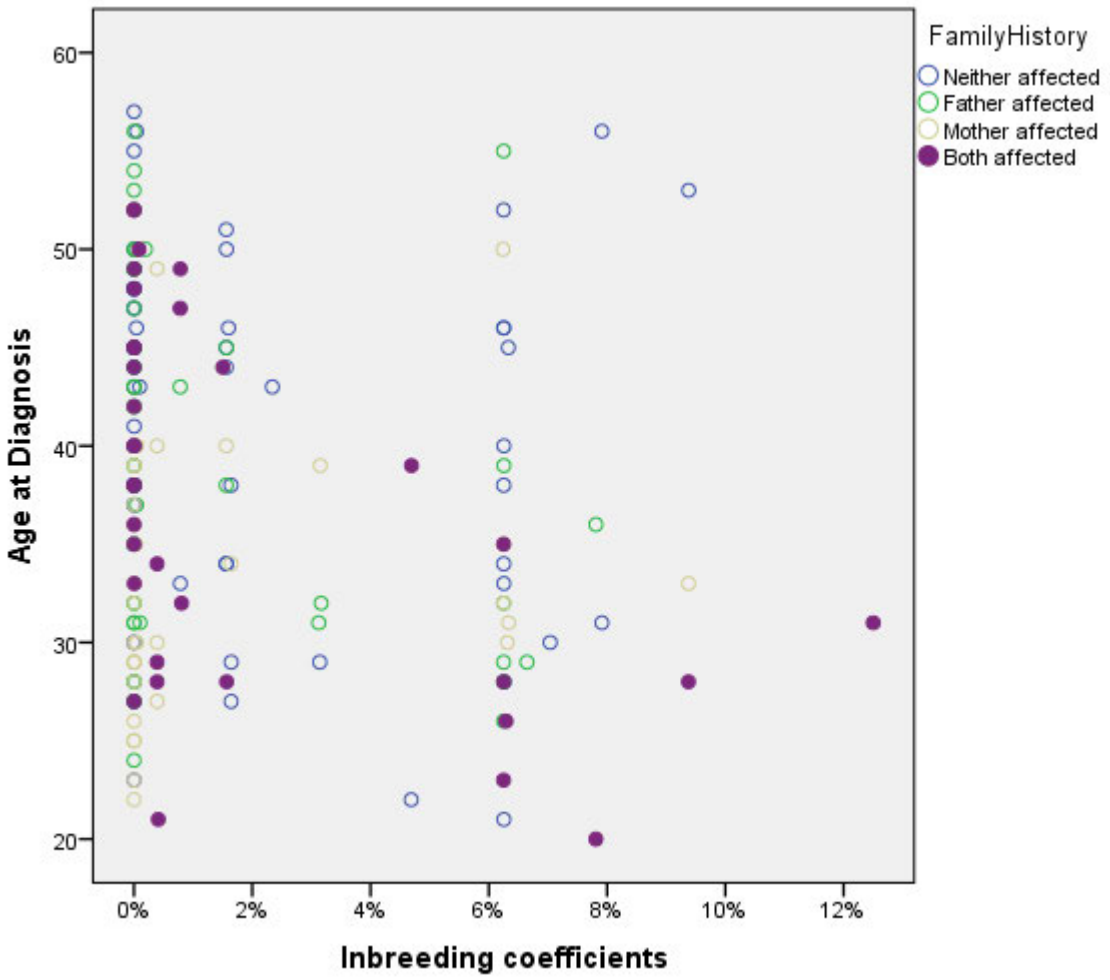


Figure 5.22: Association between age at diagnosis and inbreeding coefficients categorized according to parental history of diabetes.

Interpretation:

Mean age of diagnosis with type 2 diabetes appears to be higher when there is no parental history of the disease. The mean age at diagnosis in the no parental history group is 40.26 years compared to 36.94 years in the group that reported both parents affected with the disease. Additionally, the lowest mean of age at diagnosis with diabetes is 34.51 years which is reported in the group that reported only maternal history of diabetes. The later statement might indicate a stronger association of diagnosing at younger age when there is a maternal history of type 2 diabetes compared to the paternal history of the disease.

As indicated in Table 5.28, all Spearman's correlation coefficients in all groups were not significant except for one association. The association between inbreeding coefficients and age at diagnosis in families, which reported both parents as being affected is -0.502 (P value: 0.003). The correlation is stronger than that reported earlier correlating inbreeding coefficients with age of diagnosis in the whole sample. This higher correlation might indicate that those who have higher inbreeding coefficients and have both parents affected with diabetes are associated with development of the disease at a younger age compared to other individuals in the remaining groups.

For development of recessive illnesses, both parents must be either affected or carriers of genetic risk markers of the illness of interest. Observing a stronger inverse association between age at diagnosis and higher level of consanguinity in offspring where both parents are affected with type 2 diabetes might be attributed to possible recessive effect. Stratifying parents according to parental history eliminates possible source of confounding on age at diagnosis caused by parental history of diabetes.

5.2.3.3 Association between type 2 diabetes risk factors and inbreeding coefficients

Aim: To establish if higher inbreeding coefficients are associated with an increase of BMI, waist circumference and FBG.

Analytic Methods: Simple linear regressions to assess the associations between inbreeding coefficients and BMI, waist circumference and FBG. Multiple linear regression to account for environmental variables.

Variables:

- Dependent variables: BMI, waist circumference and FBG.
- Independent variables:
    - Inbreeding coefficients (percentages).
    - Age.
    - Average daily calorific intake.
    - Number of weekly hours spent performing sedentary activities.
    - Number of weekly hours spent performing moderate activities.
    - Number of weekly hours spent performing vigorous activities.

Subjects: 183 healthy participants for BMI, 182 for waist circumference and 153 participants for fasting blood glucose.

Results:

No statistically significant associations were detected between inbreeding coefficients and BMI (β: 0.113, p-value: 0.476), waist circumference (β: 0.207, p-value: 0.607) or fasting blood glucose (β: -0.007, p-value: 0.991). Associations remained non-significant after accounting for environmental variables and age.

5.2.3.4 Correlations of variables between healthy sibling pairs in each family

Aim: To assess resemblance between healthy siblings of several measured variables. We were interested in investigating possible shared genetic or environmental effects on resemblance of several measured traits.

Analytic Methods: In this study, we recruited 153 participants from 51 families. In these families, three participants were recruited from each family where one participant was a patient and the remaining two were healthy participants. Therefore, we have 102 healthy participants where each two participants represent a familial unit. We studied the magnitude of correlation of several variables between healthy sibling pairs in each familial unit. Pearson's correlation coefficients were calculated for variables, which were normally distributed and Spearman's rank correlation coefficients were calculated for variables with non-normal distributions.

Variables:

In each healthy sibling pair, we calculated correlations of the following variables.

- Age.
- Average daily calorific intake.
- Number of weekly hours spent performing sedentary activities.
- Number of weekly hours spent performing moderate activities.
- Number of weekly hours spent performing vigorous activities.
- Weight.
- Height.
- BMI.
- Waist circumference.
- FBG.

Subjects: 51 pairs of healthy participants for all variables except FBG (42 pairs).

Results:

Table 5.29: Correlation of several variables between healthy siblings in each familial unit:

Variables	Test Type	Correlation	P-value	Number of Familial Units
Correlation of age between healthy siblings	Pearson's correlation	0.611	0.000	51
Correlations between healthy siblings calorific intake	Pearson's correlation	0.459	0.001	51
Correlations between healthy siblings sedentary activity	Pearson's correlation	0.227	0.109	51
Correlation between healthy siblings moderate activity	Spearman's correlation	0.182	0.202	51
Correlation between healthy siblings vigorous activities	Spearman's correlation	0.161	0.258	51
Correlation between healthy siblings weight	Pearson's correlation	0.185	0.194	51
Correlation between healthy siblings height	Spearman's correlation	0.277	0.049	51
Correlation between healthy siblings BMI	Pearson's correlation	0.260	0.065	51
Correlation between healthy siblings waist circumference	Pearson's correlation	0.206	0.147	51
Correlation between healthy siblings FBG	Pearson's correlation	0.317	0.040	42

Interpretation:

Correlation of age between healthy siblings in each family is 0.611. This indicates that siblings are reasonably within similar age range. Having similar age range allows us to assume that age is less likely to affect the association of any traits detected between siblings.

There is a significant association of eating habits in siblings of each family member. Pearson's correlation coefficient for association between the calorific intake of each sibling is 0.459. This gives a partial explanation of environmental effect of associations of any trait between siblings of each family. However, none of the parameters used to measure physical activity were significantly correlated between healthy siblings of each family member. This might indicate that unlike obvious familial effect on food intake, familial effect on level of physical activity is less likely to be significant in this sample.

Two significant associations of healthy siblings' anthropometric measures were detected. Height of healthy siblings was significantly correlated. Additionally, correlation between FBG of siblings of each family was significantly associated (r: 0.317, p-value: 0.04). However, associations of waist circumference and body weight between healthy siblings were not statistically significant and correlation between siblings' BMI indicated marginal significance (0.065).

5.2.3.5 Effect of consanguinity on correlations of BMI, waist circumference and FBG between healthy sibling pairs in each family

Aim: To study the effect of consanguinity on direction of correlation of BMI, waist circumference and FBG between sibling pairs. We were interested in observing if families with higher inbreeding coefficients are associated with higher correlations between siblings of each family.

Analytic Methods: The effect of consanguinity was assessed by grouping inbreeding coefficients into four categories, where Pearson's correlation coefficients were calculated in each category. In first stage, all families were included. The second stage comprised removal of all families with zero inbreeding coefficients. Thirdly, we removed all families where the only history of inbreeding coefficients was based on mating between couples from the same tribe or subtribe. In the final stage, we removed families with inbreeding history lower than 1.5%.

Variables:

In each pair, we calculated correlations of the following variables.

- BMI.
- Waist circumference.
- FBG.

Subjects: 51 pairs of healthy participants for all variables except fasting blood glucose (42 pairs).

Results:

Calculating correlation coefficients under a restricted range of consanguinity revealed an effect on correlations of FBG (Table 6 and Figure 5). No similar effect was observed on correlations of BMI and waist circumference.

Table 5.30: Studying the effect of inbreeding on association of FBG between siblings of each family:

Inclusion Criteria	Pearson's Correlation Coefficient	Significance	Number of familial units
All healthy siblings	0.317	0.040	42
Only families with history of consanguinity	0.419	0.015	33
Only families with inbreeding coefficients larger than 0.04%	0.584	0.003	23
Only families with inbreeding coefficients larger than 1.5%	0.699	0.001	18

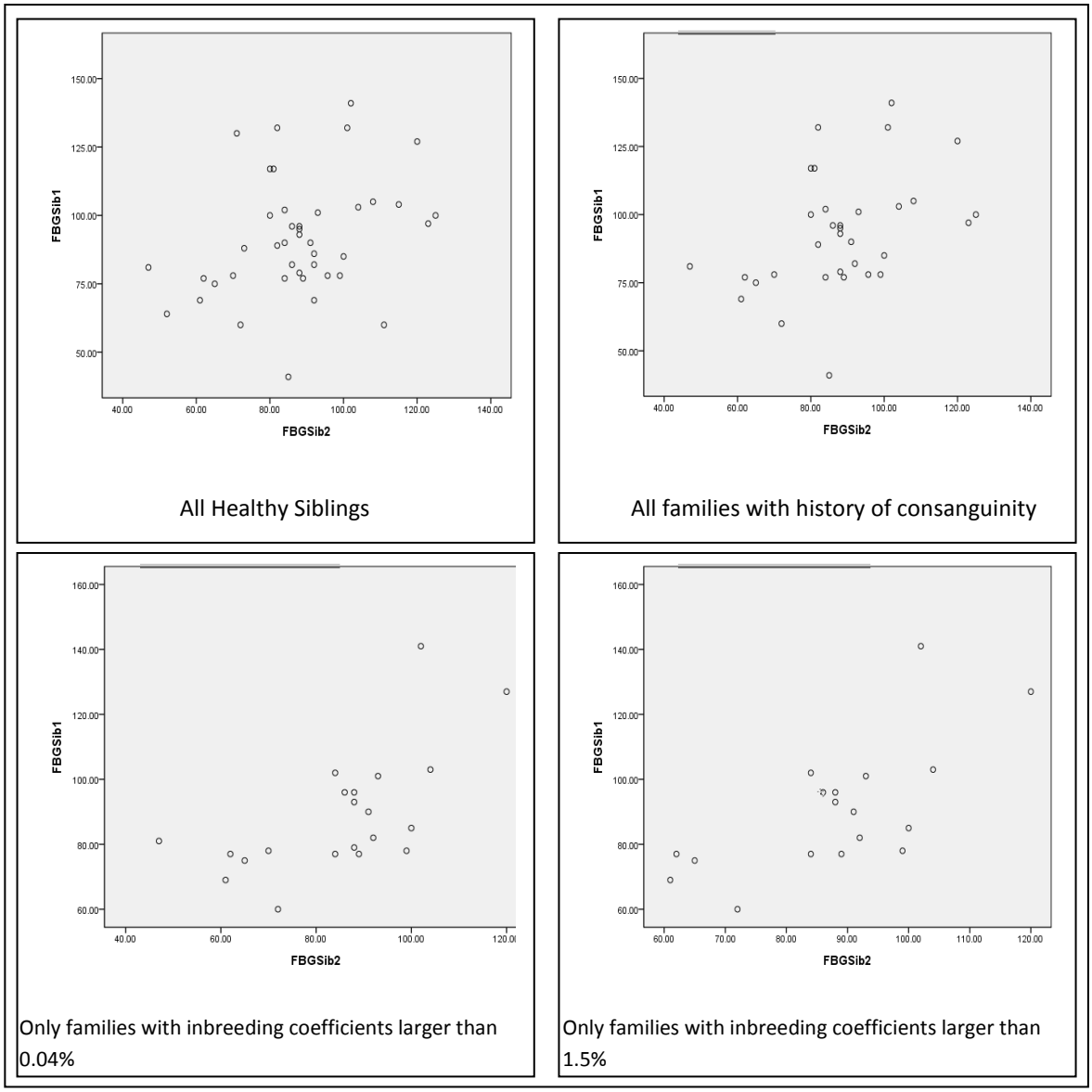


Figure 5.23: Scatter plots illustrating correlations of fasting blood glucose in different levels of consanguinity.

Interpretation:

Estimating association of healthy siblings' FBG according to inbreeding coefficients showed a significant trend. There is an increase in correlations from 0.317 between siblings in all families to 0.699 when restricting the association in families with inbreeding coefficients higher than 1.5%. This trend suggests that association of FBG between siblings is higher in families with higher inbreeding coefficients compared to families with lower inbreeding coefficients. This might give a partial explanation of the intensified genetic effect with increased inbreeding coefficients.

5.2.3.6 Effect of consanguinity on variation of FBG between healthy sibling pairs in each family

Aim: In the previous analysis, we observed higher correlations of FBG between siblings of each family with increased inbreeding coefficients. In this section, we are interested in estimating the effect of consanguinity on reducing the difference of FBG between each pair and to account for the differences of other environmental variables and age.

Analytic Methods: Difference of FBG was calculated within each pair. A simple linear regression was preformed to assess the direction of the difference of FBG with increased inbreeding coefficients. Differences in age and other environmental variables were all calculated and fitted in a multiple regression line to account for the effect of variation in these factors on the association between consanguinity and difference in FBG.

Variables:

- **Dependent variable:** Difference of FBG within pairs of each family.
- **Independent variables:**
  - Inbreeding coefficients (percentage).
  - Difference of age of siblings of each pair.
  - Difference of daily calorific intake of siblings of each pair.
  - Difference of weekly sedentary hours of siblings of each pair.
  - Difference of weekly moderate activity hours of siblings of each pair.
  - Difference of weekly vigorous activity hours of siblings of each pair.

Subjects: 42 pairs of healthy siblings.

Results:

Inbreeding coefficients were inversely associated with difference of FBG (β: -1.8, p-value: 0.039). Association between inbreeding coefficients and difference in FBG remained significant after accounting for difference in age and other environmental factors and none of the variables was significantly associated with difference in FBG. Figure 5.24 indicates the normal probability plot and residuals distribution. The distribution indicated in this figure did not improve by log transformation of the variables.

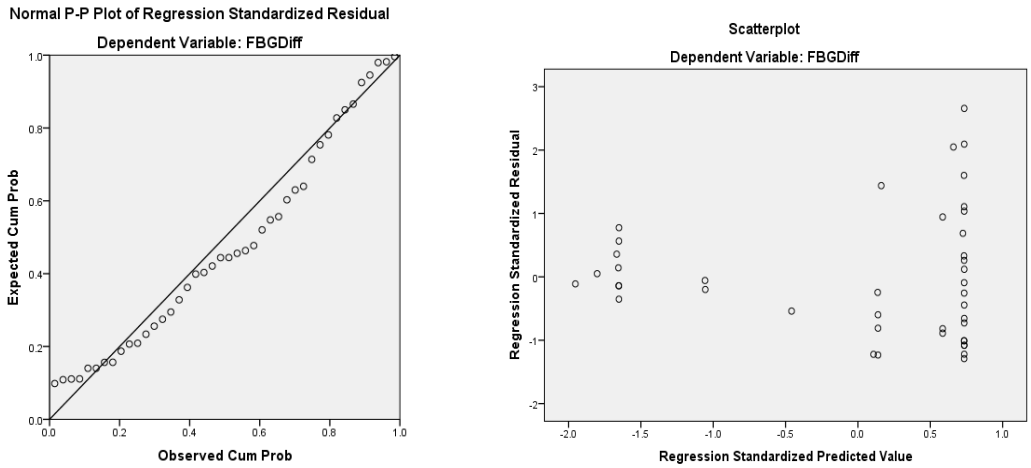


Figure 5.24: Normal probability plot and scatter plot of standardized residuals against standardized predicted values of association between difference in FBG and inbreeding coefficients.

Interpretation:

As indicated in Table 5.30, the effect of inbreeding coefficients on association of FBG between siblings of each family is still significant after accounting for several other variables. These variables include difference of age, calorific intake and level of physical activity between healthy siblings of each family. The minus sign of the slope of association between inbreeding coefficients and difference of FBG between healthy siblings indicates that in families with higher inbreeding coefficients, the difference of FBG between siblings becomes smaller. Controlling of other environmental differences indicates the significant genetic effect of this association, which in this case could be largely explained by the effect of inbreeding.

5.2.3.7 Assessing the effect of measured environmental factors on age at diagnosis

Aim: To study the association between measured environmental variables and age at diagnosis. Assessing the association might give an indication of performance of measured variables as predictors of age at diagnosis.

Analytic Methods: Pearson's correlation coefficients for variables with normal distributions and Spearman's correlation coefficients were calculated for others with skewed distributions.

Variables:

The association was measured between:

- Age at diagnosis.
- And:
    - Average daily calorific intake.
    - Number of weekly hours spent performing sedentary activities.
    - Number of weekly hours spent performing moderate activities.
    - Number of weekly hours spent performing vigorous activities.

.

Subjects: 179 patients.

Results:

Table 5.31: Associations between age at diagnosis and other environmental variables:

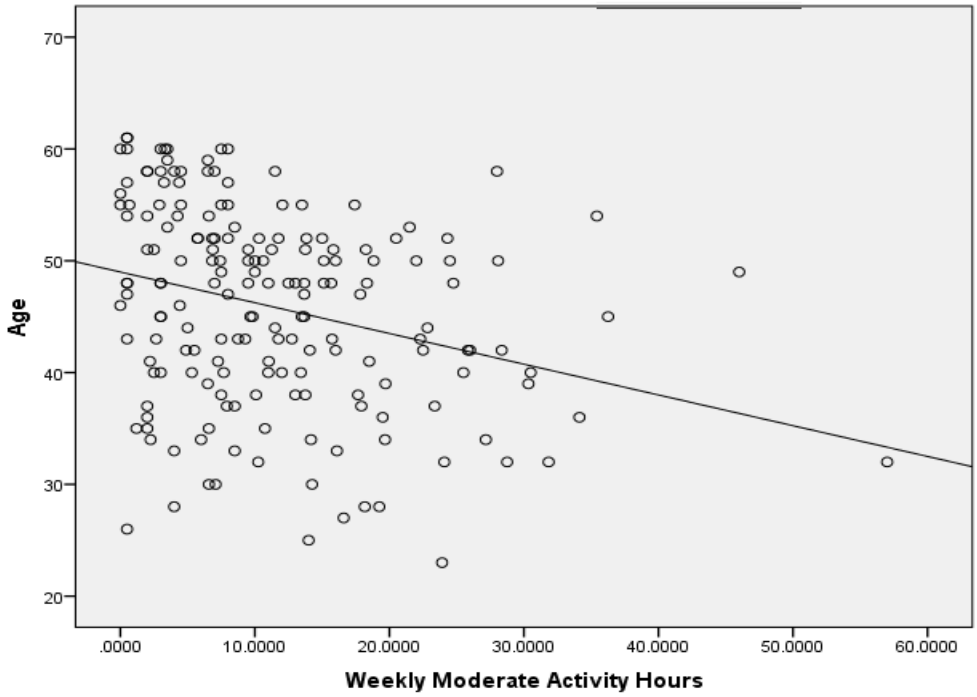
Association	Test Type	Coefficient	p-value
Age at diagnosis and average caloric intake	Pearson's coefficient	-0.114	0127
Age at diagnosis and sedentary activity	Pearson's coefficient	-0.250	0.001
Age at diagnosis and moderate activity	Spearman's coefficient	-0.150	0.045
Age at diagnosis and vigorous activity	Spearman's coefficient	-0.209	0.005

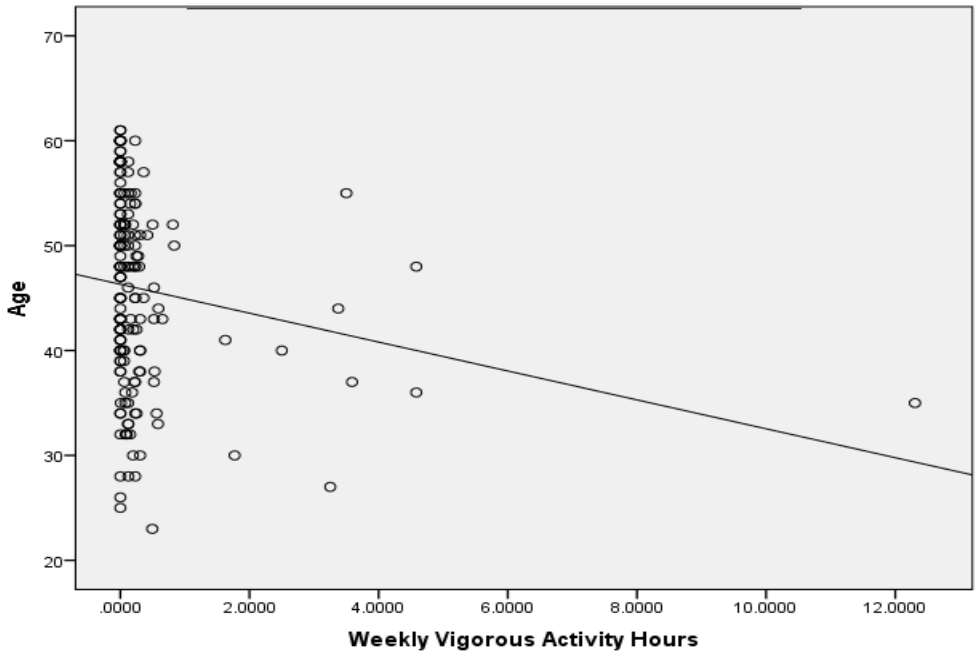
Interpretation:

Table 5.31 shows no statistically significant association between calorific intake and age at diagnosis. This could be largely explained by the tendency of patients to report having a healthier lifestyle based on their GPs recommendations. This remark was observed when comparing calorific intake in patients with the intake in healthy siblings. Healthy siblings were apparently not following any dietary restriction and therefore were likely to report having higher calorific intake compared to patients.

Parameters of physical activity witnessed significant association with age at diagnosis. The inverse association between level of sedentary activity and age at diagnosis is contradicting the observations of association between age at diagnosis and moderate and vigorous activity. The detected inverse association between age at diagnosis and sedentary activities might be biologically plausible. Those who are likely to engage in a relatively sedentary lifestyle are likely to develop the disease earlier compared to those engaged in a more active one. However, correlations with moderate and vigorous activity suggest a tendency for those who are active to have a higher risk of earlier diagnosis of the disease. However, this contradicting association could be explained by observing the association between levels of moderate and vigorous activities and age. Figure 5.25 suggests that younger patients are more likely to report engaging in physical activities compared to older patients.

Association between age and level of physical activity found in our study was detected by another recent study conducted in Saudi Arabia. A study by Amin et al. measured the level of physical activity in a sample of 2176 adult Saudi subjects. In this study, they calculated the number of minutes spent performing moderate or vigorous activities in addition to a third category of those with low levels of activity. In this study, they showed a statistically significant reduction in the level of vigorous activities with increased age (p-value 0.001) **(87)**.





Figures 5.25: Correlations of moderate and vigorous activates with age in patients (N = 179)

5.2.3.8 Assessing the association between measured environmental factors and anthropometric variables

Aim: To study the association between measured environmental variables and BMI, waist circumference and FBG. Assessing the association might give an indication of performance of measured environmental variables as predictors of anthropometric variables.

Analytic Methods: Simple linear regression was firstly conducted to assess the association between dependent and independent variables on an individual basis. Then, multiple linear regression was utilized to assess the association between measured anthropometric variables and all measured environmental variables and age.

Variables:

- Dependent variables: BMI, waist circumference and FBG.
- Independent variables:
    - Age.
    - Average daily calorific intake.
    - Number of weekly hours spent performing sedentary activities.
    - Number of weekly hours spent performing moderate activities.
    - Number of weekly hours spent performing vigorous activities.

.

Subjects: 183 healthy subjects for BMI and waist circumference and 154 for fasting blood glucose.

Results:

Table 5.32: Simple linear regressions of associations between environmental factors and anthropometric variables:

Associations	BMI	Waist circumference	FBG
Age	β: -0.02, p-value: 0.675	**β: 0.21, p-value: 0.04**	**β: 0.686, p-value: 0.000**
Average daily calorific intake	**β: 0.001, p-value: 0.047**	β: 0.00, p-value: 0.624	**β: -0.004, p-value: 0.005**
Weekly sedentary hours	β: -0.018, p-value: 0.446	β: -0.029, p-value: 0.565	**β: -0.208, p-value: 0.007**
Weekly moderate activity hours	β: -0.027, p-value: 0.551	β: 0.024, p-value: 0.809	β: -0.099, p-value: 0.478
Weekly vigorous activity hours	**β: -0.305, p-value: 0.037**	**β: -1.104, p-value: 0.000**	β: -0.887, p-value: 0.170

Multiple linear regression revealed several associations. The association between BMI and independent variants revealed only a single statistically significant association with level of vigorous activity (β: -0.510, p-value: 0.006) (Figure 5.26). Similarly, association between waist circumference and independent variants only showed statistically significant association with level of vigorous activity (β: -1.519, p-value: 0.001) (Figure 5.27). In both situations, age was not significantly associated with BMI or waist circumference. In contrast, FBG was not associated with any of the measured environmental variables and was only associated with age (β: 0.584, p-value: 0.001) (Figure 5.28).

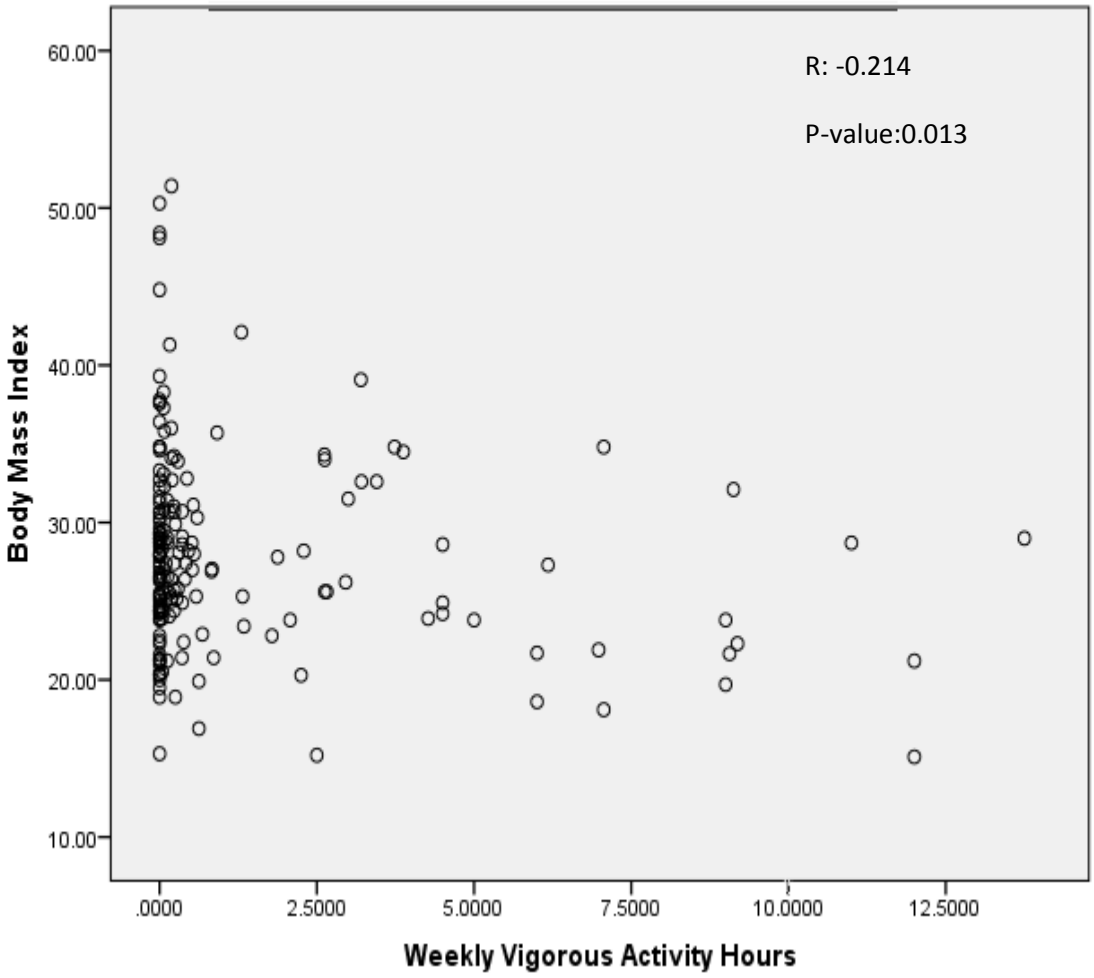


Figure 5.26: Association between number of hours spent performing vigorous activities and BMI.

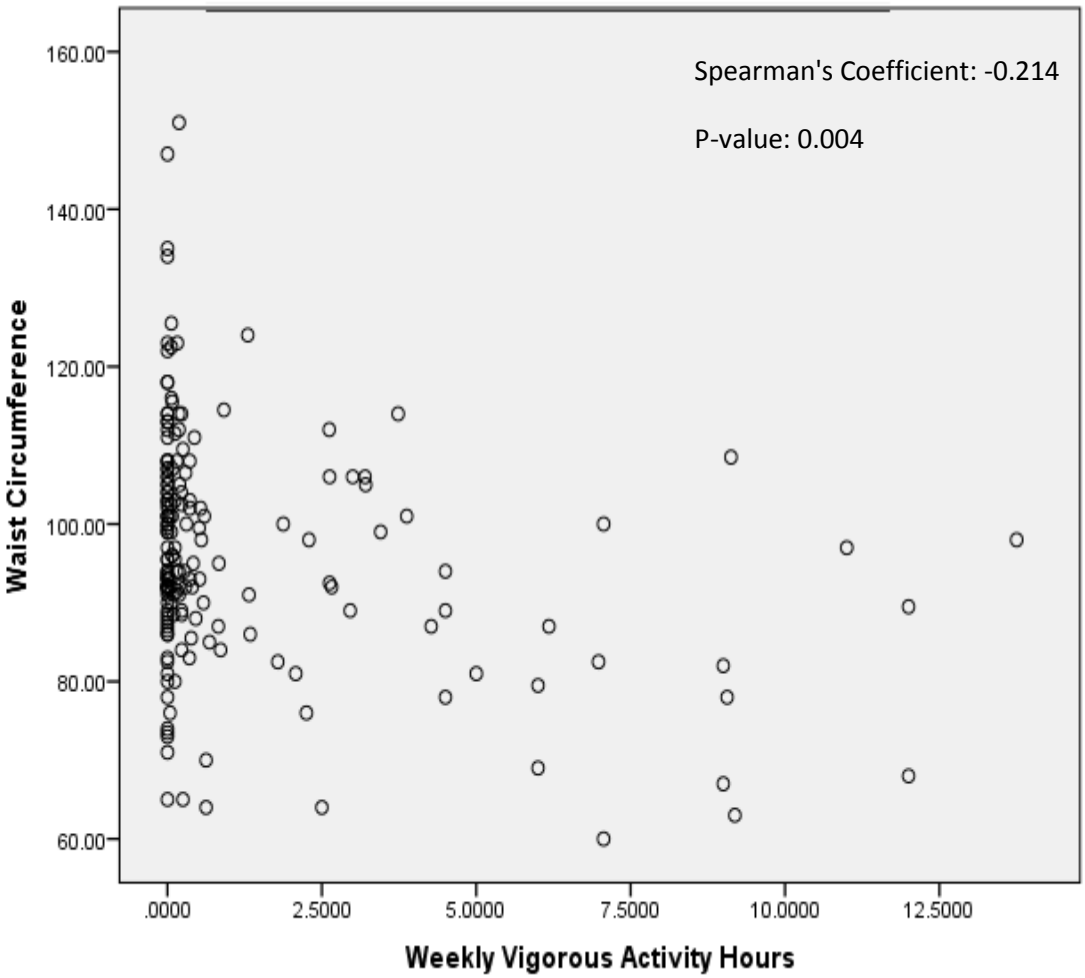


Figure 5.27: Association between number of hours spent performing vigorous activities and waist circumference.

