

**Nutrient profiling and nutritional genomic influence on
metabolic risk of cardiovascular diseases and diabetes:**

The Airwave Health Monitoring Study

Rebeca Gonzalez Eriksen

**A dissertation submitted in partial fulfillment of the requirements for the
degree of PhD**

Imperial College London Department of Medicine

2017

Table of content

Statement of personal contribution	5
Copyright declaration	6
Abstract	7
Acknowledgement	8
List of tables	9
List of figures	11
Abbreviations	12
CHAPTER 1: BACKGROUND	13
1.1. Metabolic syndrome	13
<i>1.1.1. Definition</i>	<i>13</i>
<i>1.1.2. Prevalence and association with other chronic diseases</i>	<i>14</i>
<i>1.1.3. Aetiology</i>	<i>15</i>
1.2. Diet quality association with metabolic risk factors	16
<i>1.2.1. Single nutrients association with metabolic risk factors</i>	<i>16</i>
<i>1.2.2. Overall diet quality association with metabolic risk factors</i>	<i>18</i>
<i>1.2.3. Limitations of published diet quality scores</i>	<i>21</i>
1.3. Diet quality definition and measurement	22
<i>1.3.1. Nutrient profiling</i>	<i>23</i>
<i>1.3.2. UK Nutrient profile model</i>	<i>24</i>
<i>1.3.3. Dietary assessment bias</i>	<i>27</i>
1.4. Genetics and metabolic syndrome risk factors	29
<i>1.4.1. Structure and function of the human genome</i>	<i>29</i>
<i>1.4.2 Genetic pathways of metabolic syndrome risk factors</i>	<i>30</i>
<i>1.4.3 Type 2 diabetes aetiology</i>	<i>31</i>
<i>1.4.4 Type 2 diabetes and environmental risk factors</i>	<i>33</i>
<i>1.4.5 Type 2 diabetes and genetic risk factors</i>	<i>33</i>
<i>1.4.5.1 Genetic susceptibility of type 2 diabetes</i>	<i>33</i>
<i>1.4.5.2 Epigenetics and Type 2 diabetes</i>	<i>35</i>
<i>1.4.5.3 Genetic and environmental interaction in relation to Type 2 diabetes</i>	<i>36</i>
<i>1.4.5.4 Genetic and diet interaction in relation to Type 2 diabetes</i>	<i>37</i>
RESEARCH HYPOTHESIS, AIMS AND OBJECTIVES	40
<i>Hypothesis</i>	<i>40</i>
<i>Aims</i>	<i>40</i>
<i>Objectives</i>	<i>40</i>

CHAPTER 2: STUDY METHODS	41
2.1 Study design	41
2.1.1 <i>Study population</i>	41
2.1.2 <i>Sampling</i>	42
2.1.3 <i>Ethics</i>	42
2.1.4 <i>Data collection</i>	42
2.2 Dietary assessment quality control	44
2.2.1 <i>Operational manual and protocol</i>	46
2.2.2 <i>Dietary data audit</i>	46
2.2.3 <i>Underreporting of energy</i>	48
2.3. Diet quality definition and measurement	48
2.3.1 <i>Diet quality definition</i>	48
2.3.2 <i>Diet quality measurement</i>	51
2.4 Clinical definitions	53
2.5 Genetics	54
2.5.1. <i>Selection of genetic markers</i>	54
2.5.2 <i>Systematic literature review data management and cleaning</i>	56
2.5.3. <i>Data quality control</i>	58
2.5.4 <i>Polygenetic risk score modelling</i>	60
2.6 Statistical analysis	60
2.6.1 <i>Descriptive analysis</i>	61
2.6.2 <i>Correlation tables</i>	61
2.6.3 <i>Diet quality analysis</i>	61
2.6.4 <i>Genetic risk score analysis</i>	63
CHAPTER 3: RESULTS DIET QUALITY AND CARDIO-METABOLIC RISK	66
3.1 Descriptive statistics	66
3.1.1 <i>Demographic and lifestyle characteristics of the sample</i>	68
3.1.2 <i>Prevalence of metabolic outcomes</i>	69
3.2 Dietary data and quality control	69
3.2.1 <i>Code errors analysis</i>	69
3.2.2 <i>Energy underreporting</i>	69
3.2.4 <i>Dietary data average daily intake</i>	71
3.2.5 <i>Correlation table between dietary variables and cardio-metabolic markers</i>	72
3.3 Diet quality analysis	75
3.3.1 <i>DRV index score</i>	75
3.3.2 <i>Demographic and lifestyle characteristics across DRV index quartiles</i>	75
3.3.3 <i>Association between DRV index and nutritional components</i>	77

3.3.5 Association between NP model and nutritional components	79
3.3.6 Association between nutritional components incorporated into the diet scores and cardio-metabolic markers	81
3.3.7 Association between diet scores and cardio-metabolic risk markers	82
3.3.8 Association between diet scores and cardiovascular risk and diabetes	83
3.3.9 Association between diet scores and metabolic outcomes	84
3.4 Sensitivity analysis	86
3.4.1 Excluding participants with energy under-reporting	86
3.4.2 Excluding participants with chronic diseases	88
3.5 Discussion	90
CHAPTER 4: RESULTS GENETIC RISK ON DIABETES AND INTERACTIONS	98
4.1 Genetic risk score	98
4.2 Characteristics across quartiles of genetic risk score	98
4.3 Phenotypic trait HbA1c association with genetic risk score	99
4.4 Phenotype T2D and pre-diabetes association with genetic risk score	102
4.5 Genetic interactions with diet quality and single dietary variables on HbA1c	104
4.6 Alcohol interaction with genetic risk score on HbA1c	106
4.7 BMI interaction with genetic risk score on HbA1c	107
4.8 Effect modification of diet quality and genetic risk score on T2D risk outcomes	108
4.1 Discussion	109
CHAPTER 5: STRENGTHS, LIMITATIONS AND CONCLUSIONS	119
5.1 Summary of findings	119
5.2 Strengths and limitations	119
5.3 Future work	123
5.4 Conclusion	124
REFERENCES	125
Appendix	135
Appendix 1: UK Nutrient profile model algorithm	135
Appendix 2: Illumina Infinium genotyping protocol and summary report	137
Appendix 3: Airwave diary coding operational manual and protocol	140
Appendix 4: Systematic literature review databases	172
Appendix 5: PRISMA flowchart of systematic literature review	176
Appendix 6: Final list of 87 genetic markers included in genetic risk score	177
Appendix 7: Histogram of the DRV score	186
Appendix 8: Histogram of the nutrient profile score	187
Appendix 9: Histogram of the genetic risk score	188

Statement of personal contribution

I declare that this dissertation is my own work and that where any material could be construed as the work of others, it is fully cited and referenced, and/or with appropriate acknowledgement given.

The author was responsible of all data processing, data handling, and data analyses. The author validated and calculated the diet quality scores and genetic risk scores for the participants. All critical appraisal and interpretation of results was the opinion of the author.

Name and signature of author of PhD thesis: Rebeca Gonzalez Eriksen



Date: 19/10/2017

We confirm that this is an accurate representation of this work.

Name and signature of supervisor: Professor Gary Frost



Date: 19/10/2017

Name and signature of supervisor: Professor Paul Elliott



Date: 19/10/2017

Name and signature of supervisor: Dr Queenie Chan



Date: 19/10/2017

Copyright declaration

‘The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution None Commercial No Derivatives license. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the license terms of this work’.

Abstract

Background: Cardiovascular disease (CVD) is one of the leading causes of death worldwide, driven by various metabolic risk factors and diabetes. Diet is a key modifiable component in the development of these risk factors. This study aimed to investigate cardio-metabolic risk factors associated with diet. Secondly, dietary interaction with genetic risk of diabetes.

Methods: A cross-sectional analysis was conducted on 5,864 participants' baseline data from the Airwave Health Monitoring Study. Participants completed a 7-day food diary and a health screening including blood sample, blood pressure and anthropometrics. Diet quality was assessed according to UK Dietary Reference Values (DRV) using the novel DRV index and the established UK nutrient profile model. The nutritional genomic association with diabetes and HbA1c% was studied using a polygenetic risk score (GRS).

Results: A higher DRV score was associated with a lower waist circumference (β -0.56 $p<0.0001$), total cholesterol (β -0.06, $p<0.0001$) and glycated haemoglobin A1c (β -0.02, $p=0.003$). Individuals with a higher DRV score (top quartile vs bottom quartile) were less likely to have diabetes (OR 0.75, $p=0.04$), elevated blood sugar (OR 0.80, $p=0.01$) and abdominal obesity (OR 0.64, $p<0.0001$). GRS was associated with a higher HbA1c (β 0.03, $p<0.0001$) and higher risk of diabetes (OR 1.13, 95%CI 1.04-1.24). Interactions between GRS and higher diet quality modified the genetic effect on HbA1c (β -0.017, $p_{interaction}=0.02$) in high-risk individuals. High alcohol intake ($>16\text{g/day}$) modified the genetic effect on HbA1c (β 1.22, $p=0.002$). Obesity ($\text{BMI} >30\text{kg/m}^2$) increased the effect on HbA1c by $\beta=1.90$ $p_{interaction}=0.0006$.

Conclusion: Individuals consuming a high quality diet (high in complex carbohydrates, fibre, fruit and vegetables and wholegrain) aligned with the UK guidelines were associated with fewer metabolic risk factors for CVD and diabetes. A high quality diet also reduced the genetic effect on HbA1c and diabetes in high-risk individuals. These findings support the importance of encouraging a healthy diet, along with moderate alcohol consumption, as part of public health advice for a lower risk of CVD and diabetes, especially in high genetic risk groups.

Acknowledgement

Firstly, I would like to express my gratitude to all my supervisors: Professor Gary Frost, Professor Paul Elliott, and Dr Queenie Chan for their invaluable advice, insightful comments and for challenging me to go further and beyond my expectations. And I like to thank Dr Evangelou Evangelos and his team for their contribution to the genotyping analyses of the Airwave Health Monitoring study data.

I also wish to take this opportunity to thank to all members of the Airwave Health Monitoring Study research team, and in particular Katie Lamb and Yvonne McMeel for their incredible work on the diet analysis. And to Rachel Gibson my research partner she was always there for me, offering advice and sharing her expertise in the field, which helped me throughout the research process and made writing of this thesis much less stressful.

Finally, I would like to thank my dearest friend Karim Sanbar for encouraging me to pursue this doctorate and supporting me in every step of the way.

The Airwave Study is supported by The Home Office (contract number CS780) with additional support from The National Institute for Health Research (NIHR) and Biomedical Research Centre at Imperial College Healthcare NHS Trust and Imperial College London.

List of tables

Table 1: Previous criteria proposed for clinical diagnosis of metabolic syndrome

Table 2: Unified international criteria for metabolic syndrome

Table 3: UK Dietary Reference Values associated with metabolic risk factors findings from the Committee on Medical Aspects and the Scientific Advisory Committee on Nutrition

Table 4: Summary of previous research investigating the relation of diet quality scores and cardio-metabolic risk factors

Table 5: **Summary of meta-analysis of Mediterranean diet protective role on cardio-metabolic risk factors**

Table 6: UK Nutrient profile model scoring system

Table 7: Summary of UK nutrient profile model validation studies to healthiness of diet and health outcomes

Table 8: Goldberg cut-offs to determine acceptable energy reporting

Table 9: UK Dietary Reference Values for macronutrient and food recommendations for cardiovascular health findings from the Committee on Medical Aspects and the Scientific Advisory Committee on Nutrition

Table 10: Dietary Reference Value index nutritional scoring system based on mean intake per day

Table 11: Metabolic syndrome risk factors definition used in this study

Table 12: Cardiovascular disease and diabetes definition used in this study

Table 13: Search words for systematic literature review

Table 14: Search criteria systematic literature review

Table 15: Demographic and lifestyle characteristics of the sample

Table 16: Prevalence of metabolic risk factor outcomes

Table 17: Energy under-reporters and their associated lifestyle characteristics

Table 18: Comparison of dietary intake across men and women

Table 19: Spearman rank correlation coefficients for dietary variables adjusted for sex, the Airwave Health Monitoring Study

Table 20: Partial spearman correlation coefficients for dietary variables and cardio-metabolic markers

Table 21: Demographic and lifestyle characteristics across Dietary Reference Value index quartiles

Table 22: Association between Dietary Reference Value index quartiles and nutritional components

Table 23: Association between Nutrient Profile model quartiles and nutritional components

Table 24: Association between cardio-metabolic risk markers and single nutritional components included in the diet scores (Nutrient Profile and Dietary Reference Value)

Table 25: Association between diet scores (Dietary Reference Value, Nutrient Profile and Dietary Approach to Stop Hypertension) and cardio-metabolic risk markers

Table 26: Association between diet scores (Dietary Reference Value, Nutrient Profile and Dietary Approach to Stop Hypertension) and cardiovascular diseases and diabetes

Table 27: Association between Dietary Reference Value quartiles and metabolic syndrome risk factor classifications

Table 28: Association between Nutrient Profile quartiles and metabolic syndrome risk factors classifications

Table 29: Sensitivity analysis (excluding energy under-reporters) Dietary Reference Value index association with metabolic risk factors

Table 30: Sensitivity analysis (excluding energy under-reporters) Dietary Reference Value index association with metabolic risk factors

Table 31: Sensitivity analysis (excluding ill people) cardiovascular risk factors according to adherence to the Dietary Reference Value index

Table 32: Sensitivity analysis (excluding ill people) likelihood of having metabolic risk factors across the Dietary Reference Value index

Table 33: Socio-demographic and lifestyle characteristics across genetic risk score quartiles

Table 34: Multiple regression models for HbA1c % per 1 (SD) unit increase in genetic risk score

Table 35: ANCOVA model HbA1c % mean across genetic risk score quartiles

Table 36: The genetic risk score association with pre-diabetes and type 2 diabetes

Table 37: Likelihood of having pre-diabetes and type 2 diabetes associated with genetic risk score quartiles

Table 38: ANCOVA model for HbA1c across genetic risk score quartiles interacting with dietary variables

Table 39: ANCOVA model for HbA1c across alcohol categories interacting genetic risk score

Table 40: ANCOVA model for HbA1c across BMI categories interacting genetic risk score

Table 41: Effect modification of diet quality on genetic risk on the likelihood of having type 2 diabetes and pre-diabetes

List of figures

Figure 1: Logic pathway of metabolic syndrome

Figure 2: Diet quality definition and measurement

Figure 3: Proposed mechanism in type 2 diabetes aetiology

Figure 4: Epigenetic mechanism pathways

Figure 5: Proposed nutritional genomic interactions on phenotype

Figure 6: Proposed nutritional genomic pathways associated with type 2 diabetes risk

Figure 7: Sampling of datasets

Figure 8: Schematic of quality control audit cycle

Figure 9: Definition of UK Dietary Reference Values

Figure 10: Systematic literature review flowchart

Abbreviations

BMI	Body mass index
BP	Blood pressure
CHO	Carbohydrate
CVD	Cardiovascular disease
DASH	Dietary approach to stop hypertension
DBP	Diastolic blood pressure
DIAGRAM	DIABetes Genetics Replication And Meta-analysis consortium
DNA	Deoxyribonucleic acid
DRV	Dietary reference values
EPIC	European Prospective Investigation into Cancer and Nutrition
FFA	Free fatty acids
FFQ	Food frequency questionnaire
FSA	Food Standard Agency
GLM	General linear models
GRS	Genetic risk score
GWAS	Genome wide association studies
HbA1c	Glycated haemoglobin A1c
HDL	High density lipoprotein
HEI	Healthy eating index
INTERMAP	International Collaborative Study on Macro-/Micronutrients and Blood Pressure
LD	Linkage disequilibrium
LDL	Low density lipoprotein
MAF	Minor allele frequencies
MAGIC	Meta-Analyses of Glucose and Insulin-related traits Consortium
MDS	Mediterranean diet score
MetS	Metabolic syndrome
MET	Metabolic equivalent of task
NDNS	National Diet and Nutrition Survey
NHGIR	National Human Genome Institute Research
NP	Nutrient profile
SBP	Systolic blood pressure
SNP	Single nucleotide polymorphism
T2D	Type 2 Diabetes
TG	Triglycerides

CHAPTER 1: BACKGROUND

1.1. Metabolic syndrome

1.1.1. Definition

Metabolic syndrome (MetS) is a cluster of risk factors associated with an increased risk of cardiovascular disease (CVD) and type 2 diabetes (T2D) (1, 2). The risk factors include elevated blood pressure (BP), elevated blood glucose insulin resistance, high triglycerides (TG), low high-density lipoprotein (HDL) cholesterol and abdominal obesity. MetS was first classified in 1988 in a study, which showed that these risk factors were highly interlinked. It was suggested that insulin resistance was the leading risk factor associated with metabolic abnormalities (3, 4). In 1998 the World Health Organisation (WHO) defined MetS by the presence of insulin resistance along with two other risk factors e.g. elevated waist circumference and BP (3) (Table 1). Later, in 2001 the National Cholesterol Education Program Adult Treatment Panel III (NCEP ATPIII) suggested individuals with any three of these risk factors were classified as MetS (4). This definition was backed by the International Diabetes Federation (IDF) (5). The IDF definition emphasized abdominal obesity as a prerequisite for the diagnosis of MetS and added population-specific waist circumference to their criteria (6) (Table 1).

Table 1: Previous criteria proposed for clinical diagnosis of metabolic syndrome

Risk factor	WHO (1998) (3)	NCEP (2001) (4)	IDF (2005) (5)
Abdominal obesity	Male: waist-hip ratio >0.90: female: waist-hip ratio >0.85 and/or BMI≥30	WC >102 cm in men or >88cm in women†	Population specific
Hypertension	BP ≥ 140/90 mm Hg	BP≥130/85 mm Hg	BP≥130/85 mm Hg or on BP treatment
Hyperglycaemia/ Insulin resistance	Impaired insulin resistance, glucose tolerance or T2D	Plasma glucose > 5.6mmol L or T2D	Plasma glucose > 5.6mmol L or T2D
Dyslipidaemia	TG ≥150 mg/dL and/or HDL-C <35 mg/dL in men or <39 mg/dL in women	TG≥150 mg/dL and/or HDL-C ≤40 mg/dL in men or 50 mg/dL in women	TG≥150 mg/dL and/or HDL-C ≤40 mg/dL in men or ≤50 mg/dL in women or treatment
Micro albuminuria	Urinary albumin:creatinine ratio ≥30:299 mg/g	N/A	N/A

Abbreviations: BMI, body mass index; WC, waist circumference; BP, blood pressure; T2D, type 2 diabetes; TG, triglycerides; HDL-C, high-density lipoprotein cholesterol.

In 2010, the IDF, the American Heart Association/National Heart, Lung, and Blood Institute joined with the World Heart Federation, International Atherosclerosis Society and International Association for the Study of Obesity to develop one unified international definition. The definition classifies individuals with MetS if three or more of the five risk factors are present (Table 2) as published by Alberti et al. (1).

Table 2: The Unified International Criteria for metabolic syndrome adapted from Alberti et al. (1)

Risk factor	Cut –off measures	Other criteria
Hyperglycaemia	Fasting blood glucose ≥ 100 mg/dL	Receiving drug therapy for hyperglycaemia
Hypertension	BP $\geq 130/85$ mm Hg	Receiving drug therapy for hypertension
High triglycerides	Triglycerides ≥ 150 mg/dL	Receiving drug therapy for hypertriglyceridemia
Low HDL-Cholesterol	HDL-Cholesterol ≤ 40 mg/dL in men or ≤ 50 mg/dL in women	Receiving drug therapy for reduced HDL-Cholesterol
High waist circumference	Population specific	

Abbreviations: BP, blood pressure; HDL-C, high-density lipoprotein cholesterol.

MetS criteria low cut-off measures permit early detection across all its risk factors prior to development of the actual diseases e.g. detection of elevated blood sugar before development of diabetes. This makes it a useful tool in epidemiological studies for both measuring traditional CVD risk factors as well as identifying high-risk individuals early, especially in a younger study population.

1.1.2. Prevalence and association with other chronic diseases

In Europe almost one in four adults have MetS; in the UK this number ranges from 20% in men (7) to 29% in women (8). This prevalence has been increasing parallel to the rise in obesity (9), although not all obese individuals are at risk of MetS (6). A recent study revealed

a proportion of 10-25% obese individuals remained metabolically healthy due to preserved insulin sensitivity (10).

A person who has MetS is twice as likely to develop CVD as someone who does not (6). Worldwide CVD is the number one cause of death accounting for 31% of deaths in 2012 (17.5 million) and predicted to be 23.4 million deaths by 2030 (11). A recent meta-analysis of 43 cohorts (n= 172,573 individuals) found MetS conveyed a relative risk for CVD events and death of 1.78 (2). The risk was still associated with the syndrome after adjusting for traditional CVD risk factors such as smoking, obesity, hypertension and dyslipidaemia (relative risk 1.54) suggesting that risk accompanying the MetS cannot be explained entirely by the traditional risk factors alone (2, 6).

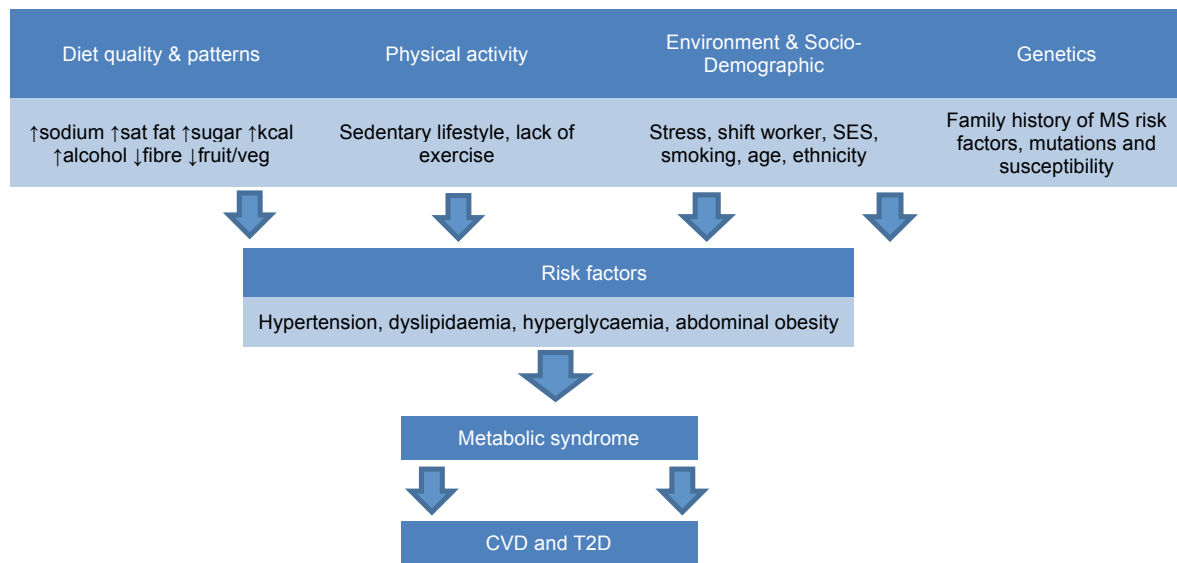
Furthermore, a person who has MetS is five times more likely to develop T2D compared to a person without MetS (6). People with T2D have a five-fold risk of having CVD and are 1.6 times more likely to die prematurely, compared with those without T2D (12). Over 382 million people worldwide are estimated to have T2D, of which 3 million live in UK (13, 14). This figure is estimated to double in the next 20 years, making it the fastest growing disease worldwide (14). In England, 1 in every 3 people are pre-diabetic (elevated blood glucose levels above the normal range, but not high enough for a diabetes diagnosis) (13, 15). A recent UK study found that the incidence of T2D over a 3-year period was 15 times higher in pre-diabetic individuals (HbA1c between 6.0–6.4%) at the beginning of the study, compared with those with a baseline HbA1c <5.0% (16). These findings prove the importance of identifying pre-diabetic individuals, as early intervention at this critical stage may halt the development of T2D.

Evidence suggests that MetS promotes the development of CVD and T2D at multiple levels. Therefore, identifying individuals who are at high-risk of MetS and effective control of its risk factors is key to overcoming life-threatening chronic diseases.

1.1.3. Aetiology

MetS is multi-factorial: obesity, poor quality diet and physical inactivity are the driving forces behind MetS risk factors (17), but a number of other contributing factors are also involved in its pathway (18). These include aging, alcohol, smoking, endocrine disorders and genetic risk factors (12, 18) illustrated in Figure 1.

Figure 1: Proposed pathway of development of metabolic syndrome risk factors



Findings from the US Third National Health and Nutrition Examination Survey showed how the likelihood of having MetS increased with BMI. Prevalence was 4.6%, 22.4%, and 59.6% of normal-weight, overweight, and obese men, respectively, and a similar distribution was observed in women. Participants who reported low physical inactivity were 40% more likely to have MetS compared to those conducting regular exercise (17). Furthermore the study also supports the findings from other studies that MetS prevalence is highly associated with a poorer quality diet, smoking and an older age (7, 19, 20).

The focus of this thesis is to investigate these metabolic risk factors for CVD and diabetes and their association with diet quality and genetics. However, other contributing factors, e.g. physical activity, alcohol and smoking, will also be investigated and adjusted for in the analysis.

1.2. Diet quality association with metabolic risk factors

1.2.1. Single nutrients association with metabolic risk factors

Diet modification is a well-established key component for preventing and treating many of the MetS risk factors. Extensive evidence from epidemiological studies and randomised controlled trials (RCTs) has established independent relationships between single nutrients in the diet and BP, adiposity markers, blood sugar control and dyslipidaemia (3, 12, 21, 22). The WHO report “Diet, nutrition and the prevention of chronic diseases” summarised a number of

RCTs and epidemiological studies on the association of nutrients and food groups with metabolic risk factors for T2D and CVD (12). A high intake of total fat, in particular saturated fat, was linked with increased risk of all MetS risk factors. High intake of sugar was linked with risk of obesity and high intake of sodium was associated with high BP. A high intake of dietary fibre, fruit and vegetables was shown to reduce the risk of all MetS risk factors (12).

In the UK, the Committee on Medical Aspects (COMA) of Food Policy published in 1991 the “Dietary Reference Values for Food Energy and Nutrients” (DRV) (23). The DRV are estimated recommended dietary intakes for energy, nutrients and food groups for optimal nutrition for a UK population. In 1994, COMA published the report “Nutritional Aspects of Cardiovascular Disease” (21), which summarised evidence from a number of RCTs and epidemiological studies on key groups of the DRV in relation to prevention of CVD. The report highlighted seven nutritional components and six nutrients: saturated fat, non-milk extrinsic sugar, sodium, total carbohydrates, fibre, and one food group fruit/vegetables as important for cardiovascular health in the UK population. The Scientific Advisory Committee on Nutrition (SACN) later revised the population recommendations for estimated energy requirements in 2011 (24), and in 2015 modified fibre intake and added free sugars to the recommendations (25). These UK dietary guidelines were more recently reviewed by Public Health England (*Levy et al. 2017*) in relation to CVD prevention (26). The review highlighted the importance of six nutrients and two food groups fruit/vegetables and fish (including oily fish) for optimal cardio-metabolic health in an adult population (Table 3). These UK guidelines also adhere closely with the WHO ‘Diet, nutrition and the prevention of chronic diseases’ report of 2003 (12).

Table 3: UK Dietary Reference Values associated with metabolic risk factors. Findings from the Committee on Medical Aspects and the Scientific Advisory Committee on Nutrition (23-27)

Nutritional component	Dietary goal*	Metabolic risk factors
Total fat	<35 % of energy intake	Obesity Hypertension Hyperglycaemia Dyslipidaemia Diabetes
Saturated fat	<11 % of energy intake	
Total carbohydrates	>50 % of energy intake	
Fruit & vegetables	>400 g/day	
Fish (including oily)	> 2 portion/week	
Fibre	>30 g/day	
Free sugar**	< 5 % of energy intake	
Sugar (non-milk extrinsic)***	< 11 % of energy intake	
Sodium	<2400 mg/day	

* Based on healthy adult British population mean energy intake 2,000 kcal for women and 2,500 kcal for men

** SACN definition free sugars are those added to food or those naturally present in honey, syrups and unsweetened fruit juices, but exclude lactose in milk and milk products.

*** SACN definition non-milk same as free sugar and extrinsic also includes 50% of the fruit sugars from dried, stewed or canned fruit.

The DRV for single nutrients and food groups have since been incorporated into UK's dietary guidelines in various forms such as nutrient profiles for food labels and a visual illustration the "Eat-well Guide", which represents the proportions of different food groups that contribute to a healthy and well-balanced diet. These guidelines reflect different aspect of diet quality and patterns important to for optimal health in a general adult UK population.

1.2.2. Overall diet quality association with metabolic risk factors

Traditionally, nutritional epidemiology has focused on the impact of single nutrients, but this approach has been linked with a number of limitations (28, 29). Studies have shown a holistic approach to dietary assessment such as diet quality scores are good tools to capture diet patterns as well as the combined effect of nutrients in a person's diet, which are important for cardio-metabolic risk (30-34). The main argument is that people do not consume single nutrients but meals with combination of different nutrients and foods, which are likely to be interactive or synergistic. Many nutrients are highly correlated, making it difficult to examine their separate effects. Furthermore, analyses of single nutrients may be confounded by the effect of dietary patterns (19, 29, 32, 33).

Studies have shown that overall diet quality is a stronger predictor for cardio-metabolic risk factors than single nutrients (19, 32, 33). Tucker *et al.*, also showed an additive protective effect of high fruit and vegetable intake with low saturated fat intake in relation to CVD risk and mortality (33). Similar clinical trials have found measuring overall diet quality rather than single diet components is more effective in lowering BP (32).

A key factor to studying overall diet quality versus single nutrients may be explained by the complex interactions amongst nutrients e.g. food matrix and nutrient bioavailability.

However, dietary factors contributing to the effect of the overall index cannot be ascertained without examining components included in the index such as single nutritional variable analyses (35), and exploring multi-collinearity of the nutritional variables included (28, 29).

There are many different diet indexes e.g. Mediterranean Diet Score (MDS), Healthy Eating Index (HEI), Dietary Approaches to Stop Hypertension (DASH) (35). An a priori diet index assesses overall diet quality in relation to specific nutritional components e.g. compliance with certain dietary recommendations and disease risk. The many indexes vary based on food groups and nutrients included, country-specific recommendations, cut-off values, and gender-specific values.

An overview of the evidence supporting an association between overall diet quality and MetS risk factors has been summarised in systematic literature reviews and meta-analysis from over 30 epidemiological studies and 35 randomised control trials (20, 36-38).

A systematic review summarized results from 17 large epidemiological studies (cross-sectional, RCT, cohorts) that examined the relationship between diet quality indexes and MetS risk factors. The review included dietary indexes derived from various countries (different population groups, ethnicity, age, social classes). It found that 11 of the 17 studies reported significant associations between various dietary indexes and cardio-metabolic risk (obesity, BP, dyslipidemia, hyperglycemia), CVD and cancer incidence (35). Table 4 summarises five large epidemiological studies investigating the relation of diet quality indexes and cardio-metabolic risk factors and outcome. They demonstrate significant associations between adherence to various dietary indexes and lower BMI, T2D and a lower CVD incidence, across different population groups.

Table 4: Summary of previous research investigating the relation of diet quality scores and cardio-metabolic risk factors

Study type	Author and date	Population and sample size	Dietary quality measure and data assessment	Outcome measure(s)	Main results	Significance
American Cancer Society Cancer Prevention Study II Nutrition Cohort 4-year follow up	Seymour et al.(2003)(38)	115,833 American adults	^a DQI from 68-item FFQ	CVD and stroke risk factors	Inversely associated with all circulatory diseases. Mortality rate ratio were 1.81 for the lower diet quality compared to only 1.00 in the highest group	P 0.003 (multivariate adjusted)
US National Health and Nutrition Examination Survey	Nicklas et al.(2012)(37)	18,988 American adults	^b HEI from 24 hour diet recall	CVD, MetS risk factors	Associated with BMI, WC, BP and cholesterol (HDL,LDL)	P <0.001 (multivariate adjusted)
US Nurses' Health Study and the Health Professionals Follow-up Study Cohort 20-year follow up	Chiuve et al. (2011) (39)	104,666 US adults	^b HEI, ^c MDS, DQI from 138-item FFQ	CVD,T2D, BMI	Inversely associated with risk reduction in DM type II, CVD and lower BMI	P <0.01 (multivariate adjusted)
Cross-sectional study	Schroder et al (2004)(40)	3,162 Spanish adults	^c MDS from 165-item food frequency questionnaire	BMI, obesity	Inversely associated with BMI and obesity risk	P 0.007, P 0.010 (multivariate adjusted)
NutriNet-Sante Cohort Study	Lassale et al (2013) (20)	7,902 French adults	^d PNNS-GS from 24 hour diet record	MetS, adiposity markers	Inversely associated with MetS: BP, TG, WC and BMI	P <0.001 (multivariate adjusted)

Abbreviations: DQI, diet quality index; HEI, healthy eating index; MDS, Mediterranean diet score; PNNS-GS, Programme National Nutrition Santé Guideline Score; BMI, body mass index; WC, waist circumference; BP, blood pressure; T2D, type 2 diabetes; TG, triglycerides; HDL, high-density lipoprotein; LDL, low-density lipoprotein.

^a total fat, saturated fat, cholesterol, fruit/veg, bread/cereal/legumes, protein, sodium, calcium

^b fruit, veg, dark green, orange veg/legumes, grains, milk, meat/beans, oils, saturated fat, sodium, kcal from fat, alcohol, sugar

^c vegetables, legumes, fruit, nuts, whole-grain, fish, monounsaturated to saturated fat ratio, alcohol, red and processed meats

^d saturated fat, mono/poly unsaturated fat, alcohol, fruit/veg, meat, legumes, dairy

^e fruit/veg, whole grain products, dairy, meat, eggs, fish, seafood: oils, salt, alcohol, sugar, physical activity

Similar results were found in a meta-analysis (random effects) of 15 epidemiological studies and 35 randomised control trials representing data from over half a million subjects (41). The study showed that adherence to MDS was associated with a reduction of various cardio-metabolic risk factors (adjusted for physical activity, age, gender, and smoking). Table 5 shows results as mean difference between highest versus lowest category of MDS (95% confidence) e.g. participants in the highest (healthiest) MDS have a - 0.42 cm (95%CI -0.82 to -0.02) lower reduction in waist circumference compared to participants in the lowest (least healthy) MDS.

Table 5: Summary of meta-analysis of Mediterranean diet protective role on cardio-metabolic risk factors (41)

Metabolic risk factors	Cross-sectional studies (n=531,470) Mean difference (95% CI)†	Randomised control trials (n=3,436) Mean difference (95% CI)††
Waist circumference (cm)	-5.78* (-7.26 to -4.31)	- 0.42 * (-0.82 to -0.02)
HDL (mg/dl)	2.25 * (1.15 to 3.34)	1.17 * (0.38 to 1.96)
TG (mg/dl)	-9.88 * (-16.05 to -3.71)	- 6.14 * (-10.35 to -1.93)
BP Systolic / Diastolic (mmHg)	-1.81 (-4.83 to 1.21) / -2.36 (-6.12 to 1.39)	-2.35 * (-3.51 to -1.18) / -1.58 (-2.02 to -1.13)
Plasma glucose (mg/dl)	-3.93 * (-6.25 to -1.61)	-3.89 * (-5.84 to -1.95)
Insulin resistance	-0.87 * (-1.15 to -0.59)	-0.45 * (-0.74 to - 0.16)

Abbreviations: BP, blood pressure; T2D, type 2 diabetes; TG, triglycerides; HDL, high-density lipoprotein

†A high adherence to Mediterranean diet compared to a low Mediterranean diet adherence

†† Mediterranean diet compared to a control diet

* Significant associations $p < 0.05$

1.2.3. Limitations of published diet quality scores

There no dietary index or pattern analysis that is regarded as better than all of the others; there is little consensus on which method should be applied for any given purpose. As previously mentioned, each dietary index is derived from different nutritional variables: a varying number, range, and type of variables estimated from a variety of dietary measurement methods, which are then entered into the analysis procedure.

The main limitation of the different dietary indexes is reproducibility and validity, which has been discussed in various papers (28, 29, 35, 42, 43). The use of dietary indexes may depend

on the nature of the dietary assessment measures (food frequency questionnaire, diet recalls, records), health outcomes and population groups of interest (28, 29, 35, 42, 43).

The combinations of foods and nutrients included in the dietary index are intended to represent the total diet or key factors in the diet of its particular study population and health outcomes of interest. These dietary variables and cut-off values are often derived from country-specific dietary recommendations and different populations groups, who may differ in social demographics (ethnicity, age, social class). Thus, dietary indexes depend on the operational definitions and cut-off values of nutritional components in the index, limiting interpretability and comparability across studies.