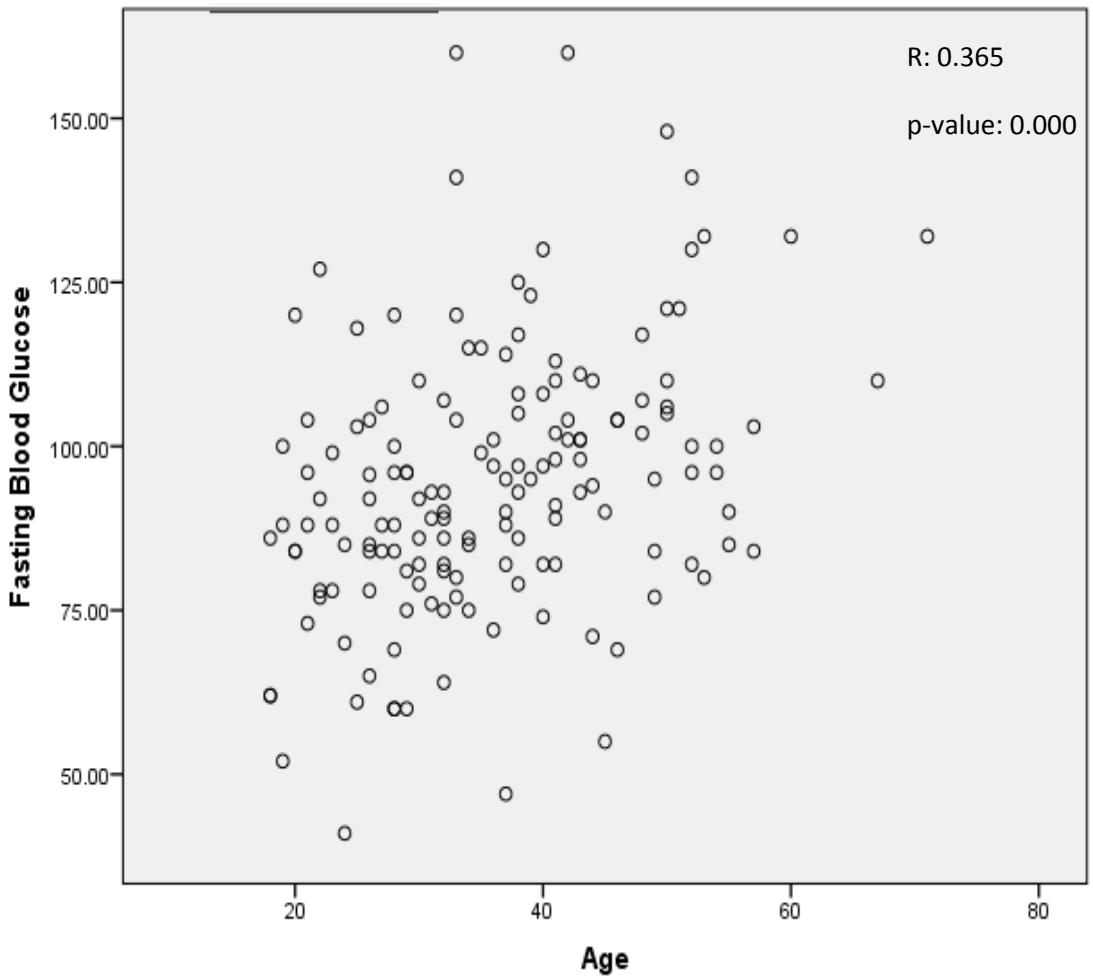


Figure 5.28: Association between age and FBG.

Interpretation:

This analysis indicates that measured calorific intake did not provide a significant influence on anthropometric measurement. Explanations for a lack of this association could involve related difficulties in estimating portion size or lack of power of this sample. However, the vigorous activity coefficients reach statistical significance with BMI and waist circumference. Failure to find statistically significant association with sedentary and moderate activities might indicate that, unlike vigorous activity, less intense activities are not strong enough to cause variation in BMI or waist circumference in this sample. FBG witnesses a different direction where multiple regression indicated that only age was significantly correlated and none of the environmental variables contributed to the variation of FBG.

5.2.3.9 Correlations between anthropometric variables

Aim: Increased BMI and waist circumference are known to increase the risk of type 2 diabetes. We are aiming to investigate whether an increase in BMI and waist circumference is associated with an increase in FBG in the study's sample.

Analytic Methods: Pearson's correlation coefficient.

Variables:

- BMI.
- Waist circumference.
- FBG.

Subjects: 183 healthy subjects for BMI and waist circumference and 154 for FBG.

Results:

Table 5.33: Association between anthropometric variables:

Association	Pearson's Correlation Coefficient	Significance	Sample Size
BMI and FBG	0.028	0.729	154
Waist circumference and FBG	0.093	0.252	154
BMI and waist circumference	0.927	0. 000	182

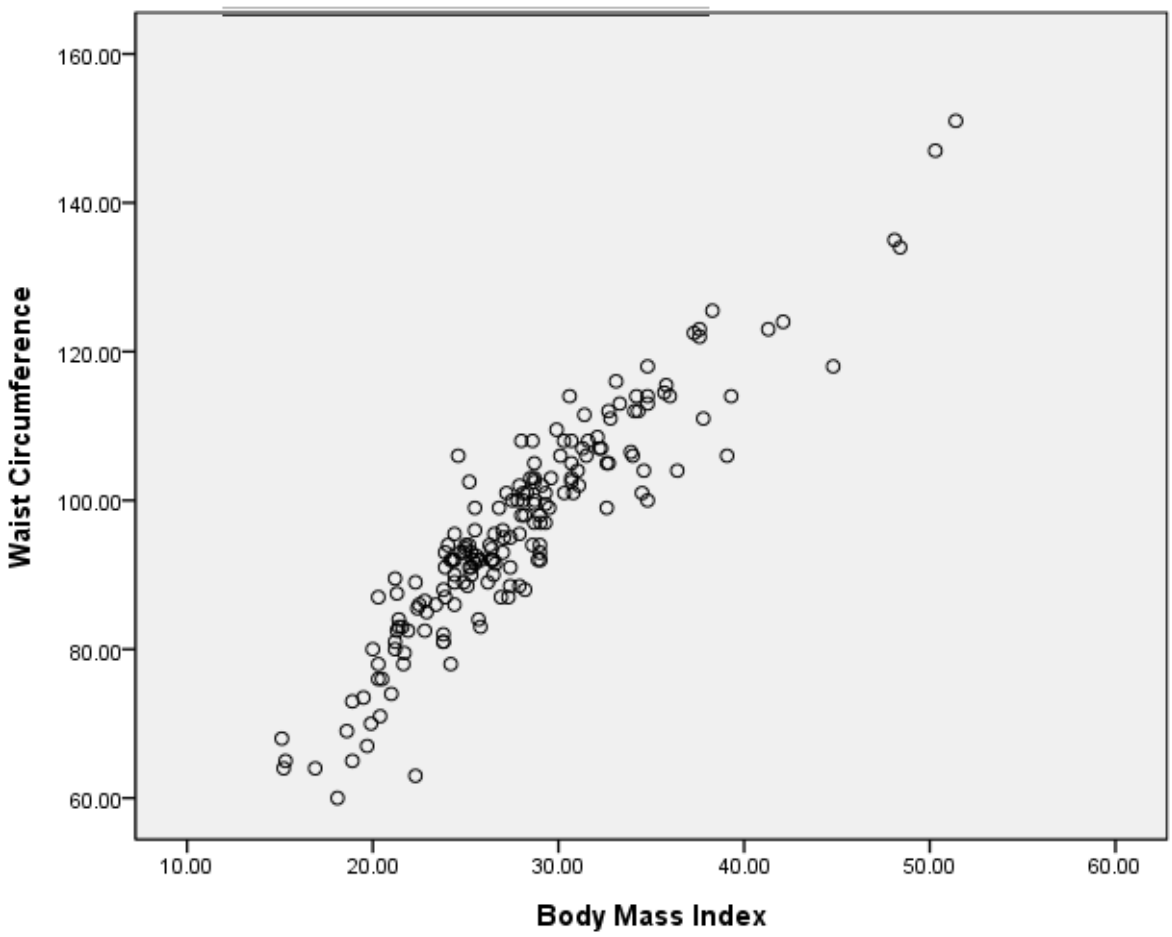


Figure 5.29: Scatter plot of the association between body mass index and waist circumference.

Interpretation:

No significant association was detected between FBG and waist circumference or BMI. However, a strong correlation was detected between BMI and waist circumference.

5.2.3.10 Risk of aggregation of T2D risk alleles according to parental history of diabetes

Aim: To investigate whether having both parents affected with type 2 diabetes is associated with increased risk of detecting loci homozygous for risk alleles or overall number of detected risk alleles. We assume that offspring with both parents affected with diabetes are likely to be associated with increased aggregation of risk alleles compared to offspring with no parental history.

We used two categories of parental history to investigate the possible effect of utilizing two extreme ends of parental history of diabetes on aggregation of risk alleles. Additionally, we are assuming that genetic effect contributed when both parents are affected might be stronger compared to having only one parent affected. Finally, having both parents affected is more likely to induce a possible recessive effect compared to having only one parent affected.

Analytic Methods: In the first stage, multiple linear regression was utilized to study the association between parental history of diabetes and number of loci identical for risk alleles and number of all detected risk alleles. Secondly, logistic regression was utilized to assess if parental history of diabetes has an effect on increasing aggregation of risk alleles of individual SNPs. Parental history of diabetes was categorized into having both parents affected with the disease or neither parents affected.

Variables:

- Dependent variables:
  - Number of loci homozygous for risk alleles.
  - Number of all detected risk alleles.
- Independent variables:
  - Both parents affected with diabetes.
  - Neither parent affected with diabetes.

Subjects: 159 patients and their healthy siblings.

Results:

There was no significant association between number of loci homozygous for risk alleles or number of all risk alleles and parental history of diabetes (Table 3.10.1).

Table 5.34: Association between parental history of diabetes and aggregation of risk alleles:

	Number of loci homozygous for risk alleles	Number of all risk alleles
Both parents affected (n = 54)	β: 0.242, P-value: 0.432	β: 0.588, P-value: 0.246
Neither parents affected (n = 105)	β: 0.143, P-value: 0.561	β: - 0.002, P-value: 0.995

Logistic regression showed only one statistically significant increased risk of detected loci identical for risk alleles at one SNP when both parents are affected in addition to another SNP, which had a marginal statistical significance. Similarly, only two significant associations witnessed increased risk of detecting a higher number of risk alleles in offspring with both parents affected (Table 5.35).

Table 5.35: Odds of inheriting homozygote risk alleles based on parental history of type 2 diabetes:

SNP	Frequency		Homozygote alleles				OR [95% CI]	P value	All alleles				OR [95% CI]	P value
	Effect	Other	Effect		Other				Effect		Other			
			B*	N**	B*	N**			B*	N**	B*	N**		
rs2191349	59.4%	40.5%	25	40	5	26	3.2[1.1-9.5]	0.002	75	121	35	35	1.6[1.01-2.67]	0.04
rs2237895	49.8%	50.1%	16	22	12	39	2.3[0.9-5.8]	0.06	62	91	54	125	1.5[1.01-2.48]	0.04

***Both parents reported with history of type 2 diabetes.**

**** Neither of the parents reported with history of type 2 diabetes**

Interpretation:

This analysis indicates higher odds of inheriting loci identical for risk alleles or higher number of risk alleles when both parents are reported to be affected in two SNPs. However, as 23 SNPs were tested and 23 hypothesis tests were conducted, we would expect 5% of these to be significant under the null hypothesis. A study which investigated the risk of inheriting risk alleles of 16 SNPs related to type 2 diabetes concluded that there is a tendency for subjects with both parents affected with diabetes to inherit a higher number of risk alleles compared to those with no history of parental diabetes or where only one parent was affected (p trend = 0.0004) **(278)**. Although we only observed the effect of paternal history of diabetes on two SNPs, the findings of this study are consistent with ours indicating significant effect of parental history on inheriting risk alleles of SNPs and eventually increasing the risk of type 2 diabetes.

5.2.3.11 Hardy-Weinberg Equilibrium (HWE) test

Aim: We made an effort to investigate whether allelic frequencies of selected SNPs fall under HWE or not. If there were a violation of the equilibrium and deviation toward having higher frequency of homozygote genotypes, this would suggest that genotypes frequency could be affected by inbreeding observed in the Saudi population.

Analytic Methods:

HWE states that alleles and genotype frequencies in a population will remain constant between generations if mating between subjects is random and there are no other forces such as selection, mutation and migration affecting the population. Based on allelic frequency in any population, genotypes frequency under HWE can be calculated. If we assume that two alleles, A and B, with allelic frequency of a and b, then we would conclude that under equilibrium, genotype frequency of AA genotype is a^2, frequency of genotype AB is 2ab, and genotype frequency of BB is b^2. If the expected frequency of genotypes under HWE are different than observed frequencies, then we would conclude that the population of interest is not exhibiting HWE and there is a force deviating allelic and genotypes frequencies (110).

HWE calculator was utilized to assess maintenance of equilibrium (Michael H. Court (2005-2008) HWE calculator (279)) .If the p-value of difference between observed and expected frequency is less than 0.05, allelic frequency is then presumed to have deviated from HWE. Otherwise, if the p-value is higher than 0.05, then we would conclude that allelic frequency of the SNP of interest is maintaining HWE.

Variables:

- Frequency of homozygote genotypes for risk alleles.
- Frequency of homozygote genotypes for none-risk alleles.
- Frequency of heterozygote genotypes.

Subjects: 339 patients and their healthy siblings.

Results:

Table 5.36: HWE of selected alleles of the whole sample:

No.	SNP	Observed frequency	Expected frequency	P value
1	**rs10946398**	(A:A) **188** (C:A)**122** (C:C) **24**	185.6 126.7 21.6	0.49
2	**rs10811661**	(T:T) **223** (T:C) **102** (C:C) **13**	222.1 103.8 12.1	0.75
3	**rs1111875**	(C:C) **223** (T:C) **83** (T:T) **26**	210 107 13	**0.000032**
4	**rs13266634**	(C:C) **255** (T:C) **68** (T:T) **7**	253.1 71.8 5.1	0.33
5	**rs7754840**	(C:C) **24** (G:C) **125** (G:G) **190**	22.1 128.9 188.1	0.58
6	**rs560887**	(C:C) **199** (T:C) **116** (T:T) **20**	197.2 119.7 18.2	0.57
7	**rs2191349**	(T:T) **130** (T:G) **130** (G:G) **68**	115.9 158.1 53.9	**0.0012**
8	**rs10830963**	(G:G) **20** (G:C) **102** (C:C) **206**	15.4 111.3 201.4	0.13
9	**rs7034200**	(A:A) **132** (C:A) **152** (C:C) **45**	131.5 153 44.5	0.905

10	**rs174550**	(T:T) **226** (T:C) **91** (C:C) **7**	227.5 88 8.5	0.53
11	**rs11708067**	(A:A) **257** (A:G) **74** (G:G) **2**	259.6 68.9 4.6	0.17
12	**rs4607517**	(A:A) **29** (A:G) **102** (G:G) **199**	19.4 121.2 189	**0.0039**
13	**rs7903146**	(T:T) **50** (T:C) **148** (C:C) **136**	46 155.9 132	0.35
14	**rs1799884**	(T:T) **34** (T:C) **115** (C:C) **190**	24.7 133.6 180.7	**0.01**
15	**rs5219**	(T:T) **23** (T:C) **94** (C:C) **218**	14 110.7 209.6	**0.005**
16	**rs7756992**	(G:G) **22** (A:G) **135** (A:A) **179**	23.8 131.3 180.8	0.607
17	**rs780094**	(C:C) **143** (T:C) **131** (T:T) **59**	130.5 155.9 46.5	**0.003**
18	**rs2237897**	(C:C) **328** (T:C) **7** (T:T) **1**	327.1 8.9 0.1	**0.000**
19	**rs2237895**	(C:C) **90** (T:C) **151** (T:T) **91**	113.2 104.6 24.2	**0.000**
20	**rs1387153**	(T:T) **31**	29.7	0.73

		(T:C) **136** (C:C) **163**	138.6 161.7	
21	**rs231362**	(G:G) **116** (A:G) **153** (A:A) **58**	113.3 158.4 55.3	0.54
22	**rs7957197**	(T:T) **253** (T:A) **71** (A:A) **10**	249.2 78.6 6.2	**0.077**
23	**rs757210**	(A:A) **91** (A:G) **140** (G:G) **96**	107.6 106.9 26.6	**0.000**

Interpretation:

By observing Table 5.36, we can detect eight SNPs deviating from HWE with p-values lower than 0.05. With exception of five SNPs, all remaining 18 SNPs witness a reduction of observed number of heterozygote genotypes compared to expected frequencies. This might indicate that in our sample there is an overall higher tendency of deviation from heterozygote genotypes to homozygote genotypes, which is in agreement with the observed rates of consanguinity in the Saudi Arabian population.

HWE exists at the population level and is frequently conducted as a quality control for genome-wide association studies to assess population stratification and genotyping errors. However, in this study, we were trying to investigate potential deviation from equilibrium induced by the high levels of consanguinity observed in the Saudi Arabian population. Testing for HWE in this study revealed deviation towards increased level of homozygosity among several SNPs. Although a random sample of the population may have a chance of including siblings, this study, however, has selectively sampled siblings from several families.

Although we observed a deviation towards increased level of homozygosity, this deviation could be due to the fact that the sample included siblings who have a chance of having similar genotypes. However, since subjects were recruited from different families, we would expect the skew in the distribution of genotypes due to sampling through siblings to be balanced.The fact that siblings have been systematically recruited means that the sample is not random and hence the distribution of the chi-square statistic under the null is compromised. Moreover, the sample contains pairs of relatives which are correlated by genotype, and hence we do not have full independence in the sample owing to the sibling clustering.

To remove the clustering induced by the sibling sampling, we performed HWE tests on patients who are not related. We did not do the same with controls because, unlike the patients group, several healthy participants were siblings as we recruited two healthy siblings from 51 families. Fewer SNPs show a statistically significant deviation from HWE in the cohort restricted to patients only compared with the cohort including all subjects. The reduction of statistically significant deviations will be due in some part to the smaller sample size of recruited unrelated patients. Among the unrelated patients cohort and apart from a few SNPs, there is an overall tendency to observe a reduced number of expected heterozygote genotypes. Based on these observations, it is possible to assume that deviation towards homozygosity in the sample of related individuals is similar to the observed deviation among non-related patients.

5.2.3.12 Direction of homozygosity

Aim: In the previous section, we were able to observe a tendency for our sample to deviate toward higher frequencies of homozygote genotypes compared to expected frequencies calculated by the HWE test. In this analysis we are investigating whether there is a higher tendency in our sample to observe higher frequency of loci homozygous for risk alleles compared to loci homozygous for non-risk alleles.

Analytic Methods: Mean and standard deviations and t-test to assess presence of statistically significant difference.

Variables:

- Number of all homozygote loci.
- Number of loci homozygous for risk alleles.
- Number of loci homozygous for non-risk alleles.

Subjects: 316 patients and their healthy siblings.

Results:

Table 5.37: Means and standard deviations of number of homozygote loci:

	Mean	Std. Deviation
Number of all homozygote loci	14.99	2.821
Number of loci homozygous for risk alleles	9.05	1.946
Number of loci homozygous for non-risk alleles	5.94	2.189

Table 5.38: T-test to measure difference of means between number of loci identical for risk alleles and non-risk alleles:

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Loci homozygous for risk alleles	82.665	315	.000	9.051	8.84	9.27
Loci homozygous for none- risk alleles	48.221	315	.000	5.937	5.69	6.18

Interpretation:

This analysis suggests a deviation toward observing higher frequencies of homozygote genotypes for risk alleles compared to homozygote genotypes for non-risk alleles. In this sample, we observed a deviation of HWE toward observing higher frequency of homozygote genotypes. We can argue that inbreeding in this sample, in addition to increasing overall homozygosity, could have contributed to increased frequency of genotypes homozygous for risk alleles.

Inbreeding does not selectively increase frequency of genotypes homozygote for risk alleles. Inbreeding would increase frequency of homozygote genotypes despite the risk status of alleles of interest. However, we might argue that the Saudi Arabian population, as explained earlier in the background chapter, might have witnessed a thrifty genotype effect leading to higher frequencies of risk alleles due to selective forces. If we assume that selection could be responsible for passing certain risk alleles for type 2 diabetes to the following generations, we can argue that inbreeding of the subsequent generations might have further contributed to increased frequency of loci homozygous for selected risk alleles.

In this analysis, we combined all patients and healthy siblings. This might seem to violate the concept of measuring frequencies of risk genotypes assuming that patients are likely to have higher frequencies of risk alleles compared to their healthy siblings. However, we conducted a t-test to measure the difference in means of risk alleles number and there was no statistical difference between patients and healthy subjects. We further examined if there is a systematic difference in risk alleles frequencies between patients and healthy siblings of each family and we also observed no significant difference between siblings. We must acknowledge that failing

to find a significant difference of risk alleles frequencies between siblings is expected hence siblings are likely to share half of their alleles including type 2 diabetes risk alleles. Additionally, healthy subjects at the time of the interview could be at a risk of developing the disease later in their lives.

5.3 Conclusion

- The study sample might have not been powered sufficiently to detect statistically significant cumulative recessive effect of measured SNPs on age at diagnosis and other type 2 diabetes risk factors.
- Individuals with higher inbreeding coefficients are likely to be associated with earlier age at diagnosis. This might indicate a possible cumulative recessive effect, which could not be detected by measured SNPs.
- The significant association between inbreeding coefficients and age at diagnosis was observed to be influenced by parental history of diabetes. Among patients with positive paternal and maternal history of the disease, patients with higher inbreeding coefficients were likely to be diagnosed earlier with the disease compared with patients with lower inbreeding coefficients.
- No statistically significant associations were detected between inbreeding coefficients and BMI, waist circumference and fasting blood glucose.
- Siblings are likely to share similar eating habits but are less likely to witness similar physical activity levels.
- Siblings with higher inbreeding coefficients are likely to have similar FBG compared to siblings with lower inbreeding coefficients.
- Measured environmental variables are less likely to have an influence on age at diagnosis.
- Levels of vigorous activity are more likely to influence anthropometric traits compared to other levels of physical activities and calorific intake.
- The HWE test of our sample suggests a deviation toward a higher level of observed homozygote genotypes compared to expected frequencies.
- There was a higher tendency to observe loci homozygote for risk alleles compared to loci homozygote for non-risk allele.

6 Chapter Six: Discussion

6.1 Association between consanguinity and type 2 diabetes

Synergism between consanguinity and type 2 diabetes has been suggested by several studies. The main aim of this study was to explain the mechanism behind this association. We hypothesized that higher levels of consanguinity might be associated with higher aggregation of loci homozygous for risk alleles and subsequently higher overall aggregation of risk alleles. Furthermore, we aimed to study the possible cumulative recessive effects of measured loci by studying associations between the number of loci homozygous for risk alleles and age at diagnosis, BMI, waist circumference or FBG.

Our investigation was not able to find any recessive effect of measured SNPs on severity or risk factors of type 2 diabetes. Associations between the number of loci homozygous for risk alleles and variables related to severity or risk factor of the disease were not statistically significant (age at diagnosis: β: -0.106, p-value: 0.762; BMI: β: 0.221, p-value: 0.438; waist circumference: β: -0.278, p-value: 0.658; FBG: β: 0.448, p-value: 0.652]. If there really is a causal association between inbreeding coefficients and severity of type 2 diabetes, then this suggests there may be many more type 2 diabetes risk loci all acting with weak recessive effects. Inbreeding alone cannot induce an association with diabetes risk if the genetic effect is additive only. Although we had selected the 23 SNPs with the strongest marginal effects (detected through GWAS) our study lacked power to detect the cumulative evidence of these weak recessive effects. However, as we see an association between the inbreeding coefficients and age at diagnosis this suggests many more recessively acting loci may contribute to the type 2 diabetes risk.

Additionally, we examined the association between the number of detected risk alleles (contributed from homozygote loci for risk alleles or heterozygote loci) and traits related to type 2 diabetes. However, no significant associations were detected between the overall number of risk alleles and other traits. This suggests that our sample and measured SNPs, in addition to not having the power to detect recessive effect, were not sufficient to detect any additive effect of risk alleles on traits related to type 2 diabetes.

Only one association between age at diagnosis and total number of risk alleles became marginally significant after accounting for parental history of diabetes and inbreeding coefficients (β: -0.399, p-value: 0.052). This might indicate that measured SNPs were not sufficient to detect significant genetic effect on age at diagnosis and considering other genetic factors such as parental history of the disease and consanguinity history were important elements.

Our study suggests that with every increase in the number of inherited risk alleles there is a tendency toward earlier diagnosis with type 2 diabetes. These finding are consistent with the findings of another study suggesting increased odds of developing type 2 diabetes when inheriting six or more risk alleles compared to those who inherit four or fewer risk alleles (OR: 2.05 95%, CI: 1.50–2.80). Furthermore, the same study also suggested that with every increase in the number of inherited risk alleles, there is a reduction in age at diagnosis β= −0.46 (−0.80 to −0.11) years, p-value = 0.009 (198).

Failure to find an association between the aggregation of risk alleles and risk factors such as BMI, waist circumference and FBG might stem from two reasons. Firstly, selected SNPs were reported to increase the odds of developing type 2 diabetes, to increase levels of fasting blood glucose, or to affect insulin function. However, not being able to find a significant association between aggregation of risk alleles and risk factors of the disease might be due to the stronger effect of unmeasured alleles which might be more inclined to independently effect those traits rather than affecting type 2 diabetes. Secondly, the low number of healthy participants (n = 159) did not enable us to detect significant associations.

The sample size for the first and third components of the investigation was smaller than the planned one. In the first and third questions, achieved powers were 69% and 60% respectively. Owing to changing inclusion criteria of patients, the power of the second component of the investigation was 81%. Given the achieved sample size of the study, the study would have been able to detect correlation coefficients of 0.14 for the first question, and 0.19 for the second and third questions, assuming a two-sided type 1 error of 5% and 80% power to detect a true correlation. However, the effects detected for all questions were very small compared with the assumed effects. Detected effect sizes varied between - 0.036 and 0.06 for the three main questions where assumed coefficients varied between 0.12 and 0.2.

P values and 95% Confidence intervals of detected effects are illustrated as below:

1st question: Spearman rho: 0.01 p-value 0.856 (95% CI: - 0.104 – 0.117).

2nd question: Pearson coefficient: 0.024 p-value: 0.762 (95% CI: - 0.212 – 0.160).

3rd question: BMI: Pearson coefficient: 0.062 p-value: 0.438 (95% CI: - 0.130 –0.243).

WC: Pearson coefficient: - 0.036 p-value:0.658 (95% CI: - 0.207 – 0.128).

FBG: Pearson coefficient: 0.039 p-value:0.652 (95% CI: - 0.156 – 0.220).

We can observe that all the confidence intervals cross the zero value, indicating no detected association between measured variables. Additionally, the wide range of the confidence intervals is due to the small size of the study sample. Apart from the first and third components of the third question, correlation coefficients assumed during estimation of sample size lie outside the detected confidence intervals. For example, the assumed effect size for the second question was 0.2 and the detected 95% confidence intervals of the study were-0.212to0.160.Based on this evidence, a true population value is not likely to be as high as 0.2 and is likely to be equal to or less than 0.160.Therefore, it might be possible to argue that true population correlation coefficients are likely to be smaller than coefficients assumed for sample size calculation,and based on that a larger sample size will be needed to detect statistically significant smaller effects.

Although 23 SNPs with the strongest and most replicable associations in different populations were chosen, their combined effect was not detectable in this study. It is possible to argue that utilization of a larger number of type 2 diabetes risk SNPs with higher frequency in the targeted populationmay have resulted in a better detectable combined effect and larger sample size would have higher power to detect smaller effect sizes.

Most of risk alleles reported to increase the risk of type 2 diabetes have been revealed to increase risk of the disease in an additive effect. This might seem to contradict our investigation where we utilized 23 of SNPs reported to act in an additive manner to investigate a possible recessive effect. Although this study failed to detect statistically significant recessive effects of measured SNPs on traits related to type 2 diabetes, there might be explanations behind this failure.

A recent study by McQuillan et al, published in 2012 revealed evidence of a recessive effect on height. This study recruited about 35,000 participants from 21 European populations and used genome wide homozygosity to investigate recessive effect on height. Despite observed heterogeneity of association between height and genome-wide homozygosity among different populations, there was a statistically significant inverse association suggesting a reduction of height with increased level of homozygosity.

Height is a complex continuous trait affected by genetic and environmental variables. However, its polygenic nature is similar to the nature of type 2 diabetes risk traits which our study investigated. The concept of the study by McQuillan et al. is similar to ours where homozygosity was utilized to assess its effect on continuous traits. However, this study was able to find this effect by observing between 295,000 to 318,000 SNPs where in our study only 23 SNPs incurring a risk of type 2 diabetes were utilized. Type 2 diabetes genetic risk markers,

which have been reported by GWAS studies to act additively, only explain about 10% of diabetes heritability. This might indicate that there could be a possible recessive effect and finding this effect in our study might have required a larger sample and a larger number of SNPs.

Although the study by McQuillan et al. was able to find a statistically significant inverse association between height and homozygosity, we must acknowledge an important point. This study only suggested a significant inverse association with height based on an observed number of homozygote loci. However, unlike our study, the observed thousands of homozygote genotypes are not necessarily known to have an effect on height. This study did not investigate whether there is a possible recessive effect of SNPs, reported by previous GWAS studies, to affect human height in an additive nature. This point could imply that this study might have investigated the effect of unknown risk markers of human height by combining thousands of homozygote genotypes in a single analysis instead of analysing the effect of SNPs individually as in commonly used additive models. This study was able to suggest an inverse association between genome-wide homozygosity but could not simply identify which loci are acting in a recessive manner to affect height.

6.2 Key findings from secondary analysis

6.2.1 Consanguinity and age at diagnosis

Our study provided several items of evidence suggesting a significant effect of inbreeding on an increased risk of type 2 diabetes. A Spearman rank correlation coefficient of -0.187 (p value: 0.012), measuring the association between inbreeding coefficients and age at diagnosis, indicates a significant effect of the degree of consanguinity on the risk of developing type 2 diabetes at a young age. Additionally, our analysis suggests a significant effect of the magnitude of consanguinity on correlations of FBG between siblings.

Significant negative correlations between age at diagnosis and inbreeding coefficients suggest that inbreeding history might increase the chance of developing the disease at a young age. Additionally, observing a marginal significant effect of measured detected type 2 diabetes risk alleles on reducing at age diagnosis only when controlling for parental history of the disease and inbreeding history, might indicate that inbreeding coefficients might have accounted for aggregation of unmeasured risk alleles which might have a recessive effect on age at diagnosis.

These findings are consistent with other studies suggesting an increased risk of type 2 diabetes with increased degree of inbreeding (245, 246, 280). However, these studies only measured the odds of the disease using consanguinity as binary value. Our study developed more in-depth analysis of the association between continuous measurement of consanguinity history and other variables related to type 2 diabetes with adjustment for environmental factors.

6.2.2 Dominance effect on correlation of FBG

Our study detected significant association of FBG between siblings (r = 0.317 P value 0.040). Since we are comparing between full siblings, this correlation could be attributed to correlation due to similar inherited alleles (additive effect), correlation due to similar inherited genotypes (dominance effect) or because of correlation due to shared environmental factors. However, once we measured correlation of FBG between siblings based on a limited range of inbreeding coefficients, we noticed an increase in correlation with increased values of inbreeding coefficients.

An increased correlation of FBG with increased level of inbreeding coefficients might indicate a dominance effect. Higher levels of inbreeding might increase the chance of inheriting similar alleles at a particular loci. Additionally, even when we controlled for the effect of environmental variation between siblings, the effect of inbreeding coefficients on the variation of FBG was still significant (β: - 1.8 P-value: 0.039). The negative regression coefficient indicates that variance between siblings' FBG is reduced with increased level of inbreeding even when we considering the variance of the environmental effect of siblings.

6.2.3 Impact of inbreeding on genotypes frequencies

We made an attempt to compare alleles and genotype frequencies in our sample to frequencies reported in other populations. The table in appendix 13 illustrates allelic frequencies variations in several populations reported by association studies. By observing allelic frequencies of different populations, we can observe a tendency for East Asians populations to have almost similar frequencies of major and minor alleles. There is a tendency for allelic frequencies in our sample to have allelic frequencies more related to European populations compared to East Asians.

The aim of constructing this table was to study possible tendency of increase homozygote genotypes of our sample compared to other populations. Observing the in appendix 13 is not sufficient to conclude whether the genotype frequencies are due to deviations of allelic frequencies toward one direction or because of other factors. In our situation, we are interested to observe if there is a tendency for genotypes of SNPs of interest to deviate toward homozygosity. However, the level of homozygosity under HWE is mainly dependent on the magnitude of deviation toward one allele. For example, if allele A has a higher frequency than allele B, then we would expect the population of interest to have higher frequency of genotype AA compared to BB. Additionally, the frequency of heterozygote genotype AB will be reduced compared to the overall level of homozygote genotypes (AA and BB).

To decide whether the increased level of homozygosity of any SNP in a population was due to deviation toward one allele or due to other forces such as inbreeding, we had to control for the relationship between level of homozygosity and allelic frequencies. Figure 6.1 illustrates the expected level of homozygosity depending on the difference between allelic frequencies under HWE. For example, if alleles A and B have a frequency of 0.5 then under HWE, we would expect frequency of genotypes AA, AB, and BB as 0.25, 0.5 and 0.25 respectively. Then we can conclude that when there is a zero difference in allelic frequencies we would expect half of the population to have homozygote genotypes and the remainder to have heterozygote genotypes.

Using this concept, we constructed a line, as observed in figure 6.1, to indicate expected level of homozygosity compared to allelic frequencies difference under HWE. It can be clearly observed that with an increased difference in allelic frequencies, there will be a tendency to deviate toward higher homozygosity. The area below the line indicates that the level of homozygosity of a given allelic frequency difference is lower than expected. Similarly, the area above the line indicates that for a given allelic frequency difference, the level of homozygosity is higher than expected and there might be a force causing an increased level of overall homozygote genotypes despite observed allelic frequencies.

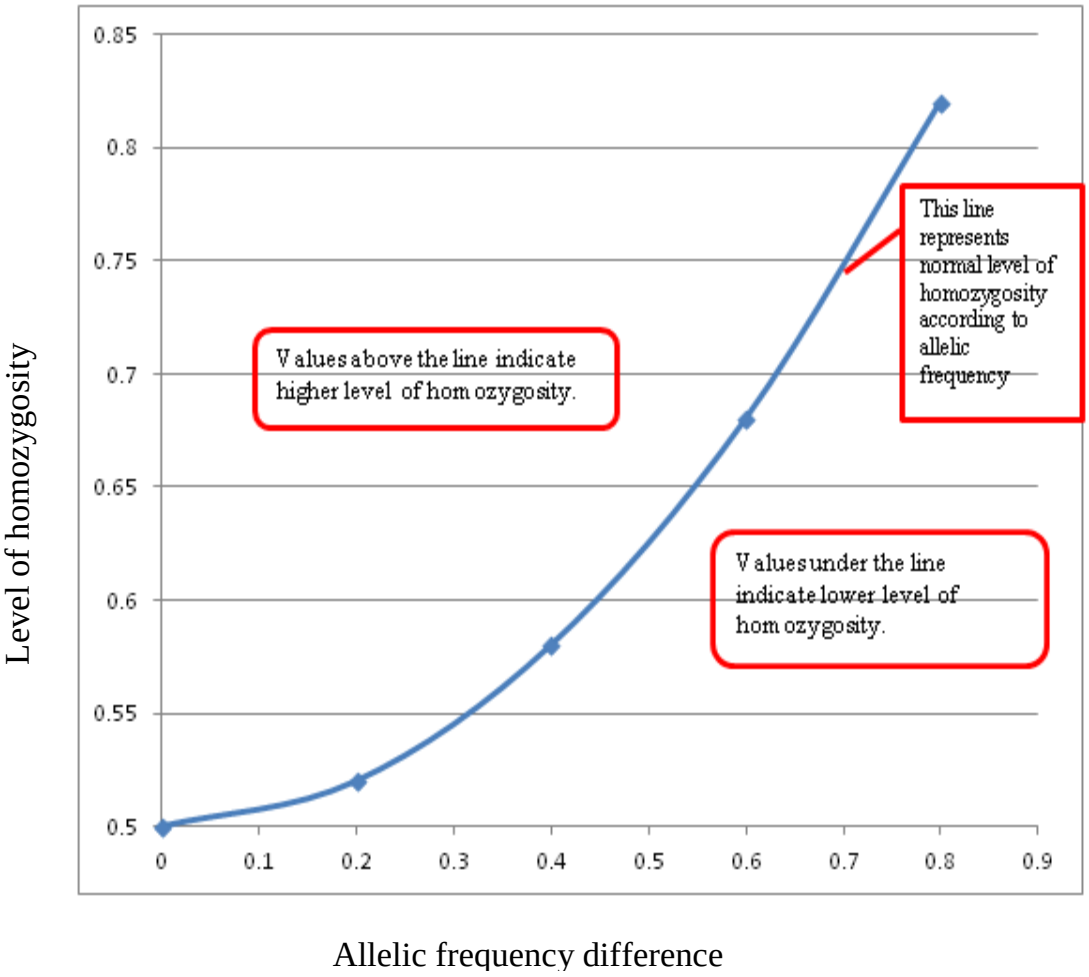


Figure 6.1: Homozygosity Index: association between difference in allelic frequencies and overall level of homozygosity.

In figure 6.2, we fitted values related to allelic frequency difference and homozygosity reported in appendix 13 in the homozygosity index. Each population was given a specific colour. Regarding disparity in allelic frequencies, there is a tendency for East Asian populations to exhibit a lower difference between allelic frequencies compared to other populations. For European populations, there is a spread of allelic frequency difference along the normality line. For Saudi Arabian subjects, there is an observed tendency to deviate from the normality line indicating higher levels of homozygosity after controlling for the effect of allelic frequency difference.

The deviation toward an increased frequency of homozygote genotypes in our sample compared to other populations is expected given the increased level of inbreeding with our population compared to others. However, once we tried to assess the effect of the degree of inbreeding on homozygosity of type 2 diabetes risk alleles of each participants, there was no significant association. There was no observed pattern in our sample to indicate that with an increased level of inbreeding, the number of homozygote genotypes of risk alleles would increase. Lack of association could be attributed by limited number of SNPs to detect an association between inbreeding and risk alleles or perhaps due to homogeneous genetic background of the Saudi Arabian population.

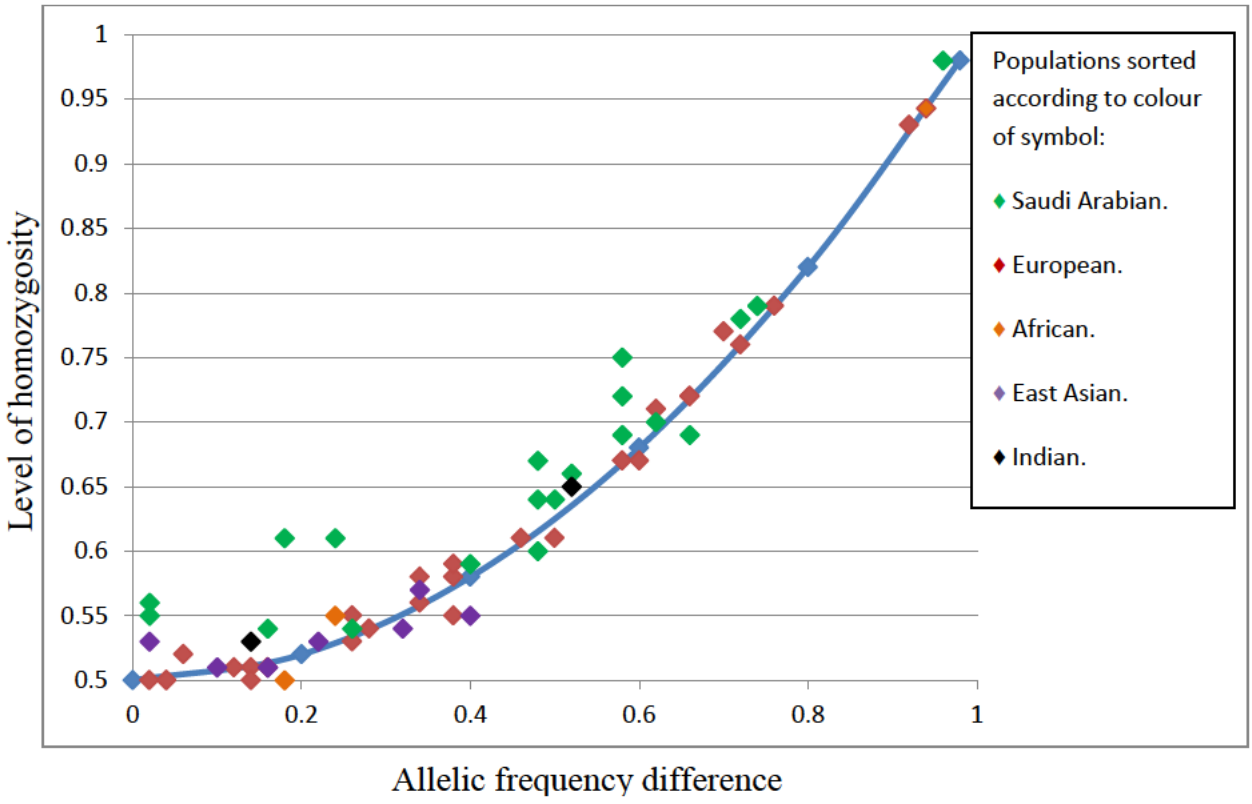


Figure 6.2: Homozygosity Index: after adding populations' frequencies.

6.3 Methodological Findings

6.3.1 Measurement of inbreeding coefficients using tribal and family tree relationships

The overall rate of consanguineous marriage in our sample is 63%. The remaining proportion of participants did not report mating of related parents. Our method of calculating inbreeding coefficients considered mating of individuals from the same tribe or from the same subtribe. Therefore, our method is more comprehensive than other methods measuring inbreeding coefficients in Saudi subjects which have mainly used categorical values of degree of consanguinity.

To precisely evaluate the validity of our method to measure inbreeding coefficients, we should compare estimated inbreeding coefficients of these individuals to genetically measured runs of homozygosity. However, this method of assessing validity was not feasible in this study. Nonetheless, we must acknowledge that our method of calculating inbreeding coefficients produced 57 different inbreeding coefficients based on 179 pedigrees which enabled us to use inbreeding values in a continuous form rather than a categorical one.

It is possible to argue that our method measuring inbreeding degree was able to detect several associations with variables relating to type 2 diabetes. However, calculated inbreeding coefficients failed to correlate with number of homozygote genotypes of selected SNPs. Failing to detect an association, though, between estimated inbreeding coefficients and measured homozygote genotypes could be attributed to several reasons.

By observing the overall distribution of homozygote genotypes in our sample, we observed a deviation toward a higher frequency of homozygote genotypes. Even when we compared the level of homozygosity of our sample to studies which measured similar SNPs in other populations, we noticed several deviations toward higher rate of homozygosity despite allelic frequencies. This might indicate that the expected long history of consanguineous marriages in our population for many generations has resulted in an increased overall frequency of homozygote genotypes.

Finally, failure of calculated inbreeding coefficients to correlate with the number of homozygote genotypes could be related to the low power of number of selected SNPs to detect a significant association. Perhaps, larger number of genome wide random SNPs would have resulted in better prediction of overall homozygosity. However, due to the financial limitation of this study, we had to opt for genotyping only 23 SNPs.

6.3.2 Type 2 diabetes and parental history of the disease

A high risk of developing type 2 diabetes in the circumstance of a family history of the disease have been documented by several studies. These studies suggest that there is an increased risk of developing type 2 diabetes when there is one parent affected. The odds of developing the disease is even higher when both parents are affected by type 2 diabetes (13).

In this study, we found a significant association between the age at diagnosis and the degree of inbreeding. However, once we assess the association in four groups according to parental history of diabetes, the association became stronger in the group which reported both parents to be affected by type 2 diabetes. This finding might not only suggest a high risk of the disease when both parents have a history of the disease but could signify the important effect of consanguinity on aggregating a higher number of risk alleles when both parents are affected by the disease.

We might argue that a higher chance of diagnosis with diabetes is merely influenced by a higher tendency of offspring of affected parents to adopt a behavioural attitude. This attitude forces them to seek earlier investigation of their blood glucose status compared to those with no family history of the disease. This point might be valid but when investigating the association between age at diagnosis and inbreeding history, we controlled for the effect of parental history of the disease by dividing subjects to four different groups. Therefore, the detected significant association between age at diagnosis and inbreeding history in the subjects with both parents affected (Spearman's rank correlation: -0.502 (P-value: 0.003)) is less likely to be biased by a tendency of siblings of parents with diabetes to be diagnosed by the disease at earlier age.

Our analysis indicated a stronger maternal effect on development of type 2 diabetes at an earlier age compared to other categories of parental history (β:-5.538, P value: 0.000). This pattern is supported by several studies suggesting stronger maternal effect of transmission of type 2 diabetes to their offspring (13, 280-283). However, other studies did not find a similar effect (284, 285). Nonetheless, the maternal effect on transmission of type 2 diabetes to offspring might not be purely genetic.

A literature review investigated possible causes leading to increased maternal transmission of type 2 diabetes. This study suggests that the maternal effect on offspring can be influenced by the intrauterine environment. It indicated the higher risk of developing diabetes in siblings who were born after maternal development of diabetes compared to those who were born before maternal development of diabetes. Despite having shared maternal genetic risk, the difference in uterine life between siblings of the same mother might explain part of the variation in the risk of developing type 2 diabetes (286).

Apart from the possible uterine effect, there might be a possible link between a maternal history of diabetes and the lifestyle of the offspring. The cultural background in Saudi Arabia dictates a strong maternal influence on the food intake behaviour of offspring since mothers – unlike fathers- are responsible for culinary activities and provision of food for the family. In other words, if the mother is exhibiting a certain lifestyle associated with increased food intake, the children are likely to exhibit similar behaviour toward higher food intake.

6. 3.3 Associations between measured phenotypes and environmental factors

Our primary aim in utilizing a food frequency questionnaire was to control for the effect of one of the environmental variations, in this case variation of calorific intake. We used a food frequency questionnaire to collect information about variations in frequency of consumption of 137 food items with additional open-ended questions to allow for other items to be added during interviews. In order to calculate calorific values of each food item, we constructed a database based on caloric values provided by a variety of nutritional databases.

Physical activity is an important environmental factor modifying the risk of developing type 2 diabetes and its related risk factors. We utilized a modified version of the EPIC physical activity questionnaire. The original EPIC physical activity questionnaire included several questions which were not applicable to a Saudi population; therefore, several questions had to be removed. The level of physical activity was measured by calculating the number of hours spent performing sedentary, moderate and vigorous physical activities per week.

Several studies have investigated the repeatability and validity of the EPIC food frequency questionnaire and physical activity questionnaires in different populations. The validity of these questionnaires was assessed by studying its construct validity against standard clinical measurements. These studies found significant correlations between estimated nutrients levels from questionnaires and nutrients measured using biochemical markers (287-289). Similarly, a study confirmed significant correlations between the level of estimated energy expenditure using the EPIC physical activity questionnaire and measured level of heart rates (290).

We had to modify the EPIC questionnaires to construct applicable versions of the questionnaires for a Saudi Arabian population. Despite the reported validity of the questionnaires, we cannot confirm whether modified versions of the questionnaires were valid to a Saudi Arabian population. Nonetheless, the association between measured dietary intake and physical activity in our sample should have several significant correlations with different anthropometric traits.

By assessing associations using continuous forms measuring environmental factors, we found significant, yet weak, association between increased calorific intake and increased BMI. Additionally, the level of vigorous level activities was significantly negatively associated with

the level of BMI and waist circumferences. These significant associations indicate that our measurement methods might have valid predictive capabilities.

Several studies have assessed associations between the level of environmental factors in continuous forms and anthropometric traits. A study has measured the level of calorific intake in a Greek population and measured its correlation with level of BMI. This study found that with every increment of 500 kcal of daily dietary intake, there was an increment of BMI of 0.14 (291). This study also found significant inverse association between the level of physical activity and BMI. These findings are consistent with our findings.

In our sample, the number of hours spent performing moderate activities was not significantly correlated with any anthropometric traits. This trend was similarly observed in a study conducted in Saudi Arabia measuring the effect of physical activity on the BMI level in a population aged between 14 and 19 years. This study was only able to find significant inverse effect of vigorous level of physical activity on BMI but not to the total level of physical activity (292). This finding is consistent with our finding indicating that measurement of vigorous activity is more likely to have a better predictive value of BMI compared to a level of moderate activity.

6.4 Strength of this study

Major strengths of this study include measures taken to reduce observer bias during the recruitment of participants and a reduction of measurement bias when answering questionnaires or collecting anthropometric data. Another strength issue is related to the handling of data using electronic forms of questionnaires. Introducing new concepts relating to pedigree constructions enabled a better estimation of inbreeding coefficients. Finally, there are several cultural issues related to the Saudi cultural background which could be considered as facilitating factors.

6.4.1 Participant recruitment and sampling

Although we targeted patients and health individuals with a family history of diabetes, the design of our study did not select participants on the basis of consanguinity history. This enabled us to have a representative sample with a reasonable range of inbreeding coefficients leading to the possibility of performing correlations with several phenotypes. Additionally, recruiting participants without prior knowledge about their inbreeding history may have reduced observer bias. Construction of pedigrees was systematic in all participants regardless of their inbreeding history.

6.4.2 Data collection

6.4.2.1 measurement of consanguinity

This study utilized a unique system of constructing extended pedigrees by developing concepts related to the mating of individuals from different tribes or subtribes. This technique enabled us to detect variability in consanguinity history that could not be detected if the shared ancestor could not clearly be identified. Among studies which found significant associations between inbreeding and risk of developing type 2 diabetes, this is the only study which utilized inbreeding coefficients rather than using a binary or categorical form of consanguinity history.

6.4.2.2 Electronic forms

Our study was mandated using questionnaires to measure environmental variations between participants. Instead of using lengthy paper-based questionnaires, we made an effort to convert our data collection tool from paper-based to electronic forms. We used Visual Basics 2010 to develop electronic questionnaires where we were able to answer questions, enter the answers and provide instant results after analysis.

This process had a great impact on reducing the time needed to transfer collected data to electronic databases. It would have been more time consuming transferring answers from paper-based questionnaires to electronic databases; this waste of time had been averted in this study. Additionally, it reduced the risk of mistakes which might have been caused by manually transferring data from paper-based questionnaires to electronic data bases.

6.4.3 Reducing measurement bias

6.4.3.1 Reducing the risk of recall bias

This study benefited from several points to tackle issues related to recall bias. From an ethical point of view, participants were informed about the main purpose of the study. However, they were not given detailed explanations of research questions such as the association between age at diagnosis and inbreeding history or the association between consanguinity and other anthropometric measures. Participants were blind to these associations in order to avoid having recall bias leading to distorted reporting of age at diagnosis or consanguinity history.

6.4.3.2 Measuring anthropometric variables

We managed to measure FBG of the majority of healthy participants. Although many subjects did not exhibit a healthy level of FBG, we only related them to being healthy hence they were not diagnosed with the disease and were not taking any medications. The main purpose of measuring blood glucose was not to avoid misclassification of healthy subjects with respect to their glycaemic index. Measuring blood glucose was mainly carried out to detect variations between individuals' blood glucose in subjects not taking any medication or not trying to impose environmental constrictions to reduce their blood glucose level. Misclassification of subjects was not a critical issue in our study hence we performed analysis in different cohorts of patients and healthy subjects.

All tools used to measure subjects phenotypes were used in a systematic manner by one observer. Measurement of environmental variation was systematic in all subjects whether they were patients or healthy. We can confidently reject overestimation of environmental risk factors in patients when filling food frequency questionnaires because, as the analysis showed, patients reported having a lower mean of average caloric intake. In theory, we would expect patients to report higher caloric intake of the disease but in reality, recall of food frequency was markedly influenced by patients' tendency to report healthier lifestyle to suggest better glycaemic control.

Interviewing all participants to answer questionnaires was an important issue aiding the avoidance of several biases. Firstly, interviewing all participants enabled us to recruit illiterate individuals. Secondly, we were able to illustrate to all participants how questionnaires are answered and give description of food items not recognized by participants. Having all subjects interviewed reduced measurement bias that might have been caused by having questionnaires filled out by participants during home visits. Thirdly, as we illustrated in the pilot study, we were able to ask specific questions relating to the frequency of consumption to avoid overestimation of intake.

Measurements of FBG, body weight, height, waist circumference were all collected by one observer. We avoided self-report bias by measuring all of these variables. We can ensure that all observations were systematically collected as the same measurement tools were utilized for all subjects. Additionally, we controlled for possible measurement bias caused by variations in successfully genotyped sample by applying cut-off criteria to exclude samples with low detected genotypes.

6.4.4 Cultural factors

The advantage of choosing to conduct this study in Saudi Arabia was due to the high prevalence of both type 2 diabetes and consanguinity. Additionally, mating usually occurs at a young age and families tend to be bigger in comparison to other populations. This increased the chance of finding subjects needed for the different components of our study.

Logistics factors were important in enhancing the data collection process. Hence the investigator comes from the region of Jazan, it was easier for him to commute to PHCs and homes of participants in several cities and villages in Jazan. It would have been more difficult for a person who did not come from Jazan region to get access to PHCs and houses without the assistance of a local individual.

Other cultural elements were of great importance. Participants tended to ask the investigator about his origin, his education and employment history. Once participants realized that the investigator belonged to a known tribe in the Jazan region and was a graduate of Jazan University, participants were then more inclined to participate, share their information and give samples. In an additional cultural gesture, the investigator always wore national Saudi clothing whether at PHCs or during home visits.

Driving a car was the only mean of transportation between the cities and villages of Jazan. There is no other means of public transportation such as buses, trams or trains between cities and villages of Jazan. Being a male is an advantage in Saudi Arabia where females are not allowed to drive. It would have been of immense difficulty to collect data for this study if the investigator of this study had been a female.

Being a medical doctor is another cultural point. Participants were more willing to welcome the investigator in their homes once they knew the investigator was a medical doctor. As what could have been a sign of health care deprivation, several participants appreciated the visit of a medical doctor to their homes and in several situations, shared medical issues with the investigator which were not necessarily related to diabetes. Although this might initiate an ethical concern, the actions provided by the investigator were merely advisory and no medical interventions were applied.

6.5 Limitations of this study

Major limitations of this study related to the lack of data about to birth weight, the utilization of proxy variables, and the risk of misclassification of cases. The food frequency questionnaire in our sample could have been subjected to measurement bias. Other limitations were related to the possible risk of selection bias caused by an inability to recruit females and a lower recruitment rate of Khat chewers. There are several cultural and logistic limitations which increased the difficulty of recruitment. Finally, there could be limitations in interpreting the findings of this study due to issues with multiple testing.

6.5.1 Lack of clinical data

The first limitation was related to the possibility of answering questions related to the associations between birth weight, inbreeding, and risk of type 2 diabetes. Most healthy subjects recruited in our study were above 30 years old despite our aim to recruit younger healthy individuals in each family. Due to recent developments in health systems in Jazan area over the last few decades, most recruited subjects did not have a record of their birth weight.

6.5.2 Use of proxy variables

Another limitation was related to using age at diagnosis as a proxy for the development of type 2 diabetes. It is almost impossible to accurately indicate age at development of type 2 diabetes in recruited subjects. This uncertainty is mainly caused by the nature of the disease where in several occasions patients are asymptomatic or even when they are symptomatic some patients might not seek health advice promptly.

6.5.3 Misclassification of diabetes

Although we recruited type 2 diabetes patients based on a diagnosis provided by their treating doctors, there might be a chance of misclassification of cases. We managed to avoid cases diagnosed as type 1 diabetes depending on medical diagnosis. However, due to the limited capability of PHCs in Saudi Arabia to diagnosis MODY cases, there might be a chance of misdiagnosing those cases as type 2 diabetes especially younger cases. Nonetheless, several SNPs were reported to be shared between MODY and type 2 diabetes.

6.5.4 Selection bias

Selection bias could have been introduced because several participants were less willing to participate in the study. Those participants were mainly Khat chewers or cases where healthy siblings refused to participate. However, we recruited several Khat chewer cases and their samples were successfully genotyped as we illustrated earlier. An inability to recruit females is another limitation of this study. The reasons were mainly related to cultural barriers as illustrated earlier in the methodology chapter.

6.5.5.Measurement bias

Despite utilizing validated EPIC questionnaires and showing the predictive validity of modified versions of these questionnaires, there are several points that might have introduced recall bias or reduced validity of questionnaires. Although these questionnaires ask the interviewee about food consumption or physical activity in the last year, this did not reflect true environmental variation affecting the incidence of diabetes in recruited subjects. This was clearly apparent upon interviewing patients where the assessment of environmental variation was strongly affected by their attitude toward management of type 2 diabetes. Subjects who were trying to control their glycaemic index were likely to report lower caloric intake compared to those not exhibiting a healthy lifestyle. However, the last notion was not systematically recorded during data collection and we are only suggesting this association.

Another issue is related to the application of a food frequency questionnaire in a Saudi population. As a usual procedure conducted to measure intake, an average portion size is usually estimated for each food item. Although we tried to propose a fixed portion size which was systematically applied to all participants, there is a large possibility of difference between estimated portion size and actual portion size consumed by participants.

The difficulty in estimating portion size in a Saudi Arabian population stems from the culture of serving food. As our analysis indicated, Chicken or Lamb Kabsa is the most frequently consumed food item in our sample. The most important ingredient of this food is rice. Kabsa is usually severed in one large dish as shown in figure 6.3 where consumers typically gather around the dish which is usually placed on the floor. Consumers eat directly from the large dish using hands unlike western societies where smaller quantities are usually served in plates. Therefore, it is very difficult to precisely estimate the portion size of this food items. Although consumers are sharing the same large dish, variations in portion sizes consumed by each individual are expected.

Figure 6.3: Method of serving popular food in Saudi Arabia.

6.5.6 Effect of Khat chewing on data collection

Catha Edulis or Khat (as called in Jazan) is a plant with stimulating amphetamine-like characteristics (293).Cultivation, transportation, trading, and consumption of Khat are strictly unlawful actions in Saudi Arabia. A study conducted in Jazan revealed that 37% of male students (aged between 15 and 25) were Khat chewers. (293). There is no study published regarding the consumption of Khat in older subjects, but based on the obvious trend of a higher tendency to consume Khat with age and based on our observation during data collection, we can expect, at least, half of the population in Jazan to be Khat chewers. Appendix 14 includes further background information about the pharmacogenetic, the heath impact and the social effects of Khat.

Even though we were aware of presence of Khat in the Jazan area, we were not perfectly alert about its potential effect on our recruitment plan. Although we did not attempt to systematically record the number of patients or healthy participants who were Khat chewers, we can confidently argue that the percentage of Khat chewers among those who were approached, including recruited subjects, was almost higher than 50%. Khat chewing made data collection a challenging task and it was one of the factors that rendered the principal investigator unable to achieve 100% of the targeted sample size.

Due to its social effect, certain individuals were not willing to participate in the study because it interfered with their chewing sessions. We advised approached individuals to abstain from Khat chewing until the interview had been completed and the buccal swab was taken. However, this appeared to reduce the participation rate and thus we had to stop requesting approached individuals not to chew Khat. Even though we changed our policy and started explaining to the approached individuals it was acceptable to chew Khat during the interviews, several individuals refused to participate.

In several situations, some approached individuals tended to show an interest in the study and agreed to participate giving a specific date for data collection but due to their addiction to Khat, they cancelled the appointment or did not respond when the investigator attended their address for the agreed data collection session. The cancellation of appointments or not showing up during home visits had a significant effect on the recruitment rate. The challenge to recruit individuals was aggravated by situations where one of the siblings, either patients or healthy, were not available due to Khat chewing sessions. In addition to the effects of unsocial behaviour caused by Khat consumption, insomnia and sleeping disturbances caused several approached individuals to sleep during agreed appointments.

The reason we were trying to convince approached individuals to refrain from Khat chewing before collection sessions was due to our effort to have clean DNA samples. Khat is chewed for several hours in one side of the mouth and taking a buccal swab during chewing session was very likely to contaminate the swab with Khat. Those who agreed to participate in the study while they were chewing Khat were requested to give other convenient appointments to give a buccal swab. Usually, those individuals preferred to give buccal swabs just before they go to work in the morning. In addition to Khat, other products with less impact interfered with the collection of buccal swabs such as chewing tobacco.

6.5.7 Limitations related to awareness and cultural Conditions

Type 2 diabetes patients were more willing to participate than their healthy siblings. In addition to the reduced participation caused by Khat consumption, several healthy siblings were afraid to participate due to fear of being diagnosed with type 2 diabetes. Although we explained the importance of screening for type 2 diabetes, our effort on certain occasions was fruitless. On other rare occasions, healthy siblings were willing to give all the needed information and measurements except blood glucose because they were afraid of needles.

In some approached families, familial ties between siblings were not strong and approached patients at PHCs informed us that they could not contact their siblings. Additionally, several patients did not have contact details for their siblings. In other families, siblings were having disputes and were not even speaking to each other. These issues interfered with recruitment conditions and rendered several approached individuals not suitable for recruitment.

Other culturally embraced behavioural concepts appeared to complicate the collection of selected information from certain participants. For example, it was rather embarrassing for certain participants to be asked about the names of their females siblings. If we had the impression that it would be embarrassing to ask about the names of female siblings, we tended to neglect asking about names and only asked about age and the health status of female siblings. Nonetheless, several participants did not mind revealing the names of their female siblings.

Another example is related to asking participants about household activities such as cooking and caring for children. According to the Saudi culture, these activities are mainly performed by females. Several participants were less interested in answering these questions and one participant refused to give answers regarding these questions listed in the physical activity questionnaire.

6.5.8 Limitations related to safety and logistics

The nature of this research mandates meeting hundreds of subjects in person and visiting several families in their homes. Visiting this large number of subjects in their houses revealed several issues which were not possible to observe during visits to PHCs. For example, we arranged to recruit a subject during a home visit; during this home visit, the subject informed us about the presence of drug dealers and the presence of illegal immigrants in his neighbourhood. This situation illustrates how dangerous data collection can be when meeting individuals from risky neighbourhoods.

Another incident occurred when a healthy participant was convinced that he had hepatitis B infection although, as he informed us, his screening blood tests were negative. We tried to assure him that he can always perform other blood tests in order to confirm the disease status. However, he insisted on having an abdominal examination performed. Although I tried to convince him that his living room was not an appropriate clinical environment and I was not entitled to perform a clinical examination for him, he insisted. Therefore, I had to perform an abdominal examination in an effort to examine his liver. However, it was merely a psychological manoeuvre to calm him down. This incident also signifies how difficult it can be to deal with the public in various sensitive matters.

In addition to safety issues, other logistic factors made data collection a difficult process. Firstly, several participants were living in remote and difficult to access places. We visited several houses distributed in different sectors. In certain situations, participants were living in deserts, mountains or poorly planned and crowded neighbourhoods. Figure 6.4 is collection of pictures of places visited during the data collection.

Figure 6.4: Photos of places visited during data collection.

Another issue was related to the unavailability of competent navigation systems. One of the most undeveloped issues in Saudi Arabia is domestic planning where established houses do not have clear addresses or even postal codes. To illustrate this issue, I am quoting an actual conversation made by a person trying to illustrate the address of his house:

'In order to get to our house, please proceed through first exit, turn toward the north, keep driving until you reach the third traffic signal. Once you reach the third traffic signal, please take the first exit to the right. Drive forward until you face a school. Drive alongside the school until the fence of the school ends. Once the school's fence ends, turn right. On your right hand you will find the house which is coloured brown'

This quote might seem exaggerated but it is a typical description of a house in Saudi Arabia. It was quite difficult in some occasions to find houses of participants and in several situations the description was so difficult that some participants had to meet us outside their neighbourhood just to show us the way to their houses. Extreme caution had to be taken by the investigator to memorize the houses of the participants in case there was a need for a second visit to collect buccal swabs or to test FBG.

6.5.9 Limitations due to multiple testing

Setting a significance level at 0.05 means that there is a 5% chance of rejecting the null hypothesis when it is true, and accepting the alternative hypothesis when it is false (false positive error). If multiple hypothesis testing is applied to the same sample, there will be a higher chance of detecting significant p-values by chance. When assuming a significance level of 0.05, performing 100 statistical tests on the same sample we would expect to detectfive significant tests by chance (false positive).

In multiple testing situations corrective strategies such as the Bonferroni correctionor the False Discovery Rate (FDR) are applied, but others have argued that simply accounting for how many tests were conducted and why they were done is all that is necessary (294) . By counting the number of tests performed, a total of 139 different statistical tests were utilized in this study, where 43 tests were statistically significant. In all tests, a significance level was set at 0.05, which means that out of the 139 tests conducted, we would expect to find about seven statistically significant associations to be false positive.

Bonferroni correction relies on giving a corrected p-value depending on the number of tests performed. In this scenario, corrected p-value will be equal to the detected p-value multiplied by the number of tests performed (295) (2). For example, if 100 tests were performed where one test had a p-value of 0.004, then the corrected p-value would be 0.4. The corrected p-value indicates that this test is not statistically significant.

Bonferroni correction is a stringent test and might result in failure to detect several true associations. By applying Bonferroni correction to our calculated p-values, 11 corrected p-values were below 0.05. The marked reduction in the number of statistically significant tests is expected, given the small sample size of this study. Although this test is conservative and reduces the increased chance of detecting false positive results, it might have resulted in elimination of true associations detected in this study. Bonferroni correction is a simple and a popular method, although highly criticized. It has been argued that it increases the chance of type 2error by rejecting true alternative hypothesis (294).

FDR controls the expected proportion of tests which are false positive among all significant results (296). This is a less stringent test and calculated by ordering p-values from largest to smallest. The largest p-values will be corrected by multiplying the p-value by the number of tests.The following formula is used to correct the remaining p-values according to the order of the p-value: p-value * (n/n-[number of rank - 1]) where n is the number of tests performed. For example, the corrected p-value for the fourth one will be p-value * (n/n-3).

Unlike the Bonferroni test, FDR correction is less conservative. By fixing the FDR at 0.05,applying FDR correction to the study's calculated p-values resulted in 25 corrected p-values below 0.05. Based on the FDR correction, the association between age at diagnosis and inbreeding coefficient became marginally statistically non-significant (corrected p-value:0.06). However, the association between inbreeding coefficient and age at diagnosis among patients both of whose parents were reported to be affected with type 2 diabetes remained statistically significant (corrected p-value: 0.02).Siblings FBG correlations were statistically significant when applying different ranges of consanguinity (p-values: 0.04, 0.015, 0.003, 0.001). Nonetheless, some of the correlations became statistically non-significant after FDR correction. However, after applying the FDR correction, there is a trend of increased significance level with increased inbreeding coefficients (0.13, 0.06, 0.02, 0.009) where siblings with higher inbreeding coefficients witnessed statistically significant correlations. Finally, all statistically significant deviation from WHE remained statistically significant after applying FDR correction. In conclusion, although some of the main findings of this study became marginally statistically significant, it is possible to argue that the main findings of the thesis are still valid even after applying the FDR correction.

6.6 Personal Experience

6.6.1 Learning experience: approaching and data collection

Working in the field gives you a chance to observe unexpected events and gain valuable experiences. Several points were not targeted objectives upon the start of data collection. However, noteworthy aspects of the population in Jazan have to be illustrated with relation to type 2 diabetes, orientation about health and their attitude toward medical research.

Upon approaching patients and participants, information sheets were provided. We also gave approached participants a chance to enquire about the purpose, steps and nature of the research. It was a valuable gained experience about how to approach individuals with different age ranges and educational backgrounds. Approaching illiterate older subjects was in essence different from approaching younger educated subjects. The magnitude and type of questions asked by approached participants varied greatly between individuals. Eventually, we were able to choose a suitable approach process for each individual.

Data collection was additionally complicated when a second visit had to be done to collect a buccal swab or to measure FBG. On several occasions, we had to start working at 4 AM to travel to a remote place and measure FBG for a participant or to collect a swab. On one occasion, we were frustrated to realize that one of our early-morning trips to collect a buccal swab was useless: on arriving at the participant's house, we were informed about his refusal to give the swab despite participating and giving most of the other needed information after signing the consent form. Nevertheless, this was quite a valuable lesson when applying research methods in the field and the importance of expecting limitations when dealing with hundreds of subjects especially when working in a non-clinical settings.

6.6.2 A need to raise awareness

Interviewing participants gave us a valuable chance to meet patients and their healthy siblings. In almost all home visits, the participants and even the non-participants tended to enquire about diabetes, its risk factors and complications. This gave us an indication that knowledge about diabetes and its complications seemed to be rather shallow. Although we did not attempt to quantify the magnitude of this issue, we think that a public orientation campaign is very important and urgently needed to reduce complications among patients of diabetes or to reduce the incidence of subjects at risk.

Several subjects seemed to be unaware that diabetes is a serious illness and in extreme situations, chose to ignore or even reject the idea of being diagnosed with diabetes. Several healthy participants refused to participate in the research thinking that their participation could result in them being diagnosed with diabetes. This clearly signifies that they are not aware that

earlier diagnosis of type 2 diabetes could significantly reduce their risk of developing complications related to the disease. We also met several patients who refuse to take their medication and maintained a very unhealthy lifestyle or chose Khat chewing as a remedy to cure diabetes.

6.6.3 Administrative capabilities

The complexity of this work and the collection of several continuous variables made this effort an unforgettable experience. Furthermore, all the work of coordinating appointments, conducting visits for families and performing data collection and data entry was conducted by one person within a limited time frame. Completing the data collection within the proposed time frame was faced with several doubts regarding its success. After all, achieving 90% of the target sample size signifies the assertive effort done by the principal investigator.

Collecting data appeared to be a challenging and, at certain times, an exhausting task. Data collection required substantial organizational skills and immaculate ability to deal with several participants in a timely manner. For several weeks, we worked for seven days. In addition, on some days we had to work for almost 10 hours due to the time needed to approach participants, make appointments, travel to remote areas and spend several hours to carry out the actual data collection. For several months, the principal investigator only stopped working for emergencies.

6.6.4 Emotions behind research

Due to visiting almost more than 170 houses, we were exposed to participants from different socioeconomic classes. In addition to visiting well-planned neighbourhoods with high living standards, we also visited several remote, unplanned and deprived regions with very low living standards. This study has a social and emotional perspective since we interviewed participants who were in serious need of help. On one occasion, one of the participants was affected by diabetic complications affecting his vision and a severe case of diabetic foot impacted on his mobility. Due to these complications, he lost his job and was not able to provide for his family. It was quite difficult for the investigator to separate between his emotions and the professional work of data collection. Although at that time we were not able to provide any significant help, we provided advice regarding better control of diabetes and our deepest emotions had to be shared with the participant.

Despite all difficulties, tragedy and drama witnessed during data collection, there were several cheerful and encouraging events. Several families were quite happy to participate. Some participants even went further and admitted that they were expecting our visit to their PHC since they had heard about our research from neighbouring PHCs.

For a reason, our research was somehow well-known. After few months we felt that this unplanned announcement made the recruitment process a success. Potential participants learned about the research from their friends and neighbours and when they were approached; it was easier for them to understand procedures which had to be taken and the purpose of the research. Some families were extremely cooperative and made data collection an enjoyable experience. Several educated participants showed their deep appreciation of the research and were very interested to learn about the results and recommendations of the study.