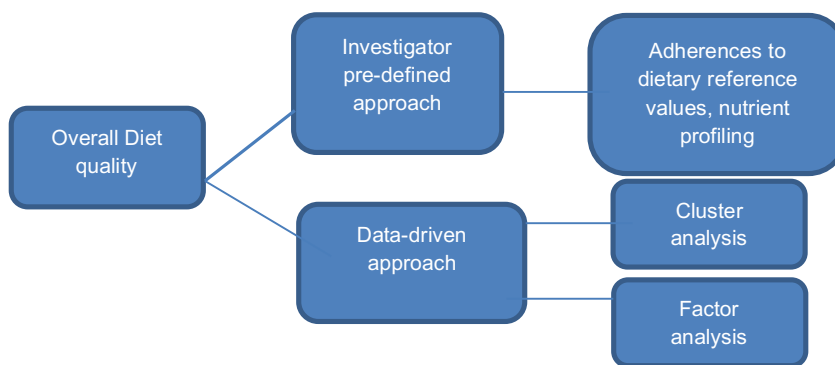
A critical review of 20 distinct pre-defined diet quality indexes (including those presented in Table 4, 5) found many differing aspects, such as cut-off values used, dietary components included and some also include alcohol (28). Evidence suggests that not all dietary indexes adequately capture dietary components related to a specific population's dietary and health outcomes (28, 35, 44). One argument is that the dietary indexes are often derived from specific populations and may not be suitable across different populations (28, 42). The European Prospective Investigation into Cancer and Nutrition (EPIC) study investigated Mediterranean diet score association with health outcomes (mortality) in 10 different European countries (42). The study found countries such as Greece and Spain had significant associations, but these were not seen in Northern countries like Germany and Netherlands suggesting that dietary indexes may need to be country specific (42).

In summary, diet quality has been found associated with MetS. Evidence supports the use of a dietary index to measure the combined effect of nutrients in a person's diet in relation to MetS risk factors. However, it suggests the need for a study-specific diet index to capture nutrients and foods essential to its study population.

1.3. Diet quality definition and measurement

There is neither an official diet quality index nor a definition of diet quality in the UK. Researchers and healthcare professionals worldwide use a variety of different diet quality definitions, often country specific, based on national dietary guidelines and set reference values (28, 31, 35). Diet quality definition and measurement methods can be classified into two main approaches as shown in Figure 2 (28, 35).

Figure 2: Diet quality definition and measurement methods



The investigator pre-defined approach has advantages over the data-driven to a specific outcome approach because it is based on existing knowledge of an optimal diet e.g. previously studied international dietary guidelines association with MetS risk factors. The pre-defined method has proven a relevant and reliable tool for defining and measuring overall diet quality in relation to MetS risk factors (28, 35). However, limitations may occur if the models incorporate nutritional components that are less sensitive to the measured health outcome of interest or they do not reflect other important dietary patterns. Whereas data-driven methods of defining dietary patterns have the advantage of not making any previous assumptions on the use of existing data to characterise total diet - hypothesis-free, so that patterns describe the eating behaviours of a population. However, the data driven method based on study-specific associations e.g. principal cluster analysis also has limitations as it does not use existing reference values (known evidence) and it may be difficult to validate and generalise the results as they are study specific. It is also prone to bias such as systematic and random errors introduced by the data analysis, collection etc. (28, 35).

1.3.1. Nutrient profiling

In 2007, a new European Commission regulation on nutrition and health claims made on foods came into force. They proposed the use of NP models to determine a food's quality based on each country's dietary guidelines (45, 46). WHO definition:

“Nutrient profiling is the science of ranking or classifying food groups according to their nutritional composition based on their relationship to preventing a disease and promoting health”(46)

Today the European Commission, along with many international governments and health departments use NP models to frame a recommended diet, promote healthier eating and regulate health claims (22, 46). A total of 119 different versions of NP models exist worldwide (47). The models vary based on the country of origin, diet and dietary recommendations.

NP models have multiple applications worldwide e.g. as a measurement tool in research for identifying healthy and less healthy foods, public health education and regulation (48, 49).

Epidemiological studies have validated many of the NP models in relation to overall diet quality and healthy eating patterns and found significant correlations (22, 46, 50). However, to date, only a few studies have validated NP models against health outcomes as a predictive validation method. The Nurses' Health Study and Health Professional Follow-up Study used the US NP model to assess diet quality. They found NP score was inversely associated with risk of total chronic disease, CVD, diabetes and all-cause mortality (P-value for trend ≤ 0.01), in both cohorts (39).

1.3.2. UK Nutrient profile model

In 2005, the UK Food Standard Agency (FSA) and British Heart Foundation research group developed a NP model for regulating the advertising of foods in the UK. Later the model was adapted by the Republic of Ireland, Australia and New Zealand for food labelling and health claim regulation (51-53). Table 6 shows how foods or drinks are assessed on a scoring system where points are allocated on the basis of the nutrient content per 100g (53). In this scoring system, nutrient thresholds are derived from the UK DRV (53, 54). One point for each nutrient corresponds to 3.75% of the DRV (45, 53, 54).

Points are scored for negative nutrients: energy, saturated fat, total sugar and sodium and for positive nutrients: fruit, vegetables and nut content, fibre and protein.

The NP scores are then summed up to a total NP score of the whole diet using a previously validated method (22, 52, 55). The total NP score interprets quality of all foods consumed: Overall score = total negative score – total positive score. The algorithm is detailed in appendix 1.

Table 6: UK Food Standard Agency Nutrient profile scoring system (53)

	Points										
	0	1	2	3	4	5	6	7	8	9	10
Negative nutrients											
Energy (kJ)	≤335	>335	>670	>1005	>1,340	>1,675	>2,010	>2,345	>2,680	>3,015	>3,350
SFA (g)	≤1	>1	>2	>3	>4	>5	>6	>7	>8	>9	>10
Total sugars (g)	≤4.5	>4.5	>9	>13.5	>18	>22.5	>27	>31	>36	>40	>45
Sodium (mg)	≤90	>90	>180	>270	>360	>450	>540	>630	>720	>810	>900
Positive nutrients											
Protein (g)	≤1.6	>1.6	>3.2	>4.8	>6.4	>8					
Fibre (g)	≤ 0.7	>0.7	>1.4	>2.1	>2.8	>3.5					
% of fruits and vegetables	≤40	>40	>60	–	–	>80					

The validation of NP has been carried out by several authors (46) (see Table 7). The UK model has been validated in line with the international criteria of the WHO and European Commission (46) prior to its implementation in 2004. Furthermore it has been evaluated in the National Diet and Nutrition Survey (NDNS) by several expert groups (22). The model showed significant accuracy in classifying healthy and unhealthy foods in several studies and correlated with the European Commission dietary guidelines (22, 45, 48, 52, 55, 56). The Whitehall Cohort Study used an amended version of the UK NP model to assess diet quality and study predictive validity on multiple health outcomes (study details and results are presented in Table 7) (57).

A number of limitations exist with the NP model as it does not aim to address overall diet quality and patterns. The nutrient composition of individual foods is not the only determinant of the overall nutrient composition of diets. Secondly, it may not reflect the variety of different foods that make up the diets and the healthiness of the diets e.g. dietary patterns.

In summary, evidence suggests the NP model is a relevant tool for a UK population to measure quality of food items contributing to an individual's diet. However, limitations may exist in its use for measuring overall "healthiness" of diets and dietary patterns. The NP model has not yet been assessed if it is associated with the UK dietary guidelines for optimal health and associated with cardio-metabolic markers in a UK population.

Table 7: Summary of UK nutrient profile model validation studies to healthiness of diet and health outcomes

Study type	Author and date	Population and sample size	Dietary data assessment/intervention assigned	Outcome measure(s)	Main results	Significance
NDNS Observational-cross-sectional study	Arambepola et al. (2007) (58)	1,117 British adults	7-days food diaries	Healthiness of diet: validated against DQI – Food categories based on “Eat Well Plate” by expert panel of 150 dietitians	Associated with healthier diets and quality of foods	P <0.001 for all associations
NDNS Observational-cross-sectional study	Scarborough et al. (2010) (56)	1,722 British adults	7-days food diaries	Healthiness of diet: validated against DQI	Associated with overall healthiness of diets and energy consumption and BMI	P <0.001 for all associations
Five large national nutrition surveys Cross-sectional study	Volatier et al. (2007) (58)	8,550 European adults from Ireland, France, Italy, Denmark, Belgium	7-days food diaries	Healthiness of diet: validated against EU dietary guidelines and against American and French nutrient profile model	All models were independently associated with ‘healthy and unhealthy diets.’ UK model was the most accurate model across all four surveys	P <0.001 for all associations
Whitehall Cohort study *	Masset (2012) (51)	7,251 British adults	122-item FFQ	Predictive validation against CVD, diabetes, and mortality. And HEI and single nutrients	Associated with healthier foods, higher intake of all vitamins & minerals, risk reduction in CHD, overall mortality	P <0.001 for all associations

Abbreviations: DQI, diet quality index; HEI, healthy eating index; MDS, Mediterranean diet score; PNNS-GS, Programme National Nutrition Santé Guideline Score; BMI, body mass index; T2D, type 2 diabetes.

* Amended version of Food Standard Agency Nutrient Profile model

1.3.3. Dietary assessment bias

Assessing a person's diet quality can be challenging due to a number of biases linked with diet recording and the complexity of measuring important nutrients and food components in a diet.

Many different methods have been developed for assessing dietary intake. These range from detailed individual weighed records collected over a number of days to food frequency questionnaires, 24-hour recall and simple food lists. Each assessment method has merit, associated errors and practical difficulties (59, 60). Food frequency questionnaires and 24 hours' recall are only a small snap shot of an individual's diet, which often does not collect sufficient detail for the investigators to assess nutrient intake. However, these methods have been evaluated to be capable of assessing and estimating intake of specific foods especially at population level and are less time consuming and more economical in large epidemiological studies (59, 60).

This thesis will use a 7-day food diary. This is often regarded as gold standard in diet recording (59, 60). This method captures a participant's diet over a 7-day period. The diary records both weekdays and weekends and any variation over these 7 days will be captured. Therefore, this increases the representativeness of the individual's diet. These have been validated in several large epidemiological studies and nutritional surveys such as EPIC and NDNS (59). However, food diaries are still linked with a number of random and systematic errors such as:

Estimating portion sizes

Estimating weights of food and drinks consumed from a 7-day food diary may introduce estimation errors by both the participants recording it and the investigator analysing it (60, 61). Ensuring standardisation of portion size requires the use of household measures e.g. cups and spoons and food photographs. The investigator then converts these estimates into weights that can be used to calculate food and nutrient intake. There are a number of methods by which portion sizes may be obtained e.g. using the relevant diaries' food photographs based on national food data bases (like FSA and NDNS) for estimated standard portion sizes (60). Similar methods and standardised protocols were applied in the EPIC and NDNS (61)

Estimating food items

By self-recording food and drink items, participants may fail to accurately describe the item, which may lead to the investigator failing to analyse the correct item. The problem of omission may be lessened by obtaining additional information from the diary's general question, which records commonly used food item such as type of bread, type of fat spread and milk.

High respondent and investigator burden

Recording diet intake for 7 days often leads to participants getting fatigue of recording all foods consumed and they may forget or skip some meals or change their eating behaviour (reporting effect). Furthermore, the diaries demand time and precision to analyse which may result in systematic errors e.g. copy/paste, incorrect quantity estimations or food item by the investigator (59).

Misreporting and energy underreporting

All dietary recording methods including 7-day food diary are prone to energy and nutrient misreporting. Most commonly observed is energy underreporting (59, 61, 62). Misreporting in dietary assessments can be spilt into two main effects: observation effect - a change in eating behaviour and second is reporting effect - misreporting their dietary intake.

The misreporting effect appears to mainly cause energy underestimation. The FSA study reported underreporting between 5-20% (62). Another article reviewing a number of studies reported an underestimation of energy as measured by doubly labelled water of 4-37% (59). The highest levels of underreporting in dietary records have been found among individuals with high BMI and in women (59-62). The misreporting effect was more marked in individuals filling out the diaries themselves compared to if it was done by an interviewer (62).

The overall misreporting is most significant in energy distribution, but studies have also shown that underreporting was seen in specific nutrients such as fat, sugar and alcoholic beverages (59). The FSA study showed that women underreported their fat intake by 12% and men underreported their alcohol intake by 13% (62).

Many methods exist to detect energy underreporting in dietary assessment including doubly-labelled water, Bland-Altman plots and the Goldberg methods (24, 59, 61). These methods

assess the relationship between energy intake and expenditure. Evidence shows a good agreement between energy intake and energy expenditure in healthy population groups (23, 24). Energy intake in healthy individuals is largely regulated by energy expenditure. It is possible to estimate the correlation between recorded energy intake and the measured energy expenditure (23, 24, 63).

In summary, dietary assessment methods are related with a number of biases challenging the interpretation of the dataset. It is important to recognise the limitations and bias in a data set so they can be measured and adjusted for in the analysis.

1.4. Genetics and metabolic syndrome risk factors

1.4.1. Structure and function of the human genome

The human genome consists of 2 x 23 chromosome pairs (one set from each parent), which contains the genetic instructions for growth and development of all human cells. Chromosomes are strings of deoxyribonucleic acid (DNA), which are building blocks of over three billion base pairs adenine (A), thymine (T), cytosine (C) and guanine (G). These strings of base pairs are wrapped around histones (proteins) creating chromatin groups. Their location on the chromosome and sequence of the base pair makes up our genes. There are approximately 20,000-25,000 different genes in human DNA, which only represent 2% of the total DNA string (exons). The remaining 98% of DNA consist of non-coding DNA (introns) and host a number of gene promoters, which are known to play a key role in regulation of gene expression and repression (64).

Most of the DNA's coding genes sequences are the same in humans with only < 1% variance (65). This variance is why we are all physiologically unique, though some variations do not make any difference to our cell functions. Variation in our genes base pair sequence can be inherited or/and occur during our life affected by environmental factors, e.g. aging, exposure to chemical/radiation. The most common form of sequence variation occurs in high frequency in the general population and involves variation at a single base pair (allele) also referred to as single nucleotide polymorphism (SNPs). SNPs are associated with both natural variations in our appearance e.g. eye and hair colour and with variation in our risk in developing certain diseases e.g. cancer and T2D. Another known type of sequence variation includes mutations, which result in a gene changing its instruction e.g. translation code of a

protein (64). The frequencies of these SNPs vary across different ethnic population groups (65).

Gene expression is a process tightly regulated in every cell nucleus via intracellular signals. Once a gene is turned on via DNA methylation it copies its DNA to a similar molecule of ribonucleic acid (RNA) also referred to as transcription. This carries the information out of the cell nucleus for translation. In the translation process the sequence of the base pair (gene code) in the intron is translated and assembled into the coded amino acid making up the final protein molecule (64).

A person's genetic make-up, also referred as genotype, can be tested via extracted DNA from any biological specimen e.g. blood or saliva test. A number of different genotyping methods exist e.g. whole genome sequencing, target regions of the DNA (specific genes of interest) or gene expression profiling. The most commonly used method is measuring the fraction of the DNA using predefined microarray chips and comparing it to a reference sequence via imputations (64, 65).

Since the development of genome-wide association studies (GWAS) in 2007, scientists have been able to study the genetic susceptibility to numerous diseases including MetS in humans. GWAS are hypothesis-free scans of the genome using predefined microarray chips to identify common (high frequency) variants in a selected general population associated with phenotype of interest. The principal of GWAS genotyping and imputation is based on linkage disequilibrium (LD) analysis. LD is defined as a non-random association between two or more alleles at different loci normally due to an interaction between genes e.g. two genes are located on same chromosome. LD can tag common SNPs using existing data, a reference panel, the HapMap project (a catalogue and location map of common genetic variants haplotype in human) (66).

1.4.2 Genetic pathways of metabolic syndrome risk factors

MetS is multifactorial and polygenic disease involving several complex and overlapping risk factors and pathways, as previously discussed in chapter 1.1.4 aetiology. GWAS have found numerous genetic variants associated with MetS risk factors: dyslipidaemia, abdominal adiposity, hyperglycaemia and elevated blood pressure. These risk factors are also multifactorial and polygenic diseases such as MetS and despite the many hundreds of genetic variants found to be associated, their effects are often modest. The genetic susceptibility (predisposition) often only represents a small fraction of the development of the disease. The

individuals' genetic susceptibility is greatly dependent on the environment, e.g. lifestyle factors, diet, smoking. The interaction between genes and environment therefore strongly determines the susceptibility of developing MetS and its risk factors. The genetic contribution in developing dyslipidaemia is between 10-30% (67, 68). A similar figure is seen for the genetic susceptibility of hyperglycaemia 6-25% (69-71) and obesity (including abdominal fat) is ~2-3% (72-75).

Although the genetic effects of these variants are modest, they do suggest that they may affect essential metabolic pathways of MetS and contribute to an increase in risk (76-80). Many of the genetic variants associated with MetS risk factors overlap with each other (79-82). Atanasovska *et al.*, reviewed over 755 SNPs encompassing 366 independent locations related to MetS risk factors and found one fifth of these locations overlapped. Harboursing SNPs associated with two or more of MetS risk factors, e.g. obesity, plasma lipids or diabetes-related traits (81).

Due to the complexity of MetS' aetiology and its multiple pathways, this thesis will focus its genetic study on only one of these pathways: elevated blood glucose (MetS hyperglycaemia), which is defined as fasting glucose >5.5 mmol/L or HbA1c $> 5.7\%$ or on glucose lowering medication (1). Several studies have supported MetS hyperglycaemia and impaired insulin resistance as the dominant risk factor in MetS (1, 3). Furthermore, as previously mentioned, T2D is the most likely (five fold) health outcome to occur in people MetS (6). Therefore, understanding the contributing risk of genetics is important for early detection of MetS and T2D.

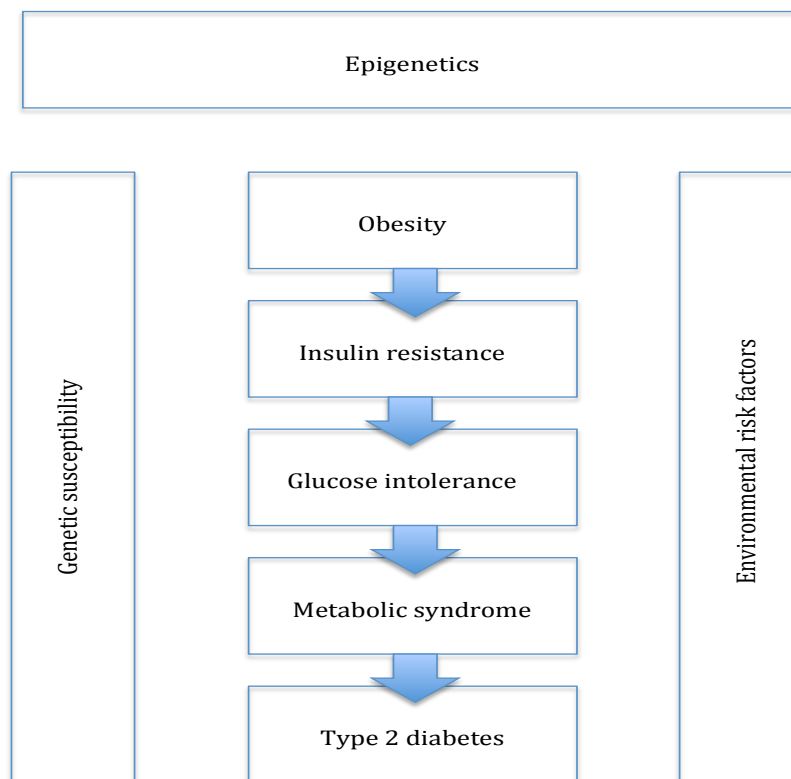
1.4.3 Type 2 diabetes aetiology

The MetS definition for hyperglycaemia captures both people who have pre-diabetes and T2D. T2D is defined as elevated blood glucose concentration when fasting (≥ 8.0 mmol/L) or elevated HbA1c levels ($\geq 6.5\%$) (83, 84). HbA1c measures the percentage of total haemoglobin that is bound by glucose produced from the non-enzymatic chemical modification by glucose of haemoglobin molecules carried in erythrocytes. Insulin is the primary hormone controlling the absorption of glucose from the blood into the body's cells. In a healthy person insulin production increases after glucose has been digested and transferred into the blood stream to initiate absorption and control healthy levels of glucose in the blood. Whereas in people with T2D the metabolic regulation is malfunctioning due to

complex interplay: insufficient insulin production in the pancreatic beta cells, poor cellular insulin and/or glucose sensitivity resulting in hyperglycaemia (83).

Pre-diabetes is a relatively new term recently defined by the American Diabetes Association as blood glucose concentration higher than normal but lower than established diabetes, defined as HbA1c between 5.7 – 6.4% (85). Pre-diabetes is a high-risk state for developing T2D and is associated with a number of metabolic abnormalities and vascular complications, similar to T2D, e.g. nephropathy, neuropathies, high BP (13, 86). Pre-diabetes aetiology follows the same pathway as T2D and they share the same risks factors such as obesity, high waist circumference, aging, and high BP (13, 15, 87). Despite its risks, pre-diabetes can be inversely impacted by lifestyle interventions such as diet, physical activity and weight loss(86, 88). Pre-diabetes and T2D are multifactorial conditions caused by complex interplay between a number of contributing factors both genetic and environmental (12, 83). These risk factors can be grouped into three main categories: environmental, genetic susceptibility and epigenetics factors.

Figure 3: Proposed mechanism in Type 2 diabetes aetiology



1.4.4 Type 2 diabetes and environmental risk factors

Several environmental (lifestyle) risk factors for T2D have been well documented, such as obesity and central obesity, physical inactivity, smoking, poor diet quality low in fibre, fruit, vegetables and high in sugar and saturated fat (12, 21, 83, 89, 90). Evidence suggests that per 1 kg/m² increase in BMI (>25 kg/m²) the risk of developing T2D rises by 8.4% and impaired fasting glucose rises by 9.5%. Per 1 cm increase in waist circumference above recommended, the risk of T2D and impaired fasting glucose increases by 3.5% and 3.2% respectively (91).

A number of larger intervention studies have demonstrated that all of these environmental factors are modifiable and can reverse the effect on T2D (12, 89, 90). The European Diabetes Prevention Study on 749 adults with impaired glucose tolerance found that lifestyle intervention (i.e. increased physical activity and consumption of a healthier diet) reduced the incidence of T2D: those with $\geq 5\%$ weight loss at one year had 65% lower T2D incidence (HR 0.35, 95% CI 0.22 to 0.56) and those who managed to maintain $\geq 5\%$ weight loss for two and three years further reduced T2D incidence (89). The Finnish Diabetes Prevention Study RCT demonstrated the efficacy of dietary and exercise intervention to prevent T2D in adults with impaired glucose tolerance: the risk of T2D was reduced by 58% ($p < 0.001$) in the intervention group, which was directly associated with changes in diet and physical activity (90).

1.4.5 Type 2 diabetes and genetic risk factors

1.4.5.1 Genetic susceptibility of type 2 diabetes

The genetic susceptibility for development of T2D and associated risk factors such as hyperglycaemia is estimated between 6-25% (69-71). Since 2007, large replicated GWAS in human populations have identified >150 SNPs highly associated with T2D (92-94) and >53 SNPs associated with T2D-related traits such as hyperglycaemia (70). The majority of the identified SNPs are high frequency SNPs commonly found across population groups of European ancestries.

The DIAbetes Genetics Replication And Meta-analysis consortium (DIAGRAM) is currently one of the largest T2D associated SNP databases collected on 12,171 cases and 56,862 controls, from a European ancestry population (92). The majority of the SNPs are associated with decreased insulin secretion and defective beta-cell function in diabetic subjects (70). To gain a better understanding of the pathophysiological processes of T2D susceptibility loci, DIAGRAM assessed the SNPs' effect on metabolic traits and their LD with other variants

associated with glycaemic traits. This analysis used data from one of the largest databases, the Meta-Analyses of Glucose and Insulin-related traits Consortium database (MAGIC). The MAGIC database includes 53 SNPs associated with fasting glucose, fasting insulin, 2 hour glucose and HbA1c derived from GWAS on 133,010 individuals of European ancestry without diabetes (70). DIAGRAM's analysis found that half of the glycaemic associated SNPs from the MAGIC database were associated with T2D. Two thirds of the T2D associated SNPs were also associated with glycaemic traits in non-diabetic individuals while one third had no association to any glycaemic traits (70). DIAGRAM also revealed the T2D and glycaemic trait associated SNPs had mechanistic heterogeneity between them, clustering them into four main metabolic pathways (69):

1. Insulin resistance
2. Insulin regulation
3. Insulin secretion and beta-cell function
4. Fasting glucose

Many large GWAS and meta-analysis studies have replicated DIAGRAM's findings confirming that T2D is a multifactorial polygenic disease involved in different pathways of T2D pathogenesis (92, 93, 95, 96).

A number of the T2D susceptibility loci overlap with known obesity-associated loci e.g. FTO and MC4R (94, 97). Many of these SNPs have been found to result in higher effect size on glycaemic trait and T2D (70, 92-94, 96, 97). Previously GWAS included these SNPs in their analysis and adjusted for BMI. However, recent evidence suggests these SNPs may still confound the association. Adipose tissue, especially abdominal tissue, is an endocrine organ affecting metabolic pathways, contributing to total glucose homeostasis (1, 18, 98). Studies have also shown that many of these SNPs that overlap with obesity variants lost their association with glycaemic trait post BMI adjustment highlighting that obesity SNPs may mediate the association via the contributing or confounding effect of adiposity tissue on T2D (69, 70, 92). Hence recent studies have found excluding these SNPs from their analysis provided a better precision of disease prediction (69, 70, 99).

The genetic susceptibility varies between individual SNPs effect size, frequency and on the number of SNPs present. Common SNPs effect size is generally low, odds ratio of T2D ranging from 1.02-1.15 per SNP (92, 93). Similar modest effect size is seen in glycaemic trait fasting glucose and HbA1c beta-coefficient ranging from 0.02-0.12 per SNP (70). However,

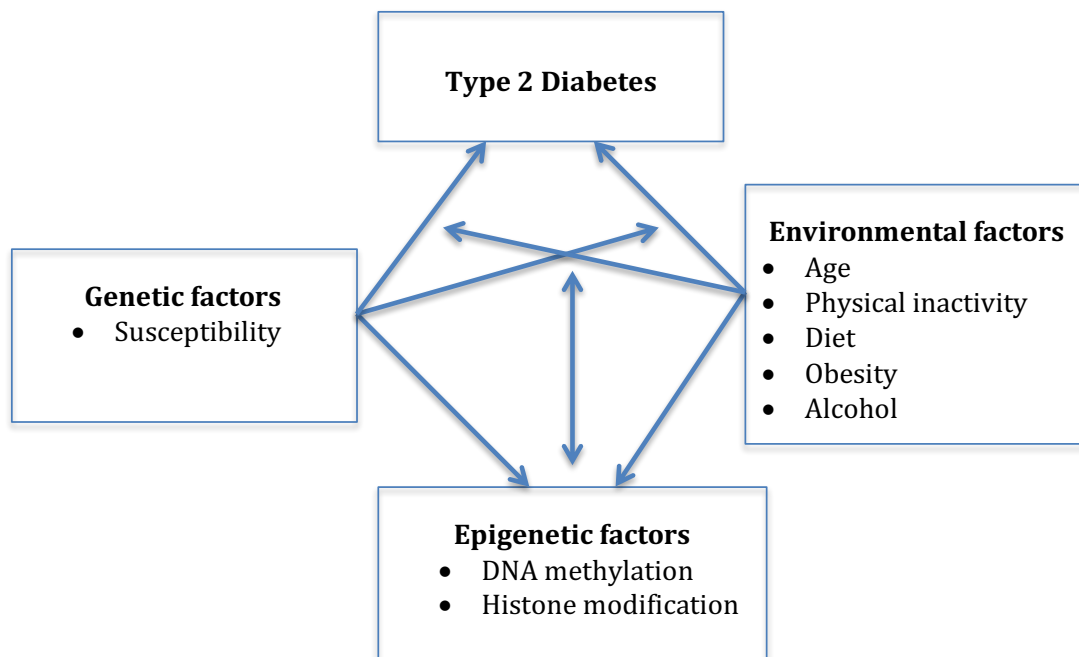
individuals with a combination of these SNPs can have an even higher risk, supporting the use of polygenetic risk scores (GRS) rather than single SNP analysis in T2D and glycaemic trait prediction models. A recent meta-analysis on T2D risk prediction using GRS's from 18 large epidemiological studies (9 cross-sectional and 10 longitudinal) reported an increased risk of 10-12% per one unit (one SD) increase in the score (99). The effect was twofold in individuals with the highest number compared to those with the lowest number of risk alleles. Similar results were found in a study comparing two polygenetic scores' influence on changes in fasting glucose. The authors reported the model containing the higher number of SNPs (46 SNPs) had a higher effect size on glucose level in 17% compared to the model with the least number of SNPs (15 SNPs) 7% (93).

T2D is a multifactorial polygenic disease and GWAS have identified more than 150 SNPs strongly associated with the disease, most of these being common variants and found in high frequency in the general population (94). The average T2D risk SNP frequency is estimated to be 54% in the general population. However, the effect size of these variants is low and only explains a small proportion of the heritability for T2D, suggesting that additional gene-environmental interaction factors remain to be identified.

1.4.5.2 Epigenetics and Type 2 diabetes

Additionally to genetic susceptibility, there is also growing evidence that T2D development is influenced by epigenetic mechanism, via a complex interplay between genetic variants and environmental factors (100). Epigenetics has been defined by Ling *et al.*, as an “*heritable changes in gene function without any change in the nucleotide sequence*” (101). Genes can be “turned on and off” in response to intracellular metabolic signals from a number of internal factors e.g. hormones, enzymes and external/environmental factors. This function is controlled by the body's cells via intracellular signals to targeted gene-regulating sites e.g. CpG sites in intron regions of DNA (Figure 4), which then stimulate gene repression via DNA methylation or gene expression via histone modification (101-103).

Figure 4: Epigenetic mechanism pathways in development of Type 2 diabetes



A replicated genome-wide DNA methylation analysis of pancreatic beta cells from diabetic and healthy individuals revealed epigenetic changes in 853 genes (1,649 individual CpG sites) (100). The changes in the genes expression were between 5-59% more frequent in diabetic compared with non-diabetic cells. Over 100 of these genes had altered expression contributing to reduced insulin production and secretion in T2D cells (adjusted for multiple covariates). Dayeh *et al.*, also found differentially methylated in regions of pancreatic beta cells in several of the candidate genes in diabetics, resulting in impaired glucose-stimulated insulin release (100).

1.4.5.3 Genetic and environmental interaction in relation to Type 2 diabetes

Epigenetic mechanisms have been found to be influenced by several external environmental factors e.g. genetic-environmental interaction. An interaction is a statistical term, where the combined effects of the genetic and environmental exposures are more than additive (94).

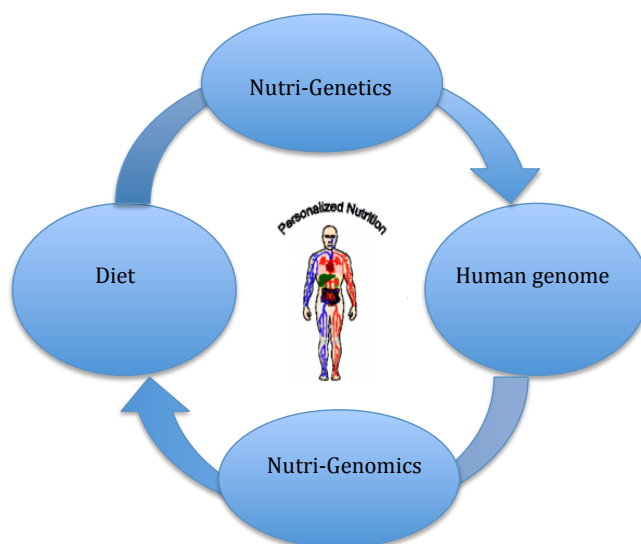
The Framingham Heart Study showed lifestyle intervention had a direct relationship on the pathophysiology of T2D due to indirect effect modification of genetic risk factors (104). This suggests that lifestyle factors may also modify the genetic effect on T2D risk, since a number of studies have shown genetic-environmental interaction in relation to phenotype T2D and hyperglycaemia. The Inter99 intervention study tested whether a GRS associates with the

incidence of T2D and with 5-year changes in glycaemic traits and whether the effects were modulated by changes in BMI and lifestyle (105). They observed an interaction between the GRS and changes in BMI, suggesting that individuals increasing their BMI may be more susceptible to the effects of T2D risk variants and individuals who lost weight significantly attenuate the effect of the genetic risk. Dayeh *et al.*, tested the impact of known environmental risk factors aging, hyperglycaemia (HbA1c) and obesity on DNA methylation (in non-diabetic subjects) and found that increasing HbA1c levels, aging and obesity affected differential DNA methylation in 92% of the CpG sites (100). Since a number of studies have found that environmental factors age, smoking, and obesity stimulate epigenetic pathways and increase the risk of T2D by activating intracellular signals to gene promoters (100-102). A recent intervention study also found protective epigenetic influence via altered DNA methylation in 21 T2D associated SNPs, in response to six months of exercise (103).

1.4.5.4 Genetic and diet interaction in relation to Type 2 diabetes

Genetic and diet interactions have also been referred as nutritional genomics (106). The nutritional genomics term covers the two main different pathways of how diet (nutrients and bioactive compounds in foods) can interact with genes and express specific phenotypic trait and health outcomes (Figure 5).

Figure 5: Proposed nutritional genomics interactions on phenotype

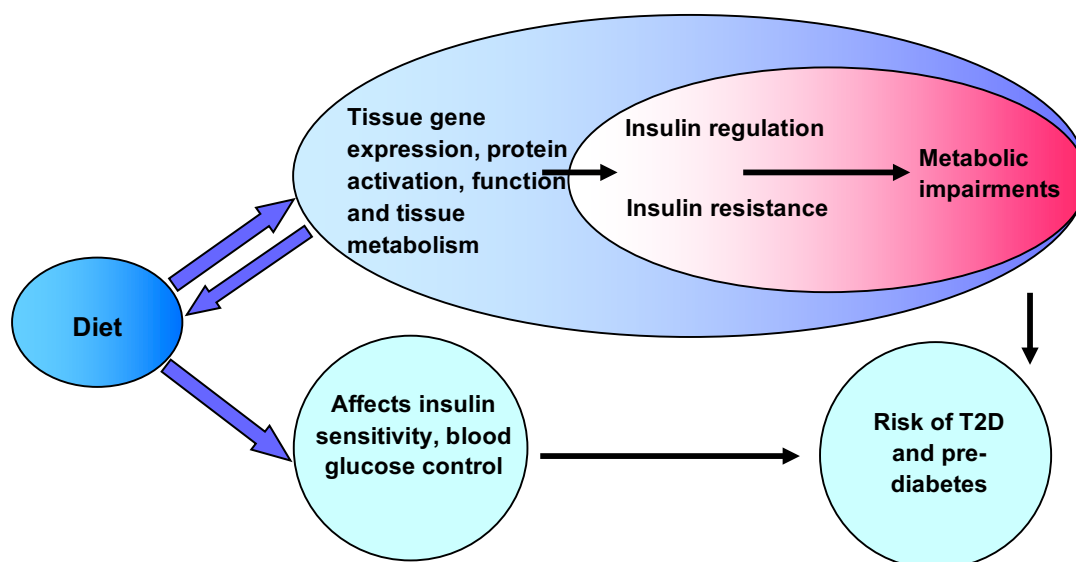


(1) Nutrigenetics: the influence of diet on genome (genetic expression) accounting for the variations in health status and disease risk (106).

(2) Nutrigenomics: the influence of genetic variants on the effect of nutritional components explaining the inter-individual variability in response to a diet (106).

A number of nutritional genomic studies have seen inter-individual variability on metabolic outcome in response to genetic and nutritional components interaction, suggesting these effect modifications may be responsible for the observed difference. A number of studies have investigated gene-diet interactions in relation to T2D risk and found that the effect modifies the risk both in positive and negative directions. The pathways affected vary depending on the genetic variant and nutrient studied. Figure 6 illustrates the reported pathways hypothesised by the findings of these studies.

Figure 6: Proposed genetic pathways associated with Type 2 diabetes risk that can be affected by diet



The US National Health and Nutrition Examination Surveys and Epidemiologic Architecture for Genes Linked to Environment Study identified genetic interaction with CHO and fibre intake as modifying the outcome on T2D. These effect modifications were observed in 6 SNPs reducing OR of T2D (107). Similar results were reported in the Nurse's Health Study amongst a European ancestors population (108). The EPIC cohort also saw interactions between wholegrain foods and SNPs associated with risk of T2D (109). Wholegrain intake was inversely associated with T2D risk among rs7903146 CC homozygote carriers (hazard ratio for 50 g portion per d=0.86: 95 % CI 0.75, 0.99).