It was quite normal to have the whole family present during data collection in addition to participating siblings. However, this welcoming attitude made it slightly difficult to collect data and more time consuming as I had to deal with several individuals in addition to the participants. Some families were quite helpful when constructing family pedigrees. One family gifted us a book written by one of the participants. The book's subject was about their family tree and a description of their ancestors back to 1400 years ago. Some families take pride in their genealogy hence they believe their family trees go back to the family of Prophet Muhammad (peace be upon him).

Other rewarding moments were experienced when we were able to be of assistance on several matters. We were able to provide advice to those who were refusing to take their medication despite being diagnosed with type 2 diabetes. Several healthy participants were having an above normal level of FBG or impaired glucose level. We provided advice to those participants regarding the importance of visiting their family doctor and of starting to change their lifestyle. On several occasions, we would measure blood glucose for those who were not participating in the study such as other siblings, parents, and friends. This had slightly increased the cost of glucose testing hence we needed to buy more glucose strips, lancets, and alcohol swabs.

In addition to measuring FBG for several individuals, adults and even children appeared to enjoy having their weight and height measured during home visits. Whenever they saw the stadiometer and the scale, they started gathering and requested the investigator to measure their heights and weights. Some non-participating individuals went further and asked the investigator to calculate their BMI and provide relevant advice. Additionally, participants frequently asked the investigator about the results of the food frequency analysis program. Based on the frequency provided, I was able to advise the participants about food items they were frequently consuming which had high calorific values.

The kindness and generosity of the participants was quite an appreciated experience. On most visits, beverages such as traditional Arabic coffee and dates were provided as a gesture of

greeting. Although our time during home visits was mostly taken up by performing data collection, we strived to appreciate their generosity and consume what they served and apologized whenever we could not, realizing that it is an offence in our culture not to accept what you are served. Several participants insisted on inviting us for another visit whereupon we had to apologize due to the limited nature of our time. One participant, and out of generosity, went further and provided us with Khat leaves. The principal investigator had to apologize to the participant explaining his attitude toward Khat chewing.

One final remark has to be mentioned due to its uniqueness to the Muslim society. The Month of Ramadan coincided with the first stage of data collection. During this month, Muslims fast during the day time which made it easier for the investigator to measure the participants' FBG. Additionally, the religious behaviour of Muslims during this month is significantly different than other months of the year. Subjects in the Jazan region were less likely to chew Khat which made the recruitment process during this month notably easier compared to other months. Several participants revealed that they had better glycaemic control during this month due to fasting and were able to reduce medicational doses.

6.7 Reflections

Revealing causal relationship between a factor and a phenotype is the essence of epidemiology. Complexity and sophistication of the human body has rendered detection of such relationships a difficult yet captivating mission. Failing or succeeding to tackle ambiguous and uncertain associations is a defining moment for any epidemiological investigation.

Determinants of any living creature are a combination between their inner propensity to survive and an outer surrounding environment which decides their chance of thriving. Humans, as any living creature, have this inner factor which determines most, if not all, of their characteristics. The variability of characteristics plays a key role in deciding whether individuals are fit to survive when placed in a similar environment.

The human genome has been estimated to be composed of more than 3 billion bases (297). In many complex illnesses, investigating the effect that a single nucleotide variation would produce is almost similar to observing the effect that a drop of water would cause upon falling in an ocean. The magnitude of complexity and ambiguousness of factors causing a complex disease such as type 2 diabetes is far from being resolved by observing SNPs acting individually or even additively.

It could be argued that the genetic effect on any trait might begin even before having an egg fertilized by a sperm. Similarities or variations between parents' genomes are solely driven by factors affecting their chance of mating. Apart from parents' ability to pass their genotypes, which is basically influenced by their capability to survive selective forces, genetic determinants of their offspring are further complicated by the chances of mating of individuals with similar or different genomes. This similarity could be driven by a similarity in phenotypes or similarity caused by a shared ancestor.

Sequencing of the human genome has revealed several nucleotides variations linked to variations of phenotypes. Most studies concerned with associating these variations with phenotypes assume an additive effect of alleles. However, this concept could be far from detecting the true causal pathway. The risk of having a particular trait is not simply caused by inheriting a single allele especially when addressing complex illnesses.

Apart from assuming an additive effect of alleles, other complex deviations from this model might be solely responsible for detecting causal pathways not revealed by additive models. Whether inheriting similar alleles at a locus would make a different effect compared to only inheriting a single allele or a single allele is dominant enough to cause an effect without having a similar copy is difficult to observe in complex phenotypes. Additionally, we could postulate whether inheriting a single allele is affecting expression of another allele in an epistatic effect or one environmental factor is affecting expression of one allele in an epigenetic effect.

Once a zygote is formed, a long journey of development begins during uterine life. Variations in intrauterine environments have a significant effect on phenotype variations despite genomic variations. Intrauterine life is a rather critical developmental stage where a single change might start a cascade of events continuing after birth. The environmental effect further continues to explain significant part of variations of complex phenotypes variations.

In addition to our study, several studies have signified the association between the non-random mating of parents and risk of the development of type 2 diabetes in their offspring. Furthermore, our study has indicated that this association is augmented when both parents are affected by the disease. This might indicate that causal pathways indicated by higher risk posed by consanguinity are different than causal pathways in those with less or no degree of inbreeding.

The complexity of type 2 diabetes as a phenotype does not only stem from its polygenic risk or multifactorial environmental factors. The effect of inheriting a similar risk allele has been shown to act in different ways. Rs7901695 near TCF7L2 has been reported to increase the risk of type 2 diabetes in non-obese individuals (205). This notion suggests that type 2 diabetes is a heterogeneous disease where different or even opposing risk factors could result in the very same outcome.

A strong example suggesting heterogeneity of type 2 diabetes is the effect of intrauterine life on the development of type 2 diabetes. Several studies have suggested a significant association between low birth weights and a high risk of type 2 diabetes. This association has been partially explained by the thrifty genotype theory. However, mothers who are pregnant while affected by type 2 diabetes are more likely to have offspring with a high birth weight (Macrosomia) and yet their offspring are at a higher risk of developing type 2 diabetes. We can observe two different and opposing risk factors (low birth weight versus macrosomia) with similar outcome (type 2 diabetes).

The importance of acknowledging heterogeneity of type 2 diabetes might have several consequences. Firstly, based on stratifying individuals according to specific criteria such as studying the causes of type 2 diabetes in lean individuals or those with higher inbreeding degree, it might be possible to identify pathways leading to development of type 2 diabetes. The second reason is related to the prediction and screening of the disease in different cohorts. For example, we might argue that those with a high degree of inbreeding and an extensive history of familial type 2 diabetes should be subjected to targeted screening and preventive measures compared to the general public.

Heterogeneity of type 2 diabetes might have an influence on how treated patients respond to given treatments. From a pharmacogenetic point of view, one therapeutic agent given to similar

type 2 diabetes patients might have a variable range of effects. This variation in response could be attributed to variation in environmental factors (such as a change in caloric intake or physical activity) or due to genomic variation. Oral glucose lowering agents have already been suggested to have different effect in monogenic forms of diabetes. Since type 2 diabetes shares several variants with monogenic form, genomic variability could be a contributing factor leading to a variable range of responses in type 2 diabetes patients.

Although this study managed to find significant associations between inbreeding history and several variables related to type 2 diabetes, failure to find significant effects of inbreeding on molecular risk factors has limited our ability to confirm a causal molecular risk related to inheriting homozygote genotypes of risk alleles. A small number of genotyped SNPs and relatively small number of subjects limited by financial and logistic circumstances might be a contributing factor leading to the failure in detecting significant effect. Having found significant association between inbreeding coefficients and age at diagnosis without finding molecular evidence might indicate the need to genotype larger number of risk alleles and patients to detect a significant association.

7 Chapter Seven: Conclusion

This study explored association between consanguinity and risk of developing type 2 diabetes. Although this study was not able to find statistically significant recessive effects of measured SNPs on traits related to severity or risk factors of type 2 diabetes, we were able to find a statistically significant association between age at diagnosis and inbreeding coefficients. Furthermore, our study suggests a possible genetic effect on FBG by observing increased siblings' correlations with increased level of consanguinity. The effect of consanguinity on age at diagnosis and correlations of FBG might give partial explanations of synergism between type 2 diabetes and consanguinity. Further investigations are required to target subjects with high consanguinity, to recruit a larger sample and to genotype a larger number of SNPs to find possible recessive effect on traits related to type 2 diabetes.

According to our knowledge, this study is the first study to measure several SNPs related to type 2 diabetes in a Saudi Arabian population. This issue was clearly apparent when we realised that research facilities in Saudi Arabia did not have the means to perform genotyping of this study and samples had to be shipped to the UK. Additionally, by comparing between genotype frequencies in our sample to frequencies reported in other populations, we were able to suggest the possible effect of increased frequency of homozygosite genotypes that could be attributed to high levels of consanguinity.

7.1 Implication for public health policies

Based on this study's finding, we can suggest that those with affected parents who are related should be targeted when implementing secondary preventive measures. However, for this suggestion to be considered, further investigations are needed to replicate this association in other region of Saudi Arabia. Furthermore, this study was restricted to males due to cultural barriers. Nonetheless, we can argue that the incidence of type 2 diabetes is alike between genders and similar causal pathways are likely to be the same in males and females.

Despite a similar prevalence of type 2 diabetes in males and females, we can argue that heritability of type 2 diabetes among Saudi female subjects could be higher than expected. Studies which measured physical activity in Saudi males and females reported that females are less likely to be physically active than males. This observation indicates the limited range of physical activity of women in Saudi Arabia and variability in incidence of type 2 diabetes could be attributed to differences in calorific intake and genetic variance.

The data collection period was a good chance to interact with hundreds of individuals. This interaction gave us several indications which were not among the aims of this study. For example, we observed a strong influence of Khat consumption on the behavioural attitude toward accepting diabetes as a disease and treatment compliance. We would argue that there might be a strong need for the health authority in the Jazan region to start investigating this issue and implement relevant educational programs to enhance overall control of the disease.

Several suggestions can be given to reduce the burden of type 2 diabetes in Saudi Arabia. We can argue that most of participants who were at risk of the disease did not have a good knowledge about the causes of type 2 diabetes. Even patients lacked good knowledge about better control of the disease and several healthy subjects were reluctant to have their blood glucose measured.

We sense a strong need for educational public health campaigns with the ability to reach individuals with limitations such as illiteracy or being resident in remote villages. Patients in Saudi Arabia need to be educated that reliance on religious factors does not interfere with utilizing medical and lifestyle interventions to control the disease. Additionally, since many health professionals working at PHC in Saudi Arabia are not native Arabic speakers, the quality of health service is likely to be compromised leading to poorer health education and reduced overall control of type 2 diabetes. This point need to be addressed on a national level where trained health professional with good Arabic speaking skills should be engaged in proper health education of individuals at risk or suffering from the disease.

A final remark is related to the benefits gained from the food frequency questionnaire program. As the program included a database comprising the calorific values of food items, we

were able to observe which food items were most commonly consumed by interviewed participants and give advice to participants about calorific intake. Several participants were not alert to the high calorific value of items they consumed on a daily basis. We could argue that using this software might allow health care professionals to comprehensively ask patients about their eating habit and decide which food items exactly can be eliminated or a reduction be made in the frequency of their consumption.

7.2 Implication for future research

Finding a statistically significant association between inbreeding coefficients and age at diagnosis with a stronger association in patients where both parents are affected, might indicate a need for future research in a specific subjects. We would argue the need to target subjects with high consanguinity and strong family history in further genetic investigations to find possible recessive effects of risk alleles. Similarly, finding higher correlations of FBG in subjects with higher levels of consanguinity might indicate a similar approach of targeting healthy individuals with high consanguinity would increase the chance of detecting genetic markers with possible recessive effect on FBG.

We hypothesized a possible intermediate effect of low birth weight on the association between consanguinity and type 2 diabetes. However, upon visiting the Directory of Health in Jazan, we were informed that the Directory was established in 1990. Therefore, the majority of participants did not have a record of birth weight. As a consequence, this study was not able to investigate this hypothesis retrospectively.

Lack of birth weight records of older generations indicates the importance of investigating the effects of consanguinity on a reduction in birth weight and possible subsequent development of type 2 diabetes in younger generations. Younger generations in Saudi Arabia are more likely to have a record of birth weight compared to older generations. Therefore, young subjects who have a record of birth weight, and not necessarily type 2 diabetes patients, could be recruited in a prospective study to investigate this hypothesis.

There are several other methodological recommendations. Firstly, Khat is less likely to interfere with DNA quality and therefore, we would argue that there should be no restriction against taking buccal swabs from subjects who are khat chewers. Secondly, measuring caloric intake and physical activity in patients was not beneficial due to bias caused by changing of life style. We would recommend measuring these environmental variables in healthy subjects to study their effect on traits related to type 2 diabetes. Thirdly, utilizing food frequency questionnaires in Saudi Arabia requires further efforts to develop better assessments of portion

sizes and to develop a Saudi Nutritional Database which should be available for the public and researchers. Finally, based on this study's findings and other relevant publications, we would argue that a vigorous activity parameter is best used to measure variation in physical activities compared to sedentary or moderate levels of physical activities.

8. References:

1. IDF. IDF Diabetes Atlas Update 2012. 2012 [cited 2013 12/02]; Available from: http://www.idf.org/diabetesatlas/5e/Update2012.
2. WHO. Diabetes. 2011 [cited 2011 25th February]; Available from: http://www.who.int/mediacentre/factsheets/fs312/en/.
3. IDF. Diabetes and Impaired Glucose Tolerance. 2010 [cited 2011 25th February]; Available from: http://www.diabetesatlas.org/content/diabetes-and-impaired-glucose-tolerance.
4. IDF. Regional Overview. 2011 [cited 2011 1st of March]; Available from: http://www.diabetesatlas.org/content/regional-overview.
5. Buse J, Polonsky K, Burant C. Type 2 Diabetes Mellitus. Williams Textbook of Endicrinology. 10th ed. Philadelphia: Saunders; 2003. p. 1427-83.
6. Harris MI, Hadden WC, Knowler WC, Bennett PH. Prevalence of diabetes and impaired glucose tolerance and plasma glucose levels in U.S. population aged 20-74 yr. Diabetes. 1987;36(4):523-34. Epub 1987/04/01.
7. Gale EA, Gillespie KM. Diabetes and gender. Diabetologia. 2001;44(1):3-15. Epub 2001/02/24.
8. Summerson JH, Spangler JG, Bell RA, Shelton BJ, Konen JC. Association of gender with symptoms and complications in type II diabetes mellitus. Womens Health Issues. 1999;9(3):176-82. Epub 1999/05/26.
9. Abdella NA, Khogali MM, Salman AD, Ghuneimi SA, Bajaj JS. Pattern of non-insulin dependent diabetes mellitus in Kuwait. Diabetes Res Clin Pract. 1995;29(2):129-36. Epub 1995/08/01.
10. Singh S, Dhingra S, Ramdath DD, Vasdev S, Gill V, Singal PK. Risk factors preceding type 2 diabetes and cardiomyopathy. J Cardiovasc Transl Res. 2010;3(5):580-96. Epub 2010/07/02.
11. Newman B, Selby JV, King MC, Slemenda C, Fabsitz R, Friedman GD. Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. Diabetologia. 1987;30(10):763-8. Epub 1987/10/01.
12. Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. Diabetologia. 1999;42(2):139-45.
13. Meigs JB, Cupples LA, Wilson PW. Parental transmission of type 2 diabetes: the Framingham Offspring Study. Diabetes. 2000;49(12):2201-7. Epub 2000/12/16.
14. Berber A, Gomez-Santos R, Fanghanel G, Sanchez-Reyes L. Anthropometric indexes in the prediction of type 2 diabetes mellitus, hypertension and dyslipidaemia in a Mexican population. Int J Obes Relat Metab Disord. 2001;25(12):1794-9. Epub 2002/01/10.
15. Bei-Fan Z. Predictive values of body mass index and waist circumference for risk factors of certain related diseases in Chinese adults: study on optimal cut-off points of body mass index and waist circumference in Chinese adults. Asia Pac J Clin Nutr. 2002;11 Suppl 8:S685-93. Epub 2003/01/22.
16. Wannamethee SG, Shaper AG. Weight change and duration of overweight and obesity in the incidence of type 2 diabetes. Diabetes Care. 1999;22(8):1266-72. Epub 1999/09/10.
17. Burke V, Zhao Y, Lee AH, Hunter E, Spargo RM, Gracey M, et al. Predictors of type 2 diabetes and diabetes-related hospitalisation in an Australian Aboriginal cohort. Diabetes Res Clin Pract. 2007;78(3):360-8. Epub 2007/05/29.
18. Misra A, Vikram NK. Clinical and pathophysiological consequences of abdominal adiposity and abdominal adipose tissue depots. Nutrition. 2003;19(5):457-66. Epub 2003/04/26.
19. Kahn SE, Hull RL, Utzschneider KM. Mechanisms linking obesity to insulin resistance and type 2 diabetes. Nature. 2006;444(7121):840-6. Epub 2006/12/15.
20. Banerji MA, Faridi N, Atluri R, Chaiken RL, Lebovitz HE. Body composition, visceral fat, leptin, and insulin resistance in Asian Indian men. J Clin Endocrinol Metab. 1999;84(1):137-44. Epub 1999/01/27.

21. Whincup PH, Kaye SJ, Owen CG, Huxley R, Cook DG, Anazawa S, et al. Birth weight and risk of type 2 diabetes: a systematic review. JAMA. 2008;300(24):2886-97. Epub 2008/12/26.
22. Onat A, Yazici M, Can G, Kaya Z, Bulur S, Hergenc G. Predictive value of prehypertension for metabolic syndrome, diabetes, and coronary heart disease among Turks. Am J Hypertens. 2008;21(8):890-5. Epub 2008/06/14.
23. Ferrannini E, Buzzigoli G, Bonadonna R, Giorico MA, Oleggini M, Graziadei L, et al. Insulin resistance in essential hypertension. N Engl J Med. 1987;317(6):350-7. Epub 1987/08/06.
24. Maskarinec G, Grandinetti A, Matsuura G, Sharma S, Mau M, Henderson BE, et al. Diabetes prevalence and body mass index differ by ethnicity: the Multiethnic Cohort. Ethn Dis. 2009;19(1):49-55. Epub 2009/04/04.
25. Pouwer F, Kupper N, Adriaanse MC. Does emotional stress cause type 2 diabetes mellitus? A review from the European Depression in Diabetes (EDID) Research Consortium. Discov Med. 2010;9(45):112-8. Epub 2010/03/03.
26. Tamayo T, Christian H, Rathmann W. Impact of early psychosocial factors (childhood socioeconomic factors and adversities) on future risk of type 2 diabetes, metabolic disturbances and obesity: a systematic review. BMC Public Health. 2010;10:525. Epub 2010/09/03.
27. CDC. 2011 National Diabetes Fact Sheet. 2011 [cited 2013 13th August]; Available from: http://www.cdc.gov/diabetes/pubs/factsheet11.htm.
28. Scanlon V, Sanders T. The endocrine system. Essentials of Anatomy and Physiology. Philadelphia: F. A. Davis Company; 2003.
29. Shier D, Butler J, Lewis R. Endocrine system. Hole's Human Anatomy and Physiology. 12th ed. New York: McGraw-Hill; 2010.
30. Sherwood L. The Peripheral Endocrine Glads. Human Physiology From Cells to Systems. 5th ed. Belmont: Brooks/Cole-Thomson; 2004.
31. Mahler RJ, Adler ML. Clinical review 102: Type 2 diabetes mellitus: update on diagnosis, pathophysiology, and treatment. J Clin Endocrinol Metab. 1999;84(4):1165-71. Epub 1999/04/13.
32. Haring HU, Mehnert H. Pathogenesis of type 2 (non-insulin-dependent) diabetes mellitus: candidates for a signal transmitter defect causing insulin resistance of the skeletal muscle. Diabetologia. 1993;36(3):176-82. Epub 1993/03/01.
33. Warram JH, Martin BC, Krolewski AS, Soeldner JS, Kahn CR. Slow glucose removal rate and hyperinsulinemia precede the development of type II diabetes in the offspring of diabetic parents. Ann Intern Med. 1990;113(12):909-15. Epub 1990/12/15.
34. Hillson R. Diabetes the complete guide. 3rd ed. London: Random House Group Ltd; 2002.
35. Watkins P. ABC of Diabetes. 5th ed. London: BMJ Publishing Group Ltd; 2003.
36. Hurst RT, Lee RW. Increased incidence of coronary atherosclerosis in type 2 diabetes mellitus: mechanisms and management. Ann Intern Med. 2003;139(10):824-34. Epub 2003/11/19.
37. Sabanayagam C, Liew G, Tai ES, Shankar A, Lim SC, Subramaniam T, et al. Relationship between glycated haemoglobin and microvascular complications: is there a natural cut-off point for the diagnosis of diabetes? Diabetologia. 2009;52(7):1279-89. Epub 2009/04/24.
38. Levy A. Complication-resistant patients. In: Raz I, Skyler J, Shafrir E, editors. Diabetes: From Research to Diagnosis and Treatment. London: Martin Dunitz; 2003. p. 55-68.
39. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group. Lancet. 1998;352(9131):837-53. Epub 1998/09/22.
40. van Leiden HA, Dekker JM, Moll AC, Nijpels G, Heine RJ, Bouter LM, et al. Risk factors for incident retinopathy in a diabetic and nondiabetic population: the Hoorn study. Arch Ophthalmol. 2003;121(2):245-51. Epub 2003/02/14.

41. WHO. Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia. 2006 [cited 2011 5th of March]; Available from: http://www.who.int/diabetes/publications/Definition%20and%20diagnosis%20of%20diabetes_new.pdf.
42. WHO. Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus. 2011 [cited 2011 5th of March]; Available from: http://www.who.int/diabetes/publications/report-hba1c_2011.pdf.
43. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. Diabetes Care. 2009;32(7):1327-34. Epub 2009/06/09.
44. Esteghamati A, Khalilzadeh O, Rashidi A, Meysamie A, Haghazali M, Asgari F, et al. Association between physical activity and insulin resistance in Iranian adults: National Surveillance of Risk Factors of Non-Communicable Diseases (SuRFNCD-2007). Prev Med. 2009;49(5):402-6. Epub 2009/09/12.
45. M K, E.D G, C A. The role of the glycemic index in the prevention and management of diabetes: A review and discussion. Can J Diabetes. 2005;29:27-38.
46. Vetter ML, Ritter S, Wadden TA, Sarwer DB. Comparison of Bariatric Surgical Procedures for Diabetes Remission: Efficacy and Mechanisms. Diabetes Spectr. 2012;25(4):200-10.
47. Leonetti F, Capoccia D, Coccia F, Casella G, Baglio G, Paradiso F, et al. Obesity, type 2 diabetes mellitus, and other comorbidities: a prospective cohort study of laparoscopic sleeve gastrectomy vs medical treatment. Arch Surg. 2012;147(8):694-700. Epub 2012/04/18.
48. Schauer PR, Kashyap SR, Wolski K, Brethauer SA, Kirwan JP, Pothier CE, et al. Bariatric Surgery versus Intensive Medical Therapy in Obese Patients with Diabetes. N Engl J Med. 2012;366(17):1567-76.
49. Leslie DB, Dorman RB, Serrot FJ, Swan TW, Kellogg TA, Torres-Villalobos G, et al. Efficacy of the Roux-en-Y gastric bypass compared to medically managed controls in meeting the American Diabetes Association composite end point goals for management of type 2 diabetes mellitus. Obes Surg. 2012;22(3):367-74. Epub 2011/09/16.
50. Pech N, Meyer F, Lippert H, Manger T, Stroh C. Complications and nutrient deficiencies two years after sleeve gastrectomy. BMC Surg. 2012;12:13.
51. Montonen J, Knekt P, Harkanen T, Jarvinen R, Heliovaara M, Aromaa A, et al. Dietary patterns and the incidence of type 2 diabetes. Am J Epidemiol. 2005;161(3):219-27. Epub 2005/01/27.
52. Orozco LJ, Buchleitner AM, Gimenez-Perez G, Roque IFM, Richter B, Mauricio D. Exercise or exercise and diet for preventing type 2 diabetes mellitus. Cochrane Database Syst Rev. 2008(3):CD003054. Epub 2008/07/23.
53. Laws A, Reaven GM. Effect of physical activity on age-related glucose intolerance. Clin Geriatr Med. 1990;6(4):849-63. Epub 1990/11/01.
54. Luo Y, Wang H, Han X, Ren Q, Wang F, Zhang X, et al. Meta-analysis of the association between SNPs in TCF7L2 and type 2 diabetes in East Asian population. Diabetes Res Clin Pract. 2009;85(2):139-46. Epub 2009/06/02.
55. Cugino D, Gianfagna F, Santimone I, de Gaetano G, Donati MB, Iacoviello L, et al. Type 2 diabetes and polymorphisms on chromosome 9p21: a meta-analysis. Nutr Metab Cardiovasc Dis. 2012;22(8):619-25. Epub 2011/02/15.
56. Chang YC, Chang TJ, Jiang YD, Kuo SS, Lee KC, Chiu KC, et al. Association study of the genetic polymorphisms of the transcription factor 7-like 2 (TCF7L2) gene and type 2 diabetes in the Chinese population. Diabetes. 2007;56(10):2631-7. Epub 2007/06/21.
57. Groves CJ, Zeggini E, Minton J, Frayling TM, Weedon MN, Rayner NW, et al. Association analysis of 6,736 U.K. subjects provides replication and confirms TCF7L2 as a type 2 diabetes susceptibility gene with a substantial effect on individual risk. Diabetes. 2006;55(9):2640-4. Epub 2006/08/29.

58. Hansen SK, Nielsen EM, Ek J, Andersen G, Glumer C, Carstensen B, et al. Analysis of separate and combined effects of common variation in KCNJ11 and PPARG on risk of type 2 diabetes. J Clin Endocrinol Metab. 2005;90(6):3629-37. Epub 2005/03/31.
59. Steinbrecher A, Morimoto Y, Heak S, Ollberding NJ, Geller KS, Grandinetti A, et al. The preventable proportion of type 2 diabetes by ethnicity: the multiethnic cohort. Ann Epidemiol. 2011;21(7):526-35. Epub 2011/04/19.
60. IDF. Proportion of people with diabetes (20-79 years). 2009 [cited 2011 23rd of March]; Available from: http://www.diabetesatlas.org/map.
61. Federation ID. Proportion of people with diabetes (20-79 years). 2009 [cited 2011 23rd of March]; Available from: http://www.diabetesatlas.org/map.
62. WHO. Prevalence of diabetes in the WHO Eastern Mediterranean Region. [cited 2011 22th of March]; Available from: http://www.who.int/diabetes/facts/world_figures/en/index2.html.
63. Bacchus RA, Bell JL, Madkour M, Kilshaw B. The prevalence of diabetes mellitus in male Saudi Arabs. Diabetologia. 1982;23(4):330-2. Epub 1982/10/01.
64. Fatani HH, Mira SA, el-Zubier AG. Prevalence of diabetes mellitus in rural Saudi Arabia. Diabetes Care. 1987;10(2):180-3. Epub 1987/03/01.
65. Anokute CC. Epidemiologic studies of diabetes mellitus in Saudi Arabia--Part I--Screening of 3158 males in King Saud University. J R Soc Health. 1990;110(6):201-3. Epub 1990/12/01.
66. El-Hazmi MA, Warsy AS, Al-Swailem AR, Al-Swailem AM, Sulaimani R, Al-Meshari AA. Diabetes mellitus and impaired glucose tolerance in Saudi Arabia. Ann Saudi Med. 1996;16(4):381-5. Epub 1996/07/01.
67. Al-Nuaim AR. Prevalence of glucose intolerance in urban and rural communities in Saudi Arabia. Diabet Med. 1997;14(7):595-602. Epub 1997/07/01.
68. Al-Nozha MM, Al-Maatouq MA, Al-Mazrou YY, Al-Harthi SS, Arafah MR, Khalil MZ, et al. Diabetes mellitus in Saudi Arabia. Saudi Med J. 2004;25(11):1603-10. Epub 2004/12/02.
69. Alqurashi KA, Aljabri KS, Bokhari SA. Prevalence of diabetes mellitus in a Saudi community. Ann Saudi Med. 2011;31(1):19-23. Epub 2011/01/20.
70. Elhadd TA, Al-Amoudi AA, Alzahrani AS. Epidemiology, clinical and complications profile of diabetes in Saudi Arabia: a review. Ann Saudi Med. 2007;27(4):241-50. Epub 2007/08/09.
71. Aljoudi AS, Taha AZ. Knowledge of diabetes risk factors and preventive measures among attendees of a primary care center in eastern Saudi Arabia. Ann Saudi Med. 2009;29(1):15-9. Epub 2009/01/14.
72. Taha AZ. Self-reported knowledge and pattern of physical activity among school students in Al Khobar, Saudi Arabia. East Mediterr Health J. 2008;14(2):344-55. Epub 2008/06/20.
73. Al-Shahri M, Al-Alamei S. Promotion of physical excercise by primary health care physicians in Riyadh city. Saudi Med J. 1998;19:67-9.
74. WHO. BMI classification [cited 2013 13th August]; Available from: http://apps.who.int/bmi/index.jsp?introPage=intro_3.html.
75. el-Hazmi MA, Warsy AS. Prevalence of overweight and obesity in diabetic and non-diabetic Saudis. East Mediterr Health J. 2000;6(2-3):276-82. Epub 2001/09/15.
76. Al-Turki YA. The prevalence of overweight and obesity amongst hypertensive and diabetic adult patients in primary health care. Saudi Med J. 2000;21(4):340-3. Epub 2001/09/05.
77. Al-Rukban MO. Obesity among Saudi male adolescents in Riyadh, Saudi Arabia. Saudi Med J. 2003;24(1):27-33. Epub 2003/02/19.
78. Al-Othaimeen AI, Al-Nozha M, Osman AK. Obesity: an emerging problem in Saudi Arabia. Analysis of data from the National Nutrition Survey. East Mediterr Health J. 2007;13(2):441-8. Epub 2007/08/10.

79. Alam AA. Obesity among female school children in North West Riyadh in relation to affluent lifestyle. Saudi Med J. 2008;29(8):1139-44. Epub 2008/08/12.
80. Al-Dossary SS, Sarkis PE, Hassan A, Ezz El Regal M, Fouda AE. Obesity in Saudi children: a dangerous reality. East Mediterr Health J. 2010;16(9):1003-8. Epub 2011/01/12.
81. Washi SA, Ageib MB. Poor diet quality and food habits are related to impaired nutritional status in 13- to 18-year-old adolescents in Jeddah. Nutr Res. 2010;30(8):527-34. Epub 2010/09/21.
82. Mahfouz AA, Shatoor AS, Khan MY, Daffalla AA, Mostafa OA, Hassanein MA. Nutrition, Physical Activity, and Gender Risks for Adolescent Obesity in Southwestern Saudi Arabia. Saudi J Gastroenterol. 2011;17(5):318-22.
83. AlQuaiz AM, Tayel SA. Barriers to a healthy lifestyle among patients attending primary care clinics at a university hospital in Riyadh. Ann Saudi Med. 2009;29(1):30-5. Epub 2009/01/14.
84. Al-Hazzaa H, Sulaimani R. Maximal oxygen uptake and daily physical activity in 7-to-12 year-old boys. Pediatric exercise science. 1993;5:357-66.
85. Al-Rafaee SA, Al-Hazzaa HM. Physical activity profile of adult males in Riyadh City. Saudi Med J. 2001;22(9):784-9. Epub 2001/10/09.
86. Al-Hazzaa HM. Physical Activity Profiles of College Male Subjects. Journal of King Saud Univeristy. 1990;2(2).
87. Amin TT, Al Khoudair AS, Al Harbi MA, Al Ali AR. Leisure time physical activity in Saudi Arabia: prevalence, pattern and determining factors. Asian Pac J Cancer Prev. 2012;13(1):351-60. Epub 2012/04/17.
88. Al-Nozha MM, Al-Hazzaa HM, Arafah MR, Al-Khadra A, Al-Mazrou YY, Al-Maatouq MA, et al. Prevalence of physical activity and inactivity among Saudis aged 30-70 years. A population-based cross-sectional study. Saudi Med J. 2007;28(4):559-68. Epub 2007/04/26.
89. Al-Hazzaa HM, Abahussain NA, Al-Sobayel HI, Qahwaji DM, Musaiger AO. Physical activity, sedentary behaviors and dietary habits among Saudi adolescents relative to age, gender and region. Int J Behav Nutr Phys Act. 2011;8:140. Epub 2011/12/23.
90. Sidawi B, Al-Hariri MT. The impact of built environment on diabetic patients: the case of Eastern Province, KIngdom of Saudi Arabia. Glob J Health Sci. 2012;4(4):126-38. Epub 2012/09/18.
91. Al-Nakeeb Y, Lyons M, Collins P, Al-Nuaim A, Al-Hazzaa H, Duncan MJ, et al. Obesity, physical activity and sedentary behavior amongst British and Saudi youth: a cross-cultural study. Int J Environ Res Public Health. 2012;9(4):1490-506. Epub 2012/06/13.
92. Cook GC. Rapid glucose absorption in Arabs in Saudi Arbia compared with that in Africans in Zambia. Br Med J. 1976;1(6011):688-9. Epub 1976/03/20.
93. El-Asrar AM, Al-Rubeaan KA, Al-Amro SA, Kangave D, Moharram OA. Risk factors for diabetic retinopathy among Saudi diabetics. Int Ophthalmol. 1998;22(3):155-61. Epub 1999/11/05.
94. Khan AR, Wiseberg JA, Lateef ZA, Khan SA. Prevalence and determinants of diabetic retinopathy in Al hasa region of saudi arabia: primary health care centre based cross-sectional survey, 2007-2009. Middle East Afr J Ophthalmol. 2010;17(3):257-63. Epub 2010/09/17.
95. Awada A, al Rajeh S. The Saudi Stroke Data Bank. Analysis of the first 1000 cases. Acta Neurol Scand. 1999;100(4):265-9. Epub 1999/10/08.
96. al-Rajeh S, Larbi EB, Bademosi O, Awada A, Yousef A, al-Freihi H, et al. Stroke register: experience from the eastern province of Saudi Arabia. Cerebrovasc Dis. 1998;8(2):86-9. Epub 1998/04/21.
97. Mitwalli AH, Al-Swailem AR, Aziz K, Paul TT, Aswad S, Shaheen FA, et al. Etiology of end-stage renal disease in two regions of saudi arabia. Saudi J Kidney Dis Transpl. 1997;8(1):16-20. Epub 1997/01/01.
98. Taha AZ, Bella H. Heart disease risk factors: prevalence and knowledge in a primary care setting, Saudi Arabia. East Mediterr Health J. 1998;4(2):293-300.

99. Al Zahrani HA, Al Bar HM, Bahnassi A, Abdulaal AA. The distribution of peripheral arterial disease in a defined population of elderly high-risk Saudi patients. Int Angiol. 1997;16(2):123-8. Epub 1997/06/01.
100. Akbar DH, Mira SA, Zawawi TH, Malibary HM. Subclinical diabetic neuropathy: a common complication in Saudi diabetics. Saudi Med J. 2000;21(5):433-7. Epub 2001/08/14.
101. Al Helali NS, Abolfotouh MA, Ghanem HM. Pattern of erectile dysfunction in Jeddah city. Saudi Med J. 2001;22(1):34-8. Epub 2001/03/20.
102. Al-Baghli NA, Al-Turki KA, Al-Ghamdi AJ, El-Zubaier AG, Al-Ameer MM, Al-Baghli FA. Control of diabetes mellitus in the Eastern Province of Saudi Arabia: results of screening campaign. East Mediterr Health J. 2010;16(6):621-9. Epub 2010/08/31.
103. Al-Elq AH. Current practice in the management of patients with type 2 diabetes mellitus in Saudi Arabia. Saudi Med J. 2009;30(12):1551-6. Epub 2009/11/26.
104. Burton PR, Tobin MD, Hopper JL. Key concepts in genetic epidemiology. Lancet. 2005;366(9489):941-51. Epub 2005/09/13.
105. Li JKY, Ng MCY, So WY, Chiu CKP, Ozaki R, Tong PCY, et al. Phenotypic and genetic clustering of diabetes and metabolic syndrome in Chinese families with type 2 diabetes mellitus. Diabetes/Metabolism Research Reviews. 2006;22(1):46-52.
106. Arfa I, Abid A, Malouche D, Ben Alaya N, Azegue TR, Mannai I, et al. Familial aggregation and excess maternal transmission of type 2 diabetes in Tunisia. Postgraduate Medical Journal. 2007;83(979):348-51.
107. Matsuda A, Kuzuya T. Diabetic twins in Japan. Diabetes Research & Clinical Practice. 1994;24 Suppl:S63-7.
108. Medici F, Hawa M, Ianari A, Pyke D, Leslie R. Concordance rate for type II diabetes mellitus in monozygotic twins: Actuarial analysis. Diabetologia. 1999;42(2):146-50.
109. Matsuda A, Kuzuya T. Relationship between obesity and concordance rate for type 2 (non-insulin-dependent) diabetes mellitus among twins. Diabetes Research & Clinical Practice. 1994;26(2):137-43.
110. Falconer D, T M. Introduction to quantitative genetics. Essex: Pearson education limited; 1996.
111. Jowett JB, Diego VP, Kotea N, Kowlessur S, Chitson P, Dyer TD, et al. Genetic influences on type 2 diabetes and metabolic syndrome related quantitative traits in Mauritius. Twin Research & Human Genetics: the Official Journal of the International Society for Twin Studies. 2009;12(1):44-52.
112. Simonis-Bik AMC, Boomsma DI, Dekker JM, Diamant M, De Geus EJC, L.M TH, et al. Genetic pattern and genetic heterogeneity of type 2 diabetes mellitus in 62 Uygur families from Xinjiang Uygur Autonomous Region. [Chinese]. Journal of Clinical Rehabilitative Tissue Engineering Research. 2007;11(39):7721-4.
113. Lehtovirta M, Pietilainen KH, Levalahti E, Heikkila K, Groop L, Silventoinen K, et al. Evidence that BMI and type 2 diabetes share only a minor fraction of genetic variance: a follow-up study of 23,585 monozygotic and dizygotic twins from the Finnish Twin Cohort Study. Diabetologia. 2010;53(7):1314-21.
114. Maes HH, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. Behavior Genetics. 1997;27(4):325-51. Epub 1997/07/01.
115. Silventoinen K, Kaprio J. Genetics of Tracking of Body Mass Index from Birth to Late Middle Age: Evidence from Twin and Family Studies. Obesity Facts. 2009;2(3):196-202.
116. Salsberry PJ, Reagan PB. Effects of heritability, shared environment, and nonshared intrauterine conditions on child and adolescent BMI. Obesity. 2010;18(9):1775-80.
117. Goode EL, Cherny SS, Christian JC, Jarvik GP, de Andrade M. Heritability of longitudinal measures of body mass index and lipid and lipoprotein levels in aging twins. Twin Research & Human Genetics: the Official Journal of the International Society for Twin Studies. 2007;10(5):703-11.

118. Bo S, Cavallo-Perin P, Scaglione L, Pagano G. Heritability of cardiovascular risk parameters in subjects with increased susceptibility to non-insulin-dependent diabetes mellitus. Acta Diabetologica. 1997;34(4):280-4.
119. Mustelin L, Silventoinen K, Pietilainen K, Rissanen A, Kaprio J. Physical activity reduces the influence of genetic effects on BMI and waist circumference: a study in young adult twins. International Journal of Obesity. 2009;33(1):29-36.
120. McCaffery JM, Papandonatos GD, Bond DS, Lyons MJ, Wing RR. Gene X environment interaction of vigorous exercise and body mass index among male Vietnam-era twins. Am J Clin Nutr. 2009;89(4):1011-8. Epub 2009/02/20.
121. Silventoinen K, Hasselbalch AL, Lallukka T, Bogl L, Pietilainen KH, Heitmann BL, et al. Modification effects of physical activity and protein intake on heritability of body size and composition. American Journal of Clinical Nutrition. 2009;90(4):1096-103.
122. Keskitalo K, Tuorila H, Spector TD, Cherkas LF, Knaapila A, Kaprio J, et al. The Three-Factor Eating Questionnaire, body mass index, and responses to sweet and salty fatty foods: a twin study of genetic and environmental associations. American Journal of Clinical Nutrition. 2008;88(2):263-71.
123. Silventoinen K, Kaprio J, Yokoyama Y. Genetic Regulation of Pre-Pubertal Development of Body Mass Index: A Longitudinal Study of Japanese Twin Boys and Girls. Behavior Genetics. 2011;41(2):234-41.
124. Ouyang F, Christoffel KK, Brickman WJ, Zimmerman D, Wang B, Xing H, et al. Adiposity is inversely related to insulin sensitivity in relatively lean Chinese adolescents: a population-based twin study. American Journal of Clinical Nutrition. 2010;91(3):662-71.
125. Liu P-H, Jiang Y-D, Chen WJ, Chang C-C, Lee T-C, Sun HS, et al. Genetic and environmental influences on adiponectin, leptin, and BMI among adolescents in Taiwan: a multivariate twin/sibling analysis. Twin Research & Human Genetics: the Official Journal of the International Society for Twin Studies. 2008;11(5):495-504.
126. Hur Y-M. Sex difference in heritability of BMI in South Korean adolescent twins. Obesity. 2007;15(12):2908-11.
127. Luo BF, Du L, Li JX, Pan BY, Xu JM, Chen J, et al. Heritability of metabolic syndrome traits among healthy younger adults: a population based study in China. Journal of Medical Genetics. 2010;47(6):415-20.
128. Luke A, Guo X, Adeyemo AA, Wilks R, Forrester T, Lowe W, Jr., et al. Heritability of obesity-related traits among Nigerians, Jamaicans and US black people. International Journal of Obesity & Related Metabolic Disorders: Journal of the International Association for the Study of Obesity. 2001;25(7):1034-41.
129. Hsu F-C, Lenchik L, Nicklas BJ, Lohman K, Register TC, Mychaleckyj J, et al. Heritability of body composition measured by DXA in the diabetes heart study. Obesity Research. 2005;13(2):312-9.
130. Rasmussen-Torvik LJ, Pankow JS, Jacobs DR, Steffen LM, Moran AM, Steinberger J, et al. Heritability and genetic correlations of insulin sensitivity measured by the euglycaemic clamp. Diabetic Medicine. 2007;24(11):1286-9.
131. Hsueh WC, Mitchell BD, Aburomia R, Pollin T, Sakul H, Gelder Ehm M, et al. Diabetes in the Old Order Amish: characterization and heritability analysis of the Amish Family Diabetes Study. Diabetes Care. 2000;23(5):595-601.
132. Zillikens MC, Yazdanpanah M, Pardo LM, Rivadeneira F, Aulchenko YS, Oostra BA, et al. Sex-specific genetic effects influence variation in body composition. Diabetologia. 2008;51(12):2233-41. Epub 2008/10/08.
133. Mathias RA, Deepa M, Deepa R, Wilson AF, Mohan V. Heritability of quantitative traits associated with type 2 diabetes mellitus in large multiplex families from South India. Metabolism: Clinical & Experimental. 2009;58(10):1439-45.
134. Bellia A, Giardina E, Lauro D, Tesauro M, Di Fede G, Cusumano G, et al. "The Linosa Study": epidemiological and heritability data of the metabolic syndrome in a Caucasian genetic isolate. Nutrition Metabolism & Cardiovascular Diseases. 2009;19(7):455-61.

135. Katoh S, Lehtovirta M, Kaprio J, Harjutsalo V, Koskenvuo M, Eriksson J, et al. Genetic and environmental effects on fasting and postchallenge plasma glucose and serum insulin values in Finnish twins. Journal of Clinical Endocrinology & Metabolism. 2005;90(5):2642-7.
136. Peeters MW, Thomis MA, Loos RJ, Derom CA, Fagard R, Vlietinck RF, et al. Clustering of metabolic risk factors in young adults: genes and environment. Atherosclerosis. 2008;200(1):168-76. Epub 2008/02/05.
137. Beardsall K, Ong KK, Murphy N, Ahmed ML, Zhao JH, Peeters MW, et al. Heritability of childhood weight gain from birth and risk markers for adult metabolic disease in prepubertal twins. Journal of Clinical Endocrinology & Metabolism. 2009;94(10):3708-13.
138. Sung J, Lee K, Song Y-M. Heritabilities of the metabolic syndrome phenotypes and related factors in Korean twins. Journal of Clinical Endocrinology & Metabolism. 2009;94(12):4946-52.
139. Rice T, Vogler GP, Perry TS, Laskarzewski PM, Province MA, Rao DC. Heterogeneity in the familial aggregation of fasting plasma glucose in five North American populations: the Lipid Research Clinics Family Study. International Journal of Epidemiology. 1990;19(2):290-6.
140. Henkin L, Bergman RN, Bowden DW, Ellsworth DL, Haffner SM, Langefeld CD, et al. Genetic epidemiology of insulin resistance and visceral adiposity. The IRAS Family Study design and methods. Annals of Epidemiology. 2003;13(4):211-7.
141. Mills GW, Avery PJ, McCarthy MI, Hattersley AT, Levy JC, Hitman GA, et al. Heritability estimates for beta cell function and features of the insulin resistance syndrome in UK families with an increased susceptibility to type 2 diabetes. Diabetologia. 2004;47(4):732-8.
142. Santos RLP, Zillikens MC, Rivadeneira FR, Pols HAP, Oostra BA, van Duijn CM, et al. Heritability of fasting glucose levels in a young genetically isolated population. Diabetologia. 2006;49(4):667-72.
143. Herbeth B, Samara A, Ndiaye C, Marteau J-B, Berrahmoune H, Siest G, et al. Metabolic syndrome-related composite factors over 5 years in the STANISLAS family study: genetic heritability and common environmental influences. Clinica Chimica Acta. 2010;411(11-12):833-9.
144. Snieder H, Boomsma DI, van Doornen LJ, Neale MC. Bivariate genetic analysis of fasting insulin and glucose levels. Genetic Epidemiology. 1999;16(4):426-46.
145. Samaras K, Nguyen TV, Jenkins AB, Eisman JA, Howard GM, Kelly PJ, et al. Clustering of insulin resistance, total and central abdominal fat: same genes or same environment? Twin Research. 1999;2(3):218-25.
146. Schousboe K, Visscher PM, Henriksen JE, Hopper JL, Sorensen TIA, Kyvik KO. Twin study of genetic and environmental influences on glucose tolerance and indices of insulin sensitivity and secretion. Diabetologia. 2003;46(9):1276-83.
147. Simonis-Bik AMC, Eekhoff EMW, Diamant M, Boomsma DI, Heine RJ, Dekker JM, et al. The heritability of HbA1c and fasting blood glucose in different measurement settings. Twin Research & Human Genetics: the Official Journal of the International Society for Twin Studies. 2008;11(6):597-602.
148. Hanson RL, Elston RC, Pettitt DJ, Bennett PH, Knowler WC. Segregation analysis of non-insulin-dependent diabetes mellitus in Pima Indians: evidence for a major-gene effect. American Journal of Human Genetics. 1995;57(1):160-70.
149. Cook JT, Shields DC, Page RC, Levy JC, Hattersley AT, Shaw JA, et al. Segregation analysis of NIDDM in Caucasian families. Diabetologia. 1994;37(12):1231-40. Epub 1994/12/01.
150. Wang JS, Zhou L, Cheng JL, Shen MY. [Studies on genetic model in familial type 2 diabetes mellitus]. Yi Chuan. 2003;25(6):637-40. Epub 2005/01/11.
151. Ouyang F, Wang W, Shen F, Wang L, Wang Z. [Genetic epidemiological study on non-insulin dependent diabetes mellitus]. Chung-Hua Yu Fang i Hsueh Tsa Chih [Chinese Journal of Preventive Medicine]. 2000;34(2):81-2.
152. McCarthy MI, Hitman GA, Shields DC, Morton NE, Snehalatha C, Mohan V, et al. Family studies of non-insulin-dependent diabetes mellitus in South Indians. 1994.

153. Candidate Gene. In: Elston RC, editor. Biostatistical Genetics and genetic Epidemiology. West Sussex: John Wiley and Sons Ltd; 2002. p. 83.
154. Mange EJ, Mange AP. Chromosome Mapping and Human Genome Project. Basic Human genetics. 2nd ed. Sunderland: Sinauer Associates In.; 1999.
155. McCarthy MI. Dorothy Hodgkin Lecture 2010. From hype to hope? A journey through the genetics of Type 2 diabetes. Diabet Med. 2011;28(2):132-40. Epub 2011/01/12.
156. Ntzani EE, Kavvoura FK. Genetic risk factors for type 2 diabetes: insights from the emerging genomic evidence. Curr Vasc Pharmacol. 2012;10(2):147-55. Epub 2012/01/14.
157. Billings LK, Florez JC. The genetics of type 2 diabetes: what have we learned from GWAS? Ann N Y Acad Sci. 2010;1212:59-77. Epub 2010/11/26.
158. Barroso I, Luan J, Middelberg RP, Harding AH, Franks PW, Jakes RW, et al. Candidate gene association study in type 2 diabetes indicates a role for genes involved in beta-cell function as well as insulin action. PLoS Biol. 2003;1(1):E20. Epub 2003/10/14.
159. Gaulton KJ, Willer CJ, Li Y, Scott LJ, Conneely KN, Jackson AU, et al. Comprehensive association study of type 2 diabetes and related quantitative traits with 222 candidate genes. Diabetes. 2008;57(11):3136-44. Epub 2008/08/06.
160. Vimaleswaran KS, Loos RJ. Progress in the genetics of common obesity and type 2 diabetes. Expert Rev Mol Med. 2010;12:e7. Epub 2010/02/27.
161. Gloyn AL, Weedon MN, Owen KR, Turner MJ, Knight BA, Hitman G, et al. Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. Diabetes. 2003;52(2):568-72. Epub 2003/01/24.
162. Florez JC, Burtt N, de Bakker PI, Almgren P, Tuomi T, Holmkvist J, et al. Haplotype structure and genotype-phenotype correlations of the sulfonylurea receptor and the islet ATP-sensitive potassium channel gene region. Diabetes. 2004;53(5):1360-8. Epub 2004/04/28.
163. Minton JA, Hattersley AT, Owen K, McCarthy MI, Walker M, Latif F, et al. Association studies of genetic variation in the WFS1 gene and type 2 diabetes in U.K. populations. Diabetes. 2002;51(4):1287-90. Epub 2002/03/28.
164. Karasik A, O'Hara C, Srikanta S, Swift M, Soeldner JS, Kahn CR, et al. Genetically programmed selective islet beta-cell loss in diabetic subjects with Wolfram's syndrome. Diabetes Care. 1989;12(2):135-8. Epub 1989/02/01.
165. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science. 2007;316(5829):1331-6. Epub 2007/04/28.
166. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet. 2008;40(5):638-45. Epub 2008/04/01.
167. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010;42(7):579-89. Epub 2010/06/29.
168. Duggirala R, Blangero J, Almasy L, Dyer TD, Williams KL, Leach RJ, et al. Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. American Journal of Human Genetics. 1999;64(4):1127-40. Epub 1999/03/26.
169. Reynisdottir I, Thorleifsson G, Benediktsson R, Sigurdsson G, Emilsson V, Einarsdottir AS, et al. Localization of a susceptibility gene for type 2 diabetes to chromosome 5q34-q35.2. American Journal of Human Genetics. 2003;73(2):323-35. Epub 2003/07/10.
170. Cauchi S, El Achhab Y, Choquet H, Dina C, Krempler F, Weitgasser R, et al. TCF7L2 is reproducibly associated with type 2 diabetes in various ethnic groups: a global meta-analysis. J Mol Med. 2007;85(7):777-82. Epub 2007/05/04.
171. Wu Y, Li H, Loos RJ, Yu Z, Ye X, Chen L, et al. Common variants in CDKAL1, CDKN2A/B, IGF2BP2, SLC30A8, and HHEX/IDE genes are associated with type 2 diabetes and impaired fasting glucose in a Chinese Han population. Diabetes. 2008;57(10):2834-42. Epub 2008/07/18.

172. Chimienti F, Devergnas S, Favier A, Seve M. Identification and cloning of a beta-cell-specific zinc transporter, ZnT-8, localized into insulin secretory granules. Diabetes. 2004;53(9):2330-7. Epub 2004/08/28.
173. Ohara-Imaizumi M, Yoshida M, Aoyagi K, Saito T, Okamura T, Takenaka H, et al. Deletion of CDKAL1 affects mitochondrial ATP generation and first-phase insulin exocytosis. PLoS One. 2010;5(12):e15553. Epub 2010/12/15.
174. Bort R, Martinez-Barbera JP, Beddington RS, Zaret KS. Hex homeobox gene-dependent tissue positioning is required for organogenesis of the ventral pancreas. Development. 2004;131(4):797-806. Epub 2004/01/23.
175. Rodriguez S, Eiriksdottir G, Gaunt TR, Harris TB, Launer LJ, Gudnason V, et al. IGF2BP1, IGF2BP2 and IGF2BP3 genotype, haplotype and genetic model studies in metabolic syndrome traits and diabetes. Growth Horm IGF Res. 2010;20(4):310-8. Epub 2010/07/16.
176. Mulder H, Nagorny CL, Lyssenko V, Groop L. Melatonin receptors in pancreatic islets: good morning to a novel type 2 diabetes gene. Diabetologia. 2009;52(7):1240-9. Epub 2009/04/21.
177. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science. 2007;316(5829):1341-5. Epub 2007/04/28.
178. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nat Genet. 2007;39(6):770-5. Epub 2007/04/27.
179. Lee YH, Kang ES, Kim SH, Han SJ, Kim CH, Kim HJ, et al. Association between polymorphisms in SLC30A8, HHEX, CDKN2A/B, IGF2BP2, FTO, WFS1, CDKAL1, KCNQ1 and type 2 diabetes in the Korean population. J Hum Genet. 2008;53(11-12):991-8. Epub 2008/11/11.
180. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science. 2007;316(5829):1336-41. Epub 2007/04/28.
181. Nemr R, Almawi AW, Echtay A, Sater MS, Daher HS, Almawi WY. Replication study of common variants in CDKAL1 and CDKN2A/2B genes associated with type 2 diabetes in Lebanese Arab population. Diabetes Res Clin Pract. 2012;95(2):e37-40. Epub 2011/11/29.
182. Lu F, Qian Y, Li H, Dong M, Lin Y, Du J, et al. Genetic variants on chromosome 6p21.1 and 6p22.3 are associated with type 2 diabetes risk: a case-control study in Han Chinese. J Hum Genet. 2012;57(5):320-5. Epub 2012/03/23.
183. Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, Furuta H, et al. Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. Nat Genet. 2008;40(9):1092-7. Epub 2008/08/20.
184. Unoki H, Takahashi A, Kawaguchi T, Hara K, Horikoshi M, Andersen G, et al. SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. Nat Genet. 2008;40(9):1098-102. Epub 2008/08/20.
185. Been LF, Ralhan S, Wander GS, Mehra NK, Singh J, Mulvihill JJ, et al. Variants in KCNQ1 increase type II diabetes susceptibility in South Asians: a study of 3,310 subjects from India and the US. BMC Medical Genetics. 2011;12:18.
186. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010;42(2):105-16. Epub 2010/01/19.
187. Ronn T, Wen J, Yang Z, Lu B, Du Y, Groop L, et al. A common variant in MTNR1B, encoding melatonin receptor 1B, is associated with type 2 diabetes and fasting plasma glucose in Han Chinese individuals. Diabetologia. 2009;52(5):830-3. Epub 2009/02/26.
188. Chambers JC, Zhang W, Zabaneh D, Sehmi J, Jain P, McCarthy MI, et al. Common genetic variation near melatonin receptor MTNR1B contributes to raised plasma glucose and increased risk of type 2 diabetes among Indian Asians and European Caucasians. Diabetes. 2009;58(11):2703-8.

189. Li S, Shin HJ, Ding EL, van Dam RM. Adiponectin levels and risk of type 2 diabetes: a systematic review and meta-analysis. JAMA. 2009;302(2):179-88. Epub 2009/07/09.
190. Richards JB, Waterworth D, O'Rahilly S, Hivert MF, Loos RJ, Perry JR, et al. A genome-wide association study reveals variants in ARL15 that influence adiponectin levels. PLoS Genet. 2009;5(12):e1000768. Epub 2009/12/17.
191. Al-Daghri NM, Al-Attas OS, Alokail MS, Alkharfy KM, Hussain T, Yakout S, et al. Adiponectin gene polymorphisms (T45G and G276T), adiponectin levels and risk for metabolic diseases in an Arab population. Gene. 2012;493(1):142-7. Epub 2011/12/14.
192. Winckler W, Weedon MN, Graham RR, McCarroll SA, Purcell S, Almgren P, et al. Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. Diabetes. 2007;56(3):685-93. Epub 2007/03/01.
193. Njolstad PR, Sovik O, Cuesta-Munoz A, Bjorkhaug L, Massa O, Barbetti F, et al. Neonatal diabetes mellitus due to complete glucokinase deficiency. N Engl J Med. 2001;344(21):1588-92. Epub 2001/05/24.
194. Hattersley AT, Ashcroft FM. Activating mutations in Kir6.2 and neonatal diabetes: new clinical syndromes, new scientific insights, and new therapy. Diabetes. 2005;54(9):2503-13. Epub 2005/08/27.
195. Horikawa Y, Iwasaki N, Hara M, Furuta H, Hinokio Y, Cockburn BN, et al. Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. Nat Genet. 1997;17(4):384-5. Epub 1997/12/17.
196. Njolstad PR, Sagen JV, Bjorkhaug L, Odili S, Shehadeh N, Bakry D, et al. Permanent neonatal diabetes caused by glucokinase deficiency: inborn error of the glucose-insulin signaling pathway. Diabetes. 2003;52(11):2854-60. Epub 2003/10/28.
197. Moller AM, Dalgaard LT, Ambye L, Hansen L, Schmitz O, Hansen T, et al. A novel Phe75fsdelT mutation in the hepatocyte nuclear factor-4alpha gene in a Danish pedigree with maturity-onset diabetes of the young. J Clin Endocrinol Metab. 1999;84(1):367-9. Epub 1999/01/27.
198. Reiling E, van 't Riet E, Groenewoud MJ, Welschen LM, van Hove EC, Nijpels G, et al. Combined effects of single-nucleotide polymorphisms in GCK, GCKR, G6PC2 and MTNR1B on fasting plasma glucose and type 2 diabetes risk. Diabetologia. 2009;52(9):1866-70. Epub 2009/06/18.
199. Franklin CS, Aulchenko YS, Huffman JE, Vitart V, Hayward C, Polasek O, et al. The TCF7L2 diabetes risk variant is associated with HbA1(C) levels: a genome-wide association meta-analysis. Annals of Human Genetics. 2010;74(6):471-8.
200. Takeuchi F, Katsuya T, Chakrewarthy S, Yamamoto K, Fujioka A, Serizawa M, et al. Common variants at the GCK, GCKR, G6PC2-ABCB11 and MTNR1B loci are associated with fasting glucose in two Asian populations. Diabetologia. 2010;53(2):299-308. Epub 2009/11/26.
201. Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, Dupuis J, et al. Common variants at 10 genomic loci influence hemoglobin A(C) levels via glycemic and nonglycemic pathways. Diabetes. 2010;59(12):3229-39. Epub 2010/09/23.
202. Pare G, Chasman DI, Parker AN, Nathan DM, Miletich JP, Zee RY, et al. Novel association of HK1 with glycated hemoglobin in a non-diabetic population: a genome-wide evaluation of 14,618 participants in the Women's Genome Health Study. PLoS Genetics. 2008;4(12):e1000312.
203. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007;316(5826):889-94. Epub 2007/04/17.
204. Zhao J, Bradfield JP, Zhang H, Annaiah K, Wang K, Kim CE, et al. Examination of all type 2 diabetes GWAS loci reveals HHEX-IDE as a locus influencing pediatric BMI. Diabetes. 2010;59(3):751-5. Epub 2009/11/26.
205. Timpson NJ, Lindgren CM, Weedon MN, Randall J, Ouwehand WH, Strachan DP, et al. Adiposity-related heterogeneity in patterns of type 2 diabetes susceptibility observed in genome-wide association data. Diabetes. 2009;58(2):505-10.

206. Bouhaha R, Choquet H, Meyre D, Abid Kamoun H, Ennafaa H, Baroudi T, et al. TCF7L2 is associated with type 2 diabetes in nonobese individuals from Tunisia. Pathologie Biologie. 2010;58(6):426-9.
207. Cauchi S, Meyre D, Dina C, Choquet H, Samson C, Gallina S, et al. Transcription factor TCF7L2 genetic study in the French population: expression in human beta-cells and adipose tissue and strong association with type 2 diabetes. Diabetes. 2006;55(10):2903-8. Epub 2006/09/28.
208. Perry JR, Voight BF, Yengo L, Amin N, Dupuis J, Ganser M, et al. Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variants in lean compared to obese cases. PLoS Genet. 2012;8(5):e1002741. Epub 2012/06/14.
209. Nelson DL, Cox MM. Hormonal Regulation and Integration of Mammalian Metabolism. In: Ahr K, editor. Lehninger Principles of Biochemistry. 2nd ed. New York: W. H. Freeman and Company; 2008.
210. Pratt CW, Cornely K. Protein Synthesis. Essential Biochemistry. Hoboken: John Wiley and Sons; 2004. p. 612-41.
211. Zubay G. RNA Synthesis and Processing. Biochemistry. 4th ed: McGraw-Hill Companies; 1998. p. 818-82.
212. Chimienti F, Devergnas S, Pattou F, Schuit F, Garcia-Cuenca R, Vandewalle B, et al. In vivo expression and functional characterization of the zinc transporter ZnT8 in glucose-induced insulin secretion. J Cell Sci. 2006;119(Pt 20):4199-206. Epub 2006/09/21.
213. Voet D, Voet J. Transport through membranes. In: Rose N, editor. Biochemistry. 2nd ed: John Wiley and Sons, Inc; 1995. p. 513-37.
214. Pratt CW, Cornely K. Regulation of Mammalian Fuel Metabolism. Essential Biochemistry. Hoboken: John Wiley and Sons; 2004. p. 500-27.
215. Nelson DL, Cox MM. Biosignaling. In: K A, editor. Lehninger Principles of Biochemistry. 2nd ed. New York: W. H. Freeman and Company; 2008.
216. Nelson DL, Cox D. Biological Membranes and Transport. Lehninger Principles of Biochemistry. 2nd ed. New York: W. H. Freeman and Company; 2008. p. 371-418.
217. Yamada T, Ishihara H, Tamura A, Takahashi R, Yamaguchi S, Takei D, et al. WFS1-deficiency increases endoplasmic reticulum stress, impairs cell cycle progression and triggers the apoptotic pathway specifically in pancreatic beta-cells. Hum Mol Genet. 2006;15(10):1600-9. Epub 2006/03/31.
218. Hattersley AT, Tooke JE. The fetal insulin hypothesis: an alternative explanation of the association of low birthweight with diabetes and vascular disease. Lancet. 1999;353(9166):1789-92. Epub 1999/05/29.
219. Freathy RM, Mook-Kanamori DO, Sovio U, Prokopenko I, Timpson NJ, Berry DJ, et al. Variants in ADCY5 and near CCNL1 are associated with fetal growth and birth weight. Nat Genet. 2010;42(5):430-5. Epub 2010/04/08.
220. Andersson EA, Pilgaard K, Pisinger C, Harder MN, Grarup N, Faerch K, et al. Type 2 diabetes risk alleles near ADCY5, CDKAL1 and HHEX-IDE are associated with reduced birthweight. Diabetologia. 2010;53(9):1908-16.
221. Morgan AR, Thompson JM, Murphy R, Black PN, Lam WJ, Ferguson LR, et al. Obesity and diabetes genes are associated with being born small for gestational age: results from the Auckland Birthweight Collaborative study. BMC Med Genet. 2010;11:125. Epub 2010/08/18.
222. Zhao J, Li M, Bradfield JP, Wang K, Zhang H, Sleiman P, et al. Examination of type 2 diabetes loci implicates CDKAL1 as a birth weight gene. Diabetes. 2009;58(10):2414-8.
223. Chimienti F, Favier A, Seve M. ZnT-8, a pancreatic beta-cell-specific zinc transporter. Biometals. 2005;18(4):313-7. Epub 2005/09/15.
224. Pound LD, Sarkar SA, Benninger RK, Wang Y, Suwanichkul A, Shadoan MK, et al. Deletion of the mouse Slc30a8 gene encoding zinc transporter-8 results in impaired insulin secretion. Biochem J. 2009;421(3):371-6. Epub 2009/05/20.

225. Ingelsson E, Langenberg C, Hivert M-F, Prokopenko I, Lyssenko V, Dupuis J, et al. Detailed physiologic characterization reveals diverse mechanisms for novel genetic Loci regulating glucose and insulin metabolism in humans. Diabetes. 2010;59(5):1266-75.
226. Wang H, Antinozzi PA, Hagenfeldt KA, Maechler P, Wollheim CB. Molecular targets of a human HNF1 alpha mutation responsible for pancreatic beta-cell dysfunction. EMBO J. 2000;19(16):4257-64. Epub 2000/08/16.
227. Beer NL, Tribble ND, McCulloch LJ, Roos C, Johnson PR, Orho-Melander M, et al. The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. Hum Mol Genet. 2009;18(21):4081-8. Epub 2009/08/01.
228. Ragvin A, Moro E, Fredman D, Navratilova P, Drivenes O, Engstrom PG, et al. Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to HHEX, SOX4, and IRX3. Proc Natl Acad Sci U S A. 2010;107(2):775-80. Epub 2010/01/19.
229. Inagaki N, Gonoi T, Clement JPt, Namba N, Inazawa J, Gonzalez G, et al. Reconstitution of IKATP: an inward rectifier subunit plus the sulfonylurea receptor. Science. 1995;270(5239):1166-70. Epub 1995/11/17.
230. Pearson ER, Flechtner I, Njolstad PR, Malecki MT, Flanagan SE, Larkin B, et al. Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. N Engl J Med. 2006;355(5):467-77. Epub 2006/08/04.
231. Pearson ER, Donnelly LA, Kimber C, Whitley A, Doney AS, McCarthy MI, et al. Variation in TCF7L2 influences therapeutic response to sulfonylureas: a GoDARTs study. Diabetes. 2007;56(8):2178-82. Epub 2007/05/24.
232. Wang Q, Curran ME, Splawski I, Burn TC, Millholland JM, VanRaay TJ, et al. Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias. Nat Genet. 1996;12(1):17-23. Epub 1996/01/01.
233. Yamagata K, Senokuchi T, Lu M, Takemoto M, Fazlul Karim M, Go C, et al. Voltage-gated K(+) channel KCNQ1 regulates insulin secretion in MIN6 beta-cell line. Biochem Biophys Res Commun. 2011;407(3):620-5. Epub 2011/03/24.
234. Takei D, Ishihara H, Yamaguchi S, Yamada T, Tamura A, Katagiri H, et al. WFS1 protein modulates the free Ca(2+) concentration in the endoplasmic reticulum. FEBS Lett. 2006;580(24):5635-40. Epub 2006/09/23.
235. Hatanaka M, Tanabe K, Yanai A, Ohta Y, Kondo M, Akiyama M, et al. Wolfram syndrome 1 gene (WFS1) product localizes to secretory granules and determines granule acidification in pancreatic beta-cells. Hum Mol Genet. 2011;20(7):1274-84. Epub 2011/01/05.
236. Barroso I, Gurnell M, Crowley VE, Agostini M, Schwabe JW, Soos MA, et al. Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension. Nature. 1999;402(6764):880-3. Epub 2000/01/06.
237. Armoni M, Kritz N, Harel C, Bar-Yoseph F, Chen H, Quon MJ, et al. Peroxisome proliferator-activated receptor-gamma represses GLUT4 promoter activity in primary adipocytes, and rosiglitazone alleviates this effect. J Biol Chem. 2003;278(33):30614-23. Epub 2003/06/05.
238. Knowler WC, Bennett PH, Hamman RF, Miller M. Diabetes incidence and prevalence in Pima Indians: a 19-fold greater incidence than in Rochester, Minnesota. American Journal of Epidemiology. 1978;108(6):497-505. Epub 1978/12/01.
239. Venkataraman R, Nanda NC, Baweja G, Parikh N, Bhatia V. Prevalence of diabetes mellitus and related conditions in Asian Indians living in the United States. American Journal of Cardiology. 2004;94(7):977-80. Epub 2004/10/07.
240. Health Do. Who gets diabetes? 2007 [cited 2011 29th of April]; Available from: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/Browsable/DH_4899956.
241. Neel JV. Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? American Journal of Human Genetics. 1962;14:353-62. Epub 1962/12/01.