A recent review of published nutritional genomic studies focusing on dietary fat intake found significant interactions between number of SNPs related to insulin resistance and T2D (110). It reported fatty acids' ability to bind specific isoforms, which activate SNPs in the PPAR gene. This binding varies according to the type of fatty acid. Increased intake of n-3 and n-6 PUFAs had a protective effect via PPAR γ activation, which regulates insulin sensitization and SFA had an opposite effect (110). The importance of type of dietary fat has also been reported in SNPs located in the SCARB1 gene. Individuals carrying this SNP showed significant increases in insulin sensitivity after a MUFA-rich diet (111). Marin *et al.*, found individuals carrying the Thr54 risk allele in FABP2 gene insulin sensitivity decreased when SFA were replaced by MUFAs and CHO (112). The Finnish Diabetes Prevention Study investigated the association of genetic risk on T2D incidence in an intervention study of 522 high-risk participants with impaired glucose tolerance (104). The study found that the incidence of T2D was lower in the intervention group and was associated with diet quality, physical activity and lower BMI. This indicates that both dietary and other environmental factors can reduce the risk conferred by genetic variants on impaired insulin secretion. Similar effect modifications were found in two other studies investigating SNPs associated with T2D located in ADIPOQ, ADIPOR1 and ADIPOR2 genes (113, 114). The studies showed a higher dietary intake of saturated fat modified the effect of insulin resistance in individuals carrying the risk alleles in SNP rs266729 and rs10920533.

Summary of the evidence from diet and environmental genetic interactions studies has demonstrated effect modifications in relation to glycaemic traits, insulin resistance and T2D risk. These modifications can affect both protective and negligent pathways of T2D development. Understanding these epigenetic mechanisms is important in relation to prevention for high-risk individuals. However, more evidence is needed to understand the nutritional genomic effects of T2D risk factors to help personalise dietetic advice in T2D prevention. While the focus of this thesis is dietary interaction, it should be noted that other environmental factors, in particular physical activity, alcohol and BMI, should also be considered.

RESEARCH HYPOTHESIS, AIMS AND OBJECTIVES

Hypothesis

1. Diet quality is associated with an inverse risk of cardio-metabolic health and T2D.
2. The genetic risk effect on HbA1c and T2D risk is modified by interactions between genes and lifestyle factors.

Aims

1. To develop a methodology for a study specific protocol for coding and quality control of dietary data from 7-day food diaries.
2. To explore methods to assess diet quality using the UK dietary guidelines and analyse its association with cardio-metabolic health amongst a British adult population.
3. To investigate the effect of genetic risk on HbA1c and T2D, and if this effect is modified by interactions between genetics and diet quality.

These aims will be explored via a cross-sectional study using clinical, genetic and dietary data available on participants from the Airwave Health Monitoring study.

Objectives

1. Develop a study-specific methodology for analysing and quality control 7-day food diaries
 - (1a) Develop an operational manual for standardisation of 7-day food diaries coding and a data quality control protocol.
2. Explore the relationship between diet quality and cardio-metabolic health and T2D.
 - (2a) Explore methods using the UK dietary guidelines to analyse diet quality and patterns in relation to cardio-metabolic health and T2D.
3. Investigate the effect of genetic risk on HbA1c and T2D, and if the effect is modified by genetics and lifestyle interactions.
 - (3a) Identify existing genetic markers for HbA1c and T2D risk by a systematic literature review and construct a genetic risk score
 - (3b) Conduct interaction studies between the genetic risk, diet and lifestyle risk factors on outcomes glycaemic trait HbA1c and risk of T2D.

CHAPTER 2: STUDY METHODS

2.1 Study design

This present study is using baseline data from the Airwave Health Monitoring study, a large cohort study investigating the health of the police force in Great Britain (115). The study was established in 2001 to evaluate possible health risks associated with the use of terrestrial trunked radio communication system amongst the police forces. The study also investigated the general health of the work force with comprehensive health screening, 7-day food diary, measurements of anthropometry, cardiovascular and cognitive functions, and performed biochemistry, genetics and metabolomics analyses. All members of the police force were eligible for inclusion. Enrolment commenced in 2004 and will continue until 2018. In January 2012, the study had recruited 42,112 participants from 28 forces throughout GB, of whom 35,199 (83.6%) had attended a health screen and 25,000 (63%) completed a 7-day food diary. The mean age of participants at baseline was 40 years, with 63% men and 94.5% white British from different socio-economic status and ranks (115).

2.1.1 Study population

The population of this PhD study included three samples purposely selected based on data availability from participants in the Airwave Health Monitoring study who took part in the health screen, genetic testing and provided dietary data.

Exclusion

Participants were excluded according to following criteria: failed to record > 1 day of their dietary intake, on meal substitution diets and gross error in their diaries and had a low energy reporting <500kcal/day.

2.1.2 Sampling

A sample 20,000 participants with clinical and dietary data (collected using 7-day food diaries) were available for this PhD study. For the purpose of this survey, three different sub-samples were selected based on data availability (Figure 7).

Sample 1 dietary data: Dietary data from 5,864 participants (7-day food diaries) were nutritionally coded. This sample was randomly selected from the main sample of 20,000 participants where both dietary and clinical data was available. The randomisation was done

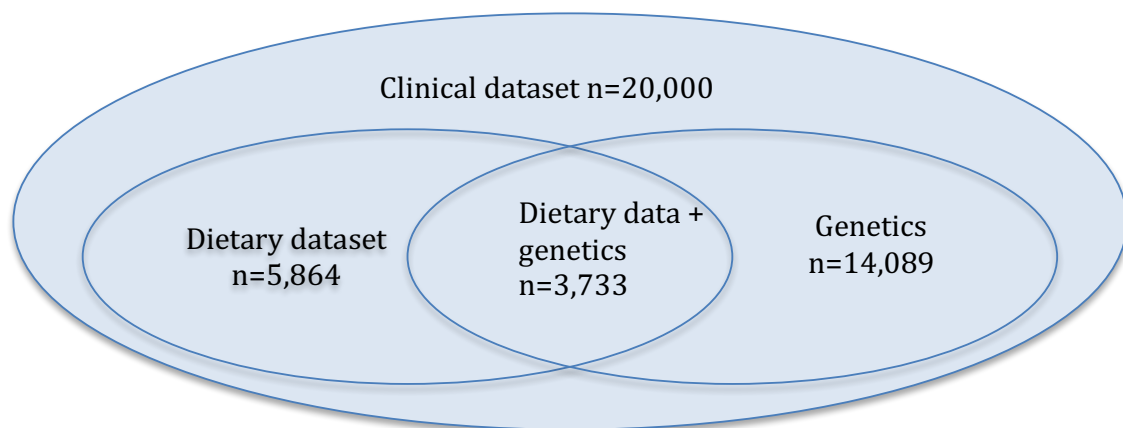
using Microsoft Excel random number generator. Due to time limitation only 5,864 participants were nutritionally coded during the three years of this survey.

Sample 2 genetic data: 14,089 participants out of the main sample had been genotyped by the time of this survey and therefore had genetic data available for use in the analyses presented here.

Sample 3 dietary and genetic data: Complete data on dietary data, clinical data and genetic data was available for 3,733 participants.

The sample's socio-demographic and lifestyle characteristics and prevalence of obesity, T2D and CVD were representative of the total cohort (115).

Figure 7: Sampling of three sub-samples from the main clinical dataset



2.1.3 Ethics

Participants provided written informed consent and the study had ethics approval from the National Health Service Multi-Site Research Ethics Committee (MREC/13/NW/0588).

2.1.4 Data collection

Apart from the data collected at the clinical examination, all data utilised from this study were based on self-administrated electronic questionnaires (115) and a 7-day food diaries (116).

2.1.4.1 Health screening and clinical examination

The health screenings were carried out in established Airwave clinics using a standard protocol. Trained nurses conducted all clinical examinations.

BP: sitting BP was taken using the Omron HEM 705-CP digital BP monitor (Omron Health Care). Three measurements were taken 30 seconds apart and the average was used.

Anthropometry: participants were measured in light clothing and without shoes or socks. Height and sitting height were measured using a Marsden H226 portable stadiometer and weight using a Marsden digital weighing scale. Two measurements were taken and the average was used. Waist circumference was measured between the lower rib margin and the iliac crest in the mid-axillary line using a Wessex-finger/joint measure tape (Seca 201, Seca). Two measurements were taken and the average was used.

Blood samples: all samples were taken non-fasting. All biochemistry tests were performed using serum except HbA1c determination, which used whole EDTA blood sample and glucose determination, which was performed using fluoride/oxalate sample tube. Blood samples were measured using an IL 650 analyser (Instrumentation Laboratory, Bedford, Massachusetts, USA).

Urine samples: spot urine samples had been collected but not processed and therefore not available to use for this thesis.

Genotyping: participants' DNA samples were processed and genotyped using the Illumina Infinium Human Exome-12v1-1 BeadChip Arrays and following the Infinium HD Assay Ultra protocol. 550,000 SNPs (350,000 common SNPs, 200,000) and 4 million SNPs are imputed in MiniMac (R2 threshold 0.9). All samples were processed in the National Phenome Center (Appendix 2). An extensive quality control was carried out on the genotyped and imputation data by researchers at Public Health Imperial College. The procedure involved checking imputation R-square correlation, genotyping call rate and filtering out samples with low call rates. The sample call rate was defined as the fraction of called SNPs per sample over the total number of SNPs in the dataset. Inclusion criteria threshold for call rate was set at <95%. Average overall call rate was >99%.

2.1.4.2 Socio-demographic, health and lifestyle questionnaire

Socio-demographic, health and lifestyle data were collected via a self-administered electronic questionnaire, which the participant filled in during their clinic visit. Variables used for this study included: age, gender, education level, ethnicity, smoking status, diagnosed diseases, medication usage and physical activity.

Physical activity: physical activity was measured using the International Physical Activity Questionnaire short version (IPAQ-SF) and the metabolic equivalent (MET) minutes per week were calculated for each participant and categorised: high (at least 60min/day of at least moderate-intensity activity), medium (at least 30min/day of at least moderate-intensity activity), low (no activity is reported or less than medium category) (117). IPAQ-SF does not permit translation to an overall physical activity level (PAL) value, hence we assigned estimated PALs of 1.4, 1.6 and 1.8 to 'low', 'moderate' and 'high' IPAQ-SF MET classification respectively based on a previous published report (117).

2.1.4.3 Dietary data

Dietary intake was assessed with a 7-day food diary (estimated weight) using instructions from food portion photographs and common household measures as well as a general question sheet to help validate the recorded intake. The diary has been previously validated in a larger UK epidemiological study (118). The diaries were completed and returned via post or given to the clinic during the respondent's health screening. They were analysed using Dietplan 6 (Forestfield Software, West Sussex, United Kingdom), which used a nutrient bank based on McCance and Widdowson's 'The Composition of Foods' published by the Food Standard Agency (FSA) UK (sixth edition 2008) (119). All coders were trained by a research dietician/nutritionist in a study-specific operational manual protocol. Portion sizes were estimated based on the validated food diary's photographs and a food portion size database based on national publications FSA 'Food Portion Sizes' (Third edition 2002) (118).

2.2 Dietary assessment quality control

To ensure high data quality of the analysed diaries, a standardised operational manual protocol was developed specifically for this study. This protocol was aimed to limit coding errors associated with use of the Dietplan programme. The protocol also served as an additional training of inputter in conjunction with a training course in Dietplan and basic nutritional information (e.g. FSA food portion size, McCance and Widdowson's database).

Furthermore, to ensure regular quality control, checks and audit were performed on the diaries and inputters.

2.2.1 Operational manual and protocol

An operational manual and protocol was developed to standardise the methodology used for analysing the diaries (Appendix 3). This was adapted from similar studies' protocols from the International collaborative study of macronutrients, micronutrients and blood pressure (INTERMAP) (120) and the EPIC study (121).

The protocol contains all stages of dietary assessment:

1. Collection of food diaries
2. Registering and setting up the assessment in Dietplan
3. Assessing and coding the diaries - selecting food codes
 - 3.1 General rules
 - 3.2 Composite dishes
 - 3.3 Cooking methods
4. Calculating portion sizes
5. Error checking
6. Logging completed diaries

The operational manual and protocol were also used for standardised training provided to all data inputters.

Two separate food databases were created in addition to the operational manual and protocol:

(1) The food portion size database is aimed for standardisation of food portion size calculation. It is based on the national publication by FSA 'Food Portion Sizes' (Third edition 2002) (122) and information from the food manufacturers.

(2) The codebook is aimed for standardisation of recipes and food items, which are not available in the Dietplan database. All items and recipes added used food portions based on

FSA ‘Food Portion Sizes’ (Third edition 2002) (122) or matched to the food manufacturers’ information. The codebook is an evolving database containing >600 supplementary codes and coding rules following some of the principles of the codebook designed by Conway *et al* for use in INTERMAP study (120).

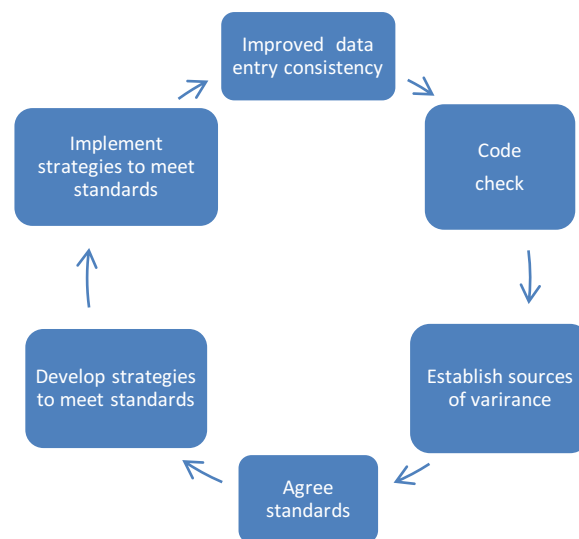
Furthermore, Mintel marketing reports for UK consumers were used to source out dietary information of the most consumed foods, e.g. cereal breakfasts, bread, fruit, vegetables, fat spread, dairy products.

A separate database is recorded for each participant with additional dietary questions to the food record such as supplements (type/dose), additional salt added to foods, cooking methods of meats, caffeinated/non-caffeinated tea/coffee and organic food products.

2.2.2 Dietary data audit

A random sample of 5% was checked every 10 weeks following a standardised audit procedure by the lead research dietitians/nutritionists (Figure 8). The sample was randomly selected using the random number generation in Microsoft Excel 2010 following protocol.

Figure 8: Schematic of quality control audit cycle



Code error check:

The diaries were assessed and scored based on type and number of coding errors using the following criteria:

1. Code selection error (food code selected in Dietplan did not match the written record)

2. Portion error (over +/- 10% difference of the protocol weight)
3. Meal code error (food item entered into incorrect meal occasion)
4. Added code error (extra food item coded which were not in the written record)
5. Missed code error (food item not coded i.e. were in the written record)

The researchers recorded the type of the errors, and missing item. Note missed food items were not included in the percentage error rate. Each diary was scored as a sum of these errors. Percentage error rate was calculated as the percentage of the total code errors possible in each diary (by multiplying the number of total food codes in the diary with the total number of possible errors i.e. 4): $(Total\ code\ errors / (total\ food\ codes\ in\ the\ diary \times 4)) \times 100\%$

Establish sources of variance:

Diaries with code error rate >10% was flagged up and investigated by the senior researcher dietitian to determine/classify type of error (knowledge gap of foods, protocol not clear, typing error, systematic error / random error). Feedback and training were then provided to the coder responsible for the diary entry.

None of the food diaries were amended to ensure that the error variance remained constant across the whole sample of diaries assessed. Correction of the 5% diaries could create a skewed dataset.

Develop strategies to meet standards:

Based upon the report of the quality checked diaries the following strategies were conducted:

- Ongoing feedback and training of the inputter team (or individual) to improve knowledge gaps and communication
- Ongoing amendments and update of the operational manual and protocol, the food portion book and codebook.

Checking gross errors and energy/macronutrient outliers

The diaries were then exported into Microsoft Excel and were screened for gross coding errors. A gross coding error is defined when the quantity of food recorded is a clear code error (123) e.g. 260 g of chocolate powder recorded rather than 260g chocolate powder made

with milk. If the quantity of a food item exceeded the set maximum portion, the diary would be flagged up and checked against the original diary and amended accordingly.

A final sample of 5,686 food diaries was further screened for extreme energy and macronutrient outliers. This was based on previously reported statistical cut-off methods checked against the mean intake in an adult UK population (61, 124). Two participants were excluded due to energy outliers (>5,000 kcal/day and <500 kcal/day), which were considered to be physiologically unsustainable. Lastly, the dietary data was checked for energy underreporting.

2.2.3 Underreporting of energy

The dietary data was checked for energy underreporting using the Goldberg individualised method (125). The principle of the Goldberg method is to identify under-reporters of energy intake based on ‘energy intake $\approx$ energy expenditure’, using the following equation for energy expenditure based on basal metabolic rate (BMR) and physical activity level (PAL) (125, 126): Energy expenditure = BMR x PAL

BMR is a measure of energy expenditure in a state of rest, and was estimated with the Schofield equation using age, sex, and body weight (23, 24, 63). PAL values were converted based on participants’ individual MET category. The Goldberg method (125, 126) applies cut-off values defined on confidence interval for each participant to determine whether the energy intake recorded is plausible (Table 8). Participants with a ratio of energy intake: BMR below the lower 95% confidence interval cut-off were classified as under-reporting and above cut-off as potentially over-reporting energy intake.

Under-reporters was not removed from the data set to avoid further distortion of the data (62). Instead the study conducted sensitivity analysis, to assess potential bias to the dietary analysis.

Table 8: Goldberg cut-offs to determine acceptable energy reporting (125, 126)

MET category	Estimated PAL assigned	Energy intake : BMR cut-off values				
			Lower		Upper	
Low	1.4	Under	0.95	Acceptable	2.05	Higher
Moderate	1.6	reporter	1.09	reporter	2.35	reporter
High	1.8		1.22		2.65	

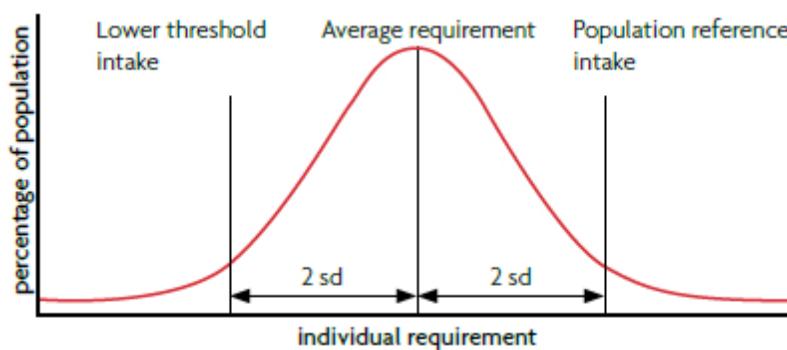
Abbreviations: MET, metabolic equivalents; PAL, physical activity level; BMR, basal metabolic rate.

2.3. Diet quality definition and measurement

2.3.1 Diet quality definition

Diet quality in this thesis is defined as adherence to the UK dietary guidelines for optimal health and prevention of CVD (26). These guidelines are based on the DRV published in the COMA reports “Dietary Reference Values for Food Energy and Nutrients” and “Nutritional Aspects of Cardiovascular Disease” (21), with the SACN updates (23-25). These reports are the most comprehensive public health reference guides produced on energy, nutrients and food groups in the UK (23-25). The reports include the health consequences of nutrient intakes and recognises the broad range of requirements for individuals within the UK adult population. The term DRV is a statistical concept based on the physiological requirements of a given nutrient for health aimed for a healthy adult UK population. It covers three terms: (1) Estimated Average Requirements (EAR); about half of a population group will need more than the EAR and half will need less: (2) Lower Reference Nutrient Intake; an intake of two standard deviation (SD) below the EAR and sufficient only for a small number of people (about 2.5%) in a group who have low needs: (3) Reference Nutrient Intake (an intake of two SD above EAR: this is enough or more than enough for at least 97.5% of people in a group, including those with high requirements).

Figure 9: Illustration of UK Dietary Reference Values definition adapted from Committee on Medical Aspects on Food Policy report (23)



Public Health England (Levy et al 2017) reviewed the evidence of the DRV from the above reports and a number of RCTs in relation to prevention of CVD in the UK population (26). The review highlights a recommended daily intake of six nutrients: total carbohydrate, fibre, sodium, added sugar, total fat, saturated fat and food groups: fruit/vegetables (Table 9). The recommendations are estimated for a healthy adult British population with a mean energy

(excluding energy from alcohol) intake 2,000 kcal for women and 2,500 kcal for men (taking into account socio-cultural and demographic factors).

Table 9: UK Dietary guidelines important for cardiovascular health (26)

Nutrient	Daily dietary goal
Total fat	<35 % of total energy intake
Saturated fat	<11 % of total energy intake
Carbohydrates	>50 % of total energy intake
Fruit & vegetables	>400 g/day
Fish (oily)	> 2 portion/week
Fibre (AOAC)	>30 g/day
Added sugar	<5 % of total energy intake
Sodium	<2400 mg/day

Abbreviations: AOAC, American Association of Analytical Chemists

Study specific amendments were applied to the current DRV due to data availability from the dietary analysis in DietPlan. DietPlan's nutrient bank is based on McCance and Widdowson's 'The Composition of Foods', which is not updated on AOAC fibre or added sugar. Therefore, NSP fibre was applied instead of American Association of Analytical Chemists (AOAC) and NMES was applied instead of added sugar. The NMES include free sugar not bound in foods, e.g. table sugar, honey and sugars in fruit juices excluding milk sugar (lactose). The only difference between NMES and free sugars is that NMES includes 50% of the fruit sugars from stewed, dried or canned fruit but free sugars includes none. For the purpose of analysis, NMES variable was created based on the nutrient database food items. Furthermore, for the purpose of this study, fish intake was assessed separately (not included in diet score) as the recommendation is based on a weekly intake rather than daily reference values as the rest of the nutrients and food groups. Rationale for this exclusion was based on not all 7-day food diaries being complete (participants with >4 days were included in the analysis).

2.3.2 Diet quality measurement

Diet quality of participants of the AIRWAVE Monitoring study was measured based on adherence to the UK dietary guidelines using two different methods: the FSA NP model (a

food-based quality score) and DRV index (overall diet quality score). The UK NP model has never been assessed in relation to adherence to the UK dietary guidelines and for the purpose of this analysis an overall diet quality DRV index was designed. Furthermore, both diet scores accuracy in capturing “healthy” foods were assessed against intake of other nutritional variables and food groups based on NDNS definitions.

1) The FSA UK NP model – healthiness of foods

The established UK NP model has been previously validated to measure healthiness of individual foods in several large UK studies including NDNS and the Whitehall Cohort (22, 51). The construction of the NP score has been described in detail (22, 52, 56) (Appendix 3). Briefly, foods and drinks (except alcohol) score points based on their content of negative nutrients: energy, saturated fat, total sugar and sodium and positive nutrients: fruit, vegetables and nut content, fibre and protein. The nutrient thresholds are derived from the UK DRV (53, 54). One point for each nutrient corresponds to 3.75% of the DRV. Each food item is given an individual score (per 100g) which is then energy adjusted (nutrient density) using previously published methods (22, 52, 55):

Each food item individual score (per 100g) $\times$ energy from the food item / TEI

A total NP score (energy adjusted) is added up for all the foods and drinks NP (energy adjusted) to provide a daily average score:

Total NP score = Σ NP (energy adjusted) scores / number of days in the diary

An additional algorithm (52) is applied to the daily average NP score to scale it from 1 to 100 points: *Scaled NP score = $(-2) \times$ Total NP score + 70*

The points are summed to calculate an overall score between 1-100 points, with a higher score indicating a diet high in food quality (“healthy” food and drinks).

2) The DRV index – healthiness of overall diet

The 14-point DRV index is an a priori score reflecting adherence to DRV (23) and UK guidelines for optimal health and prevention of CVD highlighted by Public Health England (26) (Table 9). Due to the limitations of the UK nutrient database, the DRV for non-starch polysaccharide (NSP) fibre was used rather than Association of Official Analytical Chemists

(AOAC) and added sugar replaced with non-milk extrinsic sugar (NMES). Fish intake was excluded from the DRV index based on this criterion reflecting a weekly intake rather than daily intake making it difficult to apply to diet assessments less than 7 days. The DRV index includes seven nutritional components, six nutrients: total fat, saturated fat, NMES, sodium, total carbohydrates, NSP fibre and one food group: fruit/vegetables. The construction of the DRV index was based on a previously reported method (22, 52, 127). The mean daily intakes of the seven components were assessed according to the UK DRV and scored accordingly; 1 point represents +/- two standard deviation of DRV criteria, 0 point if intake was worse than DRV criteria and 2 points if intake was better than DRV criteria (Table 10). The points are summed to calculate an overall score between 0-14 points, with a higher score indicating a more favourable diet. The score includes all foods and drinks except alcohol intake, which was analysed separately and adjusted for with other known confounders.

Table 10: The DRV index scoring system based on the UK recommended daily mean intake for optimal cardiovascular health with study adaptations (26)

Points	Fat (%) [*]	Sat fat (%) [*]	CHO (%) [*]	Sugar ^δ (%) [*]	F&V (g/day)	Fibre ^δ (g/day)	Sodium (mg/day)
0	>36.75	>11.55	<47.5	>11.55	<380	<17.10	>2520
1	<36.75	<11.55	>47.5	<11.55	>380	>17.10	<2520
	>33.25	>10.45	<52.5	>10.45	<420	<18.9	>2280
2	<33.25	<10.45	>52.5	<10.45	>420	>18.9	<2280

Abbreviations: Sat fat, saturated fat; CHO, carbohydrates; F&V, fruit and vegetables

0 point represent a worse intake than DRV criteria; 1 point represent a nutrient intake within +/- two SD of DRV criteria; 2 points represent better intake than DRV criteria

* % daily mean energy intake minus energy from alcohol

^δ Study adaptation to Levy et al values, instead of added sugar applied previous DRV for non-milk extrinsic sugar and instead of AOAC fibre applied previous DRV for non-polysaccharides fibre

Both the DRV index and NP model were assessed against the validated dietary score DASH in this study (128, 129). DASH is a dietary pattern promoted by the US National Heart, Lung, and Blood institute to prevent and control hypertension and has been previously validated in association with cardio-metabolic risk in a British population (129). The score includes nine food groups: fruit, vegetables, grains, meat, poultry, dairy, oils, added sugar. The validation and construction of the DASH score has been described in detail elsewhere (128, 129). In short, a higher DASH score indicates a better overall diet quality (0-100 points).

2.4 Clinical definitions

Cardio-metabolic risk factors and outcomes were defined by the unified international criteria for MetS (Table 11) previously described with study adaptations (1). Elevated blood pressure was defined as SBP ≥ 130 mmHg and DBP ≥ 85 mmHg or on anti-hypertensive medication. Low serum HDL cholesterol men ≤ 1.0 mmol/L, women ≤ 1.3 mmol/L or on lipid-lowering medication. Triglyceride was excluded (not available in the study). Elevated blood sugar was defined as Hb1Ac $\geq 5.7\%$ and/or on glucose-controlling medication (87). Abdominal obesity was defined on European population waist circumference men > 94 cm, women > 80 cm (1). BMI was defined as weight divided by the square of height in meters (kg/m^2). BMI categories as underweight ($\text{BMI} < 18.5$), normal (18.5-24.9), overweight (25-29.9) and obese (≥ 30) (98).

Table 11: Metabolic syndrome risk factors definition used in this study

Risk factor	Criteria
Elevated blood sugar	Hb1Ac $\geq 5.7\%$ or receiving drug therapy for hyperglycaemia
Elevated BP	$\geq 130/85$ mmHg or receiving drug therapy for hypertension
Low HDL-cholesterol	≤ 1.0 mmol/L in men or ≤ 1.3 mmol/L in women or receiving drug therapy for cholesterol lowering
Abdominal obesity	Waist circumference men > 94 cm, women > 80 cm

Abbreviations: HbA1c, glycated haemoglobin; BP, blood pressure; HDL, high density-lipoprotein

CVD and T2D outcomes were defined as per The National Institute for Health and Care Excellence (NICE) guidelines. Dyslipidaemia was defined by total cholesterol-HDL ratio cut-off > 4.5 mmol/L, diagnosed or on cholesterol-lowering medication (130). Hypertension was defined as DBP ≥ 90 mmHg and SBP ≥ 140 mmHg, diagnosed or on anti-hypertensive medication. T2D was defined as HbA1c $\geq 6.5\%$, diagnosed or on glucose-controlling medication (83, 84). Pre-diabetes is a metabolic condition defined as borderline diabetes and diagnosed by presence of blood glucose levels higher than normal. The cut-off values are set by the American Diabetes Association as HbA1c levels between 5.7-6.4% (85).

Table 12: CVD and diabetes definition used in this study

Risk factor	Criteria
Dyslipidaemia	Total cholesterol-HDL ratio cut-off > 4.5mmol/L or receiving drug therapy for cholesterol lowering (130).
Hypertension	Blood pressure SBP/DBP ≥140/90 mmHg or receiving drug therapy for hypertension (130).
T2D	Hb1Ac ≥ 6.5 % or receiving drug therapy for hyperglycaemia (85)
Pre-diabetes	Hb1Ac 5.7-6.4 % (85)

Abbreviations: HbA1c, glycated haemoglobin; BP, blood pressure; HDL, high density lipoprotein

2.5 Genetics

2.5.1. Selection of genetic markers

A combined systematic literature review on published genome wide association studies (GWAS) was conducted using four search databases (1) PubMed, (2) DIABetes Genetics Replication And Meta-analysis consortium (DIAGRAM), (3) Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) and (4) National Human Genome Institute Research (NHGIR) GWAS database. In-depth description of database and search criteria is provided in Appendix 4. The review aimed to find published GWAS on common SNPs associated phenotype T2D and glycaemic trait HbA1c.

Figure 10 shows the process of the systematic literature review. In stage 1 each database was searched systematically using search key words defined in table 13. The output of the search produced 180 papers, which were reviewed based on a set of selected inclusion/exclusion criteria (table 14) and reduced to 37 (20 duplicated) papers. From the reviewed papers a list of 161 SNPs represented different loci found to have a genome-wide statistical significance level ($p < 5 \times 10^{-8}$) associated with hyperglycaemia trait (fasting glucose, fasting insulin, HbA1c and 2 hours glucose testing) or phenotype T2D. The generated list of 161 SNPs included mapped genes, risk alleles, chromosome position, coding region, original cohort effect size and p-values and was inputted into Microsoft Excel for further data cleaning (Flowchart for exclusion and inclusion of SNPs is in Appendix 5).

Figure 10: Systematic literature review flowchart

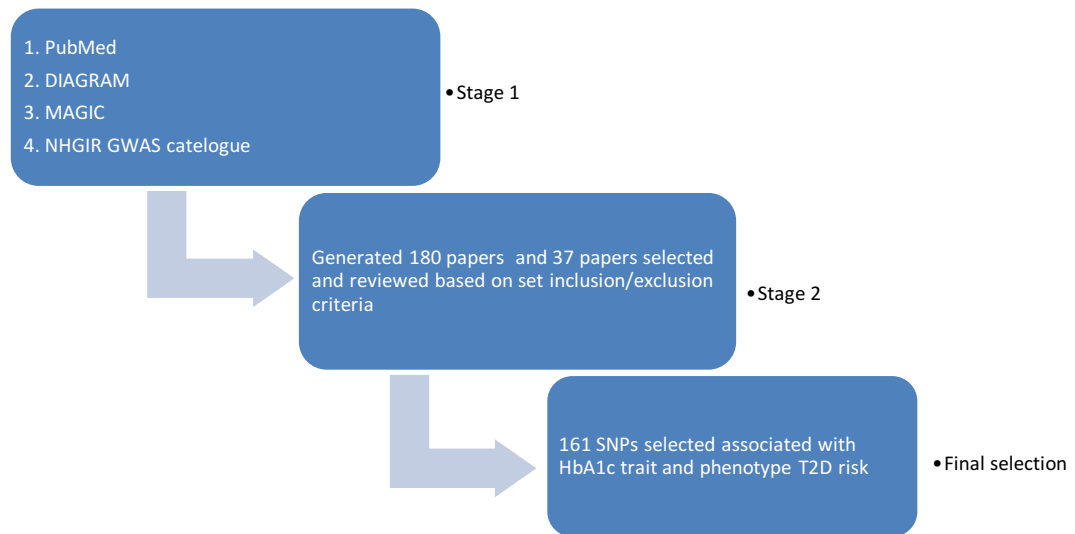


Table 13: Search key words for systematic literature review

Phenotype	Phenotypic trait	Genome wide association studies
“T2D” OR “type 2 diabetes”	“HbA1c” OR “glycated haemoglobin A1c”	“genome-wide" OR "genome AND identification" OR "genome AND association”

Abbreviations: T2D, type 2 diabetes; HbA1c, glycated haemoglobin A1c

Table 14: Search criteria systematic literature review

Inclusion criteria	<p>Papers published in English</p> <p>Original research articles</p> <p>Human studies: Caucasian population (European ancestry)</p> <p>GWAS papers including replicated studies and/or meta-analysis</p> <p><u>Additional SNP criteria</u></p> <p>Known and common disease-associated SNPs of functional significance to phenotype T2D and glycaemic trait HbA1c</p>
--------------------	--

	<p>SNPs with a genome-wide statistical significance level ($p < 5 \times 10^{-8}$)</p> <p>SNPs located inside a known gene: regulatory or coding region</p>
Exclusion criteria	<p>Publication dates > 5 years</p> <p>Non-European population groups or not stated</p> <p>Diabetes type 1 and gestational diabetes</p> <p>Novel and rare SNPs</p> <p>SNPs located in known obesity genes</p>

Abbreviations: SNP, single nucleotide polymorphism

2.5.2 Systematic literature review data management and cleaning

The list was further quality checked to avoid any overlapping with obesity SNPs and LD pairwise with other SNPs coding regions (e.g. coding for the same outcome), which may cause over emphasising SNPs effect. This was done to ensure only independent SNPs and SNPs with the highest LD, strongest association with phenotype remained (Appendix 5).

SNPs' coding regions were checked using UCSC Genome Browser on Human version GRCh38/hg38 (date Dec. 2013) (131). SNPs with overlapping regions were highlighted and their pairwise LD was calculated. Pairwise LD calculations were done in an online genetic variant-centered annotation database browser, SNIPA version 3.1 (released November 2015)(132). SNIPA estimates how often the SNPs appear in pairwise LD (distant limit set at 500 kb) tested against a reference population panel from the 1,000 Genomes (pilot), 1,000 Genome Project, Phase 3 v 5 Linkage disequilibrium data (sample of Utah residents with ancestry from northern and western Europe). r^2 threshold between the SNPs was determined as high correlation if $r^2 > 0.8$ and D-prime threshold was set at >0.8 , which determine if SNPs LD are co-inherited $>80\%$ of the time (92, 133). Those with the lowest r^2 value were excluded to ensure the SNPs with the strongest reported correlation with phenotype were used. Further LD pruning was conducted to SNPs on the same chromosome (not in same coding regions) to determine if they were in close LD with each other. The check excluded 64 SNPs. Additional 6 SNPs (rs9939609, rs1421085, rs8090011, rs8050136, rs11642841, rs12970134) were positioned within obesity-associated genes FTO and MC4R and were

therefore excluded given the well-documented primary association with BMI, which mediates the effect on T2D risk (92, 97, 134).

Protocol for linkage disequilibrium pruning of genetic markers

1. Removed SNPs within the same coding region. Checked SNPs coding region on database UCSC Genome Browser
2. Removed SNPs in close LD. Calculated LD using database SNIPA

Input SNP's rs-ID (text entry) and set the database search criteria:

- a. Distance limit 500 kb
- b. Population panel: CEU (European ancestry)
- c. Correlation between the SNPs r^2 threshold > 0.8 and D-prime > 0.8
- d. 1000 Genomes (pilot) variant dataset was used for computation of allele frequencies for the selected variant set. This was double-checked with other variant datasets 1,000 Genome Project and Phase 3 v 5 Linkage disequilibrium.

The remaining 91 SNPs were checked against the SNPs in the Airwave Health Monitoring study's genetic dataset of which the majority were found genotyped, with exception of 1 SNPs rs11603334 not genotyped. A further 3 SNPs were excluded due to chromosome position: rs5945326 positioned in x chromosome X:152899922 and no SNPs corresponded with 3:49455330, 1:158585415.

The final list of 87 SNPs consisted of 44% (n=38) HbA1c or other glycaemic trait associated SNPs and 56% (n=49) T2D associated SNPs.

The list was exported into a PED file and merged together with the Airwave Health Monitoring study's genotyping dataset and further quality controlled. All quality control was performed in PLINK version 1.07 (10. Oct. 2010) (135). PLINK is an open-source whole genome association analysis programme designed by MIT/Harvard University to perform a range of quality control and large-scale analyses.

2.5.3. Data quality control

Study design and technical errors in genotyping may introduce systematic bias, leading to an increase in the number of false-positive and false-negative associations. To help avoid or limit many such errors, a thorough quality control is conducted additionally to the genotyping control. All genetic data is cleaned and analysed in PLINK based on published protocol (105, 133) and based on previous published papers quality control (133, 136).

Minor Allele Frequencies (MAF)

SNPs were tested for MAF. MAF is the frequency at which the minor allele (or less abundant) of a SNP is present in a population based on comparisons of estimated frequencies from a sample population e.g. The International HapMap Project (applied in PLINK). The MAF for a SNP to be considered common is usually $>1\%$ (66, 92). All SNPs passed the MAF frequency test $p < 0.01$.