242. Helgason A, Palsson S, Thorleifsson G, Grant SF, Emilsson V, Gunnarsdottir S, et al. Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. Nature Genetics. 2007;39(2):218-25.
243. Southam L, Soranzo N, Montgomery SB, Frayling TM, McCarthy MI, Barroso I, et al. Is the thrifty genotype hypothesis supported by evidence based on confirmed type 2 diabetes- and obesity-susceptibility variants? Diabetologia. 2009;52(9):1846-51. Epub 2009/06/16.
244. Klimentidis YC, Abrams M, Wang J, Fernandez JR, Allison DB. Natural selection at genomic regions associated with obesity and type-2 diabetes: East Asians and sub-Saharan Africans exhibit high levels of differentiation at type-2 diabetes regions. Human Genetics. 2011;129(4):407-18.
245. Bener A, Zirie M, Al-Rikabi A. Genetics, obesity, and environmental risk factors associated with type 2 diabetes. Croat Med J. 2005;46(2):302-7. Epub 2005/04/26.
246. Anokute CC. Suspected synergism between consanguinity and familial aggregation in type 2 diabetes mellitus in Saudi Arabia. J R Soc Health. 1992;112(4):167-9. Epub 1992/08/01.
247. Abu-Amero KK, Hellani A, Gonzalez AM, Larruga JM, Cabrera VM, Underhill PA. Saudi Arabian Y-Chromosome diversity and its relationship with nearby regions. BMC Genetics. 2009;10:59. Epub 2009/09/24.
248. Abu-Amero KK, Larruga JM, Cabrera VM, Gonzalez AM. Mitochondrial DNA structure in the Arabian Peninsula. BMC Evol Biol. 2008;8:45. Epub 2008/02/14.
249. Abu-Amero KK, Gonzalez AM, Larruga JM, Bosley TM, Cabrera VM. Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. BMC Evol Biol. 2007;7:32. Epub 2007/03/03.
250. el-Hazmi MA, al-Swailem AR, Warsy AS, al-Swailem AM, Sulaimani R, al-Meshari AA. Consanguinity among the Saudi Arabian population. J Med Genet. 1995;32(8):623-6. Epub 1995/08/01.
251. El Mouzan MI, Al Salloum AA, Al Herbish AS, Qurachi MM, Al Omar AA. Consanguinity and major genetic disorders in Saudi children: a community-based cross-sectional study. Annals of Saudi Medicine. 2008;28(3):169-73.
252. El-Mouzan MI, Al-Salloum AA, Al-Herbish AS, Qurachi MM, Al-Omar AA. Regional variations in the prevalence of consanguinity in Saudi Arabia. Saudi Medical Journal. 2007;28(12):1881-4.
253. Wong SS, Anokute CC. The effect of consanguinity on pregnancy outcome in Saudi Arabia. Journal of the Royal Society of Health. 1990;110(4):146-7.
254. Mumtaz G, Tamim H, Kanaan M, Khawaja M, Khogali M, Wakim G, et al. Effect of consanguinity on birth weight for gestational age in a developing country. American Journal of Epidemiology. 2007;165(7):742-52.
255. Kulkarni ML, Kurian M. Consanguinity and its effect on fetal growth and development: a south Indian study. Journal of Medical Genetics. 1990;27(6):348-52.
256. Shami SA, Qadeer T, Schmitt LH, Bittles AH. Consanguinity, gestational period and anthropometric measurements at birth in Pakistan. Annals of Human Biology. 1991;18(6):523-7.
257. Magnus P, Berg K, Bjerkedal T. Association of parental consanguinity with decreased birth weight and increased rate of early death and congenital malformations. Clinical Genetics. 1985;28(4):335-42.
258. Rudan I, Rudan D, Campbell H, Carothers A, Wright A, Smolej-Narancic N, et al. Inbreeding and risk of late onset complex disease. J Med Genet. 2003;40(12):925-32. Epub 2003/12/20.
259. Leutenegger AL, Labalme A, Génin E, Toutain A, Steichen E, Clerget-Darpoux F, et al. Using Genomic Inbreeding Coefficient Estimates for Homozygosity Mapping of Rare Recessive Traits: Application to Taybi-Linder Syndrome. American Journal of Human Genetics. 2006;79(1):62-6.

260. Li MH, Strandén I, Tiirikka T, Sevón-Aimonen ML, Kantanen J. A Comparison of Approaches to Estimate the Inbreeding Coefficient and Pairwise Relatedness Using Genomic and Pedigree Data in a Sheep Population. PLoS One. 2011;6(11).
261. Falconer DS. Small Populations: Pedigreed Populations and Close Inbreeding. Introduction to Quantitative Genetics. 2nd ed. New York: Longman Scientific and Technical 1981. p. 75-93.
262. MRC. Dietary Assessment- Recals. [cited 2013 16/02]; Available from: http://dapa-toolkit.mrc.ac.uk/dietary-assessment/methods/recalls/index.html.
263. Nelson M, Erens B, Bates B, Church S, Boshier T. Low Income Diet and Nutrition Survey. 2007 [cited 2013 16/02]; Available from: http://www.food.gov.uk/multimedia/pdfs/lidnssummary.pdf.
264. NCI. ASA24 (Automated Self-Adminstrated 24-hour Recal). NCI (National Cancer Institute); 2012 [cited 2013 16/02]; Available from: http://riskfactor.cancer.gov/tools/instruments/asa24/background/.
265. MRC. Dietary Assessment- Weight Food Diary. [cited 2013 16/02]; Available from: http://dapa-toolkit.mrc.ac.uk/dietary-assessment/methods/weighed-food-diary/index.html
266. MRC. Dietary Assessment-Food Frequency Questionnaire. [cited 2013 16/02]; Available from: http://dapa-toolkit.mrc.ac.uk/dietary-assessment/methods/food-frequency-questionnaire/index.html.
267. Yang YJ, Kim MK, Hwang SH, Ahn Y, Shim JE, Kim DH. Relative validities of 3-day food records and the food frequency questionnaire. Nutr Res Pract. 2010;4(2):142-8. Epub 2010/05/13.
268. Brunner E, Stallone D, Juneja M, Bingham S, Marmot M. Dietary assessment in Whitehall II: comparison of 7 d diet diary and food-frequency questionnaire and validity against biomarkers. Br J Nutr. 2001;86(3):405-14. Epub 2001/09/26.
269. Marks GC, Hughes MC, van der Pols JC. The effect of personal characteristics on the validity of nutrient intake estimates using a food-frequency questionnaire. Public Health Nutr. 2006;9(3):394-402. Epub 2006/05/11.
270. Ocke MC, Bueno-de-Mesquita HB, Pols MA, Smit HA, van Staveren WA, Kromhout D. The Dutch EPIC food frequency questionnaire. II. Relative validity and reproducibility for nutrients. International Journal of Epidemiology. 1997;26 Suppl 1:S49-58. Epub 1997/01/01.
271. Welch AA, Luben R, Khaw KT, Bingham SA. The CAFE computer program for nutritional analysis of the EPIC-Norfolk food frequency questionnaire and identification of extreme nutrient values. J Hum Nutr Diet. 2005;18(2):99-116. Epub 2005/03/25.
272. JBS. Heart2005 10th of November 2011 10th of November 2011]; 9:[20 p.]. Available from: http://www.bcs.com/download/651/JBS2final.pdf.
273. Zhou X, Pang Z, Gao W, Wang S, Zhang L, Ning F, et al. Performance of an A1C and fasting capillary blood glucose test for screening newly diagnosed diabetes and pre-diabetes defined by an oral glucose tolerance test in Qingdao, China. Diabetes Care. 2010;33(3):545-50. Epub 2009/12/17.
274. Machin D, Campbell M, Tan S, Tan S. The Correlation Coefficient. Sample Size Tables for Clinical Studies. 3rd ed. Chichester: Wiley-Blackwell; 2009.
275. al-Abdulkareem AA, Ballal SG. Consanguineous marriage in an urban area of Saudi Arabia: rates and adverse health effects on the offspring. Journal of Community Health. 1998;23(1):75-83.
276. BNF. Men, How much shoud I be eating? 2012 [cited 2013 06/02]; Available from: http://www.nutrition.org.uk/healthyliving/lifestages/men?start=1.
277. Trichopoulou A, Gnardellis C, Lagiou A, Benetou V, Trichopoulos D. Body mass index in relation to energy intake and expenditure among adults in Greece. Epidemiology. 2000;11(3):333-6. Epub 2000/04/28.
278. Anand SS, Meyre D, Pare G, Bailey SD, Xie C, Zhang X, et al. Genetic Information and the Prediction of Incident Type 2 Diabetes in a High-Risk Multi-Ethnic Population: The EpiDREAM Genetic Study. Diabetes Care. 2013. Epub 2013/04/23.

279. MH C. HWE Calculator. 2005-2008 [cited 2013 10th February]; Available from: http://www.tufts.edu/~mcourt01/Documents/Court%20lab%20-%20HW%20calculator.xls.
280. Bener A, Yousafzai MT, Al-Hamaq AO, Mohammad AG, Defronzo RA. Parental transmission of type 2 diabetes mellitus in a highly endogamous population. World J Diabetes. 2013;4(2):40-6. Epub 2013/04/20.
281. De Silva SN, Weerasuriya N, De Alwis NM, De Silva MW, Fernando DJ. Excess maternal transmission and familial aggregation of Type 2 diabetes in Sri Lanka. Diabetes Res Clin Pract. 2002;58(3):173-7. Epub 2002/11/05.
282. Karter AJ, Rowell SE, Ackerson LM, Mitchell BD, Ferrara A, Selby JV, et al. Excess maternal transmission of type 2 diabetes. The Northern California Kaiser Permanente Diabetes Registry. Diabetes Care. 1999;22(6):938-43. Epub 1999/06/18.
283. Sheu WH, Song YM, Lee WJ, Yao YE, Lin CJ, Tseng LN, et al. Family aggregation and maternal inheritance of Chinese type 2 diabetes mellitus in Taiwan. Zhonghua Yi Xue Za Zhi (Taipei). 1999;62(3):146-51. Epub 1999/05/01.
284. Mitchell BD, Kammerer CM, Reinhart LJ, Stern MP, MacCluer JW. Is there an excess in maternal transmission of NIDDM? Diabetologia. 1995;38(3):314-7. Epub 1995/03/01.
285. Viswanathan M, McCarthy MI, Snehalatha C, Hitman GA, Ramachandran A. Familial aggregation of type 2 (non-insulin-dependent) diabetes mellitus in south India; absence of excess maternal transmission. Diabet Med. 1996;13(3):232-7. Epub 1996/03/01.
286. Dabelea D. The predisposition to obesity and diabetes in offspring of diabetic mothers. Diabetes Care. 2007;30 Suppl 2:S169-74. Epub 2008/02/27.
287. Kroke A, Klipstein-Grobusch K, Voss S, Moseneder J, Thielecke F, Noack R, et al. Validation of a self-administered food-frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water, urinary nitrogen, and repeated 24-h dietary recall methods. Am J Clin Nutr. 1999;70(4):439-47. Epub 1999/09/29.
288. Pisani P, Faggiano F, Krogh V, Palli D, Vineis P, Berrino F. Relative validity and reproducibility of a food frequency dietary questionnaire for use in the Italian EPIC centres. International Journal of Epidemiology. 1997;26 Suppl 1:S152-60. Epub 1997/01/01.
289. Boeing H, Bohlscheid-Thomas S, Voss S, Schneeweiss S, Wahrendorf J. The relative validity of vitamin intakes derived from a food frequency questionnaire compared to 24-hour recalls and biological measurements: results from the EPIC pilot study in Germany. European Prospective Investigation into Cancer and Nutrition. International Journal of Epidemiology. 1997;26 Suppl 1:S82-90. Epub 1997/01/01.
290. Wareham NJ, Jakes RW, Rennie KL, Mitchell J, Hennings S, Day NE. Validity and repeatability of the EPIC-Norfolk Physical Activity Questionnaire. International Journal of Epidemiology. 2002;31(1):168-74. Epub 2002/03/27.
291. Lagerros YT. Body mass index in relation to energy intake and expenditure among adults in Greece. Epidemiology. 2001;12(1):136-7. Epub 2001/01/04.
292. Al-Hazzaa HM, Abahussain NA, Al-Sobayel HI, Qahwaji DM, Musaiger AO. Lifestyle factors associated with overweight and obesity among Saudi adolescents. BMC Public Health. 2012;12:354. Epub 2012/05/18.
293. Elsanosi R, Bani I, Milaat W, El-Najjar M, Makeen A, Yagob U. Socio-Medical Problem of the Habituation of Khat Chewing in Jazan Region in Southern Saudi Arabia. European Journal of Scientific Research. 2011;63(1):122-33.
294. Perneger TV. What's wrong with Bonferroni adjustments. Bmj. 1998;316(7139):1236-8. Epub 1998/05/16.
295. S. H. A Simple Sequentially Rejective Bonferroni Test Procedure. Scandinavian Journal of Statistics. 1979;6:65 -70.
296. Y B, Y H. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society 1995(57):289 -300.

297. NCBI. The Human Genome. Garland Science; 2002 [cited 2013 30/04]; Available from: http://www.ncbi.nlm.nih.gov/books/NBK21134/.
298. Wagner R, Dudziak K, Herzberg-Schafer SA, Machicao F, Stefan N, Staiger H, et al. Glucose-raising genetic variants in MADD and ADCY5 impair conversion of proinsulin to insulin. PLoS One. 2011;6(8):e23639. Epub 2011/09/03.
299. Kan MY, Zhou DZ, Zhang D, Zhang Z, Chen Z, Yang YF, et al. Two susceptible diabetogenic variants near/in MTNR1B are associated with fasting plasma glucose in a Han Chinese cohort. Diabet Med. 2010;27(5):598-602. Epub 2010/06/12.
300. dbSNP. Short Genetic Variations. 2013 [cited 2013 10/03]; Available from: http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=7957197.
301. dbSNP. Short Genetic Variations. 2013 [cited 2013 10/04]; Available from: http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=757210.
302. Halbach H. Medical aspects of the chewing of khat leaves. Bull World Health Organ. 1972;47(1):21-9. Epub 1972/01/01.
303. Tulloch AD, Frayn E, Craig TK, Nicholson TR. Khat use among Somali mental health service users in South London. Soc Psychiatry Psychiatr Epidemiol. 2012;47(10):1649-56. Epub 2012/01/18.
304. Corkery JM, Schifano F, Oyefeso A, Ghodse AH, Tonia T, Naidoo V, et al. Overview of literature and information on "khat-related" mortality: a call for recognition of the issue and further research. Ann Ist Super Sanita. 2011;47(4):445-64. Epub 2011/12/24.
305. Al-Motarreb A, Baker K, Broadley KJ. Khat: pharmacological and medical aspects and its social use in Yemen. Phytother Res. 2002;16(5):403-13. Epub 2002/08/31.
306. Al-Sanosy RM. Pattern of khat abuse and academic performance among secondary school and college students in jazan region, kingdom of saudi arabia (ksa). J Family Community Med. 2009;16(3):89-95. Epub 2009/09/01.
307. Brenneisen R, Fisch HU, Koelbing U, Geisshusler S, Kalix P. Amphetamine-like effects in humans of the khat alkaloid cathinone. Br J Clin Pharmacol. 1990;30(6):825-8. Epub 1990/12/01.
308. Hassan NA, Gunaid AA, El-Khally FM, Murray-Lyon IM. The effect of chewing Khat leaves on human mood. Saudi Med J. 2002;23(7):850-3. Epub 2002/08/14.
309. Hassan NA, Gunaid AA, El Khally FM, Murray-Lyon IM. The subjective effects of chewing Qat leaves in human volunteers. Ann Saudi Med. 2002;22(1-2):34-7. Epub 2007/01/30.
310. Saif-Ali R, Al-Qirbi A, Al-Geiry A, M AL-H. Effect of Catha edulis on plasma glucose and C-peptide in both type 2 diabetics and non-diabetics. J Ethnopharmacol. 2003;86(1):45-9. Epub 2003/04/11.
311. Hassan F, AL-Serafy T, AL-Qiary A, Radman A. Effect of Khat Ingestion on some Glucogenic Hormones Patients. Biochemical Society Transactions. 2000;28(5).

Appendix 1: Directory of Health in Jazan ethical approval

Kingdom of Saudi Arabia
ministry of health
Directorate of health affaires jazan
PHCC ADMINISTRATION
training department
CMI and CLINICAL SUPERVISION

وزارة الصحة
Ministry of Health

المملكة العربية السعودية
وزارة الصحة
المديرية العامة للشئون الصحية بجازان
إدارة المراكز الصحية
قسم التدريب والإشراف الاكلينيكى

17/07/2011

To Whom it May Concern

This is to confirm that Dr.Ibrahim Metaan Gosadi has been given the ethical approval to conduct his investigation on the potential effect of inbreeding on type 2 diabetes susceptibility in the Saudi society. He has been given the permission to recruit type 2 diabetes patients attending primary health care centers in Jazan area and to recruit their family members. He has also been given an access to the participants' medical records.

Best regards

[redacted]

Dr Mousa M AL therwi
Consultant family medicine
Primary health care administration
Al therwi-1420@hotmail.com
Mobile no 00966506739833

Fax—00966073221087---------- KSA ---PHC JAZAN

Appendix 2: School of Health and Related Research Ethical Approval

The University Of Sheffield.

Cheryl Oliver
Ethics Committee Administrator

Regent Court
30 Regent Street
Sheffield S1 4DA
Telephone: +44 (0) 114 2220871
Fax: +44 (0) 114 272 4095 (non confidential)
Email: ca.oliver@sheffield.ac.uk

Our ref: 0544/CAO

18 June 2012

Ibrahim Gosadi
ScHARR

Dear Ibrahim

Investigating the potential effect of inbreeding on type 2 Diabetes susceptibility in a Saudi population

Thank you for submitting the above research project for approval by the ScHARR Research Ethics Committee. On behalf of the University Chair of Ethics who reviewed your project, I am pleased to inform you that on 18 June 2012 the project was approved on ethics grounds, on the basis that you will adhere to the documents that you submitted for ethics review.

The research must be conducted within the requirements of the hosting/employing organisation or the organisation where the research is being undertaken.

If during the course of the project you need to deviate significantly from the documents you submitted for review, please inform me since written approval will be required. Please also inform me should you decide to terminate the project prematurely.

The Ethics Committee appreciate the amount of work done on the application and would like to congratulate you and your supervisors.

Wishing you the best of luck with your research.

Yours sincerely

Cheryl Oliver
Ethics Committee Administrator

Appendix 3: University Insurance

The University Of Sheffield.

Finance and Commercial

To Ibrahim Gosadi

Your ref

Date Issued 19.4.12

Certificate of Insurances (non clinical trial)

Trial Number NCT 11/30

Department SoHARR

Title of Trial

Investigating the Potential Effect of Inbreeding on Type 2 Diabetes Susceptability in a Saudi Arabian Population

Name of Investigators As stated

Commencement Date 01/07/2012

The University has in place insurance against liabilities for which it may be legally liable and this cover includes any such liabilities arising out of the above research project/study

[redacted]

Colin Rose MA ACII
Risk2Value Ltd, Insurance Adviser to the University of Sheffield

M:\fi_share\Insurance\Insurances\CTRIALS\NCT TRIALS\nct certificate.docx

Appendix 4: Participants Information sheets

Investigating the Potential Effect of Consanguinity on Type 2 Diabetes Susceptibility in a Saudi Population

Patients Information Sheet

You are being invited to take part in a research project. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this information.

What is the project's purpose?

I am conducting this research as a partial requirement for my PhD degree at University of Sheffield, UK. The purpose of this research is to study genetic and lifestyle risk factors for diabetes in a Saudi population. One of the factors we are measuring is marriage between relatives. This study will be very helpful in advancing knowledge related to genetic causes of diabetes.

Why have I been chosen?

We are inviting patients with diabetes and their healthy brothers and sisters to take part in this study. To identify genetic and lifestyle risk factors that are associated with diabetes , we need to study DNA (the substance which comprises the body's genetic code) and lifestyle of individuals with diabetes. A total of 400 participants (150 diabetes patients and 250 healthy brothers or sisters of patients) from Jazan region are required to take part in this research.

Do I have to take part?

It is up to you to decide whether or not to take part. You will be given this information sheet to keep and be asked to sign a consent form once you agree to participate. You can still withdraw at any time without it affecting any health care that you are entitled to in any way. You do not have to give a reason. If you consent to participate, it is entirely up to you to decide what information to provide.

What will happen to me if I take part?

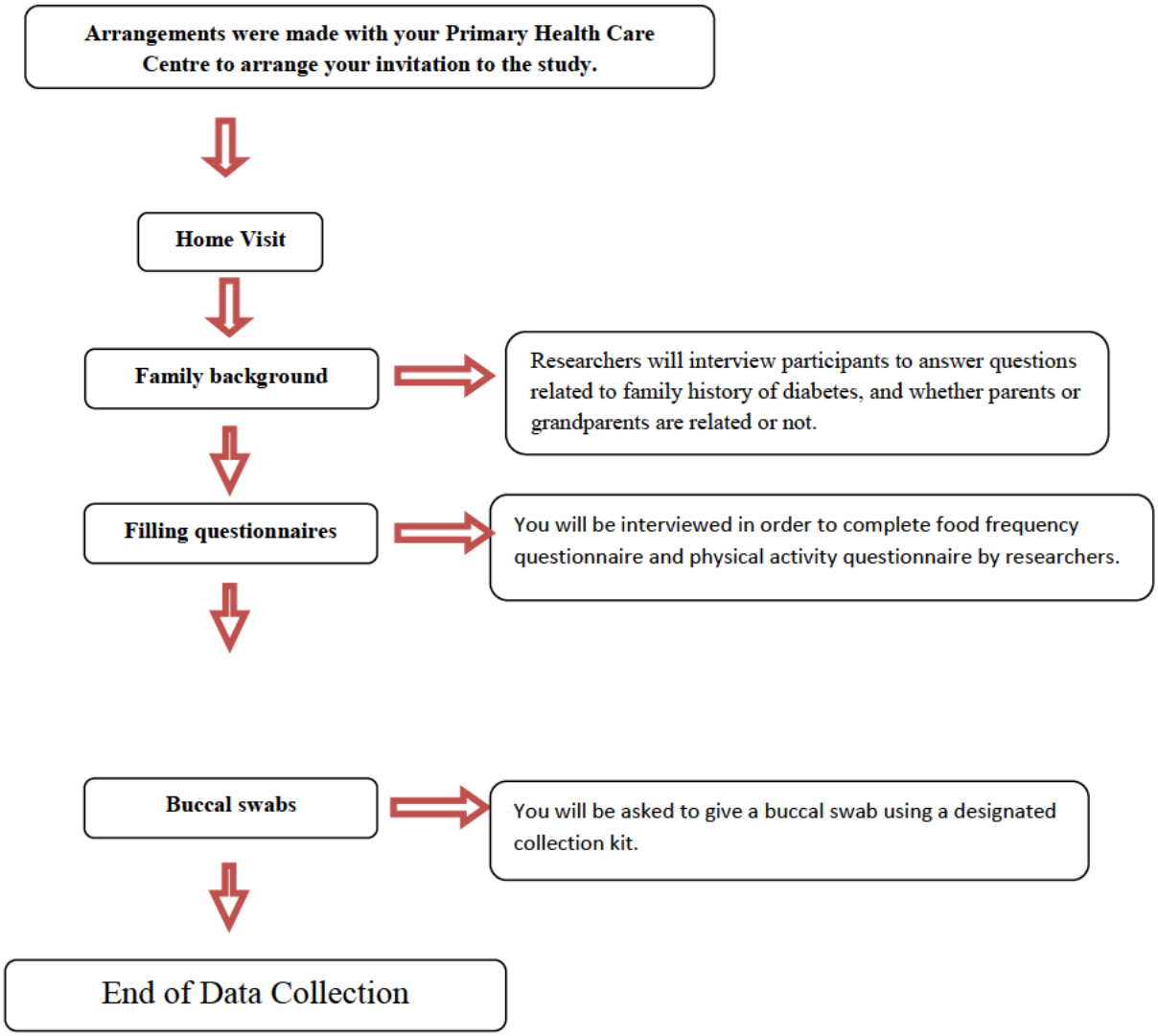
Figure 1 is a flow chart explaining the research procedure and what you would be expected to do. Data and samples will be collected during a visit to your home. You will be interviewed once in your home. During the interview, you will be asked about family tree, physical activity and how frequently you eat various foods. The approximate time period of Interview and buccal swabs collection will be 1 hour. The interview and measurement sessions will be with the principal investigator (me) for male participants and a female medical practitioner for female participants.

What do I have to do?

The interviewer will ask questions about your diet and physical exercise. Otherwise, there will be no other restriction on lifestyle or general activities.

You will be asked to give a buccal swab using a designated collection kit. Swabs will be taken one hour after eating or drinking and mouth rinsed prior to taking sample. You will be given a swab to insert into your mouth and rub firmly against the inside of your cheek for 1 minute. Afterwards, the investigator will l place the swab in a designated tube with preservative capsules .At the end of data collection, these kits will be sent to a specialist company for DNA extraction and genetic analysis.

Figure 1:Research Process and Guidelines:



What are the possible disadvantages and risks of taking part?

There are no foreseeable disadvantages and risks linked to research participation other than time consumption. However, it might be inconvenient for some participants to give buccal swabs. Giving your DNA samples for analysis will not impose any risk of identifying other medical or genetic conditions as we are looking only at genes which may be involved in diabetes risk. Collected swabs will be used to perform a DNA extraction and genetic analysis for one time only. This study will not preserve DNA samples once the analysis has been performed.

What are the possible benefits of taking part?

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will be very helpful in designing intervention programs for future prevention of diabetes.

What if something goes wrong?

If you have any query/complaint you can contact me without hesitation at my given contact number. However, if you feel that I could not handle your query/complaint appropriately then you can contact my supervisor at *(m.d.teare@sheffield.ac.uk*).

Will my taking part in this project be kept confidential?

All the information that you will provide during the session will be kept strictly confidential under my custody. Data will be analysed anonymously and individual name/identity of any participant will not be disclosed in any paper/report or publication.Buccal swabs will not be labelled by any identifiable data and researchers will only use anonymised codes to link between buccal swabs and other collected data.

What will happen to the results of the research project?

Results of the study will be submitted to The University of Sheffield in January 2014. Findings of the study will be sent to reputable scientific journals for publication. If you are interested in knowing the findings of this study, you can contact the principal investigator.

Who is organising and funding the research?

This study is partially sponsored by Saudi Cultural Bureau in London, UK and the remaining financial requirements are met by the principal investigator .

Who has ethically reviewed the project?

The Ethics Review Committee of School of Health and Related Research (ScHARR) at University of Sheffield, UK and Directory of Health in Jazan region, Saudi Arabia managed the ethics review process.

Finally …

Participants will be given a copy of the information sheet and a signed consent form to keep.

If you need further information you can contact at given below number. I will be very thankful for your cooperation.

Best Wishes

Dr. Ibrahim Gosadi

Principal Investigator, postgraduate research student student

School of Health and Related Research (ScHARR)

University of Sheffield, Sheffield, UK

Contact No. +966 562447123

Email: igosadi@ksu.edu.sa

Investigating the Potential Effect of Consanguinity on Type 2 Diabetes Susceptibility in a Saudi Population

Healthy Participant Information Sheet

You are being invited to take part in a research project. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this information.

What is the project's purpose?

I am conducting this research as a partial requirement for my PhD degree at University of Sheffield, UK. The purpose of this research is to study genetic and lifestyle risk factors for diabetes in a Saudi population. One of the factors we are measuring is marriage between relatives. This study will be very helpful in advancing knowledge related to genetic causes of diabetes.

Why have I been chosen?

We are inviting patients with diabetes and their healthy brothers and sisters to take part in this study.To identify genetic and lifestyle risk factors that are associated with diabetes , we need to study DNA (the substance which comprises the body's genetic code) and lifestyle of individuals who have a brother or sister with diabetes. A total of 400 participants (150 diabetes patients and 250 healthy brothers or sisters of patients) from Jazan region are required to take part in this research.

Do I have to take part?

It is up to you to decide whether or not to take part. You will be given this information sheet to keep and be asked to sign a consent form once you agree to participate. You can still withdraw at any time without it affecting any health care that you are entitled to in any way. You do not have to give a reason. If you consent to participate, it is entirely up to you to decide what information to provide.

What will happen to me if I take part?

Figure 1 is a flow chart explaining the research procedure and what you would be expected to do. Data and samples will be collected during a visit to your home. You will be interviewed once in your home. During the interview you will be asked about your family tree, physical activity and how frequently you eat various foods. The approximate time period of Interview, physical measurements, and buccal swabs collection will be 1 hour. The interviews and measurement sessions will be with the principal investigator (me) for the male participants and a female medical practitioner for female participants

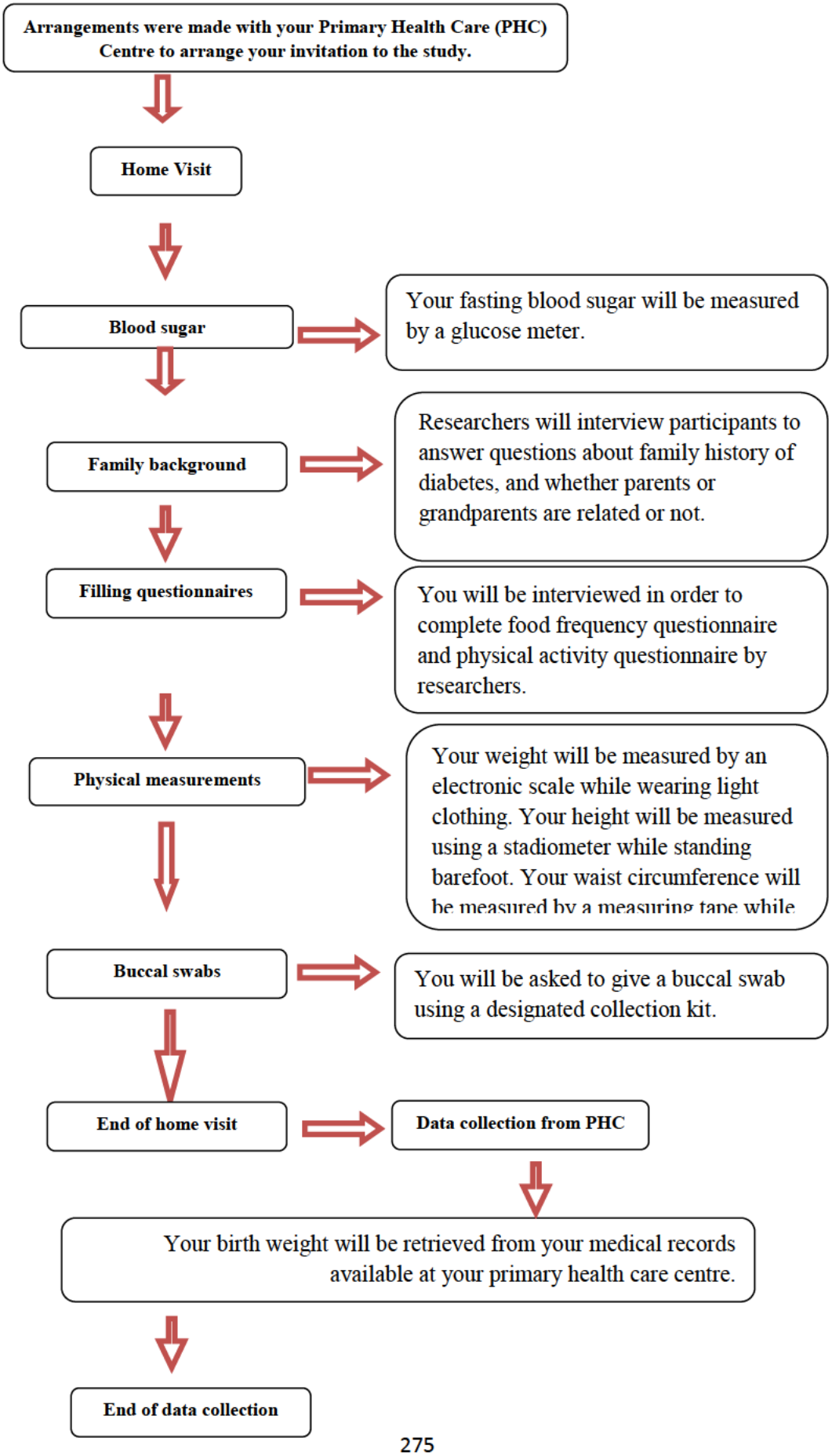
What do I have to do?

You will be kindly requested not to eat for 8 hours prior to the visit to allow for measurement of fasting blood sugar. This measurement will be done using a glucose meter during the home visit. The procedure involves pricking a finger to obtain a blood drop. The drop will be applied to a glucose meter to provide a reading of the glucose level.

The interviewer will ask questions about your diet and physical exercise. Otherwise, there will be no other restriction on lifestyle or general activities.

You will be asked to give a buccal swab using a designated collection kit. Swabs will be taken one hour after eating or drinking and mouth rinsed prior to taking sample. You will be given a swab to insert into your mouth and rub firmly against the inside of your cheek for 1 minute. Afterwards, the investigator will place the swab in a designated tube with preservative capsules .At the end of data collection, these kits will be sent to a specialist company for DNA extraction and genetic analysis.

Figure 1:Research Process and Guidelines:



What are the possible disadvantages and risks of taking part?

You might experience discomfort due to 8 hours fasting needed to measure blood sugar. A slight pain might be felt while finger pricking needed to obtain blood drop. The process of taking glucose reading will be proceeded by investigators wearing new gloves when taking the glucose reading of each participant, applying alcohol wipes at the area of finger pricking, using a new lancet for each participant.

Certain healthy participants might exhibit an above normal level of blood sugar. However, this measurement of blood glucose is not confirmatory and it does not mean that you have diabetes. If a healthy participant appears to have an above normal level of blood sugar, they will be advised to seek medical advice from their family physician. However, having an elevated level of blood sugar does not interfere with your participation in this study.

It might be inconvenient for some participants to give buccal swabs. Giving your DNA samples for analysis will not impose any risk of identifying other medical or genetic conditions as we are looking only at genes which may be involved in diabetes risk. Collected swabs will be used to perform a DNA extraction and genetic analysis for one time only. This study will not preserve DNA samples once the analysis has been performed.

What are the possible benefits of taking part?

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will be very helpful in designing intervention programs for the future prevention of diabetes.

What if something goes wrong?

If you have any query/complaint you can contact me without hesitation at my given contact number. However, if you feel that I could not handle your query/complaint appropriately then you can contact my supervisor at *(m.d.teare@sheffield.ac.uk*).

Will my taking part in this project be kept confidential?

All the information that you will provide during the session will be kept strictly confidential under my custody. Data will be analysed anonymously and the individual name/identity of any participant will not be disclosed in any paper/report or publication.Buccal swabs will not be

labelled by any identifiable data and researches will only use anonymised codes to link records between Buccal swabs and other collected data.

What will happen to the results of the research project?

The results of the study will be submitted to The University of Sheffield in January 2014. Findings of the study will be sent to reputable scientific journals for publication. If you are interested in knowing the findings of this study, you can contact the principal investigator.

Who is organising and funding the research?

This study is partially sponsored by Saudi Cultural Bureau in London, UK and the remaining financial requirements are met by the principal investigator .

Who has ethically reviewed the project?

Ethics Review Committee of School of Health and Related Research (ScHARR) at University of Sheffield, UK and Directory of Health in Jazan region, Saudi Arabia managed the ethics review process.

Finally …

Participants will be given a copy of the information sheet and a signed consent form to keep.

If you need further information you can contact at given below number. I will be very thankful for your cooperation.