Hardy-Weinberg equilibrium

All SNPs were tested for the Hardy-Weinberg equilibrium, which detects if a genotype can be due to a sequencing error or batch effect. It tests the frequency of different alleles and genotypes in a population, which are in genetic equilibrium. Genetic equilibrium is defined as remain constant from generation to generation in the absence of evolutionary influence e.g. mutations. The null hypothesis test if the sample is the same as expected i.e. the probability of the population being in Hardy-Weinberg equilibrium at that genotype. SNPs > 0.001 will be rejected and removed from the dataset (133). 1 SNP failed the Hardy-Weinberg equilibrium test and was excluded from this study.

Missingness test

Missingness rates per SNP were checked to exclude SNPs with more than 1% missing genotype data. None of the selected SNPs failed the missingness test criteria threshold < 0.01 (133).

Missingness rates per person were also checked to only include participants with over 99% of the tested SNPs. The total genotyping rate was 0.98 resulting in 282 participants being removed due to failing the criteria $p < 0.01$ (133).

Other quality control tests

Population stratification (controlling for confounding by ethnicity and ancestry) was not applicable in this quality control as the dataset only consists of white British population of European ancestry. Hence, assumptions were made that minimum variation or genetically distinct differences existed within subgroups of this study population (135). Furthermore, the SNPs applied were extracted from replicated GWAS datasets where population stratification was applied in their quality control.

Heterozygosity (controlling for proportion of heterozygosity within the SNPs) was not applicable either in this dataset as it was assumed that the structure of the sample population consisted of individuals from different families, i.e. no first-degree relatives. The individuals in the Airwave Health Monitoring study population were expected to have a low proportion of heterozygosity within the markers (no outlier > 3SD from the mean). Heterozygosity is sensitive to population structure because allele frequency distribution (*heterozygosity*) can differ between different populations (105, 133).

In summary, the quality control of the genetic data excluded 284 participants based on genotyping rate < 98% and 1 SNP was removed due to failed the Hardy-Weinberg equilibrium test > 0.001. The remaining sample consisted of 14,089 individuals with clinical data of which 3,733 also had dietary data. The final list of 87 SNPs, which passed the quality checked was exported and prepared for GRS calculations. The list included SNPs' rs ID, risk alleles and chromosome position (Appendix 6).

2.5.4 Polygenetic risk score modelling

Due to the small effect size per SNP on phenotypes T2D and HbA1c trait a combined GRS was generated based on previously reported methods (69, 137-140). Participants were scored and grouped into a GRS based on the number of risk alleles present (un-weighted score) from all 87 SNPs. The SNPs represent different loci associated with hyperglycaemia trait (fasting glucose, fasting insulin, HbA1c and 2-hours glucose testing) to phenotypes impaired glucose tolerance and T2D. The GRS was constructed in PLINK for each participant by summing the number of risk alleles present: 0 (non-carrier), 1 (heterozygous), 2 (homozygous). The sum was then divided by the number of effectively genotyped SNPs creating an average ratio:

$$GRS = \sum \text{risk alleles} / \sum \text{genotyped SNPs}$$

Appropriate scaling of the GRS was required for clinical interpretation (e.g. logistic regression analysis)(141). The GRS was standardised against the sample population's SD of the GRS using a previously reported method (140, 142). The standardised (centred) score was constructed using the following equation:

$$\text{Standardised (centred) score} = \text{GRS} / \text{GRS SD}$$

PLINK processed the inputted BED file from the genotyped data provided and analysed the entire sample (Appendix 2). A final output file was produced and exported into Microsoft excel (csv file) and imported into SAS for further GWAS with phenotypes and interaction studies.

2.6 Statistical analysis

SAS version 9.3 (SAS Institute Inc. VX, Cary, NC, USA) was used for all analyses and statistical significance was set at $p < 0.05$.

All continuous variable (dietary, genetics, anthropometric and clinical) distributions were assessed for normality. Alcohol was highly skewed and therefore categorised based on UK latest recommended allowances 16 g alcohol (2 unit) per day for women and men (category 1: no alcohol consumption, category 2: below or within the UK recommended alcohol allowance, category 3: above the UK recommended alcohol allowance) (27). Other categorical variables included gender (male, female), smoking (never, current, former), and physical activity (MET categories high, medium, low).

Possible confounders adjusted for in this thesis' analyses were selected on the basis of similar studies from the literature review in the introduction (7, 19, 20, 35). As described in the introduction, CVD and diabetes development is multi-factorial. The pathways between the metabolic risk factors and diet quality are associated with a number of known contributing factors and possible confounders (Figure 1). The identified confounders from the literature review include: age, sex, BMI, physical inactivity, smoking, alcohol and outcome specific treatment which is adjusted for in this thesis analyses. Socio-economic status and ethnicity, which are also known to be confounding factors in dietary habits and health outcomes were not included in this analysis as the study population in this sample was homogenous in ethnicity ~ 95% white British and a descriptive summary of the sample's educational levels

and household income showed there was not a wide diversity in participants' socio-economic status.

2.6.1 Descriptive analysis

Baseline socio-demographic and lifestyle characteristics of participants were compared between gender of study sample using Student t-tests and chi-square as appropriate. P-values for Student t-test (continuous variables) or Mantel-Haenzel chi-square test (categorical variables) were reported.

2.6.2 Correlation tables

Correlations between specific variables were explored via partial Spearman correlation models. Multi-collinearity between variables was assessed by multivariate analysis. This may course bias to the results via overestimation of standard errors of the coefficients.

2.6.3 Diet quality analysis

DRV index was analysed as quartiles. Quartiles were defined based on a statistical distribution. Stratified analysis by gender was examined for all analyses and found no differences in observed associations between DRV index, mean dietary intake (single nutrients and food groups) and metabolic markers. To obtain better statistical power in the analyses, men and women were combined in the study.

Socio-demographic and lifestyle characteristics of participants were compared between DRV index quartiles of study sample using general linear models (GLM) adjusted for sex and mean energy intake. P for trend was reported.

2.6.3.1 Diet quality scores correlation with other nutritional variables

Correlations with other nutritional components and between diet scores were conducted using GLM adjusted for sex and mean energy intake, testing linear trend across DRV quartiles. Analysis of variance (ANOVA) was used to compare mean nutrient, food group intake across DRV quartiles. The F-test in the ANOVA tested the significance of inter-group differences. Cross-adjusted sums of squares (SS) (Type III SS in SAS) measured the effect of DRV index adjusted for covariates in the model. P for trend was reported.

2.6.3.2 Diet quality scores association with metabolic markers

Diet quality scores association with metabolic markers: HbA1c, total cholesterol, HDL-cholesterol, BP and adiposity markers: waist circumference, BMI were analysed as continuous variables via multivariate GLM.

ANOVA was used to compare means of HbA1c, total cholesterol, HDL-cholesterol, BP, waist circumference and BMI across DRV quartiles.

All analyses were adjusted for covariates: age, gender, BMI (replaced with height in analysis with adiposity markers), mean energy intake, mean alcohol intake, smoking, physical activity, dependent variable specific diagnosis and treatments.

To analyse which of the nutrients within the DRV index were driving associations with metabolic markers, the analysis was repeated with the selected single nutrients mean intake: sodium, saturated fat, fibre, sugar, fruit and vegetables and energy intake.

2.6.3.3 Diet quality scores association with metabolic outcomes

Diet quality scores association with metabolic outcomes (MetS classification) and CVD were analysed via logistic regression models. All analyses were adjusted for covariates: age, gender, BMI (replaced with height in analysis with adiposity markers), mean energy intake, mean alcohol intake, smoking, physical activity.

2.6.3.4 Sensitivity analysis

Sensitivity analysis was tested on DRV index association with metabolic risk excluding participants on weight loss diets (n=427) and energy under-reporters (n=2,720 ~50%). Furthermore, sensitivity analysis was tested on healthy population excluding participants (n=335) self-reporting previous/current diagnosis: cancer, liver disease, thyroid, stroke, angina, heart attack and other heart diseases.

2.6.4 Genetic risk score analysis

The GRS was analysed as both a continuous variable and as quartiles. Quartiles were defined based on a statistical distribution. The GRS did not differ across gender, hence to obtain better statistical power in the analyses, men and women were combined in the study.

Socio-demographic and lifestyle characteristics of participants were compared between GRS quartiles of study sample using GLM (unadjusted). P for trend across GRS quartiles was reported.

2.6.4.1 Genetic risk score association with glycaemic trait

GRS association with phenotypic trait HbA1c % were analysed as continuous variables via multivariate GLM adjusted for covariate variables age, sex and BMI, smoking, diabetic or on drug treatment. ANOVA was also used to compare means (Bonferroni adjusted) of HbA1c % across GRS quartiles. P for trend across GRS quartiles was reported.

2.6.4.2 Genetic risk score association with phenotype outcomes

GRS association with phenotype T2D and pre-diabetes were analysed via logistic regression models, adjusted for covariate variables: age, sex, BMI, diabetes diagnosis and treatment. Diabetes diagnosis and treatment was not adjusted for in models with T2D, as it was included in their outcome definitions.

2.6.4.3 Genetic risk score interaction and effect modification studies

The interaction analyses were explored both as a continuous GRS and across its quartiles to test differences in interactions between genetic susceptible groups and its impact on HbA1c. The majority of the interaction analyses were examined across GRS quartiles as the presence of interaction may depend on the risk difference scale. Interaction can be present on one scale and absent on another (143). Definitions for interaction and effect modification terms were applied based on published papers.

Interaction: "A statistical significant difference of the observed joint effects (crossover interaction term) of the exposure (x) and third variable (z) differs from that expected on the basis of their independent effects"(143).

Effect modification: "A statistical association between an exposure (y) and a health outcome (x) differ according to a specific scale the levels of a third variable (z) – the effect modifier"(144).

Interactions and effect modification analyses were explored via GLM and GENMOD models based on SAS Knowledge Base coding instructions (144, 145) and bio-statistical lectures from McGills (146) and Penn State University (147).

2.6.4.5 Diet quality interactions on phenotypic trait HbA1c

The diet and GRS interactions' effect on HbA1c% were explored across GRS quartiles to test for risk difference scales and magnitude interactions effect across and within quartiles. Interactions were tested separately with various diet quality variables (continuous): DRV index and single dietary variables (mean intake) sodium, total fat, saturated fat, fibre, sugar, whole grains, fruit/ vegetables and energy intake.

Their interaction effects on HbA1c were analysed via multivariate GLM:

$$HbA1c = GRS_{quartiles} + interaction\ variable + (GRS_{quartiles} * interaction\ variable) + covariate$$

The parameter estimate of the interaction term (GRS*interacting variable) reported was estimates for the slope of each $GRS_{quartiles}$. Note GLM's automatically use an intercept and parameters the last category (e.g. GRS quartile 4) as nil. Hence all models were coded using “no intercept solution” to remove intercept and re-parameterising estimates for each GRS quartile. P-values for interaction reported were the interaction term's $P > |t|$ within quartiles and p-value for linear trend across GRS quartiles.

P-values were not corrected for multiple comparisons test such as Bonferroni correction as only one main effect (one interaction term) was tested per analysis. The same method was applied as in similar diet-gene interaction studies (73, 76, 148, 149).

All analyses were adjusted for covariate variables: age, sex and BMI, smoking, mean energy intake, alcohol, physical activity, diabetic or on drug treatment. Whole grains analyses were additionally adjusted for overall diet quality (DRV index). These adjustments were based on previously published studies of similar interaction analyses (73, 76, 148, 149).

2.6.4.5 Alcohol interactions on phenotypic trait HbA1c

GRS continuous score interaction with alcohol consumption (categories) modification on HbA1c was analysed via multivariate GLM crossover included interaction term. This allowed measuring the risk difference scales and magnitude effects of alcohol categories.

$$HbA1c = GRS + alcohol_{categories} + (GRS * alcohol_{categories}) + covariates$$

The parameter estimate of the interaction term ($GRS * alcohol_{categories}$) reported was based on the relationship or the parameter estimate of the slope of each $alcohol_{categories}$. P-values for interaction reported were the interaction term's $Pr > |t|$ within category and p-value for linear trend across categories.

All analyses were adjusted for covariate variables: age, sex and BMI, smoking, mean energy intake, physical activity, diabetic or on drug treatment.

2.6.4.6 BMI interactions on phenotypic trait HbA1c

GRS interaction with BMI (categories) modification on HbA1c was analysed via multivariate GLM crossover included interaction term. This allowed measuring the risk difference scales and magnitude effects of BMI.

$$HbA1c = GRS + BMI_{categories} + (GRS * BMI_{categories}) + covariates$$

The parameter estimate of the interaction term ($GRS * BMI_{categories}$) reported was based on the relationship or the parameter estimate of the slope of each interacting $BMI_{categories}$. P-values for interaction reported were the interaction term's $Pr > |t|$ within category and p-value for linear trend across categories. All analyses were adjusted for covariate variables: age, sex and alcohol smoking, mean energy intake, physical activity, diabetic or on drug treatment.

2.6.4.7 GRS and diet quality interaction and effect modification on T2D risk outcomes

Effect modifications of the interactions terms ($GRS * DRV$ index) on T2D risk outcomes were explored as continuous variables nested within the modifying variable (DRV index). The interaction effect of diet quality (DRV index) was analysed via a nesting approach where the predictor (GRS) random effect is nested within fixed level of the modifier (DRV index) using a GENMOD model (144). GENMOD contrast estimate results create parameter estimates (OR) of the two variables and also form a regression model that is nested within the model as an independent variable - their interaction (cross-over term). The models provide a likelihood ratio test of the interaction term (contrast L'Beta Estimate) including OR and 95% CI for both GRS (without interaction term) and interaction term $GRS * DRV$ score. $Pr > Chi-Square$ was reported.

CHAPTER 3: RESULTS DIET QUALITY AND CARDIO-METABOLIC RISK

The first objective of this study was to explore the relationship between diet quality (using 7-day food diary) and cardio-metabolic risk on 5,848 participants from the Airwave Heath Monitoring study. Two different diet quality scores were developed according to the UK national dietary guidelines.

3.1 Descriptive statistics

3.1.1 Demographic and lifestyle characteristics of the sample

Descriptive summary of 5,848 participants (2,352 women and 3,496 men) is presented by gender in Table 15. The study population average age was 41 years. Ethnicity was predominately British white (95.26%). The sample represented a wide range of different education levels from GCSE (30%), A levels (32%), to bachelor degrees (20%), and only 4% of the participants had no formal qualifications. Prevalence of obesity was 20% and almost half of the sample was overweight (48%). There were significant differences between genders across all lifestyle characteristics such as alcohol consumption and physical activity. Women had a lower prevalence of adiposity markers, obesity and waist circumference, compared to men. Prevalence of cardiovascular diseases was also lower in women. Hypertension was 39% in men and 16% in women and dyslipidaemia was 37% in men and 11% in women. T2D prevalence was 4% for both men and women. Gender differences in medication usage were reflected in the above figures. Use of BP lowering drugs was 5% in men and 3% in women and lipid-controlling drugs was 4% in men and 1% in women. No differences were seen in glucose-controlling drugs, both men and women were >1%.

Table 15: Demographic and lifestyle characteristics presented by sex of the Airwave Health Monitoring Study (N=5,848)

Variable	Women ^a	Men ^a	P-value ^b	All ^a
N	2,352	3,496		5,848
Sex (women) %				40.00
Age (years)	39.81 (9.56)	42.56 (8.91)	<0.0001	41.19
Ethnicity (British white) % *	96.21	94.61	0.01	95.26
Education level %			0.0003	
Post graduate	7.06	5.84		6.33
Bachelor degree	21.39	19.17		20.06
A-level	32.48	32.25		32.34
Vocational	7.36	7.24		7.29
Variable	Women ^a	Men ^a	P-value ^b	All ^a
GCSE/ O level	28.19	30.79		29.74
No formal qualification	3.53	4.72		4.24
Alcohol g/day	10.61 (12.59)	17.10 (17.42)	<0.0001	14.51
No alcohol %	25.3	17.68	<0.0001	20.75
Within guidelines %	50.47	41.42		45.06
Above guidelines %	24.23	40.90		34.20
Cigarette smoking %			0.002	
Never	67.86	70.14		69.22
Former	22.24	22.91		22.64
Current	9.91	6.95		8.14
Physical active %			<0.0001	
Low (<600min/week) %	12.76	8.98		48.80
Moderate (>600min/week) %	43.15	39.04		40.70
High (>3000min/week) %	44.09	51.97		10.50
Demographic area %**			<0.0001	
North west England	19.52	21.28		20.57
Eastern England	36.69	29.09		32.15
London	2.51	4.03		3.42
Wales	11.44	11.24		11.32
Scotland	14.75	19.16		17.39
BMI (kg/m²)	25.91 (4.70)	27.80 (3.65)	<0.0001	27.04
Normal (18.5-24.99)%	49.83	21.6	<0.0001	32.95
Overweight (25-29.99)%	34.48	55.21		46.87
Obese (>30) %	15.69	23.2		20.18
Waist circumference (cm)	82.04 (11.35)	93.97 (9.81)	<0.0001	89.17
Cardiovascular diseases and T2D % ***				
Hypertension	16.02	39.25	<0.0001	40.33
Dyslipidaemia	11.05	37.40	<0.0001	35.81
Type 2 Diabetes	3.63	4.00	0.60	3.58
Medication usage %				
Blood pressure lowering	3.40	5.01	0.003	4.36
Lipid lowering	1.23	3.72	<0.0001	2.72
Glucose lowering	0.60	0.74	0.50	0.68

^a Presented as mean (standard deviation) or percentage

^b P-values for Student t-test (continuous variables) or Mantel-Haenzel chi-square test (categorical variables) comparing between differences between gender

* Minority ethnic groups not presented included: Asian, African, Chinese and mixed background

** Minor demographic areas (< 2%) were not presented: South East, South West, Midland West, non-regional

*** Self-reported diagnosed or on treatment

3.1.2 Prevalence of metabolic outcomes

Prevalence of metabolic health outcomes (based on MetS definitions) in the total sample was approximately 50% for elevated BP (Table 16). Men had over twice as high prevalence of elevated BP compared to women, 66% and 30% respectively. 50% of the total sample was abdominally obese and 41% have elevated blood sugar, which was predominately represented by women. No gender differences were seen in having low HDL-cholesterol and its prevalence was 15%.

Table 16: Prevalence of metabolic outcomes based on the international criteria for metabolic syndrome classifications of the Airwave Health Monitoring Study (N=5,848)

	Women		Men		P-value	All	
	n	%	n	%		n	%
<i>n</i>	2,352		3,496			5,848	
Elevated blood pressure ¹	703	29.89	2,312	66.13	<0.0001	3,015	51.56
Low HDL cholesterol ²	352	14.97	520	14.87	0.92	872	14.91
Abdominal obesity ³	1,231	52.34	1,703	48.71	0.007	2,934	50.17
Elevated blood sugar ⁴	1,045	44.43	1,363	38.99	<0.0001	2,408	41.18

¹ Diastolic blood pressure $\geq$ 80 mmHg and systolic blood pressure $\geq$ 130 mmHg or on treatment

² High density lipoprotein (HDL) cholesterol < 1.03 mmol/L for men, < 1.29 mmol/L for women or on treatment

³ Waist circumference > 95cm for men, > 80 cm for women

⁴ Glycated haemoglobin 1Ac > 5.7 % or on treatment

3.2 Dietary data and quality control

3.2.1 Code errors analysis

The results from the dietary data audit cycles from the food diaries produced a mean code error rate of 3.7% (SD 3.2%). The range of the mean error was 2.8% (SD 3.5%) to 5.9% (SD 4.1%). Only one diary (out of 5,849) was excluded due to exceeding the 10% code error limit. Analysis of the audit found that the most frequent coding error in the diary was entering incorrect portion (weight) of food item (55% of total errors). Second most frequent code error was wrong code for food item (31% of total code errors).

3.2.2 Energy underreporting

The Goldberg equation estimated 50% (n=2,720) of the sample under-report energy intake (Table 17). The analysis excluded participants on weight loss diet (n=427) and participants considered to be over-reporting their energy intakes (n=15). Lifestyle characteristics of under-reporters showed men were more likely to underreport (53%) compare to women (47%) (p<0.0001). Demographics of under-reporters also showed they had slightly lower education levels (less likely to be post graduates or had bachelors) compared to acceptable energy reporters. There was also a difference between acceptable energy reporters and under-reporters from the different ranks (p<0.0001); 37% acceptable energy reporters compared to 28% energy under-reporters were police staff and 54% acceptable energy reporters compared to 60% energy under-reporters were police constable or sergeant. Energy under-reporters had a higher mean BMI and waist circumference than acceptable energy reporters. Under-reporters also reported to exercise more and drink less alcohol than acceptable reporters (p<0.0001). No differences were seen in age or smoking.

Table 17: Energy under-reporters and acceptable reporters (defined by the Goldberg equation) and their associated lifestyle characteristics of the Airwave Health Monitoring Study (n=5,420)

	Acceptable reporters ^a	Under-reporters ^a	P-value ^b
n	2,701	2,719	
Sex (women) %	59.34	40.66	<0.0001
Age (years)	41.41	41.54	0.08
Education level %			0.01
Post graduate	6.92	5.45	
Bachelor degree	21.36	18.80	
A-level	31.84	32.49	
Vocational	7.29	7.47	
GCSE/ O level	28.92	31.13	
No formal qualification	3.67	4.67	
Demographic area %*			<0.0001
North west England	22.95	17.62	
Eastern England	34.95	29.09	
London	3.33	3.71	
Wales	10.22	12.43	
Scotland	17.92	17.36	
Rank %			<0.0001
Police staff	36.87	28.72	
Police constable/Sergeant	53.71	60.44	
Inspector/Chief inspector	5.85	7.21	
Superintendent or above	0.76	1.90	
Other	2.80	1.73	
Alcohol %			
No alcohol	22.60	17.59	<0.0001
Within guidelines	46.45	43.47	
Above guidelines	30.94	38.95	
Cigarette smoking %			0.57
Never	69.55	68.94	
Former	22.49	22.78	
Current	7.96	8.28	
Physical active (METs) %			<0.0001
Low (<600min/week)	16.24	5.18	
Moderate (>600min/week)	41.19	40.26	
High (>3000min/week)	42.57	54.57	
BMI (kg/m²)	27.82 (4.15)	25.96 (3.87)	<0.0001
Normal (18.5-24.99) %	42.89	23.71	<0.0001
Overweight (25-29.99) %	44.21	49.36	
Obese (>30) %	12.90	26.94	
Waist circumference (cm)	86.35 (11.17)	91.68 (12.10)	<0.0001

The Goldberg equation: energy intake=basal metabolic rate and physical activity level

^a Presented as mean (standard deviation) or percentage

^b P-values for Student t-test (continuous variables) or Mantel-Haenzel chi-square test

(categorical variables)

* Minor demographic areas < 2 % were not presented: South East, South West, Midland West, non-regional

3.2.4 Dietary data average daily intake

Table 18 shows comparisons of dietary intake across men and women (adjusted for energy). There were significant differences across all nutrient and food groups except sweet and sugary beverages ($p=0.76$). Overall women reported a higher intake of healthier nutrients (fibre, CHO) and food groups (fruit/vegetables, wholegrain, fish, oily fish and low fat dairy products) compared to men. Women also reported a lower intake of un-favourable nutrients (total fat, saturated fat, sodium and alcohol) and food groups (processed meat and red meat) compared to men. Total sugar (note include sugar from fruit and juices) and added sugar was the only un-favourable nutrients women reported having a higher intake than men. Mean energy intake was lower in women compared to men, 1,589 kcal/day and 1,940 kcal/day, respectively.

Table 18: Comparison of dietary intake across men and women of the Airwave Health Monitoring Study (n=5,848)

	Men n=3,496		Women n=2,352		P
Nutrients	Mean (SD)*				
Mean energy intake (kcal/day)	1,940	(457)	1,589	(370)	<0.0001
Alcohol (% EI)	5	(9)	3	(7)	<0.0001
Protein (% EI)	19	(6)	21	(6)	<0.0001
Total fat (% EI)	36	(6)	36	(6)	0.006
Saturated fat (% EI)	13	(3)	13	(3)	0.02
Carbohydrate (% EI)	49	(6)	50	(70)	0.05
Total sugar (% EI)	19	(6)	20	(6)	<0.0001
NME sugar (% EI)	15	(5)	16	(5)	0.01
Sodium (% EI)	1,583	(319)	1,557	(340)	0.0009
	Median (QR)†				
Food groups					
Fruit and vegetable (g/1000kcal)	152	(110)	186	(135)	<0.0001
Whole grains (g/1000kcal)	20	(34)	20	(30)	0.01
Fish (g/1000kcal)	9	(19)	10	(21)	0.006
Oily fish (g/1000kcal)	1	(11)	3	(13)	0.0006
Low fat dairy (g/1000kcal)	98	(88)	191	(95)	0.006
Red meat (g/1000kcal)	38	(28)	31	(30)	<0.0001
Processed meat (g/1000kcal)	19	(18)	14	(18)	<0.0001
Sweet and sugary beverages (g/1000kcal)	0	(43)	0	(46)	0.76

% EI: percentage of total energy intake; NME: Non-milk extrinsic sugars

* Student t-test compared mean (standard deviation) between gender

† Mann-Whitney U-test compared median (quartiles range) between gender

3.2.5 Correlation table between dietary variables and cardio-metabolic markers

Table 19 shows the Spearman rank correlations of dietary variables, which were incorporated in the two dietary scores DRV index and NP model. It shows carbohydrate intake had a (weak-moderate) positive correlation with NMES, fibre and fruit intake. Saturated fat had a negative correlation with carbohydrate and fibre. Vegetable intake was only correlated with fibre intake.

Table 20 shows the correlations between dietary variables (which were incorporated in the DRV index and NP model) and metabolic risk markers (adjusted for age and sex). Mean energy intake was negatively associated with all metabolic markers total cholesterol, SBP, DBP, BMI and waist circumference. Mean energy intake was only positively correlated with HbA1c. No significant correlation was seen with HDL-cholesterol. CHO was negatively correlated with all metabolic markers except HbA1c (non-significant). Similar correlations were seen for total sugar and added sugar as reported for CHO. Fibre and fruit/vegetable group show similar correlations. Both groups were negatively correlated with HbA1c, total cholesterol, waist circumference DBP and BMI. Sodium was positively correlated with SBP, DBP, BMI and waist circumference.

Table 19: Spearman rank correlation coefficients for dietary variables adjusted for sex, the Airwave Health Monitoring Study (n = 5,849)

	Mean energy intake	% TEI saturated fat	% TEI protein	% TEI carbohydrate	% TEI NME sugar	NSP fibre g/1000kcal	Sodium mg/1000kcal	Fruit g/1000kcal	Vegetables g/1000kcal
Mean energy intake									
% TEI saturated fat	0.20								
% TEI protein	*-0.32	-0.20							
% TEI carbohydrate	-0.07	*-0.35	-0.22						
% TEI NME sugar	0.20	0.01	*-0.43	*0.50					
NSP fibre g/1000kcal	-0.17	*-0.36	0.20	*0.37	-0.20				
Sodium mg/1000kcal	-0.20	0.02	0.23	-0.07	-0.24	0.08			
Fruit g/1000kcal	0.01	-0.26	0.00	*0.37	0.13	*0.46	-0.14		
Vegetables g/1000kcal	-0.17	-0.22	0.28	0.01	-0.18	*0.47	0.10	0.20	

Abbreviations: %TEI, percentage of total energy intake; NME: Non-milk extrinsic sugars; NSP: Non-starch polysaccharides

* $r \pm \geq 0.30$ and significant test p-value < 0.05

Table 20: Sex-age adjusted partial spearman correlation coefficients between dietary variables and cardio-metabolic markers, the Airwave Health Monitoring Study (n=5,848)

	HbA1c (%) ^a	HDL cholesterol (mmol/L) ^b	Total cholesterol (mmol/L) ^b	WC (cm)	DBP (Hgmm) ^c	SBP (Hgmm) ^c	BMI (kg/m ²)
Mean daily energy intake	0.03*	0.01	-0.04*	-0.05*	-0.08*	-0.04*	-0.06*
Carbohydrate %TEI	-0.01	-0.08*	-0.10*	-0.09*	-0.05*	-0.02*	-0.08*
Total sugar % TEI	-0.02	-0.05*	-0.09*	-0.14*	-0.09*	-0.03*	-0.12*
Total Fat % TEI	0.04*	0.08*	0.10*	0.06*	0.04*	-0.02	0.03
Saturated fat %TEI	0.04*	0.05*	0.12*	0.03*	0.02	-0.003	-0.001
Fibre g/1000kcal	-0.05*	0.02	-0.09*	-0.07*	-0.04*	0.002	-0.05*
Sodium mg/1000kcal	-0.03	0.01	0.01	0.10*	0.06*	0.03*	0.11*
Fruit, vegetable g/1000kcal	-0.05*	0.03*	-0.07*	-0.11*	-0.04*	0.02	-0.08*

Abbreviations: HbA1c, glycated haemoglobin; SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; HDL, high density lipoprotein; WC, Waist circumference; %TEI, percentage of total energy intake.

* p-value <0.05.

^a additionally controlled for self-reported diagnosis of diabetes and treatment.

^b additionally controlled for self-reported diagnosis of dyslipidaemia and treatment

^c additionally controlled for self-reported diagnosis of hypertension and treatment.

3.3 Diet quality analysis

3.3.1 DRV index score

The histogram of DRV index shows the score distributed slightly skewed to the left (Appendix 8). Test for normality (non-parametric) Kolmogorov-Smirnov D-value was 0.13 ($p < 0.01$). The DRV index score ranged from 0-14 and the sample's mean was 4.17 (SD 2.88) and median was 4.00. The DRV quartiles were based on a normal distribution of the score. Quartiles one (0-3.49) represented the population group with “less favourable” diets, quartile two (3.50-6.99), quartiles three (7.00-10.49) and quartiles four (10.50-14.00) represented the group with the “most favourable” diet.

3.3.2 Demographic and lifestyle characteristics across DRV index quartiles

Table 21 shows the descriptive statistics of the sample across the DRV quartiles. The characteristics of healthy eaters were more likely to be older and have a healthier lifestyle: drink less alcohol and exercise more. General health characteristics also showed that the healthier eaters had a lower BMI, waist circumference and prevalence of dyslipidaemia (unadjusted). No differences were observed across educational levels and DRV quartiles.

Table 21: Demographic and lifestyle characteristics across Dietary Reference Values index quartiles, the Airwave Health Monitoring Study (n=5,848)

Variable	Q1^a	Q2^a	Q3^a	Q4^a	P trend^b
Range	(0-3.49)	(3.50-6.99)	(7.00-10.49)	(10.50-14)	
	'least healthy'		'healthiest'		
n	1,075	1,669	1,800	1,304	
sex (female)	32.09	38.89	42.06	46.09	<0.0001
Age (years)	41.05	41.34	41.47	41.91	0.02
Education level %					0.17
Post graduate	3.35	4.38	4.61	4.29	
Bachelor degree	29.67	30.34	28.94	30.14	
A-level	8.09	8.51	6.78	5.75	
Vocational	34.23	32.13	33.39	29.60	
GCSE/ O level	18.05	19.18	20.00	22.93	
No formal qualification	6.60	5.46	6.28	7.29	
Diet quality					
UK Nutrient Profile	52.79	54.75	57.06	60.71	<0.0001
Alcohol (g/day)	15.83	16.71	13.75	11.69	<0.0001
No alcohol %	18.07	17.35	21.59	26.15	<0.0001
Within guidelines %	45.88	42.67	44.85	47.48	
Above guidelines %	36.05	39.98	33.55	26.38	
Cigarette smoking %					<0.0001
Never	68.56	65.97	70.56	72.09	
Former	21.02	24.63	21.61	21.61	
Current	10.42	9.41	7.83	7.83	
Physical active %					<0.0001
Low (<600min/week)	13.21	11.26	8.83	9.59	
Moderate (>600min/week)	41.49	40.92	42.94	36.66	
High (>3000min/week)	45.30	47.81	48.22	53.76	
BMI (kg/m ²)	27.36	27.16	27.02	26.65	<0.0001
Normal (18.5-24.99) %	21.77	21.33	20.61	36.07	0.0008
Overweight (25-29.99) %	48.00	47.39	45.28	47.37	
Obese (>30) %	30.23	34.11	20.53	16.56	
Waist circumference (cm)	90.80	89.81	88.93	87.34	<0.0001
Cardiovascular diseases %*					
Hypertension	31.07	32.29	31.50	30.67	0.68
Dyslipidaemia	26.88	26.42	24.28	23.21	0.02
Diabetes type II	3.91	4.49	3.00	3.30	0.10

Abbreviations: DRV, dietary reference values; Q1, first quartile; Q2, second quartile; Q3, third quartile; Q4, fourth quartile; BMI, body mass index

^a Values are means or prevalence – unadjusted

^b P-value linear model (continuous variable) Mantel-Haenzel chi-square test (categorical variables)

* self-reported or on specific health outcome treatment

3.3.3 Association between DRV index and nutritional components

Table 22 shows that across the DRV index quartiles there was a strong trend for a greater consumption of 19 other “favourable” nutritional components including both single nutrients and food groups, with exception of total sugar (which also include sugar from fruit and fruit juices). The DRV score also showed a significant trend with a higher UK nutrient profile score ($p < 0.0001$) and DASH score ($p < 0.0001$).

Table 22: Association between Dietary Reference Values index quartiles and mean intake of nutritional components (adjusted for sex and mean energy intake), the Airwave Health Monitoring Study (n=5,848)

DRV quartiles	Q1	Q2	Q3	Q4	
Range	(0-3.49)	(3.50-6.99)	(7.00-10.49)	(10.50-14)	
	'least healthy'			'healthiest'	
n	1,075	1,669	1,800	1,304	
	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)	P trend
Nutrients					
Energy intake (kcal/day) *	1,921 (12.9)	1,769 (10.3)	1,725 (9.9)	1,691 (11.6)	<0.0001
Alcohol (g/day)	15 (0.5)	16 (0.4)	13 (0.4)	11 (0.4)	<0.0001
Protein (g/day)	82 (0.7)	80 (0.6)	80 (0.6)	79 (0.7)	0.03
Fat (g/day)	87 (0.6)	77 (0.5)	67 (0.5)	56 (0.5)	<0.0001
Saturated fat (g/day)	33 (0.3)	29 (0.2)	25 (0.2)	19 (0.2)	<0.0001
Mono-unsaturated fat (g/day)	29 (0.1)	27 (0.1)	24 (0.1)	21 (0.2)	<0.0001
Poly-unsaturated fat (g/day)	14 (0.1)	13 (0.1)	12 (0.1)	11 (0.1)	<0.0001
Carbohydrate (g/day)	218 (1.8)	205 (1.5)	218 (1.4)	234 (1.6)	<0.0001
Fibre NSP (g/day)	11 (0.1)	12 (0.1)	13 (0.1)	16 (0.1)	<0.0001
Total sugar (g/day)	77 (0.6)	79 (0.6)	93 (0.6)	107 (0.7)	<0.0001
Sodium (mg/day)	3,094 (22.4)	2,829 (17.9)	2,628 (17.1)	2,502 (20.1)	<0.0001
Food groups					
Fruit and vegetable (g/day)	208 (5.5)	265 (3.5)	338 (3.4)	474 (4.0)	<0.0001
Whole grains (g/day)	30 (1.4)	37 (1.1)	49 (1.0)	70 (1.2)	<0.0001
Fish (g/day)	19 (0.8)	22 (0.6)	23 (0.6)	28 (0.7)	<0.0001
Oily fish (g/day)	10 (0.6)	12 (0.5)	13 (0.5)	18 (0.5)	<0.0001
Low fat dairy (g/day)	172 (4.1)	171 (3.3)	197 (3.1)	231 (3.7)	<0.0001
Red meat (g/day)	79 (1.2)	73 (0.9)	61 (0.9)	51 (1.1)	<0.0001
Processed meat (g/day)	42 (0.8)	38 (0.6)	31 (0.6)	25 (40.6)	<0.0001
Sweet and sugary beverages (g/day)	72 (3.5)	58 (2.8)	62 (3.1)	56 (3.0)	0.001
Dietary scores					
DASH score	21 (0.1)	23 (0.1)	25 (0.1)	28 (0.1)	<0.0001
UK Nutrient Profile score	53 (0.1)	55 (0.1)	57 (0.1)	61 (0.1)	<0.0001

Abbreviations: DRV, dietary reference values; Q1, first quartile; Q2, second quartile; Q3, third quartile; Q4, fourth quartile; DASH, dietary approach to stop hypertension; NP, nutrient profile; NSP, non-starch polysaccharides Englyst fibre

* only adjusted for sex

3.3.4 NP model score

The histogram of NP model shows the score was normally distributed (Appendix 9). Test for normality (non-parametric) Kolmogorov-Smirnov D- value was 0.015 $p < 0.01$. The score ranged from 39.98-72.98 and the sample's mean was 56.43 (SD 4.42) and median was 56.34. Quartiles one (range 0-24.99) represented the population group with “less favourable” diets, quartile two (range 25-49.99), quartiles three (range 50-74.99) and quartiles four (range 75-100) represented the group with the “most favourable” diet.

3.3.5 Association between NP model and nutritional components

Table 23 shows that across the NP quartiles there was a strong trend for a greater consumption of other “favourable” nutritional components including both single nutrients and food groups, with exception of alcohol. The NP score also showed a significant trend with a higher overall diet quality (DRV score) ($p < 0.0001$) and DASH score ($p < 0.0001$).