Best Wishes

Dr.Ibrahim Gosadi

Principal Investigator, postgraduate research student

School of Health and Related Research (ScHARR)

University of Sheffield, Sheffield, UK

Contact No. +966 562447123

Email: igosadi@ksu.edu.sa

Appendix 5: Participants consent form

Participants Consent Form

Title of Research Project: Investigating the Potential Effect of Consanguinity on Type 2 Diabetes Susceptibility in a Saudi Population

Name of Researcher: Dr. Ibrahim Gosadi

Participant Identification Number for this project:

initial box

1. I confirm that I have read and understand the information sheet dated *26/04/2012* explaining the above research project and I have had the opportunity to ask questions about the project. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without there being any negative consequences. In addition, should I not wish to answer any particular question or questions, I am free to decline. ☐
   Contact number of lead investigator : 0562447123

3. I understand that my responses will be kept strictly confidential. I give permission for members of the research team to have access to my anonymised responses. I understand that my name will not be linked with the research materials, and I will not be identified or identifiable in the report or reports that result from the research. ☐

4.I agree for the data collected from me to be used in the above research project. ☐

5. I understand that sections of my medical records may be looked at by the principle investigator. I give permission for the Principal investigator to access my records. ☐

6. I agree to my DNA samples used for the above research project. ☐

7. I agree to take part in the above research project. ☐

_____ _____

_____

Name of Participant Date Signature

(*or legal representative*)

_____ _____

_____

Name of person taking consent Date Signature

(*if different from lead researcher*)

To be signed and dated in presence of the participant

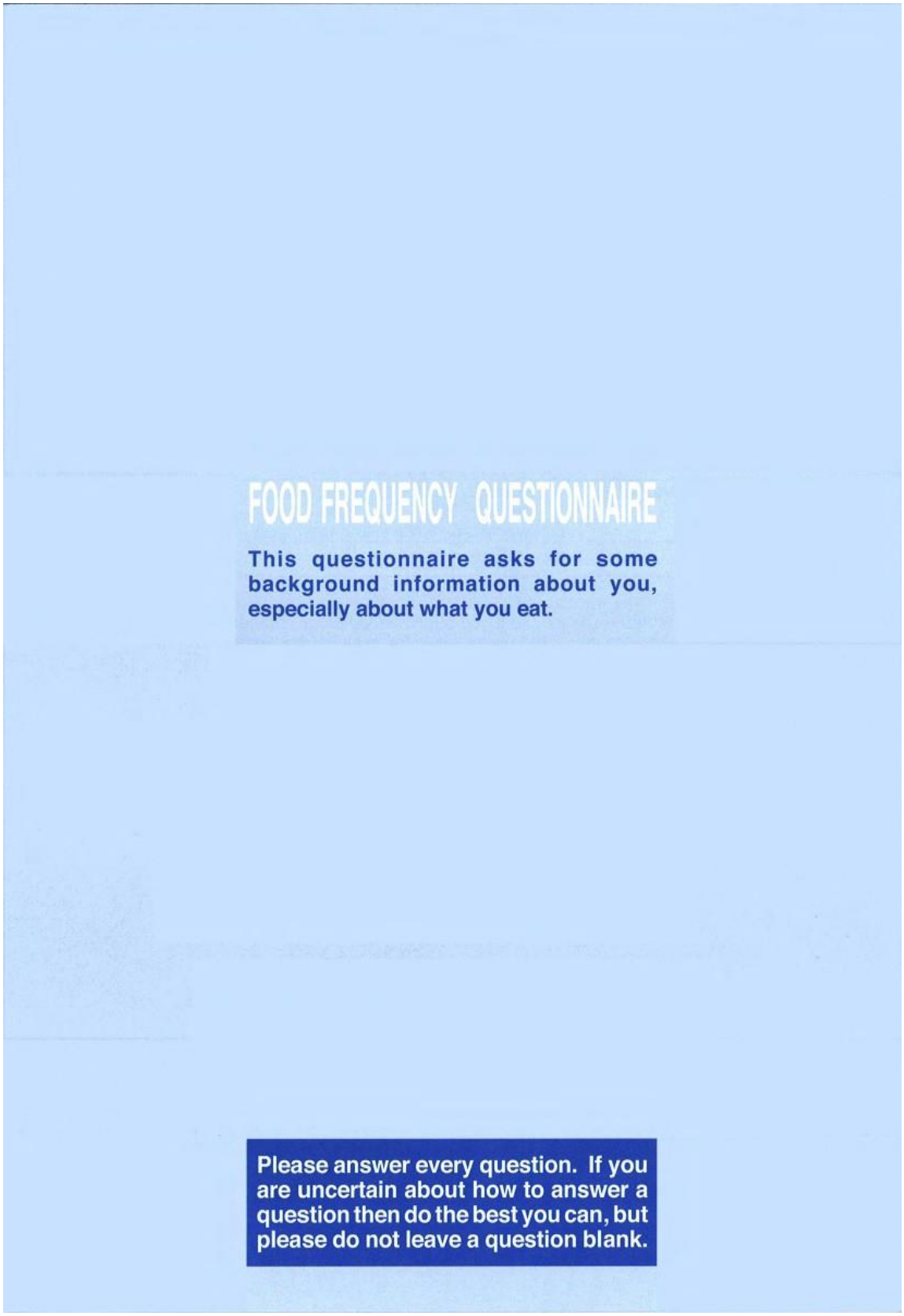
_____ _____

_____

Lead Researcher Date Signature

To be signed and dated in presence of the participant

Appendix 6 : EPIC- Food Frequency Questionnaire



FOOD FREQUENCY QUESTIONNAIRE

This questionnaire asks for some background information about you, especially about what you eat.

Please answer every question. If you are uncertain about how to answer a question then do the best you can, but please do not leave a question blank.

1. **YOUR DIET LAST YEAR**

For each food there is an amount shown, either a "medium serving" or a common household unit such as a slice or teaspoon. Please put a tick (✓) in the box to indicate how often, **on average**, you have eaten the specified amount of each food **during the past year**.

EXAMPLES:

For white bread the amount is one slice, so if you ate 4 or 5 slices a day, you should put a tick in the column headed "4-5 per day".

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
BREAD AND SAVOURY BISCUITS (one slice or biscuit)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
White bread and rolls								✓	

For chips, the amount is a "medium serving", so if you had a helping of chips twice a week you should put a tick in the column headed "2-4 per week".

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
POTATOES, RICE AND PASTA (medium serving)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Chips				✓					

For very seasonal fruits such as strawberries and raspberries you should estimate your average use when the fruits are in season, so if you ate strawberries or raspberries about once a week when they were in season you should put a tick in the column headed "once a week"

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
FRUIT (1 fruit or medium serving)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Strawberries, raspberries, kiwi fruit			✓						

Please estimate your average food use as best you can, and please answer every question – do not leave ANY lines blank. PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
MEAT AND FISH (medium serving)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Beef: roast, steak, mince, stew or casserole									
Beefburgers									
Pork: roast, chops, stew or slices									
Lamb: roast, chops or stew									
Chicken or other poultry eg. turkey									
Bacon									
Ham									
Corned beef, Spam, luncheon meats									
Sausages									
Savoury pies, eg. meat pie, pork pie, pasties, steak & kidney pie, sausage rolls									
Liver, liver paté, liver sausage									
Fried fish in batter, as in fish and chips									
Fish fingers, fish cakes									
Other white fish, fresh or frozen, eg. cod, haddock, plaice, sole, halibut									
Oily fish, fresh or canned, eg. mackerel, kippers, tuna, salmon, sardines, herring									
Shellfish, eg. crab, prawns, mussels									
Fish roe, taramasalata									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
BREAD AND SAVOURY BISCUITS (one slice or biscuit)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
White bread and rolls									
Brown bread and rolls									
Wholemeal bread and rolls									
Cream crackers, cheese biscuits									
Crispbread, eg. Ryvita									
CEREALS (one bowl)									
Porridge, Readybrek									
Breakfast cereal such as cornflakes, muesli etc.									
POTATOES, RICE AND PASTA (medium serving)									
Boiled, mashed, instant or jacket potatoes									
Chips									
Roast potatoes									
Potato salad									
White rice									
Brown rice									
White or green pasta, eg. spaghetti, macaroni, noodles									
Wholemeal pasta									
Lasagne, moussaka									
Pizza									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
DAIRY PRODUCTS AND FATS	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Single or sour cream (tablespoon)									
Double or clotted cream (tablespoon)									
Low fat yogurt, fromage frais (125g carton)									
Full fat or Greek yogurt (125g carton)									
Dairy desserts (125g carton)									
Cheese, eg. Cheddar, Brie, Edam (medium serving)									
Cottage cheese, low fat soft cheese (medium serving)									
Eggs as boiled, fried, scrambled, etc. (one)									
Quiche (medium serving)									
Low calorie, low fat salad cream(tablespoon)									
Salad cream, mayonnaise (tablespoon)									
French dressing (tablespoon)									
Other salad dressing (tablespoon)									
The following on bread or vegetables									
Butter (teaspoon)									
Block or hard margarine, eg. Stork, Krona (teaspoon)									
Polyunsaturated margarine, eg. Flora, sunflower, soya spreads (teaspoon)									
Soft margarines,including olive oil based and dairy spreads, eg. Blue Band, Olivio/ Bertolli, Clover (teaspoon)									
Low fat spreads (less than 60% fat), eg. Outline, Gold (teaspoon)									
Very low fat spread (less than 30% fat) (teaspoon)									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
SWEETS AND SNACKS (medium serving)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Sweet biscuits, chocolate , eg. digestive (one)									
Sweet biscuits, plain, eg. Nice, ginger (one)									
Cakes eg. fruit, sponge, home baked									
Cakes eg. fruit, sponge, ready made									
Buns, pastries eg. scones, flapjacks, home baked									
Buns, pastries eg. croissants, doughnuts, ready made									
Fruit pies, tarts, crumbles, home baked									
Fruit pies, tarts, crumbles, ready made									
Sponge puddings, home baked									
Sponge puddings, ready made									
Milk puddings, eg. rice, custard, trifle									
Ice cream, choc ices									
Chocolates, single or squares									
Chocolate snack bars eg. Mars, Crunchie									
Sweets, toffees, mints									
Sugar added to tea, coffee, cereal (teaspoon)									
Crisps or other packet snacks, eg. Wotsits									
Peanuts or other nuts									
SOUPS, SAUCES, AND SPREADS									
Vegetable soups (bowl)									
Meat soups (bowl)									
Sauces, eg. white sauce, cheese sauce, gravy (tablespoon)									
Tomato ketchup (tablespoon)									
Pickles, chutney (tablespoon)									
Marmite, Bovril (teaspoon)									
Jam, marmalade, honey (teaspoon)									
Peanut butter (teaspoon)									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
DRINKS	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Tea (cup)									
Coffee, instant or ground (cup)									
Coffee, decaffeinated (cup)									
Coffee whitener, eg. Coffee-mate (teaspoon)									
Cocoa, hot chocolate (cup)									
Horlicks, Ovaltine (cup)									
Wine (glass)									
Beer, lager or cider (half pint)									
Port, sherry, vermouth, liqueurs (glass)									
Spirits, eg. gin, brandy, whisky, vodka (single)									
Low calorie or diet fizzy soft drinks (glass)									
Fizzy soft drinks, eg. Coca cola, lemonade (glass)									
Pure fruit juice (100%) eg. orange, apple juice (glass)									
Fruit squash or cordial (glass)									
FRUIT **For seasonal fruits marked *, please estimate your average use when the fruit is in season**									
Apples (1 fruit)									
Pears (1 fruit)									
Oranges, satsumas, mandarins (1 fruit)									
Grapefruit (half)									
Bananas (1 fruit)									
Grapes (medium serving)									
Melon (1 slice)									
* Peaches, plums, apricots (1 fruit)									
* Strawberries, raspberries, kiwi fruit (medium serving)									
Tinned fruit (medium serving)									
Dried fruit, eg. raisins, prunes (medium serving)									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
VEGETABLES Fresh, frozen or tinned (medium serving)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Carrots									
Spinach									
Broccoli, spring greens, kale									
Brussels sprouts									
Cabbage									
Peas									
Green beans, broad beans, runner beans									
Marrow, courgettes									
Cauliflower									
Parsnips, turnips, swedes									
Leeks									
Onions									
Garlic									
Mushrooms									
Sweet peppers									
Beansprouts									
Green salad, lettuce, cucumber, celery									
Watercress									
Tomatoes									
Sweetcorn									
Beetroot									
Coleslaw									
Avocado									
Baked beans									
Dried lentils, beans, peas									
Tofu , soya meat, TVP, Vegeburger									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

YOUR DIET LAST YEAR, continued

2. Are there any **OTHER** foods which you ate more than once a week? Yes ☐ No ☐
If **YES,** please list below

Food	Usual serving size	Number of times eaten each week

3. What type of milk did you most often use?
Select one only
Full cream/whole ☐ Semi-skimmed ☐
Skimmed ☐ Channel Islands, gold ☐
Dried milk ☐ Soya ☐
Other, specify ☐ None ☐

4. How much milk did you drink each day, including milk with tea, coffee, cereals etc?
None ☐ Three quarters of a pint ☐
Quarter of a pint ☐ One pint ☐
Half a pint ☐ More than one pint ☐

5. Did you usually eat breakfast cereal (excluding porridge and Ready Brek mentioned earlier)?
Yes ☐ No ☐

If **YES,** which brand and type of breakfast cereal, including muesli, did you usually eat?
List the one or two types most often used

Brand e.g. Kellogg's	Type e.g. cornflakes

6. What kind of fat did you most often use for frying, roasting, grilling etc?
Select one only
Butter ☐ Solid vegetable fat ☐
Lard/dripping ☐ Margarine ☐
Vegetable oil ☐ None ☐
If you used vegetable oil, please give type eg. corn, sunflower ☐

7. What kind of fat did you most often use for baking cakes etc?
Select one only
Butter ☐ Solid vegetable fat ☐
Lard/dripping ☐ Margarine ☐
Vegetable oil ☐ None ☐
If you used margarine, please give name or type eg. Flora, Stork ☐

8. How often did you eat food that was fried at home?

Daily ☐ 1-3 times a week ☐ 4-6 times a week ☐
Less than once a week ☐ Never ☐

9. How often did you eat fried food away from home?

Daily ☐ 1-3 times a week ☐ 4-6 times a week ☐
Less than once a week ☐ Never ☐

10. What did you do with the visible fat on your meat?

Ate most of the fat ☐ Ate as little as possible ☐
Ate some of the fat ☐ Did not eat meat ☐

11. How often did you eat grilled or roast meat? ☐☐ times a week

12. How well cooked did you usually have grilled or roast meat?

Well done /dark brown ☐ Lightly cooked/rare ☐
Medium ☐ Did not eat meat ☐

13. How often did you add salt to food while cooking?

Always ☐ Rarely ☐
Usually ☐ Never ☐
Sometimes ☐

14. How often did you add salt to any food at the table?

Always ☐ Rarely ☐
Usually ☐ Never ☐
Sometimes ☐

15. Did you regularly use a salt substitute (eg LoSalt)? Yes ☐ No ☐

If **YES**, which brand? _____

16. During the course of last year, on average, how many times a week did you eat the following foods?

Food type	**Times/week**	**Portion size**
Vegetables (not including potatoes)	☐☐	medium serving
Salads	☐☐	medium serving
Fruit and fruit products (not including fruit juice)	☐☐	medium serving or 1 fruit
Fish and fish products	☐☐	medium serving
Meat, meat products and meat dishes (including bacon, ham and chicken)	☐☐	medium serving

17. Have you taken any vitamins, minerals, fish oils, fibre or other food supplements during the past year?

- ☐ Yes
- ☐ No
- ☐ Sometimes
- ☐ Don't know

If **YES or SOMETIMES,** please complete the table below.

If you have taken more than 8 types of supplement please put the most frequently consumed brands first.

Example: If you take one tablet of vitamin C two times a day, please write '2' in the amount-column and tick (✓) the 'once a day' box. Most supplements mention a strength value (in our example 500mg), please write this information in the table.

Supplements				Average frequency for the past year Tick (✓) **ONE** box per line to show how often on average you took the amount consumed as mentioned in 'amount' column.					
Brand	Name	Strength (strength of the supplement for each tablet or capsule)	Amount (number of tablets, capsules or teaspoons taken in one day)	Never or less than once a month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day
Boots	High strength vitamin C	500mg	2 tablets	EXAMPLE					✓

Thank you for your help

Appendix 7:EPIC Physical Activity Questionnaire

THE QUESTIONNAIRE IS DIVIDED INTO 3 SECTIONS

- **Section A** asks about your physical activity patterns in and around the house.
- **Section B** is about travel to work and your activity at work. It may be skipped by people who have not worked at any stage during the last 12 months.
- **Section C** asks about recreations that you may have engaged in during the last 12 months.

What is your date of birth? day / month / year

What is today's date? day / month / year

Your sex *(Please tick (✓) appropriate box)*? Male ☐ Female ☐

Section A HOME ACTIVITIES

GETTING UP AND GOING TO BED
*Please put a time in **each** box*

	Average over the past year	
	At what time do you normally get up?	At what time do you normally go to bed?
On a weekday		
On a weekend day		

GETTING ABOUT — Apart from going to work

Which form of transport do you use **most often** apart from your journey to and from work?

*Please tick (✓) one box **ONLY** per line*

Distance of journeys	Usual mode of transport			
	Car	Walk	Public transport	Cycle
less than one mile				
1–5 mile(s)				
More than 5 miles				

Page 3

TV OR VIDEO VIEWING

*Please put a tick (✓) on **every** line*

Hours of TV or Video watched per day	Average over the last 12 months					
	None	less than 1 hour a day	1 to 2 hours a day	2 to 3 hours a day	3 to 4 hours a day	More than 4 hours a day
On a weekday before 6 pm						
On a weekday after 6 pm						
On a weekend day before 6 pm						
On a weekend day after 6 pm						

STAIR CLIMBING AT HOME

*Please put a tick (✓) on **every** line*

Number of times you climbed up a flight of stairs (approx 10 steps) each day at home	Average over the last 12 months					
	None	1 to 5 times a day	6 to 10 times a day	11 to 15 times a day	16 to 20 times a day	More than 20 times a day
On a weekday						
On a weekend day						

ACTIVITIES IN AND AROUND THE HOME

*Please put a tick (✓) on **every** line*

Approximate number of hours each week	Average over the last 12 months						
	None	Less than 1 hour a week	1 to 3 hours a week	3 to 6 hours a week	6 to 10 hours a week	10 to 15 hours a week	More than 15 hours a week
Preparing food, cooking and washing up							
Shopping for food and groceries							
Shopping and browsing in shops for other items (e.g. clothes,toys)							
Cleaning the house							
Doing the laundry and ironing							
Caring for pre-school children or babies at home (not as paid employment)							
Caring for handicapped, elderly or disabled people at home (not as paid employment)							

Section B ACTIVITY AT WORK

Please answer this section **only** if you have been in paid employment at any time during the last 12 months or you have done regular, organised voluntary work.

If not please go to page 9

TYPES OF WORK DURING THE LAST TWELVE MONTHS

- We would like to know what full or part-time jobs you have done in the last 12 months.
- You may have held a single job or have held two jobs at once.
- If you have changed jobs with the same employer, you should enter it as a change of job **only** if it entailed a substantial change in physical effort.

EXAMPLE

Someone who worked full-time for 6 months, then retired, rested for 3 months and then started a voluntary job for 6 hours a week, would complete the questions as follows.

	Job 1	Job 2
Name of occupation	nurse	shop work
How many hours ~~per week~~ did you usually work?	38	6
For how many ~~months~~ in the last 12 ~~months~~ did you do this work?	6	3

EXAMPLE

ACTIVITY LEVELS AT YOUR WORK

Now we would like you to take the total number of hours you worked per week in each job and divide them up according to your activity level.

Please complete EACH line

	Job 1			Job 2		
	No	Yes	Hours per week	No	Yes	Hours per week
Sitting — light work e.g. desk work, or driving a car or truck		✓	6	✓		
Sitting — moderate work e.g. working heavy levers or riding a mower or forklift truck	✓				✓	2
Standing — light work e.g. lab technician work or working at a shop counter		✓	30		✓	4
Standing — light/moderate work e.g. light welding or stocking shelves		✓	2	✓		

EXAMPLE

The number of hours in each activity should add up to the number of hours that you worked in each job e.g. 6+30+2=38 (nurse)

What jobs have you held in the last 12 months, and how many months in the year did you do them?

Please complete EACH line

	Job 1	Job 2
Name of occupation		
How many hours **per week** did you usually work?		
For how many months in the last 12 months did you do this work?		

ACTIVITY LEVELS AT YOUR WORK

Now we would like you to take the total number of hours you worked per week in each job and divide them up according to your activity level.

Please complete EACH line

	Job 1			Job 2		
	No	Yes	Hours per week	No	Yes	Hours per week
Sitting — light work e.g. desk work, or driving a car or truck						
Sitting — moderate work e.g. working heavy levers or riding a mower or forklift truck						
Standing — light work e.g. lab technician work or working at a shop counter						
Standing — light/moderate work e.g. light welding or stocking shelves						
Standing — moderate work e.g. fast rate assembly line work or lifting up to 50 lbs every 5 minutes for a few seconds at a time						
Standing — moderate/heavy work e.g. masonry/painting or lifting more than 50 lbs every 5 minutes for a few seconds at a time						
Walking at work — carrying nothing heavier than a briefcase e.g. moving about a shop						
Walking — carrying something heavy						
Moving, pushing heavy objects objects weighing over 75lbs						

STAIR OR STEP CLIMBING AT WORK

Please put a tick (✓) on EACH line where appropriate

Number of times you climbed up a flight of stairs (10 steps) at work	AVERAGE OVER THE LAST 12 MONTHS					
	None	1 to 5 times a day	6 to 10 times a day	11 to 15 times a day	16 to 20 times a day	More than 20 times a day
Job 1						
Job 2						

Please put a tick (✓) on EACH line where appropriate

Number of times you climbed up a ladder at work	AVERAGE OVER THE LAST 12 MONTHS					
	None	1 to 5 times a day	6 to 10 times a day	11 to 15 times a day	16 to 20 times a day	More than 20 times a day
Job 1						
Job 2						

KNEELING AND SQUATTING AT WORK IN JOB 1

In an average working day in Job 1 did you

kneel for more than one hour in total? No ☐ Yes ☐ Don't know ☐

squat for more than one hour in total? No ☐ Yes ☐ Don't know ☐

get up from kneeling or squatting more than 30 times? No ☐ Yes ☐ Don't know ☐

KNEELING AND SQUATTING AT WORK IN JOB 2

In an average working day in Job 2 did you

kneel for more than one hour in total? No ☐ Yes ☐ Don't know ☐

squat for more than one hour in total? No ☐ Yes ☐ Don't know ☐

get up from kneeling or squatting more than 30 times? No ☐ Yes ☐ Don't know ☐

TRAVEL TO AND FROM WORK

JOB 1

Please complete EVERY line

Roughly how many miles was it from home to Job 1?	
How many times a week did you travel from home to Job 1?	

Please tick (✓) one box ONLY per line

How did you normally travel to Job 1?	Always	Usually	Occasionally	Never or rarely
By car				
By works or public transport				
By bicycle				
Walking				

JOB 2 *(if appropriate)*

Please complete EVERY line

Roughly how many miles was it from home to Job 2?	
How many times a week did you travel from home to Job 2?	

Please tick (✓) one box ONLY per line

How did you normally travel to Job 2?	Always	Usually	Occasionally	Never or rarely
By car				
By works or public transport				
By bicycle				
Walking				

Section C RECREATION

The following questions ask about how you spent your leisure time.

Please indicate how often you did each activity on average over the last 12 months.

For activities that are seasonal, e.g. cricket or mowing the lawn, please put the average frequency during the season when you did the activity.

Please indicate the average length of time that you spent doing the activity on each occasion.

EXAMPLE

If you had mowed the lawn every fortnight in the grass cutting season and took 1 hour and 10 minutes on each occasion.

If you went walking for pleasure for 40 minutes once a week.

You would complete the table below as follows:

Please give an answer for the AVERAGE TIME you spent on each activity and the NUMBER OF TIMES you did that activity in the past year.

	Number of times you did the activity in the last 12 months								Average time per episode	
	None	Less than once a month	Once a month	2 to 3 times a month	Once a week	2 to 3 times a week	4 to 5 times a week	Every day	Hours	Mins
Mowing the lawn				✓					1	10
Walking for pleasure					✓					40

EXAMPLE

Now please complete the table on pages 10 and 11

Please give an answer for the NUMBER OF TIMES you did the following activities in the last 12 months and the AVERAGE TIME you spent on each activity.

Please complete EACH line

	Number of times you did the activity in the last 12 months								Average time per episode	
	None	Less than once a month	Once a month	2 to 3 times a month	Once a week	2 to 3 times a week	4 to 5 times a week	6 times a week or more	Hours	Mins
Swimming — competitive										
Swimming — leisurely										
Backpacking or mountain climbing										
Walking for pleasure — you should not include walking as a means of transportation as this was included in Sections A & B										
Racing or rough terrain cycling										
Cycling for pleasure — you should not include cycling as a means of transportation										
Mowing the lawn — during the grass cutting season										
Watering the lawn or garden in the summer										
Digging, shovelling or chopping wood										
Weeding or pruning										
DIY e.g. carpentry, home or car maintenance										
High impact aerobics or step aerobics										
Other types of aerobics										
Exercises with weights										
Conditioning exercises e.g. using an exercise bike or rowing machine										

Please continue on the next page

Please complete EACH line

	Number of times you did the activity in the last 12 months								Average time per episode	
	None	Less than once a month	Once a month	2 to 3 times a month	Once a week	2 to 3 times a week	4 to 5 times a week	6 times a week or more	Hours	Mins
Floor exercises e.g. stretching, bending, keep fit or yoga										
Dancing e.g. ballroom or disco										
Competitive running										
Jogging										
Bowling — indoor, lawn or 10 pin										
Tennis or badminton										
Squash										
Table tennis										
Golf										
Football, rugby or hockey (during the season)										
Cricket (during the season)										
Rowing										
Netball, volleyball or basketball										
Fishing										
Horse-riding										
Snooker, billiards or darts										
Musical instrument playing or singing										
Ice-skating										
Sailing, wind-surfing or boating										
Martial arts, boxing or wrestling										

You have finished the questionnaire — *Thank you*

Appendix 8: Changes made for EPIC- food frequency questionnaire

Location in the EPIC questionnaire	Food Item(s)	Action Taken	Reason of action
Page 3: Meat and Fish.	Pork, Bacon, Ham, Corned Beef, Spam, Pork Pie, Steak and Kidney Pie, Fish Cakes, Fish Roe and Taramasalta.	Deletion	Not available in Saudi Arabia
Page 4: Bread and Savoury Biscuits	White bread and rolls	Example Addition: Tamees	Popular food in Saudi Arabia. Not Available in the UK.
Page 4: Bread and Savoury Biscuits	Marsa	Item Addition	Popular in Saudi Arabia. White or brown bread is cut into small pieces and mixed with banana, honey, and butter.
Page 4: Bread and Savoury Biscuits	Maasoub	Item Addition	Popular in Saudi Arabia. Brown bread is cut into small pieces and mixed with banana, honey, butter, cream, dates and sugar.
Page 4: Bread and Savoury Biscuits	Asseda	Item Addition	Popular food in Saudi Arabia. Made from white wheat, butter, and sugar. Vegetable oil and honey are added upon serving.
Page 4: Bread and Savoury Biscuits	Haisya	Item Addition	Popular food in Saudi Arabia. Made from home made whole wheat. Meat and gravy are added upon convenience.
Page 4: Bread and Savoury Biscuits	Crispbread	Example Replacement: Ryvita with bread rusk (white, brown or whole wheat)	Popular food in Saudi Arabia.
Page 4: Cereals	Porridge, Readybrek	Deletion	Not available in Saudi Arabia
Page 5: Dairy Products and Fats	Single or sour cream	Replacement with Cream	Sour Creams are not available in Saudi Arabia
Page 5: Dairy Products and Fats	Dairy desserts	Deletion	Not Available in Saudi Arabia
Page 5: Dairy	Cottage cheese	Replacement with	Salted cheese is common in Saudi

Products and Fats		Salted cheese	Arabia unlike the cottage cheese
Page 5: Dairy Products and Fats	Butter and margarine spread on bread and vegetables	Grouped into two items, margarine or butter on bread.	Butter and margarine are commonly used while cooking. Uncommon as a spread on bread
Page 6:Sweets and Snacks	Sponge puddings	Deletion	Not available in Saudi Arabia
Page 6:Sweets and Snacks	Baqlawa, Kenafa, Basboosa, Lugmat Alghadi, and Mushabak	Items Addition	Popular Arabic sweets in Saudi Arabia. These different items share similar ingredients including wheat, butter, cream, and sugar.
Page 6: Soups, Sauces, and Spreads	Sauces, eg. White sauce, cheese sauce, and gravy. Marmite, Bovril	Deletion	Not available in Saudi Arabia.
Page 7:Drinks	Arabic Coffee	Item Addition	Popular in Saudi Arabia.
Page 7:Drinks	Wine, Beer, Liqueurs, And Sprites	Deletion	Not available in Saudi Arabia.
Page 7:Fruits	Guava, Dates, Hard Melon, Fig, and Pomegranate.	Items Addition	Popular fruits in Saudi Arabia.
Page 8:Vegetables	Brussels sprout, Leeks, Beansprouts, Watercress, Beetroot, Tofu, Soya met and TVP	Deletion	Not available in Saudi Arabia.
Page 8:Vegetables	Green onion, and Baqel.	Items Addition	Popular vegetables in Saudi Arabia usually eaten uncooked.
Page 9: Milk	Milk types: Channel Islands: Gold, and Soya	Deletion	Not available in Saudi Arabia

Appendix 9: Changes made for EPIC- physical activity questionnaire

Location in the EPIC questionnaire	Activity	Action Taken	Reason of action
Page 3: Getting about- Apart from going to work.	Using public transport	Deletion	Public Transport is not available in Saudi Arabia
Page 3: Getting about- Apart from going to work.	Using Cycle	Deletion	Cycling is not a common activity among adults in Saudi Arabia
Page 8: Travel to and from work	Using public transport	Deletion	Public Transport is not available in Saudi Arabia
Page 8: Travel to and from work	Using Cycle	Deletion	Cycling is not a common activity among adults in Saudi Arabia
Page 10:Recreation activities	Mowing the lawn, Cycling for Pleasure, Dancing, Squash, Golf, Cricket, Rowing, Musical instrument playing or singing, Martial arts, Boxing, and Wrestling	Deletion	These activities are never or rarely performed in Saudi Arabia.
Page 10:Recreation activities	Sailing, Wind Surfing, Boating	Replacement with fishing	Fishing is a common recreational activity in Saudi Arabia which mostly mandates boating

Appendix 10: Nutritional Database

Source	Portion size in grams	Food Item	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
			0	0.07	0.14	0.43	0.79	1	2.5	4.5	6
Ingredients:200 g of stewed lamb, and 200 g of long grain rice cooked, 40 g of oil	400g	Lamb Kabsa	0	77.7	155.4	477.3	876.9	1110	2775	4995	6660
Ingredients:200 gm of chicken stew and 200 g of rice, and cookig oil	400g	Chicken Kabsa	0	60.2	120.4	369.8	679.4	860	2150	3870	5160
Chicken, broiler, back, meat and skin, batter dipped, frie, CNF	200g	Fried Chicken	0	46.34	92.68	284.66	522.98	662	1655	2979	3972
Chicken breast tenders, cooked in conventional oven, CNF	200g	Grilled Chicken, Mandi or Hanith	0	41.02	82.04	251.98	462.94	586	1465	2637	3516
wait of serving has been adjusted to resemble chicken, Lamb, New Zealand, loin, lean and fat, 6mm (1/4") trim, broiled / CNF	200g	Grilled Lamb, Mandi, Hanith	0	44.24	88.48	271.76	499.28	632	1580	2844	3792
Lamb, liver, pan-fried, USDA	90g	Lamb Liver	0	12.46	24.92	76.54	140.62	178	445	801	1068
Chicken, stewing, meat and skin, stewed, CNF	200g	Chicken Stew	0	39.9	79.8	245.1	450.3	570	1425	2565	3420
http://recipes.sparkpeople.com/recipe-calories.asp?recipe=367697	200g	Lamb Stew or Borma	0	56.84	113.68	349.16	641.48	812	2030	3654	4872

http://recipes.sparkpeople.com/recipe-calories.asp?recipe=198278 items included, lamb, anion , oil, and herbs	100g	Mogalgal, Lamb	0	13.93	27.86	85.57	157.21	199	497.5	895.5	1194
200 gm of lamb stew and 100g of brown bread	300g	Hysia	0	50.19	100.38	308.31	566.43	717	1792.5	3226.5	4302
http://recipes.sparkpeople.com/recipe-calories.asp?recipe=189697	200g	Fried Fish	0	21.7	43.4	133.3	244.9	310	775	1395	1860
raw sea bass, 30 kcals from oil, anion and herbs, tomatos http://recipes.sparkpeople.com/recipe-calories.asp?recipe=1339055	200g	Fish Stew	0	21.91	43.82	134.59	247.27	313	782.5	1408.5	1878
Sea bass, grilled, mixed herbs , amount doubled, CNF	200g	Grilled Fish	0	17.36	34.72	106.64	195.92	248	620	1116	1488
Tuna, white, canned with oil, drained, unsalted, CNF	75g	Tuna	0	9.8	19.6	60.2	110.6	140	350	630	840
Shrimp, mixed species, boiled or steamed , CNF	135g	Grilled Shrimp	0	9.31	18.62	57.19	105.07	133	332.5	598.5	798
USDA http://ndb.nal.usda.gov/ndb/foods/show/7629	135g	Fried Shrimp	0	27.79	55.58	170.71	313.63	397	992.5	1786.5	2382
Nutrient data for 18413, Bread, pita, white, unenriched, USDA	100 g	whitebread	0	19.25	38.5	118.25	217.25	275	687.5	1237.5	1650
Ingredients: wheat and oil calories	100g	Zalba, germish,	0	33.88	67.76	208.12	382.36	484	1210	2178	2904
Ingredients :100g of wheat, 5ml oil, butter,	150g	Marsa	0	81.3	162.	499.6	917.9	116	2905	5229	6972

honey, and two banana				4	68	6	8	2			
100g wheat, 40 oil, butter, two banana, 5 dates, and 100 g cream	150g	Maasoub	0	106.75	213.5	655.75	1204.75	1525	3812.5	6862.5	9150
Ingredients: wheat, butter and honey	150g	Asseda	0	73.64	147.28	452.36	831.08	1052	2630	4734	6312
Danish pastry, cinnamon CNF	100g	Pastries ready made	0	28.21	56.42	173.29	318.37	403	1007.5	1813.5	2418
Ingredients: wheat and spinach	150	Spinach pie		25.41	50.82	156.09	286.77	363	907.5	1633.5	2178
Ingredients: Wheat and cheese	150	Cheese Pie		38.64	77.28	237.36	436.08	552	1380	2484	3312
Ingredients: wheat and labnah	150	Labnah Pie		30.31	60.62	186.19	342.07	433	1082.5	1948.5	2598
Ingredients: cheese, labnah, honey,	150	Ush Albulbul		46.76	93.52	287.24	527.72	668	1670	3006	4008
Ingredients: wheat, labnah, honey	150	Labnah with Honey		35.63	71.26	218.87	402.11	509	1272.5	2290.5	3054
Ingredients: wheat and meat	150	Meat Pie		39.55	79.1	242.95	446.35	565	1412.5	2542.5	3390
Ingredients: banana, egg, wheat, oil for usual cook	175	Banana Pie		40.53	81.06	248.97	457.41	579	1447.5	2605.5	3474
Ingredients: chicken and wheat	150	Chicken Pie		35.56	71.12	218.44	401.32	508	1270	2286	3048

Ingredients: Zaater and wheat	150	Zaatar pie		35	70	215	395	500	1250	2250	3000
Fast foods, pizza, cheese and vegetables, regular crust ,CNF	one slice 139g	Pizza	0	23.24	46.48	142.76	262.28	332	830	1494	1992
Ingredients: tomato, cucumber, and onion , CNF	100	green salad	0	1.89	3.78	11.61	21.33	27	67.5	121.5	162
Ingredients:50 g of yogurt and 50g os cucumber	100g	cucumber with laban	0	2.66	5.32	16.34	30.02	38	95	171	228
Arugula, raw, CNF	100g	Jargeer	0	1.75	3.5	10.75	19.75	25	62.5	112.5	150
Radish, white icicle, raw, CNF	one, 17g	Baqel	0	0.14	0.28	0.86	1.58	2	5	9	12
Onion, spring (green) or scallion (includes tops and bulb), raw ,CNF	100g	Green onion	0	2.24	4.48	13.76	25.28	32	80	144	192
Hummus, commercial ,CNF	100g	Hummus	0	11.62	23.24	71.38	131.14	166	415	747	996
http://recipes.sparkpeople.com/recipe-calories.asp?recipe=72323	100	Fatoush	0	28.77	57.54	176.73	324.69	411	1027	1849.5	2466
http://www.food.com/recipe/dolma-stuffed-grape-leaves-198071	105	stuffed vine leaves	0	15.01	30.03	92.235	169.45	214	536.2	965.25	1287
http://caloriecount.about.com/calories-sabra-babaganoush-i163033	100	Baba Ganoush	0	19.95	39.9	122.55	225.15	285	712.5	1282.5	1710
40 g of bulgar, parsley, etc http://recipes.sparkpeople.com/recipe-calories.asp?recipe=357	100g	taboulah		19.24	38.48	118.20	217.17	274.	687.2	1237.0	1649
Coleslaw (cabbage salad), with dressing, homemade, CNF	100g	Coleslaw	0	5.46	10.92	33.54	61.62	78	195	351	468