Table 23: Association between nutrient profile quartiles and mean dietary intake adjusted for sex and mean energy intake, the Airwave Health Monitoring Study (n=5,848)

NP quartiles	Q1	Q2	Q3	Q4	P trend
Range	(0-24.99)	(25-49.99)	(50-74.99)	(75-100)	
	'least healthy'			'healthiest'	
n	1,075	1,669	1,800	1,304	
	Mean				
	(SE)				
Nutrients					
Energy intake (kcal/day)*	1958	1801	1725	1477	<0.0001
Alcohol (g/day)	13	14	15	14	0.23
Protein (g/day)	81	81	80	79	0.03
Fat (g/day)	87	75	67	54	<0.0001
Saturated fat (g/day)	34	28	24	18	<0.0001
Mono-unsaturated fat (g/day)	28	26	24	22	<0.0001
Poly-unsaturated fat (g/day)	13	13	13	12	<0.0001
Carbohydrate (g/day)	231	219	216	206	<0.0001
Fibre (g/day)**	12	13	14	15	<0.0001
Total sugar (g/day)	95	87	86	82	<0.0001
Added sugar (g/day)	82	72	67	58	<0.0001
Sodium (mg/day)	2996	2806	2689	2470	<0.0001
Food groups					
Fruit and vegetable (g/day)	256	294	333	395	<0.0001
Whole grains (g/day)	34	39	50	61	<0.0001
Fish (g/day)	18	21	25	28	<0.0001
Oily fish (g/day)	10	11	14	17	<0.0001
Low fat dairy (g/day)	171	182	194	214	<0.0001
Red meat (g/day)	71	69	64	57	<0.0001
Processed meat (g/day)	42	35	32	25	<0.0001
Sugary drinks (g/day)	71	71	61	44	<0.0001
Dietary scores					
DASH score	21	23	25	27	<0.0001
DRV Index	2	3	4	7	<0.0001

Abbreviations: DRV, dietary reference values; Q1, first quartile; Q2, second quartile; Q3, third quartile; Q4, fourth quartile; DASH, dietary approach to stop hypertension; NP, nutrient profile

* only adjusted for sex

** non-starch polysaccharides Englyst fibre

3.3.6 Association between nutritional components incorporated into the diet scores and cardio-metabolic markers

Table 24 shows the single nutritional components (incorporated into the DRV index and NP model) driving diet scores in association with metabolic markers. A lower HbA1c was associated with a diet with higher levels of fibre, fruit and vegetables and lower level of saturated fat. DBP was negatively associated with a high fibre intake and positively associated with intake of total fat intake. Waist circumference was negatively associated with intakes of CHO, fibre, fruit and vegetable and positively associated with intakes of saturated fat and sodium.

Table 24: Association between cardio-metabolic risk markers and single nutritional components included in the diet scores, the Airwave Health Monitoring Study (n=5,848)

Variable	CHO		Fibre†		F&V		Sugar††		Total fat		Sat fat		Sodium	
	β	SE	β	SE	β	SE	β	SE	B	SE	β	SE	β	SE
HbA1c (%)	0.000	0.000	*-0.006	0.002	*-0.010	0.004	0.000	0.000	*0.002	0.001	*0.003	0.001	-0.027	0.012
SBP (mmHg)	0.009	0.006	0.008	0.044	* 0.025	0.100	0.011	0.007	-0.021	0.015	-0.024	0.029	-0.400	0.310
DBP (mmHg)	0.002	0.004	*-0.071	0.030	-0.001	0.068	-0.002	0.005	*0.023	0.010	0.036	0.019	-0.026	0.211
BMI (kg/m ²)	*-0.012	0.002	*-0.062	0.014	*-0.002	0.031	*-0.012	0.002	*0.012	0.005	0.008	0.009	*0.680	0.096
WC (cm) δ	*-0.030	0.005	*-0.239	0.034	*-0.600	0.077	*-0.046	0.006	*0.055	0.012	*0.072	0.022	*1.600	0.237
HDL-C (mmol/L)	*-0.001	0.000	0.000	0.001	-0.002	0.003	*-0.001	0.000	*-0.001	0.000	*0.003	0.001	0.011	0.008
Total C (mmol/L)	*-0.003	0.000	*-0.026	0.003	*-0.040	0.007	0.000	0.001	*0.009	0.001	*0.019	0.002	-0.027	0.022

Abbreviations: β, beta-coefficient; Sat fat, saturated fat; CHO, carbohydrates; F&V, fruit and vegetables; HbA1c, glycated haemoglobin; SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; HDL, high density lipoprotein; WC, Waist circumference; C, cholesterol

δ Adjusted for height instead of BMI

Multiple regression adjusted for age, sex, BMI, alcohol, energy intake, smoking, physical activity, diagnosis and treatment for specific outcome

The regression models provide regression coefficients (β) in outcome variables for each 1 g increase of nutrients and 100 g increase of F&V

† non-starch polysaccharides fibre ††Non-milk extrinsic sugar

* p-value<0.05

3.3.7 Association between diet scores and cardio-metabolic risk markers

A higher DRV score was inversely associated with HbA1c (β -0.02, p 0.003), total cholesterol (β -0.06, p <0.0001), HDL-cholesterol (β -0.01, p 0.005), BMI (β -0.15, p <0.0001) and waist circumference (β -0.54, p <0.0001). The NP score was associated with HbA1c (β -0.01, p <0.0001), total cholesterol (β -0.33, p <0.0001) and an increase in BMI (β 0.06, p 0.01) (Table 25). A higher DASH score was associated with a lower waist circumference (β =-0.57 p <0.0001), BMI (β =-0.17, p <0.0001), total cholesterol (β -0.03, p <0.0001), DBP (β -0.18, p 0.0003), and HbA1c (β -0.01, p 0.001).

Table 25: Association between diet scores Dietary Reference Values index, Nutrient Profile and Dietary Approach to Stop Hypertension score and cardio-metabolic risk markers

	DRV score			NP score			DASH score		
	β^a	SE	<i>P</i> value	β^a	SE	<i>P</i> value	β^a	SE	<i>P</i> value
<i>n</i> =5,848									
HbA1c (%)	-0.02	0.004	0.003	-0.01	0.003	<0.0001	-0.01	0.00	0.001
SBP (mmHg)	0.22	0.118	0.01	0.12	0.076	0.17	0.03	0.12	0.70
DBP (mmHg)	-0.05	0.080	0.56	-0.01	0.052	0.72	-0.18	0.08	0.0003
HDL cholesterol (mmol/L)	-0.01	0.003	0.005	-0.003	0.002	0.10	0.01	0.03	0.69
Total cholesterol (mmol/L)	-0.06	0.008	<0.0001	-0.33	0.050	<0.0001	-0.03	0.01	<0.0001
Waist circumference (cm)*	-0.54	0.095	<0.0001	0.05	0.060	0.41	-0.57	0.10	<0.0001
BMI (kg/m ²)**	-0.15	0.036	<0.0001	0.06	0.02	0.01	-0.17	0.04	<0.0001

Abbreviations: DRV, Dietary Reference Values; NP, nutrient profile; DASH, Dietary Approach to Stop Hypertension; HbA1c, glycated haemoglobin; SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; HDL, high density lipoprotein; β , beta-coefficient; SE, standard error.

^a Multivariate linear regression models provide regression coefficients (β) in outcome variables for 2-points increase in diet scores adjusted for age, sex, energy intake, smoking, alcohol, physical activity, BMI, diagnosis and treatment for specific outcome

* Adjusted for height not BMI

** not adjusted for BMI

3.3.8 Association between diet scores and cardiovascular risk and diabetes

Results in Table 26 shows that a higher DRV score was associated with a lower risk of T2D OR 0.94 (p=0.03). NP score was not associated with any CVD or T2D. DASH score showed an inverse association with dyslipidaemia OR 0.98 (p=0.003).

Table 26: Association between diet scores (DRV, NP and DASH score) and cardiovascular risk and diabetes in the Airwave Health Monitoring Study (n=5,848)

	cases/ subcohort	DRV score			NP score			DASH score		
		OR ^a	95% CI	P	OR ^a	95% CI	P	OR ^a	95% CI	P
Hypertension ¹	1,840/4,008	1.01	0.99-1.03	0.26	1.01	0.99-1.03	0.07	0.99	0.98-1.00	0.10
Dyslipidaemia ²	1,471/4,377	0.99	0.97-1.02	0.59	1.01	0.99-1.03	0.07	0.98	0.76-0.99	0.003
Type 2 diabetes ³	214/5,848	0.94	0.90-0.99	0.03	0.98	0.95-1.01	0.24	0.99	0.96-1.08	0.44

Abbreviations: DRV, Dietary Reference Values UK; NP, nutrient profile; DASH, Dietary Approach to Stop Hypertension; OR, odds ratio; 95% CI, 95% confidence interval; P, p-value

^a Values are odds ratio adjusted for age, sex, BMI, energy intake, alcohol, smoking, physical activity

¹ BP 140/90 mmHg or receiving drug therapy for hypertension or self-reported diagnosis

² Total cholesterol:HDL ratio cut-off > 4.5mmol/L or receiving drug therapy for cholesterol lowering or self-reported diagnosis

³ Hb1Ac $\geq$ 6.5 % or receiving drug therapy for hyperglycaemia or self-reported diagnosis

3.3.9 Association between diet score quartiles (DRV index and NP model) and metabolic outcomes

DRV score was also associated with a reduced risk for a number of metabolic outcomes. Individuals with a better DRV score (comparing “healthiest” top DRV quartile vs “least healthy” bottom) were less likely to have obesity (OR=0.68, p=0.002) and hyperglycaemic (OR=0.80, p=0.01), abdominal obesity (OR=0.64, p<0.0001) (Table 27) based on MetS classifications.

NP score was also associated with a reduced risk for a number of metabolic outcomes. Individuals with a better NP score (comparing top quartile “healthiest” vs “least healthy” bottom) were less likely to have hyperglycaemic (OR=0.78, p=0.01) and there was an increase in elevated BP (OR=1.23, p=0.03) and low HDL-cholesterol (OR=1.48, p=0.003) (Table 28) based on MetS classifications.

Table 27: Association between Dietary Reference Values index quartiles and metabolic syndrome risk factors classifications in the Airwave Health Monitoring Study (N=5,848)

Range		Q1 ^a 0-3.49 'least healthy'	Q2 ^a 3.50-6.99	Q3 ^a 7.00-10.49	Q4 ^a 10.50-14 'healthiest'	P trend
	cases/subcohort					
Increased blood pressure ¹	3,015/2,833	<i>ref</i>	1.17 (0.98-1.38)	1.21 (1.02-1.44)	1.11 (0.92-1.33)	0.15
Low HDL cholesterol ²	872/4,976	<i>ref</i>	1.08 (0.86-1.37)	1.08 (0.86-1.36)	1.39 (1.09-1.77)	0.02
Increased blood sugar ³	2,408/3,440	<i>ref</i>	0.97 (0.82-1.14)	0.84 (0.71-0.98)	0.80 (0.67-0.95)	0.01
Obesity ^{4*}	1,180/4,669	<i>ref</i>	0.94 (0.77-1.37)	0.90 (0.74-1.09)	0.68 (0.55-0.82)	0.002
Abdominal obesity ^{5**}	2,934/2,914	<i>ref</i>	0.90 (0.77-1.06)	0.78 (0.67-0.92)	0.64 (0.54-0.77)	<0.0001

Abbreviations: DRV, dietary reference values; Q, quartiles; HDL, high density-lipoprotein; BMI, body mass index; DBP, diastolic blood pressure; SBP, systolic blood pressure; Hb1Ac, glycated haemoglobin

^a Values are odds ratio (95% CI) adjusted for age, sex, BMI, energy intake, alcohol, smoking, physical activity

* Not adjusted for BMI ** Adjusted for height not BMI

¹ DBP ≥ 85mmHg and SBP ≥ 130mmHg or on treatment

² HDL < 1.03 mmol/L for men, < 1.29 mmol/L for women or on treatment

³ Hb1Ac > 5.7 % or on treatment

⁴ BMI >30kg/m²

⁵ Waist circumference > 95cm for men, > 80 cm for women

Table 28: Association between nutrient profile quartiles and metabolic syndrome risk classifications, the Airwave Health Monitoring Study (N=5,848)

NP quartiles		Q1 ^a	Q2 ^a	Q3 ^a	Q4 ^a	P trend
Range		(0-24.99)	(25-49.99)	(50-74.99)	(75-100)	
	cases/subcohort	'least healthy'			'healthiest'	
Increased blood pressure ¹	3,015/2,833	<i>ref</i>	1.25 (0.07-1.50)	1.14 (0.95-1.33)	1.23 (1.07-1.50)	0.03
Low HDL cholesterol ²	872/4,976	<i>ref</i>	1.18 (0.96-1.52)	1.32 (1.08-1.68)	1.48 (1.20-1.85)	0.003
Increased blood sugar ³	2,408/3,440	<i>ref</i>	0.94 (0.80-1.09)	0.85 (0.73-1.00)	0.79 (0.67-0.91)	0.01
Obesity ^{4*}	1,180/4,669	<i>ref</i>	1.67 (0.87-1.92)	0.98 (0.82-1.16)	1.04 (0.80-1.21)	0.25
Abdominal obesity ^{5**}	2,934/2,914	<i>ref</i>	1.03 (0.87-1.18)	0.95 (0.80-1.10)	1.00 (0.86-1.15)	0.76

Abbreviations: NP, nutrient profile score; Q, quartiles; HDL, high density-lipoprotein; BMI, body mass index; DBP, diastolic blood pressure; SBP, systolic blood pressure; Hb1Ac, glycated haemoglobin

^a Values are odds ratio (95% CI) adjusted for age, sex, BMI, energy intake, alcohol, smoking, physical activity

* Not adjusted for BMI ** Adjusted for height not BMI

¹ DBP ≥85mmHg and SBP ≥130mmHg or on treatment

² HDL < 1.03 mmol/L for men, < 1.29 mmol/L for women or on treatment

³ Hb1Ac > 5.7 % or on treatment

⁴ BMI >30kg/m²

⁵ Waist circumference > 95cm for men, > 80 cm for women

3.4 Sensitivity analysis

3.4.1 Excluding participants with energy under-reporting

Two sets of sensitivity analyses were carried out to test if bias was introduced by either participants with chronic illnesses (self-reporting previous/current diagnosis: cancer, liver disease, thyroid, stroke, angina, heart attack and other heart diseases) or by participants under-reporting energy.

Table 29 show sensitivity analysis of DRV index with metabolic markers conducted excluding participants who were classified as energy under-reporters (n 2,720). The analyses of 2,815 participants showed a higher DRV score was associated with a reduction in HbA1c (β -0.02, $p=0.003$), total cholesterol (β -0.05, $p=0.0002$), waist circumference (β -0.94, $p<0.0001$), and BMI (β -0.30, $p<0.0001$). There was no association with HDL-cholesterol, SBP and DBP.

Table 29: Sensitivity analysis (excluding energy under-reporters): Dietary Reference Values index association with metabolic markers, the Airwave Health Monitoring Study (n 2,815)

Metabolic markers	β^a	SE	<i>P</i> value
HbA1c (mmol/L%)	-0.02	0.006	0.003
SBP (mmHg)	0.46	0.18	0.10
DBP (mmHg)	0.10	0.12	0.36
HDL cholesterol (mmol/L)	0.006	0.004	0.89
Total cholesterol (mmol/L)	-0.05	0.01	0.0002
Waist circumference (cm)*	-0.94	0.14	<0.0001
BMI (kg/m ²)**	-0.30	0.05	<0.0001

Abbreviations: β , beta-coefficient; SE, standard error; HbA1c, glycated haemoglobin; SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; HDL, high density lipoprotein.

Multivariate linear regression models provide regression coefficients (β) in outcome variables for each 2 point increase in DRV score

^a Values are beta-coefficients (standard error) adjusted for age, sex, energy intake, smoking, alcohol, physical activity, BMI, diagnosis and treatment for specific outcome

* Adjusted for height not BMI

** Not adjusted for BMI

Table 30 show sensitivity analysis of DRV index with metabolic outcomes conducted excluding participants who were classified as energy under-reporters. The analyses showed a higher DRV score was associated with a lower likelihood of MetS outcomes, increased blood sugar (OR 0.97, $p<0.0001$), abdominal obesity (OR 0.94, $p<0.0001$) and obesity (OR 0.93, $p<0.0001$). A higher was DRV score was also associated with a higher likelihood of low HDL-cholesterol (OR 1.03, $p=0.04$).

Table 30: Sensitivity analysis (excluding energy under-reporters) Dietary Reference Values index association with metabolic outcomes, the Airwave Health Monitoring Study (n 2,815)

Metabolic outcomes	cases/subcohort	OR ^a	95% CI	P value
Increased blood pressure ¹	1,462/1,353	1.02	1.00-1.04	0.06
Low HDL cholesterol ²	398/2,417	1.03	1.00-1.06	0.04
Increased blood sugar ³	1,378/1,439	0.97	0.95-0.99	<0.0001
Obesity ^{4*}	563/2,254	0.94	0.91-0.96	<0.0001
Abdominal obesity ^{5**}	1,393/1,422	0.93	0.91-0.95	<0.0001

Abbreviations: OR, odds ratio; 95% CI, 95% confidence interval; P, p-value; HbA1c, glycated haemoglobin; SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; HDL, high density lipoprotein.

^a odds ratio and p-value within dietary quartiles adjusted for age, sex, BMI, mean alcohol, smoking, physical activity

* Not adjusted for BMI

** adjusted for height not BMI

¹ DBP ≥ 85 mmHg and SBP ≥ 130 mmHg or on treatment

² HDL < 1.03 mmol/L for men, < 1.29 mmol/L for women or on treatment

³ Hb1Ac > 5.7 % or on treatment

⁴ BMI >30 kg/m²

⁵ Waist circumference > 95cm for men, > 80 cm for women

3.4.2 Excluding participants with chronic diseases

Table 31 show sensitivity analysis of DRV index with metabolic markers conducted on a healthy sample population excluding participants self-reporting previous/current diagnosis: cancer, liver disease, thyroid, stroke, angina, heart attack and other heart diseases, (n 335). The analyses showed a higher DRV score was associated with a reduction in HbA1c (β -0.01, $p=0.002$), HDL-cholesterol (β -0.01, $p=0.004$), total cholesterol (β -0.06, $p<0.0001$), waist circumference (β -0.60, $p<0.0001$), and BMI (β -0.17, $p<0.0001$). There was no association with SBP and DBP.

Table 31: Sensitivity analysis (excluding ill people*) Dietary Reference Values index association with metabolic markers, the Airwave Health Monitoring Study (n=5,529)

Metabolic markers	β^a	SE	<i>P</i> value
HbA1c (%)	-0.01	0.004	0.002
SBP (mmHg)	0.22	0.118	0.07
DBP (mmHg)	-0.10	0.080	0.24
HDL cholesterol (mmol/L)	-0.01	0.020	0.004
Total cholesterol (mmol/L)	-0.06	0.008	<0.0001
Waist circumference (cm)**	-0.60	0.095	<0.0001
BMI (kg/m ²) ***	-0.16	0.036	<0.0001

Abbreviations: β , beta-coefficient; HbA1c, glycated haemoglobin SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; HDL, high density lipoprotein.

Multivariate linear regression models provide regression coefficients (β) in outcome variables for each 2 point increase in DRV score

^a Values are beta-coefficients (standard error) adjusted for age, sex, energy intake, smoking, alcohol, physical activity, BMI, diagnosis and treatment for specific outcome

* Participants with chronic illnesses (self-reporting previous/current diagnosis: cancer, liver disease, thyroid, stroke, angina, heart attack and other heart diseases)

** Adjusted for height not BMI

*** Not adjusted for BMI

Table 32 show sensitivity analysis of DRV index with metabolic outcomes conducted on a healthy sample population. The analyses showed a higher DRV score was associated with a lower likelihood of MetS outcomes, increased blood sugar (OR 0.97, p 0.002), abdominal obesity (OR 0.96, p=0.0004) and obesity (OR 0.95, p<0.0001). A higher DRV score was also associated with a higher likelihood of low HDL-cholesterol (OR 1.04, p=0.002).

Table 32: Sensitivity analysis (excluding ill people*) Dietary Reference values Index association with metabolic outcomes, the Airwave Health Monitoring Study (n=5529)

Metabolic outcomes	cases/subcohort	OR ^a	95% CI	P value
Increased blood pressure ¹	2,854/2,736	1.02	0.99-1.04	0.06
Low HDL cholesterol ²	826/4,764	1.04	1.02-1.07	0.002
Increased blood sugar ³	2,282/3,308	0.97	0.95-0.99	0.002
Obesity ^{4**}	1,105/4,485	0.96	0.94-0.98	0.0004
Abdominal obesity ^{5**}	2,767/2,823	0.95	0.93-0.97	<0.0001

Abbreviation: OR, odd ratio; CI, confidential interval; HbA1c, glycated haemoglobin
SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index;
HDL, high density lipoprotein.

* Participants with chronic illnesses (self-reporting previous/current diagnosis: cancer, liver disease, thyroid, stroke, angina, heart attack and other heart diseases)

** Adjusted for height not BMI

*** Not adjusted for BMI

^a odds ratio and p-value within dietary quartiles adjusted for age, sex, BMI, mean alcohol, smoking, physical activity

¹ DBP ≥85mmHg and SBP ≥ 130mmHg or on treatment

² HDL < 1.03 mmol/L for men , < 1.29 mmol/L for women or on treatment

³ Hb1Ac > 5.7 % or on treatment

⁴ BMI >30 kg/m²

⁵ Waist circumference > 95cm for men, > 80 cm for women

3.5 Discussion

Summary of main findings

- Quality control audit cycle showed a mean code error rate of 3.7% per food diary from the Airwave Monitoring Study.
- Both diet scores NP and DRV were associated with twelve nutritional components essential to a healthy diet and with the DASH score.
- Dietary variables incorporated into the diet scores (NP and DRV) show a weak to modest collinearity between variables.
- NP score was inversely associated with cardio-metabolic risk: HbA1c and total cholesterol.
- DRV score was inversely associated with T2D risk, HbA1c, total cholesterol and adiposity markers. Sensitivity analyses showed these relationships association between DRV and HDL-cholesterol and BP was biased by participants' energy under-reporting.

The first objective of this study was to develop study-specific operational manual for coding 7-day food diaries, which is presented in Method chapter 2.2 and appendix 3. The diaries were coded using the Dietplan programme, which was originally developed for smaller datasets analysis. Unlike other larger nutritional studies such as EPIC-Norfolk who used study-specific diet analysis programme with built-in functions to help reduce coding errors and bias, such as flagging up gross weight errors when a food item was entered incorrectly, these functions were not available in Dietplan. Therefore, in order to address potential coding errors, an additional protocol to audit the quality of the dietary dataset of the Airwave Monitoring study was developed. The results from the audit cycle showed a low mean code error rate per food diary (3.7%). Due to lack of published data from other nutritional studies it is difficult to draw comparisons.

The second objective was to explore methods of assessing diet quality based on adherence to the UK dietary guidelines for optimal health and prevention of cardio-metabolic risk. This was explored by a single dietary quality score as well as the dietary factors contributing to the effect of the dietary score. Multi-collinearity analysis between the dietary variables included

in the calculation of diet quality scores (DRV index and NP model) demonstrated weak to moderate correlations. Collinearities were observed between the dietary variables: carbohydrates and intake of fruit and fibre (r 0.30-0.47), which are to be expected as both fruit and fibre have a high content of carbohydrates. No correlations were observed between fruit and vegetables. No correlation was found between NME sugar and fruit intake (r 0.13) as expected, NME sugar does not include sugar bound in fruit. However, there is a strong correlation between sugar intake and carbohydrate (r 0.50). A negative correlation was observed between high intake of carbohydrates and saturated fat (r -0.35); No correlations were observed between sodium and other nutrients, which may be due to measurement bias in sodium recording in the diaries. Similar collinearities have been explored in other dietary scores and single nutrients analyses (23, 28, 29, 35). According to Field (150), correlations below 0.8 are a good indication that multicollinearity is not occurring in the data. Furthermore, as described in the background, many nutrient groups are highly correlated, such as fibre, fruit and vegetables, making it difficult to examine their individual effects (29). Analyses of single nutrients may be confounded by the effect of dietary patterns. The combination of nutritional variables in the two applied dietary scores (NP model and DRV index) are similar to those incorporated in other known priori diet quality scores such as HEI, MDS, DASH, using country-specific cut-off values (28, 35). All dietary analyses conducted between diet quality scores' quartiles and nutritional variables and mean dietary intake were explored separately in men and women (not presented in thesis) but showed no differences in trends hence were analysed as one group.

Another way of exploring the dataset for multi-collinearity bias and generating dietary score empirically is a data driven analysis of the dietary dataset using principal component analysis. Rayner and Arambepola et al. (22, 58) conducted principal component analysis using the NDNS dataset as validation of the NP model and they found that a high NP score (higher intake of healthier foods) identified dietary patterns high in bread, cereals and fruit and vegetables; whereas, a lower NP score identified dietary patterns high in fatty and sugary foods. However, using principal component analysis alone is associated with limitations such as data-driven results that are mainly study specific and not based on existing evidence. Due to time limitations of this thesis study, only one of the priori diet scores was chosen. Furthermore, due to the cross-sectional nature of this study, a priori diet score was used rather than principal component analysis, as it would be difficult to draw conclusions between diet and health related outcomes. Therefore, a diet quality score derived from principal

component analysis would only be assessing diet quality according to specific dietary patterns found in this study and not accordingly to a diet recommended for a British adult population for optimal health and cardio-metabolic prevention.

The present study demonstrates that both diet scores NP model and DRV index are associated with over 19 other dietary components essential to assess a person's diet quality including components incorporated in the UK dietary guidelines. Similar results have been previously shown with the UK NP model (20, 22, 46, 58). Participants with the "healthiest diets" (defined by DRV quartile 4) consumed a mean daily energy intake of 1691 kcal, total fat 56g (30% total energy intake (TEI)), saturated fat 19g (10 % TEI), carbohydrate 234g (55 %TEI), 16g fibre and 474g fruit/vegetable, which are in line with the national dietary guidelines (23-25, 27). Whereas, the NP quartiles 4 reflected a lower mean daily energy intake 1477kcal, whilst macronutrients intake were similar; total fat 54g (33%TEI), saturated fat 18g (11 %TEI), (32 %TEI), carbohydrate 206g (56 %TEI), 15g fibre and fruit/vegetable was lower 395g. Participants in the "healthiest" diet categories are in alignment with the UK dietary guidelines. However, due to limitations of the nutrient database (Dietplan version 7.0) used to analyse the diets, study amendments were made to the existing UK guidelines on fibre and added sugar (25). Hence, the DRV index reflect the previous guidelines for fibre (NSP) and added sugar (NMES) and not the updated guidelines from SACN in 2015, which reflect new evidence showing health benefits of lowering added sugar to 5% TEI and fibre (AOAC) to 30g/day. Therefore, future studies using the DRV index should apply these updates if possible.

The results support the application of NP or DRV in epidemiological studies to capture intake of essential nutrients and food groups in a UK diet. However, these results are based on a UK population group (95% white British) who are younger (mean age 42 year) and primary men (60%). Therefore, the diet patterns identified in this study does not necessarily represent what would be identified in a random sample from the general UK population. Despite neither of the scores (NP and DRV) including intake of fish (oily), mono-and poly unsaturated fats, which are important for CVD health, results showed both scores were associated with a higher intake of these nutritional components. However, both diet scores are based on the UK guidelines and limitations may exist for use in other study populations, as previously discussed by Kant et al and Moeller et al. (29, 35). More research is needed to confirm if the scores are able to capture healthy diets in other countries such as Japan and Mediterranean

where consumption of fish, seafood, nuts, and olive oil are much higher than in the UK. Therefore, NP and DRV diet scores may be more suitable to countries consuming similar diets as in the UK such as other northern European countries e.g. Germany, Denmark and Belgium.

DRV score confirmed an inverse association with HbA1c, total cholesterol and adiposity markers (BMI and waist circumference). The same relationships were also observed for DASH score, which also showed an association with DBP. Siervo et al (129) also reported DASH was associated with total cholesterol and BP, but no significant association with HDL-cholesterol. The single nutrient analyses showed that these relationships were driven by diets higher in fruit, vegetables, fibre and lower in sugar, sodium, fat and saturated fat, which have also been previously demonstrated in other single nutrient studies and RCTs (3, 12, 21, 151). These relationships between single nutrients and these metabolic markers are in line with the WHO and NICE guidelines for prevention and management of CVD and T2D (12, 87). The CRESSIDA showed an intervention diet adhering to the UK dietary guidelines driven by sodium, total fat, saturated fat, NMES, fruit/vegetables and wholegrain lowered, BMI, BP and lipid profile (151). The intervention trial provided dietary education based on food groups from the “*Eat-Well Guide*”, which are food recommendations based on the UK dietary guidelines. The CRESSIDA study highlighted that six nutrients and one food group could be used to measure and reflect the intake of a wide range of foods from the “*Eat-Well Guide*”. Similar results were shown previously with the NP score and a wide range of foods from the “*Eat-Well Guide*” (58), suggesting that diet score composed of a small key group of nutritional components may measure a wide range of foods essential to health outcomes. Hence, such diet scores are important tools in epidemiological studies as they may ease the burden of researchers as this often requires fewer resources and time to disaggregate and analyse many different food groups.

The mechanisms behind these relationships involve many different biochemical pathways. A high intake of total fat, particular saturated fat, is directly linked to an increase in serum total cholesterol via biochemical pathways (endogenous synthesis) from saturated fat digestion (152). Furthermore, a diet high in total fat and saturated fat is also linked with increased HbA1c and T2D both directly and via mediated adiposity markers (12, 153-155). The mechanism behind high fat (saturated fat) intake and T2D risk is related to several pathways affecting obesity, insulin resistance, inflammation and elevated HbA1c (12, 153, 154).

Whereas, a high intake of dietary fibre, fruit and vegetables has an inverse relationship with these metabolic risk factors including T2D, BP, HbA1c, total cholesterol and adiposity markers ((12, 153, 154). High intake of fibre, fruit and vegetables has been shown to control blood glucose levels and reduce risk of T2D in individuals (90, 156-158). The mechanism behind this relationship is primarily related the digestion process of fibre, which is known to slow down the gastric and intestinal emptying. Together it results in a prolongation and reduction of intestinal glucose absorption and thus lowers the demand for postprandial insulin secretion, which helps blood sugar control (156-159). Fibre, fruit and vegetables inverse relationship with total cholesterol is related to decreased re-absorption of cholesterol (bile salts) and increased excretion of bile salts as it binds to the indigestible fibre in the small intestine (152).

NP score showed inverse associations with HbA1c and total cholesterol. In contrast, NP was associated with a higher BMI. However, this association may have been biased by the cross-sectional nature of the study or residual confounding of the energy adjustment (nutrient density) of the score. There exist various methods of energy adjusting diet scores such as residual and partition methods. This study chose to apply a nutrient density (energy adjusted) method previously applied in other studies of the NP model (29, 32, 33). However, it may be argued that if energy intake lies in the causal pathway between certain nutrient groups e.g. high fat and sugar and cardio-metabolic outcomes, it should be treated as a mediator rather than a confounder. Ideally, nutritional assessments should include validation of biomarkers or doubly labeled water assessment to help confirm accuracy of self-reported dietary data and to limit bias of misreporting. These findings are challenging to compare as only one other study, the Supplementation Vitamineset Mineraux Antioxydants (SUVIMAX) cohort, to our knowledge, has investigated NP model (with study adaptations to the French guidelines) in relation to cardio-metabolic risk factors in a French population (160). SUVIMAX cohort saw no association with waist circumference, fasting glucose, or blood lipids. Another study by Arambepola et al. (58) also did not report any correlation with NP model and BMI. The NP model's limitations have been discussed by several authors (22, 51, 53, 58) including the above mentioned energy-adjustment methods. Another important limitation of the NP model is that it only measures one aspect of diet quality (food quality and density) and unlike DRV index it does not measure overall nutritional intake, diet patterns or variety of foods consumed.

The DRV score inverse association with HDL-cholesterol was not significant once energy under-reporters were removed from the analysis suggesting that the association may have been biased by misreporting of dietary intake. Single nutrient analyses showed that dietary variables carbohydrate, sugar and fat (in-cooperated in the DRV and NP score) were inversely associated with HDL-cholesterol. Despite these associations neither the DRV score nor NP score showed a positive association with HDL-cholesterol. Both diet scores DRV and NP score also showed a significant lower intake of poly- and monounsaturated fats across all 4 quartiles. Suggesting that neither of them captured these essential nutrients in the diet which are positively associated with HDL-cholesterol such as poly- and monounsaturated fats (12, 21). This may explain why no positive relationship was observed, as reported for MDS, which include these nutrients (36, 128). However, DASH score incorporates fish/seafood in the score and did not report any associations either with HDL-cholesterol in this study.

This study also reported an increase in SBP with a higher DRV score. However, this association was no longer significant once energy under-reporters were removed suggesting this association may have been biased by misreporting. No association between BP and NP score was observed in this study. The SUVIMAX cohort reported a positive association between the NP score and both SBP and DBP in a French population (160). DASH score did find an inverse relationship with DBP. Suggesting that a dietary score measuring food groups high in salt may capture this relationship better than measuring a single nutrient (sodium) intake. The inconsistency in this study's findings may be biased by this study's difficulties in capturing "added" salt in participants' diets and limitations of the nutritional databases (applied in Dietplan). Due to inconsistency of participants reporting added salt in the diary means that the nutritional data only includes intrinsic salt usage, which may reflect why SBP or DBP were not associated with sodium as expected from the literature (12, 21). For reasons that are unclear fruit and vegetable food group was found to be positively associated despite a number of studies having reported an inverse association (12). This may also have been driving the association between SBP and the DRV score. The fruit and vegetable food group includes canned and processed vegetable dishes high in sodium, which may have led to an indirect association. DBP was significantly associated with fat and inversely with fibre, which is in line with other studies (12, 21).

The study showed NP score only had reverse relationship with elevated blood sugar, which was driven by its relationship with HbA1c. No other associations were observed for metabolic risk outcomes as seen with other NP models (39, 46, 160). SUVIMAX cohort

found a lower UK NP score (adapted version) was associated with MetS (OD 1.06) (160). The US NP model found a higher score (healthier food quality) was associated with a lower risk of T2D, CVD and mortality rates in participants from the Nurses' Health Study and Health Professionals Follow-Up Study (39). The DRV score demonstrated reverse relationships with several cardio-metabolic risk factors: obese, abdominal obese and T2D risk. These results were predominately driven by the DRV inverse relationship with BMI, waist circumference and HbA1c, which are in line with previous studies of other known diet quality scores (35). Nicklas et al. (37) reported similar inverse relationships, and OR with metabolic risk factors elevated blood sugar, abdominal obesity and obesity was reported in the US National Health and Nutrition Examination Survey for individuals aligned with the national dietary guidelines. Comparing DRV with DASH score associations with CVD and T2D showed that DRV was significant inversely associated with T2D, whereas DASH was only associated with dyslipidaemia. Differences in dietary scores relationships with health outcomes such CVD and T2D have been previously reported and discussed by Kant et al. (29, 35) and Moeller et al. (29, 35) who argue that diet scores different inclusions of dietary variables and use of cut-off values vary to such extent that they capture different associations with risk factors and health outcomes.

This study highlights some of the limitations of the NP model in relation to capturing diet quality and associated cardio-metabolic health and outcomes. The nutrient composition of individual foods is not the only determinant of the overall nutrient composition of diets. Secondly it may not reflect the variety of different foods that make up the diets and the healthiness of the diets e.g. dietary patterns.

In summary, evidence suggests the NP model is a relevant tool to measure quality of individual foods contributing to an individual diet in a UK population. However, our study demonstrates that a diet score (DRV index) assessing overall diet quality adherence to the national dietary guidelines performs better in capturing important nutritional components in a UK diet in relation to cardio-metabolic risk, compared to a food-based score (NP model). This supports the importance of promoting both overall dietary guidelines in public health as well as food choices such as in the *"Eat Well Guide"* and *"Traffic Light System"* labelling in relation to their beneficial roles in cardio-metabolic health.

These results further support the application of the DRV index to measure diet quality in relation to diabetes and HbA1c and studying effect modification and interactions by genetic risk.

However, as previously discussed, limitations of a “single” overall diet score relationship with health outcomes such as dietary factors contributing to the effect of the overall index cannot be ascertained without examining the single components included in the index. The DRV index and genetic interactions on HbA1c/T2D risk will be strengthened by also conducting single nutritional variable analyses to help determine the variables which are driving the pathways and interactions.

CHAPTER 4: RESULTS GENETIC RISK ON DIABETES AND INTERACTIONS

The second objective of this study is to investigate interactions between diet, lifestyle factors and the genetic effect on glycaemic trait HbA1c and risk of T2D. For this analysis, a poly-genetic risk score was calculated for 14,085 participants from the Airwave Health Monitoring Study.

4.1 Genetic risk score

The histogram of GRS (Appendix 10) shows a normal distribution. The sample's GRS mean was 0.53 (SD 0.03) and median was 0.53. The score ranged from 0.37-0.65 from which the standardised GRS was derived (range 10.77-18.63). The mean of the standardised GRS was 15.67 (SD 1.00). The GRS quartiles were based on a statistical distribution of the score: Q1, lowest genetic risk quartile ($10.77 < \text{GRS} < 14.53$); Q2, second genetic risk quartile ($14.53 \leq \text{GRS} < 15.21$); Q3, third genetic quartile ($15.21 \leq \text{GRS} < 15.73$); Q4, highest genetic risk quartile ($15.73 \leq \text{GRS} \leq 18.63$).

4.2 Characteristics across quartiles of genetic risk score

GRS was not correlated with any socio-demographic and lifestyle characteristics (unadjusted). Only differences were seen in prevalence of T2D. Participants with the highest genetic risk (quartile 4) had the highest prevalence 4.36% compared to 3.21% in participants in the lowest risk (quartile 1) P-value for trend was 0.05 (Table 33).

Table 33: Socio-demographic and lifestyle characteristics across genetic risk quartiles for men and women, the Airwave Health Monitoring Study (n=14,085)

Genetic risk quartiles	Q1^a	Q2^a	Q3^a	Q4^a	P-value^b
	<i>Lowest risk</i>			<i>Highest risk</i>	
<i>n</i>	<i>3,434</i>	<i>3,606</i>	<i>3,563</i>	<i>3,486</i>	
Sex (female)	36.64	37.71	37.62	36.98	0.81
Age (years)	40.98	41.08	40.75	40.68	0.20
Diet quality*					
DRV score	4.33	4.19	4.10	4.14	0.34
Nutrient Profile score	56.47	56.51	56.15	56.46	0.24
Alcohol*					
No alcohol %	20.02	21.23	21.22	19.47	0.68
Within guidelines %	44.56	44.77	44.00	44.55	
Above guidelines %	35.42	34.00	34.78	35.98	
Cigarette smoking %					
Never smoker	67.52	68.34	67.44	67.86	0.53
Ex-smoker	24.38	25.02	24.32	23.00	
Current smoker	8.11	6.64	8.24	9.14	
Physical active %					
Low (<600min/week)	49.69	52.11	50.15	51.41	0.25
Moderate (>600min/week)	39.92	37.89	40.45	39.56	
High (>3000min/week)	10.38	9.99	9.40	9.03	
BMI (kg/m²)	27.12	27.24	27.26	27.22	0.52
Normal (18.5-24.99) %	35.00	30.00	31.00	35.00	0.30
Overweight (25-29.99) %	47.00	50.00	49.00	45.00	
Obese (>30) %	17.71	21.00	20.02	20.00	
Waist circumference (cm)	89.67	89.83	89.90	89.93	0.85
Cardiovascular diseases and diabetes % δ					
Type 2 Diabetes	3.58	3.38	4.29	4.62	0.006
Hypertension	30.51	30.93	30.88	32.70	0.20
Dyslipidaemia	23.61	26.93	25.47	25.94	0.25

Abbreviations: Q1, lowest genetic risk quartile ($10.77 < \text{genetic risk score} < 14.53$); Q2, second genetic risk quartile ($14.53 \leq \text{genetic risk score} < 15.21$); Q3, third genetic quartile ($15.21 \leq \text{genetic risk score} < 15.73$); Q4, highest genetic risk quartile ($15.73 \leq \text{genetic risk score} \leq 18.63$); DRV, dietary reference value; BMI, body mass index

^a Values are means or prevalence - unadjusted

^b P-value for differences between quartiles linear regression (continuous variables) or Mantel-Haenzel chi-square test (categorical variables)

* dietary dataset n=3,733

δ self-reported diagnosis or on treatment

4.3 Phenotypic trait HbA1c association with genetic risk score

GRS was associated with HbA1c%. Table 34 shows three different regression models (including various covariates) testing GRS's association with HbA1c. Coefficients and p-values remained the same for all three models; there were no differences in coefficients between the model adjusted for BMI or the one with waist circumference adjustment; the

fully adjusted model (including age, sex, BMI, smoking, physical activity, diabetes diagnosis and treatment estimated) showed similar association with HbA1c % ($\beta = 0.03$, SE 0.01, $p < 0.0001$) per 1 (SD) unit increase in the GRS.

The ANCOVA model (Table 35) showed that mean HbA1c% increased across GRS quartiles (p for trend < 0.0001). Participants in the highest risk group (quartile 4) had higher HbA1c % mean 6.47 compared to participants in the lowest risk group (quartile 1) mean 6.39.

Table 34: Multiple regression models for HbA1c % per 1 (SD) unit increase in the genetic risk score, the Airwave Health Monitoring Study (n=14,085)

	Model 1 ^a			Model 2 ^b			Model 3 ^c		
	β	SE	P-value	β	SE	P-value	β	SE	P-value
HbA1c (%)	0.03	0.01	<0.0001	0.03	0.01	<0.0001	0.03	0.01	<0.0001

Abbreviations: HbA1c, glycated haemoglobin A1c, β , beta-coefficient; SE, standard error

^a adjusted for body mass index, age and sex

^b adjusted for waist circumference, age and sex

^d adjusted for age, sex, body mass index, smoking, physical activity, diabetes diagnosis and treatment

Table 35: ANCOVA model HbA1c % mean across genetic risk quartiles, the Airwave Health Monitoring Study (n=14085)

Genetic risk quartiles	Q1 ^a	Q2 ^a	Q3 ^a	Q4 ^a	P for trend
<i>n</i>	3434	3606	3563	3486	
HbA1c (%)	6.39 (0.14)	6.43 (0.14)	6.44 (0.14)	6.47 (0.14)	<0.0001

Abbreviations: Q1, lowest genetic risk quartile; Q2, second genetic risk quartile; Q3, third genetic quartile; Q4, highest genetic risk quartile;

HbA1c, glycated haemoglobin A1c

^a Values are Bonferroni adjusted means (standard error) adjusted for age, gender, body mass index, smoking, physical activity, diabetes diagnosis, treatment

4.4 Phenotype T2D and pre-diabetes association with genetic risk score

The GRS was associated with both phenotypes pre-diabetes and T2D. Table 36 presents the three different logistic regression models tested to analyse the GRS association with the likelihood of having pre-diabetes and T2D. The association with both pre-diabetes (OR 1.09, 95%CI 1.05-1.23) remained the same in model 1 (adjusted for sex, age, and BMI) and model 2 (adjusted for sex, age, and waist circumference). Similar associations were observed for T2D in model 1 and 2. Model 3 was the fully adjusted model included covariates age, sex, BMI, smoking, physical activity. This model showed OR for having pre-diabetes increased by 10% (95%CI 1.06-1.14) per 1 (SD) unit increase in the GRS ($p < 0.0001$), and OR for having T2D increased by 14% (95%CI 1.04-1.24) ($p = 0.006$).

Table 37 shows the likelihood of having pre-diabetes and T2D increased across genetic risk groups (quartiles). Participants in the highest GRS (quartile 4) were 25% more likely to have pre-diabetes compared to participants in the lowest GRS (quartile 1) (p for trend < 0.0001). Participants in the highest genetic risk group (GRS quartile 4) were 35% more likely to have T2D compared to participants in the lowest genetic risk group (GRS quartiles 1) (p for trend 0.009).

Table 36: Likelihood of having pre-diabetes and type 2 diabetes across the genetic risk score (continuous), the Airwave Health Monitoring Study (n=14,085)

	(cases/subcohort)	Model 1 ^a			Model 2 ^b			Model 3 ^c		
		OR	95% CI	P	OR	95% CI	P	OR	95% CI	P
Pre-diabetes*	5,253/8,820	1.09	1.05-1.23	<0.0001	1.09	1.05-1.23	<0.0001	1.10	1.06-1.14	<0.0001
Type 2 diabetes**	516/13,569	1.13	1.04-1.24	0.005	1.13	1.03-1.23	0.007	1.14	1.04-1.24	0.006

Abbreviations: OR, odd ratio; CI, confidential interval; P, p-value

* HbA1c > 5.7 <6.5%, **HbA1c ≥ 6.5% and/or on glucose lowering medication

^a adjusted for age, sex, body mass index

^b adjusted for age, sex, waist circumference

^c adjusted for age, sex, body mass index, smoking, physical activity. Note pre-diabetes analysis was also adjusted for diabetes diagnosis and medication

Table 37: Likelihood of having pre-diabetes and type 2 diabetes across genetic risk quartiles of the Airwave Health Monitoring Study (N=14,085)

Genetic risk quartiles		Q1 ^a	Q2 ^a	Q3 ^a	Q4 ^a	
	(cases/subcohort)		OR (95% CI)	OR (95% CI)	OR (95% CI)	P for trend
Pre-diabetes*	5,253/8,820	<i>ref</i>	1.09 (0.99-1.21)	1.20 (1.08-1.33)	1.25 (1.13-1.39)	<0.0001
Type 2 diabetes**	516/13,569	<i>ref</i>	0.93 (0.71-1.20)	1.22 (0.95-1.57)	1.35 (1.07-1.73)	0.009

Abbreviations: *ref*, reference group; Q1, lowest genetic risk quartile; Q2, second genetic risk quartile; Q3, third genetic quartile; Q4, highest genetic risk quartile; OR, odd ratio; CI, confidential interval

* HbA1c > 5.7 <6.5% ** HbA1c ≥ 6.5% and/or on glucose lowering medication

^a adjusted for age, sex, body mass index, smoking, physical activity.

Note: pre-diabetes analysis was also adjusted for diabetes diagnosis and medication

4.5 Genetic interactions with diet quality and single dietary variables on HbA1c%

Table 38 shows the interaction effect on HbA1c% (coefficient and SE) across genetic risk quartiles interacting with dietary variables (continuous). All interaction models between GRS and the nine dietary variables were conducted in nine separate models (one dietary variable at a time). The beta-coefficients represented the estimated effect on HbA1c per one unit increase in each dietary variable (continuous) nested within GRS quartiles adjusted for covariates age, sex, BMI, smoking, alcohol, energy intake, physical activity, diabetes diagnosis and treatment.

Diet quality interaction with GRS on HbA1c

With multivariable adjustment, higher diet quality (DRV score) attenuated the effect on HbA1c in genetically higher risk individuals: GRS quartile 3 (β -0.014, $p_{\text{interaction}}$ 0.02), GRS quartile 4 (β -0.010, $p_{\text{interaction}}$ 0.05) (Table 38).

Single nutrients and food group interactions with GRS on HbA1c

Interactions were observed between GRS and single nutrients and food groups (Table 38), mostly among genetically higher risk individuals (GRS quartile 3 and 4). The effect on HbA1c% in the higher genetic risk individuals was offset by intake of carbohydrates (β -0.01, $p_{\text{interaction}}$ 0.005), fibre (β -0.08, $p_{\text{interaction}}$ 0.03), fruit and vegetable (β -0.02, $p_{\text{interaction}}$ 0.02) and wholegrain (β -0.14, $p_{\text{interaction}}$ 0.002). Opposite interaction effects on HbA1c% were observed with an increased intake of total fat (β 0.03, $p_{\text{interaction}}$ 0.007), saturated fat (β 0.06, $p_{\text{interaction}}$ 0.002) and by a higher intake of mean energy (β 0.07, $p_{\text{interaction}}$ 0.03).

No significant interactions are seen between sugar (NMES) intake and GRS quartiles.

Table 38: The interaction effect on HbA1c% across genetic risk quartiles interacting with dietary variables, the Airwave Health Monitoring Study (n=3,733)

	GRS Q1^a n 981		GRS Q2^a n 943		GRS Q3^a n 921		GRS Q4^a n 888	
	β	SE	β	SE	β	SE	β	SE
Carbohydrates	<-0.001	0.004	-0.006	0.003	-0.002	0.004	** -0.007	0.004
Fibre	-0.04	0.03	-0.05	0.04	-0.06	0.04	* -0.08	0.04
Fruit and Vegetables †	-0.014	0.009	-0.007	0.009	* -0.022	0.009	-0.01	0.009
Wholegrain †	-0.001	0.03	-0.063	0.04	** -0.104	0.03	-0.037	0.03
DRV score ††	-0.007	0.006	-0.002	0.006	** -0.014	0.005	-0.011	0.005
Total Fat	* 0.024	0.009	0.007	0.009	** 0.031	0.009	0.015	0.009
Saturated fat	* 0.05	0.02	0.01	0.02	** 0.06	0.02	0.01	0.02
Total sugar	0.005	0.005	-0.005	0.005	<0.001	0.005	-0.006	0.005
Sodium †††	0.002	0.002	*-0.005	0.002	0.002	0.002	<0.001	0.002
Energy	0.006	0.004	<0.001	0.003	* 0.074	0.003	0.001	0.004

Abbreviations: GRS, Genetic risk score; Q1, lowest genetic risk quartile ($10.77 < \text{genetic risk score} < 14.53$); Q2, second genetic risk quartile ($14.53 \leq \text{genetic risk score} < 15.21$); Q3, third genetic quartile ($15.21 \leq \text{genetic risk score} < 15.73$); Q4, highest genetic risk quartile ($15.73 \leq \text{genetic risk score} \leq 18.63$); DRV, Dietary Reference Value Index; β , beta-coefficient; SE, standard error

^a Estimated interaction effect on HbA1c% per 10g increase in nutrient interacting GRS within genetic risk quartile adjusted for age, gender, smoking, alcohol, energy intake, physical activity, BMI, diabetes diagnosis and treatment.

† coefficients represent per 100 g increase; †† coefficients represent per 1 point increase in DRV score; ††† coefficients represent per 1g increase in sodium

* p-value <0.05; ** p-value <0.01

4.6 Alcohol interaction with genetic risk score on HbA1c

Table 39 shows the estimated interaction coefficient in HbA1c% per 1 (SD) unit increase in the GRS interacting with alcohol categories adjusted for age, sex, BMI, smoking, physical activity, energy intake, diabetes diagnosis and treatment. The result indicated a higher effect on HbA1c% among participants in the highest alcohol consumption group (>16g alcohol/day) β 1.22 ($p_{\text{interaction}}$ 0.002) (Table 39). No interactions were observed in individuals drinking within recommended allowance or in non-drinkers.

Table 39: The interaction effect on HbA1c% across alcohol allowance categories interacting with genetic risk score (continuous), the Airwave Health Monitoring Study (n=3733)

Category	<i>n</i>	HbA1c (%)		P interaction
		β^a	SE	
No alcohol	767	0.66	0.53	0.21
Within allowance†	1670	0.53	0.28	0.11
Above allowance††	1296	1.22	0.38	0.002

Abbreviations: β , Beta-coefficient interaction effect; SE, Standard error; HbA1c, glycated haemoglobin A1c

^a Estimated interaction effect on HbA1c% per 1 point increase in genetic risk score interacting with alcohol category adjusted for age, gender, smoking, diet quality, physical activity, BMI, diabetes and treatment

† >0 unit and <2 units of alcohol/day; †† > 2 units of alcohol /day

4.7 BMI interaction with genetic risk score on HbA1c

Table 40 shows the estimated interaction coefficient in HbA1c% per 1 (SD) unit increase in the GRS interacting with BMI categories adjusted for age, sex, smoking, physical activity, energy intake, alcohol, diabetes diagnosis and treatment. There was a significant interaction between GRS and BMI indicating a higher effect on HbA1c among obese participants (BMI $>30\text{kg/m}^2$) β 1.90 ($p_{\text{interaction}}$ 0.001) (Table 40). No interactions were seen for healthy weight or overweight individuals.

Table 40: The interaction effect on HbA1c% across body mass index categories interacting with genetic risk score (continuous), the Airwave Health Monitoring Study (n=3,733)

Category	<i>n</i>	HbA1c (%)		P interaction
		β^a	SE	
BMI (18.5- 24.99 kg/m ²)	1,273	0.63	0.38	0.10
BMI (25-29.99 kg/m ²)	1,741	0.53	0.34	0.12
BMI (> 30 kg/m ²)	711	1.90	0.56	0.001

Abbreviations: β , Beta-coefficient; SE, Standard error; HbA1c glycated haemoglobin A1c

^a Estimated interaction effect on HbA1c% per 1 point increase in genetic risk score interacting with BMI category adjusted for age, gender, smoking, alcohol, energy intake, physical activity and diabetes and treatment

4.8 Effect modification of diet quality and genetic risk score on T2D and pre-diabetes

Table 41 shows the relative risk (RR) of genetic score alone and the effect modification of diet. The results showed that diet quality offset the genetic effect on T2D (RR 0.99, 95% CI 0.99-1.00) $p_{\text{interaction}} = 0.04$. There was no significant effect modification observed for pre-diabetes.

Table 41: Estimates of relative risk effect of genetic risk score and the effect of diet quality on type 2 diabetes and pre-diabetes, the Airwave Health Monitoring Study (n=3,733)

	(cases/subcohort)	GRS		GRS*DRV		P interaction
		RR ^a	95% CI	RR ^a	95% CI	
Pre-diabetes*	1,316/2,409	1.14	1.06-1.22	1.00	1.00-1.00	0.08
Type 2 diabetes**	133/3,592	1.06	0.88-1.27	0.99	0.99-1.00	0.04

Abbreviations: GRS, genetic risk score; DRV, dietary reference value score; RR, relative risk; CI, confidential interval

* HbA1c > 5.7% < 6.5% **HbA1c ≥ 6.5% and/or on glucose lowering medication

a. Relative risk and 95% CI based on GENMOD relative risk model adjusted for age, gender, body mass index, smoking, alcohol, energy intake, physical activity

4.1 Discussion

Summary of main findings

- GRS was associated with glycaemic trait HbA1c%, pre-diabetes and T2D outcome
- GRS was not associated with any adiposity markers
- A healthier diet was associated with a reduced effect on HbA1c% in high genetic risk individuals. The interactions were driven by diets high in fibre, carbohydrates, whole grains, fruit and vegetables
- Diets high in energy, total fat, saturated fat as well as excessive alcohol consumption were associated with an increase effect on HbA1c% in high genetic risk groups
- Interactions between GRS and BMI indicated a higher effect on HbA1c among obese participants

Genetic score association with glycaemic trait HbA1c and diabetes risk

The study demonstrated a GRS based on 87 established genetic markers for T2D and glycaemic trait was highly associated with both HbA1c% and T2D risk outcomes. The direction and magnitude of these associations were in line with previous studies (70, 71, 92, 161). The results showed a modest genetic effect size on HbA1c of 0.03% per one unit (SD) increase in GRS, which is similar to effect sizes previously reported by MAGIC and NHGIR (70, 71, 92, 161) (Appendix 6). The difference in mean HbA1c% between individuals in the highest and lowest genetic risk (GRS quartiles 1 vs 4) showed a clinical significant difference between the groups as participants in GRS quartile 4 were defined as being T2D (HbA1c% 6.5) compared to GRS quartiles (HbA1c% 6.4). This difference may have been greater if GRS was studied across ten-tiles rather than quartiles, but the small sample size did not allow for this analysis. Nonetheless, even a modest increase in HbA1c of 1% has been associated with a number of cardiovascular complications and mortality (87). Meta-analysis of 10 prospective cohort studies (n=7,435) found per 1% increase of HbA1c the relative risk increased for CVD by 18% and fatal coronary heart diseases increased by 16% in individuals with T2D (162).

This study also found that the GRS was associated with phenotypes pre-diabetes and T2D. Per 1 (SD) unit increase in the GRS the OR for having T2D increases by 14% (95%CI 1.04-1.24) and OR for having pre-diabetes increases by 9% (95%CI 1.05-1.13), which are in line with other studies (93, 99, 138, 163). Previously published GRS on T2D reported OR between 1.04-1.41 (per one unit increase) in individuals of European ancestry (93, 99, 138, 163). Only one other study (to date) has tested a T2D associated GRS in relation to pre-

diabetes (164). Zyriax et al reported a higher OR 1.57 per one unit increase; however, with much wider 95%CI 1.21-2.10 and their GRS included obesity associated SNPs, which may explain the larger effect size. Compared to T2D, the information on genetic effect on pre-diabetes risk is limited and needs further studies for clarification. Nevertheless, this analysis suggests that genetic markers, which predispose individuals to T2D and hyperglycaemia, may serve as risk markers for development of pre-diabetes. These findings prove the importance of identifying pre-diabetic individuals, as early intervention at this critical stage may halt the development of T2D, MetS and CVD (16, 86, 88).

Variations in reported genetic effects on T2D risk may be due to a number of conditions such as genetic interactions, epigenetic changes, the number and types of SNP included in the GRS. Genetic association studies require a large sample size to have sufficient power to detect an association with a health outcome. This is primarily due to most of these common T2D associated SNPs being derived from GWAS based on very large sample sizes and their effect sizes on T2D are low to moderate. DIAGRAM also showed variation between SNPs' effect sizes on T2D ranging from 1.02 to 1.25 (92), hence a GRS's overall effect size on T2D may vary depending on their combination of these SNPs. This study's GRS was constructed based on only 56% (n=49) T2D associated SNPs, which only represent 62% of all the known T2D associated SNPs (n=79). This may also have contributed to an overall lower genetic effect and power on T2D. Similar variations in GRS's effect on T2D risk have also been reported by other authors (99, 161, 165). Vaxillaire et al (165) evaluated three GRS association with hyperglycaemia and T2D. GRS1 solely T2D associated SNPs (n=65): GRS2 mix T2D SNP and glycaemic trait SNPs (n=79): GRS3 solely glycaemic trait SNPs (n=24). They found the two most inclusive GRS gave a better prediction of both T2D and hyperglycaemia.

The GRS only represent a fraction of the many other genetic markers involved in the development of T2D (70, 92). This study did not investigate other genetic pathways and markers, which may carry a protective effect, related to T2D or related trait-associated variants. Therefore, it cannot rule out gene-gene interactions or other variants, which may mask the susceptibility of T2D outcome, as certain variants only contribute to the genetic heritability of T2D in the context of other "modifying" events (70, 92, 161).

Additionally, there is the possibility that epigenetic changes e.g. DNA methylation (silencing) could have occurred in some of the T2D associated SNPs, hence no association was observed with T2D. Dayeh *et al.* (100) revealed over 853 T2D related genes (1649 CpG sites) were

affected via epigenetic effects from environmental (age, physical activity, adiposity) and external (hormones) factors, resulting in modifications to hyperglycaemia (HbA1c) and glucose-stimulated insulin secretion. This could explain the argument that the effect size in this study's findings are lower as the study population is younger and the participants could carry protection by length of exposure to other clinical risk factors (e.g. physical inactivity, smoking, alcohol).

As discussed, a number of conditions could have affected the genetic effects found in this study. The modest effects of genetic risk on hyperglycaemia and T2D risk have previously been discussed by a number of authors (99, 161, 165). In a review by Lyssenko *et al.* (99) a total of 19 different studies were presented comparing GRS vs. clinical models (age, smoking, obesity etc.) for prediction of T2D risk. They found in all these studies that the clinical models gave higher prediction than the genetic models. The area receiver operating characteristic curve (AUC) on the genetic models ranged from 0.54 to 0.63, indicating limited use for prediction of risk T2D. Compared to 0.61 to 0.95 for the clinical models (including different combinations of clinical markers: age, sex, and BMI in all models and family history of diabetes and fasting glucose). A third of the clinical models reported a small but significant improvement when combining both models (GRS + clinical risk factors).

Available data to date have not yet provided convincing evidence to support use of genetic screening for prediction of T2D in clinical practise. The effect of genetic risk on hyperglycaemia and T2D is only modest due to the complex nature of the disorder. Many environmental factors are involved in its pathways such as lifestyle, keeping a healthy weight and dietary factors and their effects are much stronger, compared to genetics.

However modest the genetic risk effect, it is an important additive risk factor, which can be detected much earlier (at birth) than any other clinical risk factors. The GRS is often uncorrelated with any other traditional risk factors e.g. waist circumference, smoking, age, sex, which were also observed in this study, hence in conjunction they may have a strong clinical implication for prediction and prevention of hyperglycaemia and T2D risk. In the future, this GRS effect may change if rare or novel SNPs with high effect sizes are identified.

This study highlights the methodological challenges in studying the genetic risk associations with glycaemia control and outcome of T2D and pre-diabetes. The results support the evidence that a GRS, which is unrelated to any other risk factors, is associated with glycaemic control and risk of T2D and pre-diabetes in populations of European ancestry. A

GRS in conjunction with other important clinical markers such as BMI, family history and age, may play an important role in clinical practice by early detection of high-risk individuals and by early stage intervention for prevention of T2D.

Genetic interaction with diet and lifestyle factors

This analysis suggests individuals of at increased genetic risk of diabetes can offset their genetic risk by not being obese, not being a heavy drinker and by consuming a high-quality diet. This could have important public health applications if in future genotyping becomes commonplace and higher genetic risk individuals are identified in childhood or early adulthood, as early intervention at this critical stage may halt or delay the development of T2D and CVD in later life (16).

Healthier diets (higher DRV quality score) attenuated the genetic effect on HbA1c among high-risk individuals, which supports the benefits of a healthy diet, especially for genetically susceptible individuals. Similar findings were reported in the randomised control trial (RCT) Prevención con Dieta Mediterránea study (PREDIMED) (n=7,018) which observed interactions between a higher diet quality (Mediterranean diet) and rs7903146 SNP (TCF7L2 gene) in relation to fasting glucose (166). Showing that individuals of European ancestry at high-risk of diabetes consuming a high diet quality could eradicate their genetic risk. Another RCT Preventing Overweight Using Novel Dietary Strategies (POUNDS LOST) (148) also found participants (white Americans) with a higher GRS had a greater reduction in fasting insulin (β -0.025%) and HbA1c (β -0.02%) when consuming a healthier diet (high in carbohydrates, protein and low in energy). Dayeh *et al.* (100) also concluded these interactions were most beneficial in higher genetic risk groups. They saw effects of genetic-lifestyle interactions being higher in genetically high-risk individuals compared to low-risk, suggesting the more risk alleles present the more susceptible these individuals are in relation to effect modification from external lifestyle factors.

Due to lack of published interactions studies between dietary scores and GRS it is challenging to compare and discuss these findings. Most nutritional genomic studies published to date are conducted on single SNP and interactions with single nutrient groups, in relation to T2D and the few observational studies published based on dietary scores have been inconclusive (93, 107, 148, 163, 167).

This study found that the interaction between GRS and overall diet quality was driven by

diets high in fruit, vegetables, fibre, carbohydrates, whole grains and low in total fat, saturated fat and energy, which also aligns with NICE and WHO's dietary guidelines for prevention and management of T2D (12, 87).

The single nutrient analyses demonstrated the genetic effect on HbA1c% was offset in individuals consuming a high intake of fibre, CHO, fruit and vegetables, whole grains. Whereas, the effect of a high intake of total fat, saturated fat and energy was worsened in individuals with a high GRS. These findings are in line with previously reported studies (109, 168, 169). The European Prospective Investigation of Cancer (EPIC) study found interaction with whole grain intake, which inversely modified the association between rs7903146 (TCF7L2 gene) and T2D risk (109). The Malmö Diet and Cancer Study showed interactions between intake of dietary fibre and SNP rs7903146 (TCF7L2); T2D incidence was lower by 32% among the CC carriers with the highest fibre intake (OR: 1.24, 95% CI 1.04-1.47) compared to those with the lowest fibre intake (OR: 1.56, 95% CI 1.31-1.86), $p_{\text{interaction}} = 0.049$ (168). The same study reported a reduced effect of SNP rs10423928 (GIPR gene) on T2D incidence in individuals who had the highest intake of carbohydrate (OR 1.23, 95% CI: 5-39%) compared to those with the lowest carbohydrate intake (OR 1.69, 95% CI: 29-86%). Similar associations were seen in individuals with the lowest total fat intake compared to those with highest fat (169). However, to our knowledge no study has looked at the poly-genetic effect of these diet interactions.

Interactions between GRS and saturated fat are also reported in this thesis and support findings previously reported (110, 113, 114). Three recent studies investigating interactions between saturated fat and T2D associated SNPs showed a higher dietary intake of saturated fat increased the genetic effect on insulin resistance in individuals carrying the risk alleles (113, 114). One of the studies proposed the theory behind this mechanism was related to fatty acids ability to bind specific isoforms, which activate SNPs in the PPAR gene (110). This binding varied according to the type of fatty acid. Increased intake of n-3 and n-6 PUFAs had a protective effect via PPAR γ activation, which regulates insulin sensitization and SFA had an opposite effect

The interactions observed in this study between nutrient group, fibre, CHO and GRS could have been driven by a diet high in whole grains and fruit/vegetables, as similar protective interactions were observed in these food groups on the genetic effect on HbA1c in other studies (76, 78, 109, 137). The EPIC study also found whole grain intake inversely modified

the association between rs7903146 (TCF7L2 gene) and T2D risk (109). A diet high in fruit and vegetable intake has not been reported to interact directly with T2D risk alleles. However, both INTERHEART and FINRISK study found significant interaction between the genetic markers for CVD risk and diet (prudent diet high in fruit and vegetables) (76, 137). The effect of the genetic risk on CVD outcome was significantly lower in individuals with 'healthier diets' compared to those with a lower intake of fruit and vegetables (76, 78).

A higher sodium intake was also found to interact with GRS in the second lowest genetic risk group (quartiles 2). This finding was not possible to compare as no previous nutritional genomic studies have been published on sodium intake. Though sodium has not been reported to have a direct relationship with T2D, its known from the literature to be associated with other metabolic markers, e.g. BP which is associated with an increased risk of T2D (12, 21), which contradicts this study's findings. Multiple regression analysis between mean sodium intake and metabolic markers including HbA1c, showed no significance associations in this dataset, which may be biased by methodological issues (as previously mentioned) in recording total sodium intake. Added salt to foods not recorded, mean sodium intake in this study only represent extrinsic sodium in foods and not total intake. Therefore, any interactions observed could be due to multi-collinearity with other "favourable" food groups, such as packaged bread and cereals, which are the biggest contribution of extrinsic sodium in the UK diet (170).

Lack of relevant published genetic-diet interaction studies makes it challenge to draw comparisons and support this study's findings. Many of the discussed studies are not directly comparable to this study as they were conducted on single SNPs analysis and mainly reported on T2D outcome or other glycaemic traits. However, the studies confirm observations of interactions between the genetic risk on T2D risk and nutritional components (CHO, fibre, wholegrain, total fat and saturated fat) such as those in this thesis.

The biological mechanism behind these observed genetic-diet interactions remains unclear. However, the evidence from RCTs, which have studied the direct relationships between diet and hyperglycaemia, does support a diet high in CHO, fibre, fruit, vegetables and whole grains as linked with a better glycaemic control and reduction of T2D (12, 87, 98). A high intake of fibre has repeatedly been shown to reduce blood glucose and insulin levels in individuals with impaired glucose tolerance and T2D (156-158). The mechanism behind this relationship is primarily related the digestion process of fibre, which is known to slow down

the gastric and intestinal emptying. Together it results in a prolongation and reduction of intestinal glucose absorption and thus lowers the demand for postprandial insulin secretion, which helps blood sugar control (156-159). A similar mechanism is believed to support the increased intake of whole grains, vegetables and fruits (all high in fibre) in relation to glycaemic control and reduced risk of progression of impaired glucose tolerance to T2D (90, 156, 157). Furthermore, fruit and vegetables' high content of essential vitamins, minerals and antioxidants have been linked with protective properties pathways of T2D development: anti-inflammatory, improved insulin productivity and lowering BP (171-173). In contrast, both NICE and WHO reported that a diet high in total fat, saturated fat and energy is linked with increased blood glucose (HbA1c), insulin resistance and T2D both directly and via mediated adiposity markers (12, 87, 174). Both reports gathered evidence from a number of RCTs showing a direct relationship between reductions in total energy, total fat and saturated fat intake were linked with a reduction in HbA1c%. Barnard *et al.* (155) showed the effect of changing to a low-fat vegan diet (low in total fat and saturated fat) for 22 weeks in patients with T2D, was associated with a 1.23% reduction in HbA1c% ($p < 0.0001$). Another study found replacing saturated fats with unsaturated fats led to improved glucose tolerance and enhanced insulin sensitivity (175) as long as total fat intake did not exceed 37% of total energy. The mechanism behind high fat (saturated fat) intake and T2D risk is related to several pathways affecting obesity, insulin resistance, inflammation and elevated HbA1c (12, 153, 154). RCTs have demonstrated that high intake of total fat and saturated fat build up in the muscle cells and around the organs affecting insulin signalling (176). Also break down of saturated fat increases circulation of TG, LDL-cholesterol and free radicals, which is associated with increased inflammation, insulin sensitivity and glycaemic control (177).

This study demonstrates that diet quality attenuates the genetic effect on HbA1c among participants with a healthier diet. Even though the biological mechanism underlying the observed interactions remains unclear at this point, several lines of evidence suggest that such interactions between single nutrient groups are plausible. In this study, most of the diet interactions were observed in the highest genetic risk groups e.g. higher number of risk alleles, which suggests that diet interactions are more likely to happen in high-risk individuals. Another explanation could be that these interactions are more likely to be captured (better power) with a higher number of measured risk alleles. This supports the use of a comprehensive GRS to improve power in such interaction studies. It is important to mention that most of this thesis' results were only modest and the significance levels were modest (p -

values 0.05 and 0.01). Such analyses are prone to false-positive detections; hence such studies may need even lower threshold significance level p-values. Furthermore, these results were limited by the cross-section design of the study and replication studies and longitudinal studies are needed to fully explore such interaction between diet and genetics in relation to T2D risk.

However modest the genetic risk is on HbA1c at a clinical and individual level, the attenuation by a healthy diet is a valuable modifiable factor at a public health level. Early screening and intervention of high-risk individuals may not only halt development of T2D, but also help prevent of a number of CVD outcomes and mortalities (16, 86, 88, 162). The UK Prospective Diabetes study (n=3,642) found 1% reduction in HbA1c was associated with a reduction in relative risk of T2D (21%, 95%CI 12-24%) and number of CVD complications (37%, 95%CI 33-41%) and mortality from diabetes (21%, 95%CI 15-27%) (178).

Interaction between alcohol consumption and genetic risk on HbA1c

This study also found interactions between GRS and alcohol consumption. Individuals consuming above the recommended alcohol allowance (>16g/day for men and women) were associated with an increase in HbA1c by 22% per one unit increase in GRS. No additional increased risk was observed for individuals consuming a moderate alcohol intake ($\leq$ 16g/day). There have been no published gene-alcohol interaction studies in relation to glycaemic traits or T2D risk to help confirm these findings. However, the findings are in line with other studies, which have investigated the direct relationship between alcohol consumption, HbA1c and T2D. The evidence shows that there is an increased risk of T2D in heavy drinkers, while moderate alcohol intake may be protective, which also forms part of the WHO and NICE guidelines for prevention of T2D (12, 87).

Excessive and long-term alcohol consumption have been linked with an increased HbA1c and risk of T2D. The Finnish Twin Cohort (n=22,778) also found that high alcohol consumption (>20 g/day) was associated with an almost triple-fold increased incidence of T2D (179). A review by Emanuele *et al.* (180) on consequences of alcohol use in diabetes discussed the evidence on how long-term alcohol usage increases risk of T2D via alternative pathways such as elevated waist circumference, BP and altered lipid metabolism. A case-control study (n=81) also found chronic excessive alcohol intake (40-45 g/day) was associated with higher fasting and postprandial glucose concentrations and higher HbA1c in T2D patients (181). Though

these above studies are not directly comparable to this study's findings, it does suggest that high alcohol intake above >20-50 g/day may increase HbA1c and risk of T2D. Whereas this study's findings suggest that this alcohol threshold may need to be even lower (<16 g/day) in individuals with a high genetic risk of T2D or pre-diabetes.

In line with other studies this study did not observe an interaction (increased risk) in individuals with a moderate consumption of alcohol. Evidence suggests that moderate alcohol consumption may improve insulin sensitivity in non-diabetics and reduce their risk of T2D. A large meta-analysis including 15 prospective cohort studies (n=369,862) investigated the association between alcohol consumption and T2D risk (182). They showed a U-shaped relationship with a ~30% reduction of T2D risk in consumers of 6-48 g alcohol/day compared to non-consumers. No reduction was observed for heavy alcohol consumers (> 48g/day). The protective mechanisms underlying the relationship between alcohol consumption and T2D risk have been demonstrated in a number of RCTs and observational studies. A large meta-analysis of 14 intervention studies (n=678) showed alcohol consumption of 10-40 g/day was associated with a reduction in HbA1c and fasting insulin levels in non-diabetic individuals (183). Similar findings were reported in a prospective study (n=820), which found regular (1-50g alcohol/day) to high alcohol consumption (50-100 g/day) was associated with increased insulin sensitivity with lower plasma insulin levels and HbA1c compared to non-drinkers (184). Other plausible mechanisms playing a role in alcohol's relationship with HbA1c and T2D risk may be due to other gene-alcohol interactions. There is increasing evidence that the other gene-gene interactions impact the effect of alcohol consumption on T2D risk (185, 186). A nested case-control in the Nurses' Health study found that the relationship between alcohol consumption and T2D risk was modified (reduced) by alcohol dehydrogenase 1c (ADH1C gene) polymorphism (OR 0.45: 95%CI 0.33-0.63) (185). They found ADH1C allele slows down alcohol's oxidation rate and attenuates a lower T2D risk amongst moderate drinkers. This could suggest that the association between alcohol and T2D may be causal but mediated by ADH1C gene.