Ingredients: tomato, potato, carrot, anion, and oil	100	Vegetarian Stew	0	7.21	14.42	44.29	81.37	103	257.5	463.5	618
http://www.food.com/recipe/fool-madamas-134366	112	Fuol	0	10.71	21.43	65.833	120.94	153.	382.7	688.95	918.6
Lentils, raw anion, oil and tomato ,CNF	100g	lentils stew	0	29.26	58.52	179.74	330.22	418	1045	1881	2508
Spices, fenugreek seed, anion, oil and tomato, CNF	100g	Hulbah	0	22.61	45.22	138.89	255.17	323	807.5	1453.5	1938
Peas, green, raw, anion, oil and tomato, CNF	100g	Peas stew	0	10.01	20.02	61.49	112.97	143	357.5	643.5	858
Beans, snap, green, anion, oil and tomatoraw, USDA	100g	Beans stew	0	6.72	13.44	41.28	75.84	96	240	432	576
Squash, summer, zucchini, includes skin, raw anion, oil and tomato, USDA	small, 118g	Khousa stew	0	5.95	11.9	36.55	67.15	85	212.5	382.5	510
Okra (gumbo), raw , anion, oil and tomato, CNF	100g	Bhamia stew	0	6.72	13.44	41.28	75.84	96	240	432	576
Eggplant (aubergine, brinjal), raw, anion, oil and tomato, CNF	100g	Eggplant stew	0	6.86	13.72	42.14	77.42	98	245	441	588
This Reading was provided by inspecting a product which contains this plant	100g	Mloukiah stew	0	5.6	11.2	34.4	63.2	80	200	360	480
Beans, kidney, all types, raw, anion, oil and tomato, CNF	100g	Djer stew	0	26.74	53.48	164.26	301.78	382	955	1719	2292
http://pankaj3625.hubpages.com/hub/How-To-	100g	Qawar stew	0	7.35	14.7	45.15	82.95	105	262.	472.5	630

Count-Calories-In-Your-Vegetables									5		
Fast foods, side dish, potato, french-fried in vegetable oil, CNF	100g	French Fried	0	22.33	44.66	137.17	252.01	319	797.5	1435.5	1914
CNF	100g	mashed potatoes	0	24.78	49.56	152.22	279.66	354	885	1593	2124
http://low-cholesterol.food.com/recipe/easy-falafel-317819	100g	Flafel	0	7.7	15.4	47.3	86.9	110	275	495	660
Ingredients: bread 100g, Flafel 50g, cheese 20g	170g	Flafel Sandwitch		28.7	57.4	176.3	323.9	410	1025	1845	2460
Ingredients: bread 100g, liver 50g, cheese 20g	170g	Liver Sandwitch		31.78	63.56	195.22	358.66	454	1135	2043	2724
Ingredients: bread, lamb 50g and cheese 20g	170g	Lamb Sandwitch		39.06	78.12	239.94	440.82	558	1395	2511	3348
Ingredients: bread, fried egg 50g and cheese 20g	170g	Fried Egg Sandwitch		31.57	63.14	193.93	356.29	451	1127	2029.5	2706
Ingredients: bread, boiled egg 50g and cheese 20g	170g	Boiled Egg Sandwitch		29.75	59.5	182.75	335.75	425	1062.5	1912.5	2550
Ingredients: bread, chips 50g and cheese 20g	170g	Fries Sandwitch		35.98	71.96	221.02	406.06	514	1285	2313	3084
Ingredients: bread, tuna 50g and cheese 20g	170g	Tuna sandwitch		34.16	68.32	209.84	385.52	488	1220	2196	2928
Ingredients: bread, chicken 50g and cheese 20g	170g	chicken Sandwitch or shaorwma		35.07	70.14	215.43	395.79	501	1252.5	2254.5	3006

Ingredients: bread, jam 50g and cheese 20g	170g	Jam Sandwich		30.59	61.18	187.91	345.23	437	1092.	1966.5	2622
Grains, rice, white, long-grain, regular, cooked, CNF	150g	rice	0	13.65	27.3	83.85	154.05	195	487.5	877.5	1170
Pasta, spaghetti, enriched, cooked ,CNF	100g	pasta, eg. Spaghetti, macaroni, noodles	0	11.06	22.12	67.94	124.82	158	395	711	948
Ingredients:100 g of lamb fried, 100g of pasta, one egg, 100 ml of milk and 100 ml of cheese,	100g	Bashamil,	0	24.29	48.58	149.21	274.13	347	867.5	1561.5	2082
Ingredients: 100g wheat, frying oil, and 50g of cheese	100g	sambosa, cheese	0	37.38	74.76	229.62	421.86	534	1335	2403	3204
Ingredients: aot, tomato paste 40g, onion 40g, oil	100g	Oatmeal soup	0	33.95	67.9	208.55	383.15	485	1212	2182.5	2910
Pickles, cucumber, sour, CNF	65 g	Pickles	0	0.49	0.98	3.01	5.53	7	17.5	31.5	42
Jam type spread, Double Fruit, CNF	15 ml , 19g	Jam, (teaspoon)	0	2.94	5.88	18.06	33.18	42	105	189	252
honey, USDA	21g, 15ml	honey	0	4.48	8.96	27.52	50.56	64	160	288	384
Nuts, mixed nuts, dry roasted with peanuts, CNF	100g	Peanuts or other nuts	0	41.58	83.16	255.42	469.26	594	1485	2673	3564
ketchup, CNF	6 g, packet	Tomato ketchup (teaspoon)	0	0.42	0.84	2.58	4.74	6	15	27	36
Peanut butter, smooth type, fat and sugar added, CNF	15 ml	Peanut butter	0	6.65	13.3	40.85	75.05	95	237.5	427.5	570
Olive, ripe, canned, small to extra large, CNF	10 pieces	Olive		2.8	5.6	17.2	31.6	40	100	180	240

Milk, fluid, whole, producer, 3.7% M.F. CNF	100 ml	full fat milk	0	4.62	9.24	28.38	52.14	66	165	297	396
Milk, fluid, partly skimmed, 1% M.F.CNF	100ml	low fat milk	0	3.01	6.02	18.49	33.97	43	107.5	193.5	258
http://www.livestrong.com/thedailyplate/nutrition-calories/food/generic/fresh-laban-almarai/	100ml	Full fat laban	0	4.2	8.4	25.8	47.4	60	150	270	360
http://tracker.dailyburn.com/nutrition/almarai_almarai_vetal_laban_low_fat_calories	100ml	Low fat Laban	0	3.5	7	21.5	39.5	50	125	225	300
Cream, whipped, cream topping, pressurized, USDA	50g	Cream	0	8.995	17.99	55.255	101.515	128.5	321.25	578.25	771
Yogurt, plain, whole milk, USDA	100g	Full fat yogurt	0	4.27	8.54	26.23	48.19	61	152.5	274.5	366
Yogurt, plain, low fat, 12 grams protein per 8 ounce, USDA	100gm	Low fat yogurt	0	4.41	8.82	27.09	49.77	63	157.5	283.5	378
Cheese spread, cream cheese base, CNF	100 ml	Cheese	0	20.93	41.86	128.57	236.21	299	747.5	1345.5	1794
Cheese, feta, CNF	100ml	Salted cheese	0	11.69	23.38	71.81	131.93	167	417.5	751.5	1002
Egg, chicken, whole, fried, CNF	100g	fried egg	0	13.51	27.02	82.99	152.47	193	482.5	868.5	1158
Egg, chicken, whole, boiled in shell, hard-cooked, CNF	100g	Eggs boiled	0	9.87	19.74	60.63	111.39	141	352.5	634.5	846
Salad dressing, mayonnaise type, commercial,	15ml	mayonnaise	0	4.06	8.12	24.94	45.82	58	145	261	348

regular , CNF											
Biscuit, plain/buttermilk, commercial , CNF	100g	Sweet biscuits plain	0	25.55	51.1	156.95	288.35	365	912.5	1642.5	2190
Cake, white, dry mix, pudding-type, unenriched. USDA	100g	Cakes ready made	0	29.61	59.22	181.89	334.17	423	1057.5	1903.5	2538
Ingredients: 100 wheat, one egg, 35 g butter, frying oil, sugar	100g	Trumba	0	43.82	87.64	269.18	494.54	626	1565	2817	3756
http://www.food.com/recipe/baklava-143885	100g	Baqlawa	0	31.15	62.3	191.35	351.55	445	1112.5	2002.5	2670
http://www.food.com/recipe/konafa-342899	100g	Kunafa	0	31.08	62.16	190.92	350.76	444	1110	1998	2664
http://www.food.com/recipe/basboosa-easy-semolina-cake-254309	100g	Basboosa	0	29.26	58.52	179.74	330.22	418	1045	1881	2508
http://www.food.com/recipe/luqaimaat-arab-donuts-309634	150g	Lugmat alqadi	0	18.48	36.96	113.52	208.56	264	660	1188	1584
Ingredients:from wheat, frying oil, cornstarch, and yogurt	100g	Mushabak	0	32.41	64.82	199.09	365.77	463	1157.5	2083.5	2778
Dessert, pudding, all flavours except chocolate, instant, dry mix, prepared with 2% milk, CNF	100g	Milk puddings eg. Rice, custard, trifle	0	14.49	28.98	89.01	163.53	207	517.5	931.5	1242
Dessert, frozen, ice cream, vanilla, 11% M.F. , CNF	100g	Ice cream	0	14.56	29.12	89.44	164.32	208	520	936	1248
Candies, bars / chips, milk chocolate, plain, CNF	100g	Chocolates, single	0	37.4	74.9	230.0	422.6	535	1337	2407.	3210

		or squares		5		5	5		.5	5	
Candies, bars, Twix, chocolate fudge cookie, CNF	45g	Chocolate snack bars eg. Snickers	0	17.36	34.72	106.64	195.92	248	620	1116	1488
Candies, toffee, homemade, CNF	12g , piece	Sweets, toffees, mints	0	4.69	9.38	28.81	52.93	67	167.5	301.5	402
CNF	5ml spoon	Sugar added to tea, coffee, cereals	0	1.12	2.24	6.88	12.64	16	40	72	96
Snacks, potato chips, plain, salted, USDA	small bag 45g	Crisps or other packet snacks	0	17.01	34.02	104.49	191.97	243	607.5	1093.5	1458
CNF	100g	Tea (cup)	0	0.07	0.14	0.43	0.79	1	2.5	4.5	6
http://www.food.com/recipe/arabic-qahwa-189733	62	Arabic Coffee	0	0.322	0.644	1.978	3.634	4.6	11.5	20.7	27.6
http://www.food.com/recipe/yemeni-chai-chai-adani-199390	100g	Adani Tea	0	2.24	4.48	13.76	25.28	32	80	144	192
Coffee, instant, regular, powder ,CNF	one teaspoon , 5ml, 1g	Coffee, instant or ground (cup)	0	0.14	0.28	0.86	1.58	2	5	9	12
Coffee whitener (non dairy), powdered ,CNF	one packet , 3g	Coffee whitener, eg. Coffee-male (teaspoon)	0	1.12	2.24	6.88	12.64	16	40	72	96
Milk, evaporated, whole, canned, undiluted, 7.8% M.F.CNF	15ml	ADDED MILK		1.47	2.94	9.03	16.59	21	52.5	94.5	126
Hot chocolate, cocoa, homemade, prepared with	100ml	Cocoa, hot	0	6.93	13.8	42.57	78.21	99	247.	445.5	594

whole milk,CNF		chocolate (cup)			6				5		
CNF	355ml	Diet soft drinks (glass)	0	0.07	0.14	0.43	0.79	1	2.5	4.5	6
Carbonated drinks, cola,CNF	355 , can	sweetened soft drinks	0	9.87	19.74	60.63	111.39	141	352.5	634.5	846
Orange juice, raw,CNF	100ml	Pure fruit juice (100%) eg. Organe	0	3.29	6.58	20.21	37.13	47	117.5	211.5	282
CNF	1 fruit	Apples	0	6.65	13.3	40.85	75.05	95	237.5	427.5	570
CNF	1 fruit	Pears	0	7.21	14.42	44.29	81.37	103	257.5	463.5	618
CNF	1 fruit	Oranges	0	6.02	12.04	36.98	67.94	86	215	387	516
CNF	1 fruit	Grapefruit	0	2.87	5.74	17.63	32.39	41	102.5	184.5	246
CNF	1 fruit	Bananas	0	7.35	14.7	45.15	82.95	105	262.5	472.5	630
Grape, red or green (European type, such as Thompson seedless), adherent skin, raw, CNF	(medium serving)	Grapes	0	4.76	9.52	29.24	53.72	68	170	306	408
watermelon, raw, CNF	(water melon or had melon) (1 slice)	Melon	0	2.1	4.2	12.9	23.7	30	75	135	180
apricot,raw, CNF	3 fruits 103 g	apricots	0	3.36	6.72	20.64	37.92	48	120	216	288

peaches, raw, USDA	small fruit, 130 b	Peaches,	0	3.57	7.14	21.93	40.29	51	127.5	229.5	306
plums, USDA	one fruit, 66g	plums, (1 fruit)	0	2.1	4.2	12.9	23.7	30	75	135	180
raspberries ,raw, CNF	10 fruits, 19 g	raspberries(medium serving)	0	0.7	1.4	4.3	7.9	10	25	45	60
strawberries, raw, USDA	five fruits medium , 60g	Strawberries,	0	1.12	2.24	6.88	12.64	16	40	72	96
Kiwi fruit, raw, CNF	1 medium fruit 76 g	Kiwi	0	3.22	6.44	19.78	36.34	46	115	207	276
Guava, common, raw, CNF	one, 90 g	Guava	0	4.27	8.54	26.23	48.19	61	152.5	274.5	366
Date, domestic, natural and dry, CNF	10 fruits	Dates	0	16.38	32.76	100.62	184.86	234	585	1053	1404
pomegranate, raw ,CNF	half a fruit, 77g	Pomegranate	0	4.48	8.96	27.52	50.56	64	160	288	384
fig, raw, CNF	one medium, 50 g	Fig	0	2.59	5.18	15.91	29.23	37	92.5	166.5	222
pineapple, raw, CNF	one slice 90 g	Pineapple	0	3.15	6.3	19.35	35.55	45	112.5	202.5	270
Cherry, sweet, raw, CNF	10 fruits, 82g	Cherry	0	3.64	7.28	22.36	41.08	52	130	234	312
Tangerine (mandarin), raw, CNF	one medium, 84g	Tangerine	0	3.15	6.3	19.35	35.55	45	112.5	202.5	270
Papaya, raw, CNF	half small fruit, 76	Papaya	0	2.065	4.13	12.685	23.305	29.5	73.75	132.75	177
Raisins, seedless, USDA	small box, 43g	Dried fruit, eg.	0	9.03	18.0	55.47	101.9	129	322.	580.5	774

		Raising, prunes (medium serving)			6		1		5		
Cabbage, raw, CNF	100g	Cabbage	0	1.75	3.5	10.75	19.75	25	62.5	112.5	150
Ingredients: bread 100g, cheese 40g	140g	cheese Sandwich		30.45	60.9	187.05	343.65	435	1087.5	1957.5	2610
Ingredients: 135g Shrimp, mixed species, oil, tomato onion	225	stewed Shrimp	0	13.86	27.72	85.14	156.42	198	495	891	1188
Ingredients: wheat, spinach, cheese, meat, egg	200	mixed pie	0	38.78	77.56	238.22	437.66	554	1385	2493	3324
Ingredients: fried egg, tomato, onion	190	shakshookah	0	15.26	30.52	93.74	172.22	218	545	981	1308
Ingredients: Bread 100g, Falafel 50g, cheese 20g , boiled egg 50g, fries 50g	270g	mixed sandwitch		44.73	89.46	274.77	504.81	639	1597.5	2875.5	3834
Ingredients:100 g wheat, 25 mozzarella, 50g boiled egg	175g	egg pastry		35.84	71.68	220.16	404.48	512	1280	2304	3072
Ingredients:100 g wheat, 25 mozzarella, 50g tuna	175g	tuna pastry		37.45	74.9	230.05	422.65	535	1337.5	2407.5	3210
Fast foods, sandwiches and burgers, chicken fillet sandwich with cheese, CNF	228g	chicken burger		43.61	87.22	267.89	492.17	623	1557.5	2803.5	3738
Ingredients: 80 g of tahini seed and 50ml of sugar	100g	taheniah		35.35	70.7	217.15	398.95	505	1262.5	2272.5	3030
Ingredients:100 g bread, 20g cheddar, 15ml mayo	120g	cheese and may sand		28.91	57.82	177.59	326.27	413	1032.5	1858.5	2478

Ingredients: camel meat 250g, rice 150g, onion 40, http://www.myfitnesspal.com/food/calories/generic-camel-meat-26858616	440g	hashi kabsa		34.72	69.44	213.28	391.84	496	1240	2232	2976
Ingredients: Camel meat 250g, potato 50g, http://www.myfitnesspal.com/food/calories/generic-camel-meat-26858616		hashi stew		29.68	59.36	182.32	334.96	424	1060	1908	2544
Mango, raw, CNF	1 fruit/ 207g	mango		9.45	18.9	58.05	106.65	135	337.5	607.5	810
Ingredients: 100g of bread, and 15ml peanut butter	120g	peanut butter sand		25.9	51.8	159.1	292.3	370	925	1665	2220
Ingredients:100 g of bread, 100 g of banana	200g	bread and banana		26.6	53.2	163.4	300.2	380	950	1710	2280
Ingredients: 100g bread, 50g shakshookah, 20g cheese	170g	shakshookah sand	0	28.84	57.68	177.16	325.48	412	1030	1854	2472
Ingredients: bread and 50g fried shrimp, 20g cheese	170g	shrimp sand	0	35.14	70.28	215.86	396.58	502	1255	2259	3012

Caloric of ingredients (CNF and USDA)	
butter, for marsa, 75g	538
calories in wheat 100g	362
kcal in sunflower oil 15ml for frying	122
kcal in sunflower oil 5ml for usual cooking	40
zaatar 50g	138
mozzeralla 50	160
labnah 50g	71
lamb 50g	203
chicken 50g	146
sugar 100ml	327
tomato 50g	9
onion 40g	16
spinach 50g	11.5
bread 100g	275
flafel 50g	55
cheese cheddar 20g	80
liver 50g	99
lamb 50g	203
fried egg 50g	96
boiled egg 50g	70
Fries 50g	159
tuna 50g	93
corn starch 100g	381
oat 100g	365
jam two table spoons	82
Seeds, sesame butter, tahini, roasted and toasted kernels 100g	150g
camel meat 250g	245

Appendix 11: Custom Declaration Form

Great Britain Grande-Bretagne

CUSTOMS DECLARATION — **CN 23**

No. of item (barcode, if any) | May be opened officially | Important! See instructions on the back

DÉCLARATION EN DOUANE — Nº de l'envoi (code à barres, s'il existe) | Peut être ouvert d'office

From / De

Name: IBRAHIM GOSADI

Business: —

Street: SABYA

Postcode: — City: SABYA

Country: SAUDI ARABIA

To / A

Name: RAMI OBAID

Business: SHEFFIELD UNIVESITY

Street: FLAT K181, OPAL 3. HOYLE STREET 80

Postcode: S3 7LG City: SHEFFIELD

Country: UK

Importer's reference (if any) (tax code/VAT No./importer code) (optional)

Importer's telephone/fax/e-mail (if known): 07545454428

Detailed description of contents (1)	Quantity (2)	Net Weight (3)	Value (5)	For commercial items only: HS tariff number (7)	Country of origin of goods (8)
BUCCAL SWABS AND ORI-CAPSULES SEALD IN TUBES. HAVE BEEN USED FOR RESEARCH PURPOSE	361 USED SWAB PACKS		—		
		Total gross weight (4): 2 kG	Total value (6)	Postal charges/Fees (9)	

Category of item (10): ☐ Gift ☐ Documents ☐ Commercial sample ☑ Other — Explanation:

Office of origin/Date of posting: SABIA / JIZAN - SAUDI ARABIA 07/01/2013

Comments (11): THERAPEUTIC SUBSTANCES OF HUMAN ORIGIN FOR NON-COMMERCIAL MEDICAL/SCIENTIFIC PURPOSES: RELIEF CLAIMED. OFFICIAL SHEFFIELD UNIVERGITY LETTER ENCLOSED

I certify that the particulars given in this customs declaration are correct and that this item does not contain any dangerous article or articles prohibited by legislation or by postal or customs regulations

☐ Licence (12)	☐ Certificate (13)	☐ Invoice (14)	Date and sender's signature (15)
No(s). of licence(s)	No(s). of certificate(s)	No. of invoice	07/01/2013

Appendix 12: Shipment Cover Letter

The University Of Sheffield.

School Of Health And Related Research.

Dean
Professor Jon Nicholl

Regent Court
30 Regent Street
Sheffield S1 4DA

Telephone: +44 (0) 114 22 5201
Fax: +44 (0) 114 272 4095
Email: j.nicholl@sheffield.ac.uk

To whom it may concern.

This letter is to accompany the shipment of samples which have been collected as part of the research study '*Investigating the potential effect of inbreeding on type 2 Diabetes susceptibility in a Saudi population*' The purpose of the research is to examine the potential effect of consanguinity and specific genetic variants on the risk of developing type 2 diabetes. Recruitment to this study took place during the period July 2012 to January 2013. In order to measure the genetic characteristics of this population DNA samples are required and buccal swabs have been used to collect DNA samples from all participants. The samples were all collected in Saudi Arabia under the approval of Directorate of health affairs in Jazan, Saudi Arabia. All participants consented to provide single buccal swabs. Each used a DNA swab pack which contained 1 swab sealed in a tube with a single 'Dri-capsule' (a preservative capsule). The swab packs have been purchased from *Isohelix*, (SK-1S DNA Swab Pack and SGC Dri-Capsules) a division of *Cell Projects, UK.* Samples are now required to be imported to the UK for the purpose of extracting the DNA and performing the genetic analysis of the collected buccal swabs by *K-biosciences, UK.* This study is a research project. Due to its academic nature, this project is exempt from license by Human Tissue Authority and Department of Health in the UK.

Research Study Investigator:

Dr Ibrahim Gosadi.
Postgraduate Research Student
School of Health and Related Research
Contact: +966 562137711
Email: igosadi@ksu.edu.sa

DATE: 07/01/2013

The University Of Sheffield.
School Of Health And Related Research

Research Study Supervisors:

Dr Dawn Teare
Senior Lecturer in Genetic Epidemiology
School of Health and Related Research
email: m.d.teare@sheffield.ac.uk
telephone: 0114 222 8398

Professor Elizabeth Goyder,
Deputy Dean of ScHARR and Professor of Public Health
School of Health and Related Research
email: e.goyder@sheffield.ac.uk
telephone: 0114 222 0783

Dawn Teare

E Goyder

The Queen's Anniversary Prizes

Appendix 13: allelic and genotype frequencies in different populations

No	SNP	N	Frq 22*	Frq21*	Frq11 *	Frq Risk alleles	Frq Other Alleles	Population (ref)
1	rs10946398	2917	0.33	0.49	0.18	0.58	0.42	Han Chinese (171)
		334	0.56	0.36	0.08	0.74	0.26	Saudi Arabian
2	rs10811661	2922	0.28	0.51	0.21	0.53	0.47	Han Chinese (171)
		501	0.29	0.52	0.19	0.55	0.45	Korean (179)
		4678	0.74	0.24	0.02	0.86	0.14	Finnish (177)
		338	0.66	0.3	0.04	0.81	0.19	Saudi Arabian
3	rs1111875	501	0.08	0.45	0.47	0.30	0.70	Korean (179)
		4748	0.29	0.48	0.23	0.53	0.47	Finnish (177)
		332	0.67	0.25	0.08	0.79	0.21	Saudi Arabian
4	rs13266634	4739	0.40	0.45	0.15	0.63	0.37	Finnish (177)
		16137	0.12	0.42	0.46	0.33	0.67	European (178)
		2396	0.21	0.49	0.30	0.45	0.55	Chinese (178)
		2033	0.003	0.057	0.94	0.03	0.97	African (178)
		330	0.77	0.20	0.02	0.87	0.13	Saudi Arabian
5	rs7754840	2932	0.18	0.49	0.33	0.42	0.58	Han Chinese (171)
		500	0.14	0.52	0.35	0.40	0.60	Korean (179)
		4644	0.14	0.47	0.39	0.37	0.63	Finnish (177)
		339	0.08	0.36	0.56	0.25	0.75	Saudi Arabian
6	rs560887	2225	0.10	0.42	0.48	0.31	0.69	Dutch (198)
		335	0.59	0.34	0.07	0.76	0.24	Saudi Arabian
7	rs2191349	1779	0.31	0.49	0.20	0.56	0.44	European (298)
		328	0.39	0.39	0.22	0.59	0.41	Saudi Arabian
8	rs10830963	2295	0.55	0.39	0.06	0.75	0.25	Dutch (198)
		2238	0.18	0.49	0.33	0.42	0.58	Chinese (187)
		328	0.07	0.31	0.62	0.21	0.79	Saudi Arabian
9	rs7034200	1781	0.24	0.50	0.26	0.49	0.51	European (298)
		329	0.4	0.46	0.14	0.63	0.37	Saudi Arabian
10	rs174550	1758	0.46	0.45	0.09	0.69	0.31	European (298)
		324	0.69	0.28	0.03	0.83	0.17	Saudi Arabian
11	rs11708067	1778	0.62	0.33	0.05	0.79	0.21	European (298)
		333	0.77	0.21	0.02	0.88	0.12	Saudi Arabian
12	rs4607517	1727	0.04	0.29	0.67	0.19	0.81	European (298)
		330	0.09	0.31	0.6	0.17	0.83	Saudi Arabian
13	rs7903146	4733	0.04	0.33	0.63	0.20	0.80	Finnish (177)
		931	0.16	0.50	0.34	0.41	0.59	Moroccan (170)

		1561	0.11	0.41	0.48	0.31	0.69	Austrian (170)
		334	0.15	0.44	0.41	0.33	0.67	Saudi Arabian
14	rs1799884	2208	0.69	0.28	0.03	0.83	0.17	Dutch (198)
		339	0.11	0.33	0.56	0.26	0.74	Saudi Arabian
15	rs5219	4658	0.23	0.50	0.27	0.48	0.52	Finnish (177)
		335	0.07	0.28	0.65	0.21	0.79	Saudi Arabian
16	rs7756992	2910	0.27	0.50	0.23	0.52	0.48	Han Chinese (171)
		16158	0.54	0.39	0.07	0.73	0.27	European (178)
		2391	0.27	0.47	0.26	0.51	0.49	Chinese (178)
		1886	0.16	0.45	0.39	0.38	0.62	African (178)
		336	0.07	0.4	0.53	0.26	0.74	Saudi Arabian
17	rs780094	1759	0.32	0.49	0.19	0.57	0.43	European (298)
		333	0.42	0.39	0.19	0.62	0.38	Saudi Arabian
18	rs2237897	4842	0.43	0.46	0.11	0.66	0.34	Japanese (184)
		3168	0.45	0.43	0.12	0.67	0.33	Singaporean (184)
		9479	0.93	0.07	0.00	0.96	0.04	Dane (184)
		336	0.95	0.02	0.03	0.98	0.02	Saudi Arabian
19	rs2237895	4714	0.15	0.47	0.38	0.39	0.61	Japanese (184)
		3378	0.13	0.46	0.41	0.36	0.64	Singaporean (184)
		8181	0.18	0.50	0.32	0.43	0.57	Dane (184)
		2995	0.19	0.47	0.34	0.43	0.57	Indian (185)
		332	0.27	0.45	0.28	0.49	0.51	Saudi Arabian
20	rs1387153	3870	0.16	0.51	0.33	0.42	0.58	Chinese (299)
		330	0.09	0.41	0.5	0.3	0.7	Saudi Arabian
21	rs231362	120	0.58	0.35	0.07	0.76	0.24	Indian (185)
		327	0.37	0.46	0.17	0.58	0.42	Saudi Arabian
22	rs7957197	120	0.74	0.23	0.03	0.85	0.15	European (300)
		334	0.75	0.21	0.03	0.86	0.14	Saudi Arabian
23	rs757210	120	0.15	0.57	0.28	0.44	0.56	European (301)
		327	0.27	0.43	0.29	0.49	0.51	Saudi Arabian

*22 indicates homozygote risk alleles, 21 indicates heterozygote alleles,

and 11 indicates homozygote non risk alleles.

Appendix 14: Introduction to Khat

What is Khat?

Catha edulis or Khat (as called in Jazan) is a plant with stimulating amphetamine-like characteristics **(293)**. The act of (Takzeen), or chewing leaves of Khat, is performed in order to achieve the stimulant effect of this plant. Tender and fresh leaves are usually picked and chewed. Chewed leaves are not swallowed and kept in one side of the mouth for several hours.

This habit of Khat chewing is more prevalent in Eastern African countries and in the south western areas of the Arabic peninsula **(302)**. However, it is also consumed by immigrants to European countries, including the UK. Although its importation is banned in several western countries, it is allowed to be imported and consumed in the UK **(303)**.

Cultivation, transportation, trading, and consumption of Khat are strictly unlawful actions in Saudi Arabia. The legal action has also been supported by religious restrictions prohibiting the consumption of Khat. Those who are arrested performing activities related to Khat trading or consumption can face imprisonment. Despite all governmental restrictions, chewing Khat is widely spread social behaviour in Jazan area.

Khat is cultivated in regions with high altitudes (1500-2000 meters above sea level) **(304)**. This quality makes mountains in Jazan and borders of Yemen a suitable natural habitat for the tree of Khat. In order for stimulant effect to be maximized, Khat leaves always have to be fresh. Leaves are usually consumed within few days of harvesting if not on the same day. Khat leaves are usually kept fresh by wrapping it with a wit clothes or banana leaves. Additionally, the commercial value of the leaves are correlated with its freshness **(305)** (Figure 1).

Figure 1: Khat Leaves. *

*This photo was taken by a friend who is a Khat chewer. This photo has no Copy of Right.

A typical session of Khat chewing usually starts in the evening. Chewers usually consume a large meal before the chewing session hence once they start chewing Khat, they will no longer have the appetite to eat any food. Chewing sessions usually done in groups and those who chew

Khat are more likely to be tobacco or water-pipe smokers **(305, 306)**. Khat chewing session can last up to six hours **(306)**. In some extreme sessions, chewing can last overnight and some individuals might attend their jobs without nocturnal sleeping.

Khat in Jazan:

A study conducted in Jazan revealed that 37% of male students (aged between 15 and 25) were Khat chewers. The percentage of female Khat chewers in the same population was greatly smaller (3.7%). There is also a tendency for older subjects to chew Khat more than younger ones. Percentage of those who chew Khat under the age of 15 was 17.4% compared to 24.2% in those who are older than 20 **(293)**. This study was restricted to high school and university students. Although there is no study published regarding consumption of Khat in older population, based on the obvious trend of higher tendency to consume Khat with older age and based on our observation during data collection, we can expect that percentage of adult subjects who are Khat chewer to be around 50% if not more.

Pharmacological effect of Khat:

The effective substance in Khat is called Cathinone. It has a stimulating effect of the sympathatic nervous system leading to an increase in blood pressure and tachycardia. It also has psychological and euphorigenic action (307). This psychological action gives Khat an addictive property hence it gives the chewer a temporary sensation of activeness and happiness.

The effect of Khat chewing on health:

Although Khat chewing is associated with activeness and euphoria during the session, several drawbacks have been reported especially after the chewing session. A study suggested in increase in mood disturbances and depression in Khat chewers compared to non chewers (308). The depressive effect of Khat was reported to increase toward the end of the chewing session (305). Another study indicated that anxiety and insomnia were more prevalent in chewers than non chewers (309).

The strongest effect of Khat is achieved when fresh leaves are chewed. The concentration of Cathinone in Khat leaves depends on where the tree was cultivated and on the freshness of leaves (305). Storing the leaves for longer durations causes the effect of Cathinone to be reduced when compared to fresh ones. This effect interferes with storing Khat leaves for longer terms and therefore, Khat chewers will always be on a continuous quest to find fresh Khat.

Certain individuals exhibit daily consumption of Khat. For those individuals, finding fresh Khat is another cause of anxiety giving that Khat is banned and has to be smuggled in Jazan region.

Consumption of Khat can induce several adverse effects on several systems. Hence Khat is kept in the mouth for several hours, it can be associated with periodontal diseases. In addition, it has been claimed that it might be associated with gastrointestinal illnesses such as oesophagitis and gastritis (293). During chewing sessions, chewers are less likely to consume food which might end in anorexia and malnutrition.

Khat and diabetes:

During Data collection, several type 2 diabetes patients claimed that they were chewing Khat due to its effect on blood glucose. They were believing that Khat chewing is likely to reduce blood glucose and thus leading to better control of glucose level. In few extreme occasions, certain patients admitted stopping taking medications prescribed by their doctors and favoured to use Khat as a remedy.

Although Khat might have an impact on blood glucose, it is likely to be temporal due to the anorexic effect of Khat leading to lack of appetite and consequently temporal reduction of blood glucose level during or toward the end of chewing session. This effect is likely to be reversed once the effect of Cathinone is reduced after cessation of Khat chewing. Two studies have suggested that Khat's effect on level of blood glucose is not significant and thus unlikely to reduce it **(310, 311)**.

Socioeconomic effect of Khat consumption:

Although Khat consumption leads to a sensation of activeness and euphoria, it is chewed while individuals are resting to allow proper chewing. Chewers tend to exhibit lack of physical activities for several hours in every session. Khat consumption has to be done in private and secluded areas hence it is banned in Jazan area. Therefore, Khat chewers are almost never engaged in public or social actions while they are chewing. Psychological effect of Khat might aggravate the social isolation where in best situations, chewers tend only to meet closed circle of relatives and friends during chewing sessions.

In addition to its adverse effect on social behaviour, Khat consumption can have adverse effect on economical status of chewers. While Khat consumption causes insomnia, it leads to disturbance in sleeping time and therefore, affects level of responsiveness and activity during work hours. If we assume that half of the adult workers are Khat chewers, we can confidently argue that the quality of work provided by chewers is very likely to be incompetent affecting the overall economic potentials of a whole society.

The cost of Khat varies between countries. The cost of Khat required for one session in the UK is approximately 5 to 10 pounds. Hence it is banned in Jazan area, the cost of same amount of is approximately between 20 and 50 pounds depending on the freshness and type of Khat leaves. Although expenditure on Khat varies between magnitudes of addiction, in extreme situations, certain individuals tend to spend up to £1000 on monthly basis. Giving the elevated cost of Khat, its consumption poses a financial burden on families if more than one member is Khat chewer.