The above studies showed that a moderate alcohol intake (1-48 g/day) improves insulin sensitivity and protect against increased HbA1c and T2D risk. This study suggests consumption of alcohol may need to be even lower (<16 g/day) in individuals with a high genetic risk of T2D or pre-diabetes to obtain the protective effects of alcohol. However, it is plausible that this figure may be higher as alcohol consumption may have been under-reported.

Interaction between BMI and genetic risk on HbA1c

There was a significant interaction between GRS and BMI indicating a higher effect on HbA1c among obese participants (BMI <30 kg/m²) (β 1.9, $p_{\text{interaction}}$ 0.001), which is in line with previously reported findings. The Prevención con Dieta Mediterránea RCT (n 7,018) also reported interactions between GRS and BMI, seeing a stronger genetic effect in obese participants (108). Another RCT Inter99 (n 6,127) demonstrated an increased BMI over a 5-year period resulted in an increased genetic effect on fasting plasma glucose. The largest effect by this interaction was observed in individuals with the highest weight gain, which increased fasting glucose percent by 0.23 per one unit increase in GRS (93). In reverse, individuals who lost weight attenuated this genetic effect by -3.3%, which suggests the genetic risk may be reversible with targeted weight loss interventions.

This study cannot confirm which mechanisms or pathways may explain the observed interactions between GRS and BMI. Knowledge about the molecular mechanisms linking environmental factors and T2D still remains limited. A genome-wide DNA methylation analysis of human pancreatic islets (non-diabetic) demonstrated that increased BMI altered DNA methylation and expression at 16 CpG sites, which contribute to perturbed insulin and glucagon secretion (100). Obesity is an important contributor and predictor of T2D through several direct and mediated mechanisms including insulin resistance, inflammation and increased adiposity tissue around organs such as liver and pancreas (177). A plausible theory may be gene-gene interaction with variants in obesity genes, increase the susceptibility for T2D. Several GWAS have detected strong associations of common SNPs at obesity genes FTO and MC4R, which have shown to confer a increase of T2D risk in individuals of European ancestry (69, 70, 92, 96, 97). Though these SNPs were excluded in this study's GRS and the analysis was adjusted for BMI, it still remains unknown if the participants were carrier for other high-risk obesity genes.

These findings suggest that environment may counter the genetic risk of T2D and support a healthy diet, weight and moderate alcohol intake, as part of public health prevention of T2D, especially in high genetic risk groups. However, longitudinal replication studies are needed to support these findings and to evaluate if implementation of therapeutic dietary recommendations for pre-diabetics and high genetic risk groups require a 'personalised nutrition' approach.

CHAPTER 5: STRENGTHS, LIMITATIONS AND CONCLUSIONS

5.1 Summary of findings

Diet quality association with cardio-metabolic risk

This cross-sectional study demonstrates that the DRV index and NP model evaluate a person's overall diet quality in relation to the UK national dietary guidelines and that the DRV index is inversely associated with adiposity markers and total cholesterol, HbA1c and T2D outcome. A stronger adherence to dietary guidelines may play an important role in regulation of metabolic risk factors for CVD and diabetes. These results support a beneficial role for promoting UK dietary guidelines on metabolic and cardiovascular health in public health.

Genetic association with HbA1c and diabetes risk and interactions with lifestyle factors

The study also demonstrates a GRS based on 87 established genetic markers for T2D and glycaemic trait is highly associated with both HbA1c% and phenotypes pre-diabetes and T2D. The genetic effect on HbA1c is reduced in individuals consuming a high quality diet (high in complex carbohydrates, fibre, fruit and vegetables and wholegrain). The study also found there was a significant interaction between GRS and BMI indicating a stronger effect (almost two-fold) on HbA1c among obese participants ($\text{BMI} > 30 \text{ kg/m}^2$) compared to non-obese individuals. These findings support the importance a healthy diet, moderate alcohol consumption, and maintaining a healthy weight ($18.5\text{-}30 \text{ kg/m}^2$) as part of public health advice for lowering risk of T2D, especially in high genetic risk groups.

5.2 Strengths and limitations

DRV index association with diet quality and cardio-metabolic risk

The main strength of this study is the Airwave Health Monitoring study's extensive data collection on a relatively large sample of British adults. The data allows for in-depth insights into overall diet quality in relation to a number of metabolic risk factors, compared with usual measures of diet in cohort studies. Use of a 7-day food diary is known to limit measurement errors and provide accurate estimates of individual diet intake (61) compared to other methods such as food frequency questionnaires or 24-hr recall (187, 188). Another strength is the rigorous quality control of the dietary data: a study specific operational protocol,

standardised coder training and quality control audit cycle helped maintain a low mean code error rate per food diary (3.7%). Due to lack of published data from other nutritional studies it is difficult to draw comparisons.

The first and major limitation of this study is the use of a cross-sectional study design, which cannot provide evidence of a causal relationship between diet quality (DRV index) and metabolic risk factors. These results could be due to the reverse causality in individuals who have made dietary improvements after being informed of a medical condition, such as the absence of alcohol intake may be due to previously high consumptions and related health concerns. Another limitation lies in the non-generalisation of the results due to those who completed the food diaries may introduce selection bias, leading to underestimation of the strength of the association. Participants volunteering to complete food diaries may be more health conscious and vary in lifestyle characteristics. However, the diet and clinical data reported in this study is comparable to the general UK population. The dietary data (mean energy and macronutrient intake) is comparable to the NDNS (21). The mean daily energy intake reported in this study was 1,673 and 2,071 kcal for women and men respectively compared with 1,560 and 2,032 kcal in the NDNS. The Airwave Health Monitoring Study population is a young cohort, which may differ from the general population in health outcomes. However, the sample's prevalence of cardiovascular risk factors is both comparable with the total cohort (115), and representative of the general UK population (Health Survey for England 2012) (189): hypertension (28.0% vs Health Survey 29.0%), obesity (20.0% vs Health Survey 19.5%) and diabetes (5.0% vs Health Survey 5.8%). Only dyslipidaemia is lower in the sample 4.75%, compared to two thirds of general population (190).

Another limitation in this study is the amount of self-reported data such as physical activity, alcohol and diet, which may have introduced systematic bias of misreporting. Energy under-reporting was investigated and the figures and characteristics are in line with those reported in other larger UK epidemiological studies NDNS and Norfolk-EPIC (61, 191). The characteristics of the under-reporters show they were more likely to be participants from lower socio-economic status (lower education levels, lower job ranks). They were more likely to be men, have a higher BMI and reported a higher METs score. These traits are also similar to those reported in the NDNS. It has been suggested calculating METs using the standard resting oxygen uptake may overestimate energy expenditure in overweight and obese individuals by up to 39% (192). Systematic over-estimation of METs in obese participants

may result in misclassification of PAL category, subsequently over-estimating and biasing classification of under-reporting in obese participants. The precision of the Goldberg method has also been criticised in regard to its sensitivity for possible bias introduced by overestimation of participants energy expenditure (125). Furthermore, the Goldberg method makes no distinction between varying degrees of misreporting. Therefore, it is important to carefully investigate the dataset when using the dietary data to investigate health outcomes as it may be associated or biased by total energy intake. This limitation may be addressed by conducting sensitivity analyses and to adjust for energy intake (193, 194). The DRV index was adjusted for energy. And sensitivity analyses in this study found no differences or confounding effect of the under-reporters between the association of DRV index and metabolic risk factors. Ideally misreporting should have been investigated further by other validating methods such as doubly labelled water and urine biomarkers (e.g., sodium, potassium and nitrogen). However, these analyses were not available for this study and the urine samples of the present study had not yet been processed for use of biomarkers. Furthermore, the dietary data quality control method used in the study did not account for missed food items (codes) into the equation of assessing the total code error rate per diary. This is a limitation which should be considered for future nutritional studies to help address misreporting.

Genetic interaction with diet and lifestyle factors

The main strength of this genetic study is the extensive data collection on a large sample of white British population (of European ancestry), which provides a unique investigation of the relationship between genetics, diet and T2D risk. Its large data collection allows for sufficient statistical power to analyse a wide range of contributing risk factors for T2D. Another strength is the use of a comprehensive GRS, which has several advantages over single markers by increasing power and predictive accuracy in studies of common polygenetic metabolic disorders such as hyperglycaemia (69, 137, 140). This is especially important for interaction studies, as they require high power to detect genetic interactions (108, 163). Secondly, the strength of the GRS is that it is based on well-established common polygenic markers found to be significant with T2D and hyperglycaemic trait, replicated in studies both as single markers and in GRS. The rigorous inclusion criteria to the SNP selection and quality control ensured only independent phenotype/trait specific markers (exclusion of obesity SNPs) were included in the GRS.

The genetic markers included in the GRS were based on GWAS with HbA1c%, which represents a more stable and longer-term indicator of glycaemic status compared with fasting glucose (84, 87). Also, HbA1c represents an average blood glucose concentration over the previous 8 to 12 weeks, or the average lifespan of the erythrocyte. The heritability of HbA1c levels is relatively high (47% to 59%) when compared with fasting glucose (34% to 36%) or oral glucose tolerance test (33%) (95) and thus amenable to genetic analysis. Thus HbA1c is a good marker for studying longer term implications and associations with both phenotype T2D and genetic markers (71).

The main limitation to the Airwave Health Monitoring study is the epidemiological design. A cross-sectional study only provides a snapshot of the genetic effect related to T2D, and it is not clear from this study to what extent the genetic risk may affect changes over time. Cross-sectional data does not permit examination of the temporality of the observed associations (cause-effect), which is important when studying the longitudinal dimension of T2D development and the complexity of its progression. Therefore, these findings should be replicated in longitudinal studies and RCTs to determine the prospectively genetic prediction on T2D risk and if this risk is truly modified by interactions between genetics, diet and lifestyle factors. Such studies should ideally include advanced statistical analysis and prediction models, for example, AUC or Bayesian method to be able to assess the true effects and predictions of the GRS on T2D, and discuss the effectiveness of such tests in clinical practice.

The genetic interaction analyses with diet were not only limited by lack of published evidence, i.e. comparison studies to support this study's findings and to help understand the mechanisms behind these interactions. Limitations to the interaction analyses also exist in the genotyping method used in this study. The Illumina Infinium Human Exome Microarray chip does not measure the actual gene expression (mRNA expression analysis) and is therefore unable to test whether there is an actual expression or suppression of the SNPs. To fully assess for the effect of diet or other lifestyle factors at the transcription level, it is important to support this study findings in longitudinal studies and RCTs using genotyping technologies such as genome-wide DNA methylation profiling and microarray mRNA expression analysis. These methods allow assessment of gene expression (at the level of DNA methylation) pre- and post-exposure. However, these methods are very time consuming and expensive to conduct on a large sample size.

The representation of genetic markers in the GRS also has certain limitations. Though the GRS is one of the largest scores (n 87 SNPs) constructed to date, this still does not represent the entire genetic variation. The quality control LD pruning excluded almost 40% (n=64) SNPs, which were in high LD pairwise ($r^2 > 0.8$). This may have resulted in exclusion of important SNPs e.g. with higher effect size on T2D or glycaemic traits than those included in the GRS. Overall, this may have decreased the effect and power in this study. Sensitivity analyses with different composed GRS and testing each SNP effect size on HbA1c or T2D in this dataset could have helped to determine which SNPs held the highest association and to include in the GRS. Furthermore, the GRS excluded rare SNPs and SNPs located outside a mapped candidate gene, mapped loci or identified SNPs based on selection criteria in the literature review. Many of these markers may still have been associated with outcomes or clinical pathways of known phenotype and could have had a high effect size, explaining some of the missing heritability. Therefore, this study's findings should be supported by a replication study addressing the above-mentioned limitations, performing sensitivity analyses based on different GRS, and including rare markers and SNPs located outside known loci.

5.3 Future work

To understand the public health implication of this thesis' findings on diet quality score influence on T2D and cardio-metabolic risk, and its application in epidemiological studies further work is needed. Though the single dietary components (incorporated into the DRV index) have been studied and replicated in many longitudinal studies and RCTs, the application as an overall dietary score has only been demonstrated in this study. More replication studies are needed to validate these findings and to confirm the effectiveness of the DRV index for assessing cardio-metabolic health. Such longitudinal studies and RCTs should also aim to address previously mentioned limitations. The Airwave Health Monitoring study is currently collecting additional data and when follow up data is available, ideally the dietary and genetic analyses should be followed to permit temporal investigations in relation to cardio-metabolic health and T2D risk.

There is a need to evaluate if implementation of therapeutic dietary recommendations for groups at genetically higher risk for T2D would benefit from such a 'personalised nutrition' approach. In order to critically test such effects replication studies are needed: RCTs measuring a wider range of clinical biomarkers and genetic markers to fully understand which of the pathways are modifiable e.g. beta cell function, insulin resistance. In order to address limitations of both study design and genotyping methods, these studies need to be

longitudinal with several follow up points to capture the effects and include high-throughput, whole genome sequencing methods together with microarray mRNA analysis.

5.4 Conclusion

In conclusion, this study suggests that adherence to the UK dietary guidelines measured by the DRV index was inversely associated with a number of metabolic risk factors, adiposity markers and T2D in a UK population. Furthermore, the study supports the application of a simple population-specific dietary score in epidemiological studies to investigate dietary quality in relation to metabolic health. However, more studies, especially longitudinal studies, are needed to support these findings and to confirm the effectiveness of the use of DRV index to measure metabolic risk factors.

The comprehensive GRS was associated with glycaemic control (HbA1c) and risk of T2D and pre-diabetes in a British population of European ancestry. The findings suggest that environmental/lifestyle factors may help counter the genetic risk of T2D and lend support to policies to promote a healthy diet, healthy weight and moderate alcohol intake, as part of public health measures to prevent T2D, especially in groups genetically at higher risk. However, longitudinal replication studies are needed to support these findings.

REFERENCES

1. Alberti K, Eckel R, Grundy S, Zimmet P, Cleeman J, Donato K. Harmonizing the metabolic syndrome. A joint interim statement of the IDF Task Force on Epidemiology and Prevention; NHL and Blood Institute; AHA; WHF; IAS; and IA for the Study of Obesity. *Circulation*. 2009;120(16):1640-5.
2. Gami AS, Witt BJ, Howard DE, Erwin PJ, Gami LA, Somers VK, et al. Metabolic syndrome and risk of incident cardiovascular events and death: a systematic review and meta-analysis of longitudinal studies. *Journal of the American College of Cardiology*. 2007;49(4):403-14.
3. Sattar N. Why metabolic syndrome criteria have not made prime time: a view from the clinic. *International Journal of Obesity*. 2008;32:S30-S4.
4. Executive summary of the third report of the National Cholesterol Education Program (NCEP) expert panel on Detection, Evaluation, and Treatment of high blood cholesterol in adults (Adult Treatment Panel III). *JAMA: the journal of the American Medical Association*. 2001;285(19):2486.
5. International Diabetes Federation. A new IDF worldwide definition of the metabolic syndrome: the rationale and the results. *Diabetes Voice*. 2005;50(3).
6. Despres JP, Lemieux I. Abdominal obesity and metabolic syndrome. *Nature*. 2006;444(7121):881-7.
7. Tillin T, Forouhi N, Johnston DG, McKeigue PM, Chaturvedi N, Godsland IF. Metabolic syndrome and coronary heart disease in South Asians, African-Caribbeans and white Europeans: a UK population-based cross-sectional study. *Diabetologia*. 2005;48(4):649-56.
8. Lawlor DA, Smith GD, Ebrahim S. Does the new International Diabetes Federation definition of the metabolic syndrome predict CHD any more strongly than older definitions? Findings from the British Women's Heart and Health Study. *Diabetologia*. 2006;49(1):41-8.
9. Hossain P, Kavar B, El Nahas M. Obesity and diabetes in the developing world—a growing challenge. *New England Journal of Medicine*. 2007;356(3):213-5.
10. Blüher M. The distinction of metabolically 'healthy' from 'unhealthy' obese individuals. *Current opinion in lipidology*. 2010;21(1):38-43.
11. World Health Organisation. WHO Guidelines: Sodium intake for adults and children. Approved by the Guidelines Review Committee. Geneva: WHO; 2012.
12. World Health Organisation. Diet, nutrition and the prevention of chronic diseases. Technical report series; 2003. Contract No.: ISSN 0512-3054.
13. Mainous AG, Tanner RJ, Baker R, Zayas CE, Harle CA. Prevalence of prediabetes in England from 2003 to 2011: population-based, cross-sectional study. *British Medical Journal open*. 2014;4(6):e005002.
14. International Diabetes Federation: Diabetes atlas. 2013. Assess date: 15.05.2016. Available from: <http://www.idf.org/diabetesatlas/>
15. Morris D, Khunti K, Achana F, Srinivasan B, Gray L, Davies M, et al. Progression rates from HbA1c 6.0–6.4% and other prediabetes definitions to type 2 diabetes: a meta-analysis. *Diabetologia*. 2013;56(7):1489-93.
16. Chamnan P, Simmons RK, Forouhi NG, Luben RN, Khaw K-T, Wareham NJ, et al. Incidence of Type 2 Diabetes Using Proposed HbA1c Diagnostic Criteria in the European Prospective Investigation of Cancer–Norfolk Cohort Implications for preventive strategies. *Diabetes Care*. 2011;34(4):950-6.
17. Park YW, Zhu S, Palaniappan L, Heshka S, Carnethon MR, Heymsfield SB. The metabolic syndrome: prevalence and associated risk factor findings in the US population from the Third National Health and Nutrition Examination Survey, 1988-1994. *Archives International Medicine*. 2003;163(4):427-36.
18. Grundy SM. Metabolic syndrome: a multiplex cardiovascular risk factor. *Journal of Clinical Endocrinology & Metabolism*. 2007;92(2):399-404.
19. Babio N, Bullo M, Salas-Salvado J. Mediterranean diet and metabolic syndrome: the evidence. *Public Health Nutrition*. 2009;12(9A):1607-17.

20. Lassale C, Galan P, Julia C, Fezeu L, Hercberg S, Kesse-Guyot E. Association between Adherence to Nutritional Guidelines, the Metabolic Syndrome and Adiposity Markers in a French Adult General Population. *PLoS ONE*. 2013;8(10):e76349.
21. Department of Health. Report of the Cardiovascular Review Group, Committee on Medical Aspects of Food Policy. *Nutritional Aspects of Cardiovascular Disease*. 1994(No.46).
22. Scarborough P, Arambepola C, Kaur A, Bhatnagar P, Rayner M. Should nutrient profile models be 'category specific' or 'across-the-board'? A comparison of the two systems using diets of British adults. *European Journal of Clinical Nutrition*. 2010;64(6):553-60.
23. Committee on Medical Aspects of Food Policy. *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom: Report of the Panel on Dietary Reference Values of the Committee on Medical Aspects of Food Policy*. London: H.M. Stationery Office, 1991.
24. Scientific Advisory Committee on Nutrition. *Dietary Reference Values for Energy and Nutrients*. Public Health England. 2011.
25. Scientific Advisory Committee on Nutrition. *Carbohydrates and Health report England Public Health*, editor. 2015. Access date: 01.02.2016. Available from: <https://www.gov.uk/government/publications/sacn-dietary-reference-values-for-energy>
26. Levy L, Tedstone A, editors. *UK Dietary Policy for the Prevention of Cardiovascular Disease*. Healthcare; 2017: Public Health England.
27. Alcohol Policy Team, Department of Health. *How to keep health risks from drinking alcohol to a low level*. Government response to the public consultation. 2016.
28. Waijers PM, Feskens EJ, Ocké MC. A critical review of predefined diet quality scores. *British Journal of Nutrition*. 2007;97(02):219-31.
29. Moeller SM, Reedy J, Millen AE, Dixon LB, Newby P, Tucker KL, et al. Dietary patterns: challenges and opportunities in dietary patterns research: an Experimental Biology workshop, April 1, 2006. *Journal of the American Dietetic Association*. 2007;107(7):1233-9.
30. Conlin PR, Chow D, Miller Iii ER, Svetkey LP, Lin P-H, Harsha DW, et al. The effect of dietary patterns on blood pressure control in hypertensive patients: results from the dietary approaches to stop hypertension (DASH) trial. *American Journal of Hypertension*. 2000;13(9):949-55.
31. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Current Opinion in Lipidology*. 2002;13(1):3-9.
32. Toft U, Kristoffersen LH, Lau C, Borch-Johnsen K, Jorgensen T. The Dietary Quality Score: validation and association with cardiovascular risk factors: the Inter99 study. *European Journal of Clinical Nutrition*. 2007;61(2):270-8.
34. Sacks FM, Svetkey LP, Vollmer WM, Appel LJ, Bray GA, Harsha D, et al. Effects on blood pressure of reduced dietary sodium and the Dietary Approaches to Stop Hypertension (DASH) diet. *New England Journal of Medicine*. 2001;344(1):3-10.
35. Kant AK. Dietary patterns and health outcomes. *Journal of the American Dietetic Association*. 2004;104(4):615-35.
36. Esposito K KC, Panagiotakos DB, Giugliano D. Mediterranean diet and weight loss: meta-analysis of randomized controlled trials. *Metabolic Syndrome and Related disorders*. 2011;9(1).
37. Nicklas TA, O'Neil CE, Fulgoni VL. Diet quality is inversely related to cardiovascular risk factors in adults. *The Journal of Nutrition*. 2012;142(12):2112-8.
38. Seymour JD, Calle EE, Flagg EW, Coates RJ, Ford ES, Thun MJ. Diet quality index as a predictor of short-term mortality in the American Cancer Society Cancer Prevention Study II Nutrition Cohort. *American Journal of Epidemiology*. 2003;157(11):980-8.
39. Chiuve SE, Sampson L, Willett WC. Adherence to the Overall Nutritional Quality Index and Risk of Total Chronic Disease. *American Journal of Preventive Medicine*. 2011;40(5):505.
40. Schröder H, Marrugat J, Vila J, Covas MI, Elosua R. Adherence to the traditional Mediterranean diet is inversely associated with body mass index and obesity in a Spanish population. *The Journal of Nutrition*. 2004;134(12):3355-61.
41. Kastorini C-M, Milionis HJ, Esposito K, Giugliano D, Goudevenos JA, Panagiotakos DB. The effect of Mediterranean diet on metabolic syndrome and its components: a meta-analysis of 50

- studies and 534,906 individuals. *Journal of the American College of Cardiology*. 2011;57(11):1299-313.
42. Trichopoulou A, Orfanos P, Norat T, Bueno-de-Mesquita B, Ocké MC, Peeters PH, et al. Modified Mediterranean diet and survival: EPIC-elderly prospective cohort study. *British Medical Journal*. 2005;330(7498):991.
 43. Newby P, Tucker KL. Empirically derived eating patterns using factor or cluster analysis: a review. *Nutrition Reviews*. 2004;62(5):177-203.
 44. Park YW ZS, Palaniappan L, Heshka S, Carnethon MR, Heymsfield SB. The MetS: prevalence and associated risk factor findings in the US population from Thrid National Health and Nutrition Examination Survey, 1988-1994. *Archive International Medicine*. 2003.
 45. Quinio C, Biloft-Jensen A, De Henauw S, Gibney MJ, Huybrechts I, McCarthy SN, et al. Comparison of different nutrient profiling schemes to a new reference method using dietary surveys. *European Journal of Nutrition*. 2007;46 Suppl 2:37-46.
 46. World Health Organisation. Nutrient Profiling - Report of WHO/IASO Technical Meeting. 2010.
 47. Labonte MA M, Rayner M, L'Abbe M. Systematic review of nutrient profile models developed for nutrition-related policies and regulations aimed at noncommunicable disease prevention. National Institute for Health Research. 13 August 2015.
 48. Sacks G, Rayner M, Stockley L, Scarborough P, Snowdon W, Swinburn B. Applications of nutrient profiling: potential role in diet-related chronic disease prevention and the feasibility of a core nutrient-profiling system. *European journal of clinical nutrition*. 2011;65(3):298-306.
 49. Townsend MS. Where is the science? What will it take to show that nutrient profiling systems work? *The American journal of clinical nutrition*. 2010;91(4):1109S-15S.
 50. Tzoulaki I, Patel CJ, Okamura T, Chan Q, Brown IJ, Miura K, et al. A nutrient-wide association study on blood pressure. *Circulation*. 2012;126(21):2456-64.
 51. Masset G, Scarborough P, Rayner M, Mishra G, Brunner EJ. Can nutrient profiling help to identify foods which diet variety should be encouraged? Results from the Whitehall II cohort. *British Journal of Nutrition*. 2015;113(11):1800-9.
 52. Rayner M. Scarborough P. and Lobstein T. The UK Ofcom Nutrient Profiling Model. British Heart Foundation Health Promotion Research Group, Department of Public Health, University of Oxford, International Obesity Task Force, London. 2009. Access date: 01.12.2015. Available from: <https://www.ndph.ox.ac.uk/cnpn/files/about/uk-ofcom-nutrient-profile-model.pdf>
 53. Rayner M, Scarborough P, Williams C. The origin of Guideline Daily Amounts and the Food Standards Agency's guidance on what counts as 'a lot' and 'a little'. *Public Health Nutrition*. 2004;7(04):549-56.
 54. Rayner M, Scarborough P, Stockley L. Nutrient profiles: options for definitions for use in relation to food promotion and children's diets. Food Standard Agency, London. 2004.
 55. Volatier JL, Biloft-Jensen A, De Henauw S, Gibney MJ, Huybrechts I, McCarthy SN, et al. A new reference method for the validation of the nutrient profiling schemes using dietary surveys. *European Journal of Nutrition*. 2007;46 Suppl 2:29-36.
 56. Scarborough P, Rayner M, Boxer A, Stockley L. Application of the Nutrient profiling model: Definition of 'fruit, vegetables and nuts' and guidance on quantifying the fruit, vegetable and nut content of a processed product. British Heart Foundation Health Promotion Research Group, Department of Public Health, University of Oxford Retrieved February. 2005;7:2007.
 57. Masset G. Predictive validity of WXYfm and SAIN, LIM food nutrient profiling models in the Whitehall II cohort. University College London PhD Thesis. 2012.
 58. Arambepola C, Scarborough P, Rayner M. Validating a nutrient profile model. *Public Health Nutrition*. 2008;11(4):371-8.
 59. Coulston AM, Boushey C. Nutrition in the Prevention and Treatment of Disease: Academic Press; 2008.

60. Wrieden W, Peace H, Armstrong J, Barton K. Monitoring the Scottish Diet Action Plan Targets - review of national dietary surveys and Scottish research studies. (Available from FSA library, Aviation House, London or www.food.gov.uk) 2003.
61. Lentjes MA, McTaggart A, Mulligan AA, Powell NA, Parry-Smith D, Luben RN, et al. Dietary intake measurement using 7 d diet diaries in British men and women in the European Prospective Investigation into Cancer-Norfolk study: a focus on methodological issues. *British Journal of Nutrition*. 2013;1-11.
62. Stubbs, J. Detecting and Modelling Mis-reporting of Food Intake with Special Reference to Under-reporting in the Obese. FSA Project Code: NO8001 (Available from the FSA library, Aviation House, London or www.food.gov.uk). 2003
63. World Health Organization, the Food and Agriculture Organization of the United Nations. Energy and Protein Requirements: Report of a Joint FAO/WHO/UNU Expert Consultation: World Health Organization Technical Report Serie. 1985;724:1-206.
64. Lister Hill National Center for Biomedical Communications, US National Library of Medicine. Cells and DNA basics. National Institutes of Health. Assess date: 31.03.2016. Available from: <http://ghr.nlm.nih.gov/handbook/>.
65. Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
66. Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, et al. The international HapMap project. *Nature*. 2003;426(6968):789-96.
67. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nature Genetics*. 2008;41(1):56-65.
68. Nock NL, Pillai ALC. Dyslipidemia: genetics and role in the metabolic syndrome: INTECH Open Access Publisher; 2012. Chapter 6. P. 93-120. Access date: 31.03.2016. Available from: <https://www.intechopen.com/books/dyslipidemia-from-prevention-to-treatment/dyslipidemia-genetics-and-role-in-the-metabolic-syndrome>
69. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, Hivert M-F, et al. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes*. 2013;DB_130949.
70. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan Ja, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature Genetics*. 2012;44(9):991-1005.
71. Soranzo N. Genetic Determinants of Variability in Glycated Hemoglobin (HbA1c) in Humans: Review of Recent Progress and Prospects for Use in Diabetes Care. *Current diabetes reports*. 2011;11(6):562-9.
72. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature Genetics*. 2010;42(11):949-60.
73. Nettleton JA, Follis JL, Ngwa JS, Smith CE, Ahmad S, Tanaka T, et al. Genex dietary pattern interactions in obesity: analysis of up to 68 317 adults of European ancestry. *Human Molecular Genetics*. 2015;24(16):4728-38.
74. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics*. 2010;42(11):937-48.
75. Herrera BM, Lindgren CM. The Genetics of Obesity. *Current Diabetes Reports*. 2010;10(6):498-505.
76. Do R, Xie C, Zhang X, Männistö S, Harald K, Islam S, et al. The effect of chromosome 9p21 variants on cardiovascular disease may be modified by dietary intake: evidence from a case/control and a prospective study. *PLoS Medicine*. 2011;8(10):e1001106.
77. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nature Genetics*. 2009;41(1):56-65.

78. Mehta NN. The Effect of Chromosome 9p21 Variants on Cardiovascular Disease May Be Modified by Dietary Intake. *Circulation: Cardiovascular Genetics*. 2012;5(1):148-50.
79. Povel C, Boer J, Reiling E, Feskens E. Genetic variants and the metabolic syndrome: a systematic review. *Obesity Reviews*. 2011;12(11):952-67.
80. Garaulet M, Madrid JA. Chronobiology, genetics and metabolic syndrome. *Current Opinion in Lipidology*. 2009;20(2):127-34.
81. Atanasovska B, Kumar V, Fu J, Wijmenga C, Hofker MH. GWAS as a Driver of Gene Discovery in Cardiometabolic Diseases. *Trends in Endocrinology & Metabolism*. 2015;26(12):722-32.
82. 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 and Cardiovascular Diseases*. 2009;19(7):455-61.
83. International Expert Committee. Report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes Care*. 2009;32(7):1327-34.
84. World Health Organization. Use of glycated haemoglobin (HbA1c) in diagnosis of diabetes mellitus: abbreviated report of a WHO consultation. 2011.
85. Inzucchi S, Bergenstal R, Fonseca V, Gregg Ed, Mayer-Davis B, Spollett G, et al. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care*. 2010;33.
86. Ford ES, Zhao G, Li C. Pre-diabetes and the risk for cardiovascular disease: a systematic review of the evidence. *Journal of the American College of Cardiology*. 2010;55(13):1310-7.
87. National Institute for Health and Care Excellence. TYPE 2 DIABETES National clinical guideline for management in primary and secondary care. 2008. Access date: 01.10.2015. Available date: <https://www.nice.org.uk/guidance/cg87>
88. Schellenberg ES, Dryden DM, Vandermeer B, Ha C, Korownyk C. Lifestyle interventions for patients with and at risk for type 2 diabetes: a systematic review and meta-analysis. *Archives of International Medicine*. 2013;159(8):543-51.
89. Penn L, White M, Lindström J, den Boer AT, Blaak E, Eriksson JG, et al. Importance of weight loss maintenance and risk prediction in the prevention of type 2 diabetes: analysis of European Diabetes Prevention Study RCT. *PLoS One*. 2013;8(2):e57143.
90. Tuomilehto J, Lindstrom J, Eriksson JG, Valle TT, Hamalainen H, Ilanne-Parikka P, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *The New England Journal of Medicine*. 2001;344(18):1343-50.
91. Bombelli M, Facchetti R, Sega R, Carugo S, Fodri D, Brambilla G, et al. Impact of body mass index and waist circumference on the long-term risk of diabetes mellitus, hypertension, and cardiac organ damage. *Hypertension*. 2011;58(6):1029-35.
92. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genetics*. 2012;44(9):981.
93. Andersson EA, Allin KH, Sandholt CH, Borglykke A, Lau CJ, Ribel-Madsen R, et al. Genetic risk score of 46 type 2 diabetes risk variants associates with changes in plasma glucose and estimates of pancreatic β -cell function over 5 years of follow-up. *Diabetes*. 2013;62(10):3610-7.
94. Prasad RB, Groop L. Genetics of Type 2 Diabetes—Pitfalls and Possibilities. *Genes*. 2015;6(1):87-123.
95. Meigs JB, Panhuysen CI, Myers RH, Wilson PW, Cupples LA. A Genome-Wide Scan for Loci Linked to Plasma Levels of Glucose and HbA1c in a Community-Based Sample of Caucasian Pedigrees The Framingham Offspring Study. *Diabetes*. 2002;51(3):833-40.
96. Vassy JL, Hivert M-F, Porneala B, Dauriz M, Florez JC, Dupuis J, et al. Polygenic type 2 diabetes prediction at the limit of common variant detection. *Diabetes*. 2014;DB_131663.
97. Karaderi T, Drong AW, Lindgren CM. Insights into the Genetic Susceptibility to Type 2 Diabetes from Genome-Wide Association Studies of Obesity-Related Traits. *Current Diabetes Report*. 2015;15.
98. World Health Organisation. Report of a WHO consultation Obesity: preventing and managing the global epidemic. 2000. WHO Technical Report Serie 894: i-253.

99. Lyssenko V, Laakso M. Genetic Screening for the Risk of Type 2 Diabetes Worthless or valuable? *Diabetes Care*. 2013;36(Supplement 2):S120-S6.
100. Dayeh T, Volkov P, Salö S, Hall E, Nilsson E, Olsson AH, et al. Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genetics*. 2014;10(3):e1004160.
101. Ling C, Groop L. Epigenetics: a molecular link between environmental factors and type 2 diabetes. *Diabetes*. 2009;58(12):2718-25.
102. Bramswig NC, Kaestner KH. Epigenetics and diabetes treatment: an unrealized promise? *Trends in Endocrinology & Metabolism*. 2012;23(6):286-91.
103. Rönn T, Volkov P, Davegårdh C, Dayeh T, Hall E, Olsson AH, et al. A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue. *PLoS Genetics*. 2013;9(6):e1003572.
104. Wang J KJ, Vääntinen M, Kuulasmaa T, Lindström J, Tuomilehto J, Uusitupa M, Laakso M. Variants of transcription factor 7- like 2 (TCF7L2) gene predict conversion to type 2 diabetes in the Finnish Diabetes Pre- vention Study and are associated with im- paired glucose regulation and impaired insulin secretion. *Diabetologia*. 2007;50:1192–200.
105. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nature Protocols*. 2010;5(9):1564-73.
106. Camp KM, Trujillo E. Position of the Academy of Nutrition and Dietetics: nutritional genomics. *Journal of the Academy of Nutrition and Dietetics*. 2014;114(2):299-312.
107. Villegas R, Goodloe RJ, McClellan BE, Jr., Boston J, Crawford DC. Gene-carbohydrate and gene-fiber interactions and type 2 diabetes in diverse populations from the National Health and Nutrition Examination Surveys (NHANES) as part of the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study. *Bio Medical Central genetics*. 2014;15:69.
108. Cornelis MC, Qi L, Zhang C, Kraft P, Manson J, Cai T, et al. Joint effects of common genetic variants on the risk for type 2 diabetes in US men and women of European ancestry. *Achieves of International Medicine*. 2009;150(8):541-50.
109. Fisher E, Boeing H, Fritsche A, Doering F, Joost HG, Schulze MB. Whole-grain consumption and transcription factor-7-like 2 (TCF7L2) rs7903146: gene-diet interaction in modulating type 2 diabetes risk. *The British Journal of Nutrition*. 2009;101(4):478-81.
110. Phillips CM, Tierney AC, Roche HM. Gene–nutrient interactions in the metabolic syndrome. *Journal of Nutrigenetics and Nutrigenomics*. 2008;1(3):136-51.
111. Perez-Martinez P P-JF, Bellido C, Ordovas JM, Moreno JA, Marin C, Gomez P, Delgado-Lista J, Fuentes F, Lopez- Miranda J. A polymorphism exon 1 variant at the locus of the scavenger receptor class B type I (SCARB1) gene is associated with differences in insulin sensitivity in healthy people during the consumption of an olive oil-rich diet. *Journal of Clinical Endocrinology Metabolism* 2005(90):2297–300.
112. Marin C P-JF, Gomez P, Del- gado J, Paniagua JA, Lozano A, Cortes B, Jimenez-Gomez Y, Gomez MJ, Lopez-Mi- randa The Ala54Thr polymorphism of the fatty acid-binding protein 2 gene is as- sociated with a change in insulin sensitivity after a change in the type of dietary fat. *Journal of American Clinical Nutrition* 2005(82):196–200.
113. Ferguson JF, Phillips CM, Tierney AC, Pérez-Martínez P, Defoort C, Helal O, et al. Gene- nutrient interactions in the metabolic syndrome: single nucleotide polymorphisms in ADIPOQ and ADIPOR1 interact with plasma saturated fatty acids to modulate insulin resistance. *The American Journal of Clinical Nutrition*. 2010;91(3):794-801.
114. Phillips CM, Kesse-Guyot E, McManus R, Hercberg S, Lairon D, Planells R, et al. High Dietary Saturated Fat Intake Accentuates Obesity Risk Associated with the Fat Mass and Obesity–Associated Gene in Adults. *The Journal of Nutrition*. 2012;142(5):824-31.
115. Elliott P VA-C, Singh D, Neasham D, Spear J, Heard A. The Airwave Health Monitoring Study of police officers and staff in Great Britain: Rationale, design and methods. *Environmental Research* 2014 Sep 3;134C:280–5.

116. Gibson R, Frost G, Elliott P, Chan Q, Eriksen R, Vergnaud A, et al. Dietary assessment of British police force employees: a description of diet record coding procedures and cross-sectional evaluation of dietary energy intake reporting (the airwave health monitoring study). 2017. *British Medical Journal*.7(4).
117. International Physical Activity Questionnaire scoring protocol. 2005. <http://www.ipaq.ki.se/scoring.htm> (accessed October 2014).
118. Bingham SA, Gill C, Welch A, Cassidy A, Runswick SA, Oakes S, et al. Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. *International Journal of Epidemiology*. 1997;26(suppl 1):S137.
119. Food Standard Agency. McCance and Widdowson's the Composition of Foods: 6th Edition. Royal Society of Chemistry, UK. 2002.
120. Conway R, Robertson C, Dennis B, Stamler J, Elliott P. Standardised coding of diet records: experiences from INTERMAP UK. *British Journal of Nutrition*. 2004;91(05):765-71.
121. Dennis B, Stamler J, Buzzard M, Conway R, Elliott P, Moag-Stahlberg A, et al. INTERMAP: the dietary data—process and quality control. *Journal of Human Hypertension*. 2003;17(9):609-22.
122. Food Standards Agency. Foodbase. 2005. http://www.foodbase.org.uk/results.php?f_report_id=82 (accessed November 2013).
123. Department of Health Australia. Australian Bureau of Statistics. Australian Health Survey: Users' Guide. 2013
124. Gandy AM J, and Holdsworth M. Oxford Handbook of Nutrition and Dietetics. Oxford University Press, UK. 20 February 2012.
125. Black AE. Critical evaluation of energy intake using the Goldberg cut-off for energy intake: basal metabolic rate. A practical guide to its calculation, use and limitations. *International Journal of Obesity & Related Metabolic Disorders*. 2000;24(9).
126. Goldberg G, Black A, Jebb S, Cole T, Murgatroyd P, Coward W, et al. Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. *European Journal of Clinical Nutrition*. 1991;45(12):569-81.
127. Simon A, O'Connell R, and Stephen A. Designing a nutritional scoring system for assessing diet quality for children aged 10 years and under in the UK. *Methodological Innovation Online* 2012(7(2) 27-40).
128. Fung TT, Chiuve SE, McCullough ML, Rexrode KM, Logroscino G, Hu FB. Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. *Archives of Internal Medicine*. 2008;168(7):713-20.
129. Siervo M, Lara J, Chowdhury S, Ashor A, Oggioni C, Mathers JC. Effects of the Dietary Approach to Stop Hypertension (DASH) diet on cardiovascular risk factors: a systematic review and meta-analysis. *British Journal of Nutrition*. 2015;113(01):1-15.
130. National Collaborating Centre for Primary Care. Lipid modification: cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease. 2008.
131. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Research*. 2002;12(6):996-1006.
132. Arnold M, Raffler J, Pfeufer A, Suhre K, Kastenmüller G. SNiPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics*. 2014;btu779.
133. Genotyping and quality control of UK Biobank, a large-scale, extensively phenotyped prospective resource. Oct 2015. Available from: http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UKBiobank_genotyping_QC_documentation-web.pdf.
134. Fall T, Hägg S, Mägi R, Ploner A, Fischer K, Horikoshi M, et al. European Network for Genetic and Genomic Epidemiology (ENGAGE) consortium. The role of adiposity in cardiometabolic traits: a Mendelian randomization analysis. *PLoS Medicine*. 2013;10(6):e1001474.

135. Purcell S NB, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC, . PLINK 1.07: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*. 2007 81.
136. Eyheramendy S, Juhanson P, Gieger C, Lichtner P, Klopp N, Veldre G, et al. Genome-wide scan identifies CDH13 as a novel susceptibility locus contributing to blood pressure determination in two European populations. *Human Molecular Genetics*. 2009;18(12):2288-96.
137. Tragante V, Barnes MR, Ganesh SK, Lanktree MB, Guo W, Franceschini N, et al. Gene-centric meta-analysis in 87,736 individuals of European ancestry identifies multiple blood-pressure-related loci. *The American Journal of Human Genetics*. 2014;94(3):349-60.
138. Li S, Zhao JH, Luan Ja, Ekelund U, Luben RN, Khaw K-T, et al. Physical activity attenuates the genetic predisposition to obesity in 20,000 men and women from EPIC-Norfolk prospective population study. *PLoS Medicine*. 2010;7(8):e1000332.
139. Li S, Zhao JH, Luan Ja, Luben RN, Rodwell SA, Khaw K-T, et al. Cumulative effects and predictive value of common obesity-susceptibility variants identified by genome-wide association studies. *The American Journal of Clinical Nutrition*. 2010;91(1):184-90.
140. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genetics*. 2013;9(3):e1003348.
141. Drenos F. The use of genetic information in the prediction of Type 2 diabetes. *Personalized Medicine*. 2015;12(5):483-96.
142. Maehlen MT, Olsen IC, Andreassen BK, Viken MK, Jiang X, Alfredsson L, et al. Genetic risk scores and number of autoantibodies in patients with rheumatoid arthritis. *Annals of the Rheumatic Diseases*. 2013;annrheumdis-2013-204173.
143. VanderWeele TJ. On the distinction between interaction and effect modification. *Epidemiology*. 2009;20(6):863-71.
144. Van Ness PH, Allore HG, editors. Using SAS to investigate effect modification. *SAS Users Group International Proceedings*; 2006: Citeseer.
145. Bell BA, Smiley W, Ene M, Sherlock Jr PR, Blue GL. An Intermediate Primer to Estimating Linear Multilevel Models using SAS® PROC MIXED.
146. Lecture from McGill University by Pai M. Effect measure modification & Interaction. Assess: 01.07.2016. Available from: http://www.teachepi.org/documents/courses/fundamentals/Pai_Lecture9_Effect%20measure%20modification.pdf
147. Lecture from Penn State University Department of Statistics: Analysis of Variance and Design of Experiments. Penn State University. Assess: 01.07.2016. Available from: <https://onlinecourses.science.psu.edu/stat502/node/188>.
148. Huang T, Ley SH, Zheng Y, Wang T, Bray GA, Sacks FM, et al. Genetic susceptibility to diabetes and long-term improvement of insulin resistance and β cell function during weight loss: the Preventing Overweight Using Novel Dietary Strategies (POUNDS LOST) trial. *The American Journal of Clinical Nutrition*. 2016;ajcn121186.
149. Qi L, Liang J. Interactions between genetic factors that predict diabetes and dietary factors that ultimately impact on risk of diabetes. *Current Opinion in Lipidology*. 2010;21(1):31.
150. Field A. *Discovering statistics using IBM SPSS statistics*: Third edition. Sage; 2013. P. 176-190
151. Reidlinger DP, Darzi J, Hall WL, Seed PT, Chowienzyk PJ, Sanders TA. How effective are current dietary guidelines for cardiovascular disease prevention in healthy middle-aged and older men and women? A randomized controlled trial. *The American Journal of Clinical Nutrition*. 2015;101(5):922-30.
152. Boden G, Scapa EF, Kanno K, Cohen DE, Brosnan ME, Brosnan JT, et al. *Metabolism. Textbook of Hepatology*: Blackwell Publishing Ltd; 2008. p. 129-249.
153. Harding A-H, Sargeant LA, Welch A, Oakes S, Luben RN, Bingham S, et al. Fat Consumption and HbA1c Levels The EPIC-Norfolk Study. *Diabetes care*. 2001;24(11):1911-6.
154. Lovejoy J, DiGirolamo M. Habitual dietary intake and insulin sensitivity in lean and obese adults. *The American Journal of Clinical Nutrition*. 1992;55(6):1174-9.

155. Barnard ND, Cohen J, Jenkins DJ, Turner-McGrievy G, Gloede L, Jaster B, et al. A low-fat vegan diet improves glycemic control and cardiovascular risk factors in a randomized clinical trial in individuals with type 2 diabetes. *Diabetes Care*. 2006;29(8):1777-83.
156. Chandalia M, Garg A, Lutjohann D, von Bergmann K, Grundy SM, Brinkley LJ. Beneficial effects of high dietary fiber intake in patients with type 2 diabetes mellitus. *New England Journal of Medicine*. 2000;342(19):1392-8.
157. Mann, J. Dietary fibre and diabetes revisited. *European Journal of Clinical Nutrition*. 2001;55(11):919-21.
158. Simpson H, Lousley S, Geekie M, Simpson R, Carter R, Hockaday T, et al. A high carbohydrate leguminous fibre diet improves all aspects of diabetic control. *The Lancet*. 1981;317(8210):1-5.
159. Frost G, Wilding J, Beecham J. Dietary advice based on the glycaemic index improves dietary profile and metabolic control in type 2 diabetic patients. *Diabetic Medicine*. 1994;11(4):397-401.
160. Julia C, Fézeu LK, Ducrot P, Méjean C, Péneau S, Touvier M, et al. The Nutrient Profile of Foods Consumed Using the British Food Standards Agency Nutrient Profiling System Is Associated with Metabolic Syndrome in the SU.VI.MAX Cohort. *The Journal of Nutrition*. 2015;145(10):2355-61.
161. 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. *Nature Genetics*. 2010;42(2):105-16.
162. Selvin E, Marinopoulos S, Berkenblit G, Rami T, Brancati FL, Powe NR, et al. Meta-analysis: glycosylated hemoglobin and cardiovascular disease in diabetes mellitus. *Archives of Internal Medicine*. 2004;164(6):421-31.
163. Langenberg C, Sharp SJ, Franks PW, Scott RA, Deloukas P, Forouhi NG, et al. Gene-Lifestyle Interaction and Type 2 Diabetes: The EPIC InterAct Case-Cohort Study. *PLoS Medicine*. 2014;11(5):e1001647.
164. Zyrinx B-C, Salazar R, Hoepfner W, Vettorazzi E, Herder C, Windler E. The association of genetic markers for type 2 diabetes with prediabetic status-cross-sectional data of a diabetes prevention trial. *PloS One*. 2013;8(9):e75807.
165. Vaxillaire M, Yengo L, Lobbens S, Rocheleau G, Eury E, Lantieri O, et al. Type 2 diabetes-related genetic risk scores associated with variations in fasting plasma glucose and development of impaired glucose homeostasis in the prospective DESIR study. *Diabetologia*. 2014;57(8):1601-10.
166. Corella D, Carrasco P, Sorlí JV, Estruch R, Rico-Sanz J, Martínez-González MÁ, et al. Mediterranean Diet Reduces the Adverse Effect of the TCF7L2-rs7903146 Polymorphism on Cardiovascular Risk Factors and Stroke Incidence A randomized controlled trial in a high-cardiovascular-risk population. *Diabetes Care*. 2013;36(11):3803-11.
167. Sotos-Prieto M, Luben R, Khaw K-T, Wareham NJ, Forouhi NG. The association between Mediterranean Diet Score and glucokinase regulatory protein gene variation on the markers of cardiometabolic risk: an analysis in the European Prospective Investigation into Cancer (EPIC)-Norfolk study. *The British Journal of Nutrition*. 2014;112(1):122-31.
168. Hindy G, Sonestedt E, Ericson U, Jing X-J, Zhou Y, Hansson O, et al. Role of TCF7L2 risk variant and dietary fibre intake on incident type 2 diabetes. *Diabetologia*. 2012;55(10):2646-54.
169. Sonestedt E, Lyssenko V, Ericson U, Gullberg B, Wirfält E, Groop L, et al. Genetic variation in the glucose-dependent insulinotropic polypeptide receptor modifies the association between carbohydrate and fat intake and risk of type 2 Diabetes in the Malmö Diet and Cancer Cohort. *The Journal of Clinical Endocrinology & Metabolism*. 2012;97(5):E810-E8.
170. Brinsden HC, He FJ, Jenner KH, MacGregor GA. Surveys of the salt content in UK bread: progress made and further reductions possible. *British Medicine Journal Open*. 2013;3(6):e002936.
171. Liu S, Ajani U, Chae C, Hennekens C, Buring JE, Manson JE. Long-term β -carotene supplementation and risk of type 2 diabetes mellitus: a randomized controlled trial. *Journal of the American Medical Association*. 1999;282(11):1073-5.
172. Larsson S, Wolk A. Magnesium intake and risk of type 2 diabetes: a meta-analysis. *Journal of Internal Medicine*. 2007;262(2):208-14.

173. Li M, Fan Y, Zhang X, Hou W, Tang Z. Fruit and vegetable intake and risk of type 2 diabetes mellitus: meta-analysis of prospective cohort studies. *British Medical Journal open*. 2014;4(11):e005497.
174. World Health Organization. Cardiovascular diseases (CVDs) 2012. Access date: 01.10.2015. Available from: <http://www.who.int/mediacentre/factsheets/fs317/en/>
175. Vessby, B. Substituting dietary saturated for monounsaturated fat impairs insulin sensitivity in healthy men and women: the KANWU Study. *Diabetologia*. 2001(44):312-9.
176. Storlien L, Baur L, Kriketos A, Pan D, Cooney G, Jenkins A, et al. Dietary fats and insulin action. *Diabetologia*. 1996;39(6):621-31.
177. DeFronzo R. Insulin resistance, lipotoxicity, type 2 diabetes and atherosclerosis: the missing links. The Claude Bernard Lecture 2009. *Diabetologia*. 2010;53(7):1270-87.
178. Stratton IM, Adler AI, Neil HAW, Matthews DR, Manley SE, Cull CA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. *British Medical Journal*. 2000;321(7258):405-12.
179. Carlsson S, Hammar N, Grill V, Kaprio J. Alcohol Consumption and the Incidence of Type 2 Diabetes A 20-year follow-up of the Finnish Twin Cohort Study. *Diabetes Care*. 2003;26(10):2785-90.
180. Emanuele NV, Swade TF, Emanuele MA. Consequences of alcohol use in diabetics. *Alcohol Research and Health*. 1998;22(3):211.
181. Ben G, Gnudi L, Maran A, Gigante A, Duner E, Iori E, et al. Effects of chronic alcohol intake on carbohydrate and lipid metabolism in subjects with type II (non-insulin-dependent) diabetes. *The American Journal of Medicine*. 1991;90(1):70-6.
182. Koppes LL, Dekker JM, Hendriks HF, Bouter LM, Heine RJ. Moderate Alcohol Consumption Lowers the Risk of Type 2 Diabetes A meta-analysis of prospective observational studies. *Diabetes Care*. 2005;28(3):719-25.
183. Schrieks IC, Heil AL, Hendriks HF, Mukamal KJ, Beulens JW. The effect of alcohol consumption on insulin sensitivity and glycemic status: a systematic review and meta-analysis of intervention studies. *Diabetes Care*. 2015;38(4):723-32.
184. Kiechl S, Willeit J, Poewe W, Egger G, Oberhollenzer F, Muggeo M, et al. Insulin sensitivity and regular alcohol consumption: large, prospective, cross sectional population study (Bruneck study). *British Medical Journal*. 1996;313(7064):1040-4.
185. Beulens JW, Rimm EB, Hendriks HF, Hu FB, Manson JE, Hunter DJ, et al. Alcohol consumption and type 2 diabetes influence of genetic variation in alcohol dehydrogenase. *Diabetes*. 2007;56(9):2388-94.
186. Young-Wolff KC, Enoch M-A, Prescott CA. The influence of gene-environment interactions on alcohol consumption and alcohol use disorders: a comprehensive review. *Clinical Psychology Review*. 2011;31(5):800-16.
187. Day NE, McKeown N, Wong M-Y, Welch A, Bingham S. Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *International Journal of Epidemiology*. 2001;30(2):309-17.
188. Bingham S, Gill C, Welch A, Day K, Cassidy A, Khaw K, et al. Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records. *British Journal of Nutrition*. 1994;72(04):619-43.
189. Health and Social Care Information Centre. Health Survey for England - 2012 Trend tables. 2013. Access date: 15.03.2016. Available from: <http://content.digital.nhs.uk/catalogue/PUB13219>
190. Boon N, Boyle R, Bradbury K, Buckley J, Connolly S, Craig S, et al. Joint British Societies' consensus recommendations for the prevention of cardiovascular disease (JBS3). *Heart*. 2014;100(Suppl 2):ii1-ii67.
191. Rennie KL, Coward A, Jebb SA. Estimating under-reporting of energy intake in dietary surveys using an individualised method. *British Journal of Nutrition*. 2007;97(06):1169-76.
192. Wilms B EB, Thurnheer M et al.. Correction factors for the calculation of metabolic equivalents (MET) in overweight to extremely obese subjects. *International Journal of Obesity* 2014;38, 1383-1387.

193. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. *The American Journal of Clinical Nutrition*. 1997;65(4):1220S-8S.
194. Poslusna K, Ruprich J, de Vries JH, Jakubikova M, van't Veer P. Misreporting of energy and micronutrient intake estimated by food records and 24 hour recalls, control and adjustment methods in practice. *British Journal of Nutrition*. 2009;101(S2):S73-S85.

Appendix 1: UK Nutrient profile model algorithm

The UK NP model is also derived from the UK DRV. Table 10 and 11 indicate the points scored depending upon the amount of each nutrient in 100g of the food or drink.

The NP model's algorithm works out the overall score of a food or drink using following four steps:

1. Total "A" points

Total "A" points = (energy points) + (sat fat points) + (sugar points) + (sodium points)

Table 10: "A" points for negative nutrients

Points	Energy (kJ)	Sat Fat (g)	Total Sugar (g)	Sodium (mg)
0	≤ 335	≤ 1	≤ 4.5	≤ 90
1	>335	>1	>4.5	>90
2	>670	>2	>9	>180
3	>1005	>3	>13.5	>270
4	>1340	>4	>18	>360
5	>1675	>5	>22.5	>450
6	>2010	>6	>27	>540
7	>2345	>7	>31	>630
8	>2680	>8	>36	>720
9	>3015	>9	>40	>810
10	>3350	>10	>45	>900

2. Total "C" points

Total "C" points = (% fruit, veg & nuts points) + (fibre points) + (protein points) (56)

Table 11: "C" points for positive nutrients

Points	Fruit, Veg & Nuts (%)	NSP Fibre ' (g)	Or AOAC Fibre ' (g)	Protein (g)
0	≤ 40	≤ 0.7	≤ 0.9	≤ 1.6
1	>40	>0.7	>0.9	>1.6
2	>60	>1.4	>1.9	>3.2
3	-	>2.1	>2.8	>4.8
4	-	>2.8	>3.7	>6.4
5*	>80	>3.5	>4.7	>8.0

3. NP score

$$\text{NP score} = (\text{total "A" points}) - (\text{total "C" points})$$

Unless a food or drink scores 11 or more "A" points and less than 5 points for fruit, vegetable and nuts. Then the overall score is calculated as follows:

$$\text{NP score} = (\text{total "A" points}) - (\text{fibre} + \text{fruit, vegetable and nuts points})$$

4. NP score (energy adjusted)

The energy adjusted score ensures an equal proportional distribution of all foods and reflects the actual amount eaten in proportion to the whole diet's total energy intake (minus energy from alcohol):

$$\text{NP score per item (energy adjusted)} = \frac{\text{NP score} \times \text{energy from the item}}{\text{Total energy intake}}$$

Interpretation of total NP score

The scores are then summed to a total NP score which interprets the quality the whole diet (all foods consumed) and rate the score on a scale 1-100 (healthiest) based on previous validated method (22, 52, 55):

$$\text{Total NP score} = \sum \text{NP scores (energy adjusted)}$$

$$\text{NP score} = ((-2) * \text{total NP score}) + 70$$

Appendix 2: Illumina Infinium genotyping protocol and summary report

1) Project Summary:

Overview of genotyping results:

For this project, 2016 Human DNA samples were processed using the Illumina Infinium HumanExome-12v1-1 BeadChip Arrays and following the Infinium HD Assay Ultra protocol.

The manifest file (which describes the SNP/probe content on a BeadChip) used for this project was the HumanExome-12v1-1_A.bpm. For this project, the samples were clustered against themselves in Genome Studio.

Sample QC review: Picogreen testing has been performed by the customer and the samples were processed as they were received.

Controls review: In general, all samples performed well across all controls.

Technical errors: none.

Samples removed: None. The cut-off for the Human Exome arrays is 95% as it is a custom made array. No sample fell below this threshold, they have all successfully passed.

Sample Summary:

Average Overall Call Rate: 99.95%

Average call rate for samples above cut-off (>95%): N/A

Additional Files Provided:

Image Data: Raw data files for each BeadChip array.

Samples Table.txt: These tables show the call rates for all the samples for each batch processed.

Final Report: Report generated with all parameters from Genome Studio.

Full Data Table: Report generated with selected parameters from Genome Studio.

Zcall File: Report generated with specific parameters from Genome Studio.

Call Rate Averages.xlsx: This table shows the average call rates for all the batches as well as the number of samples with call rates below the cut-off threshold.

Infinium Genotyping controls

This appendix describes the controls used in the Illumina Infinium Assay which are checked during the data analysis QC using Illumina Genome Studio software. The controls can be split into sample dependent and sample independent categories. The controls are both useful by themselves and in combination with each other.

Sample-independent controls

Staining controls

Extension controls
controls

Target removal controls

Hybridization controls

Sample-dependent controls

Stringency controls

Non-specific binding

Non-polymorphic controls

Staining controls

These controls examine the efficiency of the staining step in both the red and green channels. The beads have varying levels of biotin or dinitrophenyl (DNP) attached: the sensitivity monitored can be compared to expected intensities which can tell us how efficient the staining step has been.

The staining controls are independent of both the hybridization and extension step of the protocol. Good staining controls should give readings around and above 5000-6000 and should be much higher than the background readings.

Extension controls

The extension efficiency of A, T, C and G nucleotides from a hairpin probe are tested in the extension controls. These controls should show a high signal (e.g. in the 1000's)

Target removal controls

These controls highlight the efficiency of the removal of the probe sequences post extension. Target removal controls should exhibit a low signal compared to the hybridization controls. These controls should only be observed in the green channel and should exhibit a low level of signal.

Hybridization controls

Synthetic targets, at three concentration levels (high, medium and low), that complement the sequence of the array are used to test the overall performance of the entire assay. These controls can only be monitored in the green channel.

Stringency controls

Mismatches are introduced into the body of the extension probe sequences to affect hybrid stability: the controls contain 0-12 mismatched nucleotides between target and probe. These controls test the stringency of the hybridization process and are only observed in the red channel. The perfect match should show high signal whilst the mismatch should show low signal.

Non-specific binding controls

Probe sequences complementary to bacterial sequences are used to monitor the specificity of the hybridization of the amplified DNA as they should not hybridize to human sequences under standard conditions. These controls are monitored in both the red and green channels: they should show low intensities which indicate there is minimal cross hybridization of human DNA to bacterial sequences.

Non-polymorphic controls

These controls query a particular base in a non-polymorphic region of the genome which allows comparison of assay performance across different samples. There should be high signal for the controls for all four nucleotides. This control is a good indicator of sample quality.

Appendix 3: Airwave diary coding operational manual and protocol

Imperial College London

Airwave Health Monitoring Study

Standard protocol for food diary coding

Developed by: Rachel Gibson, Rebeca Eriksen, Kathryn Lamb and Yvonne McMeel.

Nutrition and Dietetics Research Group

Acknowledgements

Dietary coders who have contributed to the dietary data set and code book development: Odette Albani, Shatha Alrabiah, Jessica Ayling, Andrea Carames, Zhengyu Fan, Kirsty Frost, Louise Hirichi, Zanna Hofstede, Niamh O'Sullivan, Kristina Petersen, Claudia Schrender, Manny Singh, Elizabeth Slack, Jill Twomey, Alan Wan, Jessica Ware, Yiling Zhu

Dietplan6.7 nutritional software programme (*Forestfield Software Ltd, Horsham, UK*)

Contents	Page
1. Introduction	4
2. General questions	
i) Recording the answers to the general questions	5
ii) Using the general questions to aid code selection	6
3. Registering & setting up the assessment in Dietplan6	11
4. Selecting food codes	15
5. Portion size algorithms	25
6. Error checking	31
7. Log the diary as complete	28
References	33
Appendix 1: Practice coding scenarios	34

1.0 Introduction:

The following pages provide an outline of how to input and code the food diaries from the AIRWAVE study.

The primary aim of this protocol is to reduce error and to standardise code and portion selection.

Please read through before you start to code and input the food diaries. You will also need to refer to the i) AIRWAVE code book, ii) AIRWAVE portion size document and iii) Foodbase portion size document. These documents can be found in the shared drive.

All coders should keep a weekly log of any issues regarding code / portion selection to discuss at the weekly meeting.

EXCLUSION CRITERIA for coding:

- Less than one day completed
- Meal replacement diet
- Incomplete diaries are coded but incomplete days are not e.g. if days 1-6 are complete but day 7 is does not follow the same pattern as the previous days. Therefore days 1-6 are coded but 7 is not

If a diary cannot be coded due to any of the above it needs to be logged on the food diary inventory excel file (see section 6.0).

Diaries with missing gender and/or date of birth should be placed in the missing date of birth box file.

****Please do not write on or in the food diaries****

2.0 General questions

1.1 Recording the answers to the questions in page 45-50 of the food diary

- Open the file: \\wmdi-nutritech\Dietary assessment protocol files for shared access / general questions

Study No.	Gender	Q1a	Q1b	Q1c	Q1d	Q1e	Q1f	Q1g	Q1h	Q1i	Q1j	Q1k	Q1l	Q1m	Q1n	Q1o	Q1p	Q1q	Q1r	Q1s	Q1t	Q1u	Q1v	Q1w	Q1x	Q1y	Q1z
3	56961	ES	1	2	3	9	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
4	37021	ES	1	3	3	9	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
5	56475	ES	1	3	3	9	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
6	56970	RG	1	3	3	5	9	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
7	37029	RG	1	9	1	9	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
8	56887	ES	2	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
9	56461	RG	1	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
10	56772	ES	1	2	3	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
11	56999	RG	1	3	3	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
12	56776	ES	1	3	3	9	9	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
13	56478	RG	1	3	3	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
14	56988	ES	1	3	3	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
15	56982	RG	1	3	3	9	5	2	9	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
16	56472	ES	1	3	3	9	9	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
17	56774	ES	1	3	3	9	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
18	37034	RG	1	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
19	56998	ES	1	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
20	56989	ES	1	2	5	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
21	56899	ES	1	3	3	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
22	56235	RG	1	3	3	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
23	56761	ES	2	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
24	37037	RG	1	2	3	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
25	56764	ES	1	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
26	56507	ES	1	3	3	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
27	56886	RG	1	3	3	9	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
28	56987	ES	1	3	3	9	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
29	37032	ES	2	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
30	56889	RG	1	3	3	2	9	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
31	56870	RG	1	3	3	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
32	37036	RG	1	3	3	1	1	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
33	56473	ES	1	2	9	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
34	37037	RG	1	3	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
35	37029	ES	1	3	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
36	56990	ES	1	3	3	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
37	56974	RG	1	3	3	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

- The participant responses to questions 3, 8, 12, 14 and 16 are recorded in this spread sheet.
- The responses to the questions have been coded – the codes are detailed in the top row or in the comment box.
- Only one value should be entered per cell.
- In column 'U' select the code that best fits what the participant has written in the free text box on page 50.
- In column 'V' enter the number of days that the diary has been completed for.
- In column W – quality of the dietary record
 - 0 = poor, little detail (high level of assumptions applied)/ illegible handwriting
 - 1 = good, clear details
- In column X enter the date of the first day in the food diary

2.2 Using the answers recorded in page 45-50 of the food diary to aid code selection

In the first instance select the food codes based on what the participant has written in the actual food diary. Refer to the general questions to assist when insufficient information has been provided in the food diary.

Q. 1 MILK - type

The answers recorded here detail the type of milk used by the participant. This should be the type of milk that you input unless another type is written in the daily diet entry. If this section is not completed assume UKN 12-313 'Semi-skimmed milk pasteurized' is consumed, unless otherwise stated in the food diary.

Q.2 MILK - amount

Details the amount of milk to input unless otherwise written in the daily entry (n.b. these are in grams therefore specific gravity does not need to be applied):

	A lot (g)	Average (g)	Hardly any (g)
Tea	40	30	23
Coffee	64	43	30
Cereal	175	123	91

Source: Foodbase 75th / 50th / 25th centile values 'all age groups' mixed gender. Values calculated by combining the recording portion sizes for semi-skimmed and whole milk and then divided by two and rounded to the nearest whole number.

If this section is not completed assume 'medium' amount is consumed unless otherwise stated in the food diary.

- Tea - add milk (semi-skimmed, average amount) if milk is recorded elsewhere in the diary
- Cereal - add milk (semi-skimmed, average amount) if milk is recorded elsewhere in the diary.

Exception: if there is a specific pattern to consumption, e.g. every evening 'tea' written without milk – then do not include milk.

Q.3 TYPE OF TEA /COFFEE

This should be the type of tea / coffee that you input unless another type is written in the daily diet entry. Check if the respondent has indicated when they have caffeinated drinks or not.

If they have not indicated in the food diary which type of drinks are consumed and have filled in the back of the diary – please enter as follows:

Answer:	Tea	coffee
Always decaffeinated	USF A000006	USF A000005
Sometimes decaffeinated	Use decaffeinated code when indicated in the diary, otherwise use standard default code	Use decaffeinated code when indicated in the diary, otherwise use standard default code
Never decaffeinated	Refer to the Airwave code book for standard and default codes.	Refer to the Airwave code book for standard and default codes.

If this section is not completed assume caffeinated versions are consumed, unless otherwise stated in the food diary.

Q.4 FATS

Details the type of fat to input unless otherwise written in the diary entry. If this section is not completed assume UKN 17-486 Butter, spreadable consumed unless otherwise stated in the food diary.

Q.5 BREAD

This provides details of the type of bread to enter unless otherwise written in the daily entry. If this section is not completed use code UKN 11-099 White bread, average . If the respondent has toast use the toasted code. There is no toasted code for granary bread so the bread code 11-461 has to be used.

NOTE: Brown bread is no longer available in the UK –Always code wholemeal bread

Q.6 FATS – frequency of use

States if the participant used butter, margarine or spread (as selected in Q.4) when they ate bread / toast / crackers / rolls.

Answer =	Enter as:
Always	Always add butter / spread as Q.4/Q.7 when the item is mentioned
Sometimes	Code when specified in the diary
Never	Never enter butter or spread on any of these items unless recorded in the daily entry.
Don't know	Never enter butter or spread on any of these items unless recorded in the daily entry.

If this section is not completed apply 'sometimes' option.

Q.7 FATS – quantity used on bread and crackers

This gives an indication of the amount in grams used on bread and crackers, unless the participant has stated another quantity or referred to the photo images. Refer to the portion size algorithm. NB. This refers to the serving size per slice/side of bread e.g. one bread roll = 2 sides

If participant recorded usage on a 'cracker' use 50% of the quantity

Answer =	Enter as:
thick	14g
medium	10g
thin	7g

Source: Foodbase (1) serving 25th / 50th / 75th centile

If this section is not completed assume 'medium' amount is consumed unless otherwise stated in the food diary.

Q.8 MEAT – level of cooking

This information only needs to be recorded in the Excel spread sheet.

Q.9 MEAT – amount of fat eaten

Answer =	Enter as:
Ate all of the fat	Select a code that states 'lean and fat' every time meat is eaten.
Ate some of the fat	Every time meat is eaten enter 50% of the reported weight as a code that states 'lean and fat' and 50% as a code that states 'lean' or 'meat only'. i.e. <i>sirloin beef steak fried 230g</i> enter as: 115g UKN 18-066 beef, sirloin steak, fried, lean 115g UKN 18-067 beef, sirloin steak, fried, lean and fat
Ate most of the fat	Every time meat is eaten enter 75% of the reported weight as a code that states 'lean and fat' and 25% as a code that states 'lean' or 'meat only'. i.e. <i>sirloin beef steak fried 230g</i> enter as: 57.5g UKN 18-066 beef, sirloin steak, fried, lean 172.5g UKN 18-067 beef, sirloin steak, fried, lean and fat
Ate as little as possible	Every time meat is eaten enter 25% of the reported weight as a code that states 'lean and fat' and 75% as a code that states 'lean' or 'meat only'. i.e. <i>sirloin beef steak fried 230g</i> enter as: 172.5g UKN 18-066 beef, sirloin steak, fried, lean 57.5g UKN 18-067 beef, sirloin steak, fried, lean and fat
Did not eat meat	n/a
No fat eaten	Select a code that states 'lean' every time meat is eaten.

If this section is not completed apply 'Ate some of the fat' option. **Q.10 POULTRY – amount of skin eaten**

Answer =	Enter as:
yes	Select option for 'meat and skin' every time poultry is eaten
sometimes	Every time poultry is eaten enter 50% of the reported weight as a code that states 'meat and skin' and 50% as a code that states 'without /no skin / meat only'.
no	Select option for 'without /no skin' every time poultry is eaten
Did not eat poultry	n/a

If this section is not completed apply 'sometimes' option.

Q.11 GRAVY

	Enter as:
With juices / pan residues	RCP AR00008 homemade beef gravy made with meat fat juices added
Without juices / pan residues	UKN 17-311 Gravy instant granules made up with water
Sometimes	Every time gravy is consumed enter the recorded portion as UKN 17-311
Don't know	Enter as default : UKN 17-311 Gravy instant granules made up with water

Default code if not stated use UKN 17-311 Gravy instant granules made up with water

Q12. SALT

This information only needs to be recorded in the Excel spread sheet. No salt added at the table or to cooking water should be included.

NB. Default code for all vegetables boiled – select the code stating 'unsalted water' unless option not available.

(ref: methodology as NDNS 2008/09 -2009/10 sodium intake excludes cooking and table salt).

Q.13 FRUIT

This tells you if the participant eats the skin/peel on fruits where it is edible.

Answer:	Apple	Pear
Skin eaten	UKN 14-012 Apples, eating, average, raw	UKN 14-190 Pears, average, raw
Skin not eaten	UKN 14-014 Apples, eating, average, raw, peeled	UKN 14-192 Pears, average, raw, peeled

If this section is not completed apply 'Skin eaten' option.

Q.14 Vitamins and supplements

This information only needs to be recorded in the Excel spread sheet.

Q.15 WATER

	Hot drinks	Cold drinks
Tap water (unfiltered)	ENTER: UKN 17-377 water, distilled	
Filtered water - hard water filter		
Filtered water - other		
Bottled water	ENTER: USF A000001 mineral water	

If this section is not completed assume tap water only consumed unless otherwise stated in the food diary.

Q. 16 Organic foods

This information only needs to be recorded in the Excel spread sheet.

3.0 Registering & setting up the assessment in Dietplan6

REGISTRATION SCREEN:

1. From the drop down menu select 'analyse assessments'

2. Registration number: 5 figure number under the barcode top right hand corner of food diary cover

4. Sex: as recorded on the cover of the food diary.

3. Surname: AIRWAVE

5. Date of birth: as recorded on the cover of the food diary.

The screenshot shows the Dietplan6 software interface. The 'File' menu is open, and 'Analyse Assessment' is selected. The 'Registration' tab is active in the left sidebar. The main form is titled 'Enter new registration details'. It contains fields for 'Registration No.' (with a 5-digit barcode input), 'Surname' (filled with 'AIRWAVE'), 'Forenames', 'Address', 'Telephone', 'E-mail', 'Sex' (a dropdown menu), 'Date of birth', 'Height (m)', 'Weight (kg)', 'Referred by', and 'Notes'. There are 'Accept', 'Continue', and 'Cancel' buttons. A status bar at the bottom indicates 'Diet assessments selected' and the date '29/07/2013'.

SELECTION SCREEN:

Dietplan6 Reg 36879 AIRWAVE, Assessment 1

File Edit Foods Tasks Views Data Admin Help

Analysse Assessment

Registration

No 36879 AIRWAVE Male Born 02 November 1967

Date	Days	Type	Title
29 Jul 2013	7	Diary	

New [Select this to create a new assessment](#)

Amend/confirm assessment details

Date: 29 July 2013 Type: Intake diary Title:

At assessment date

Height (m)

Weight (kg)

Target weight (kg)

Reference values: None

Created 29 July 2013
Modified 29 July 2013

Scope

Max number of days: 7

Max meals per day: 8

Max courses per meal: blank

Date of day 1 or select its weekday: 22 January 200

Enter meal times: ☒ No ☐ Yes

Food categories:

Diet assessments selected

11:10 29/07/2013

Max number of days: 7

Max number of meals per day: 8

Max courses per meal: blank

Date of day 1: as per date recorded in day one of the food diary. (The date does not always register on DietPlan, to ensure it does do not press 'entre' key always click continue and check the grid.)

Date: to show the date of diary input.

Reference values
'none'

Type: select intake diary from drop down screen

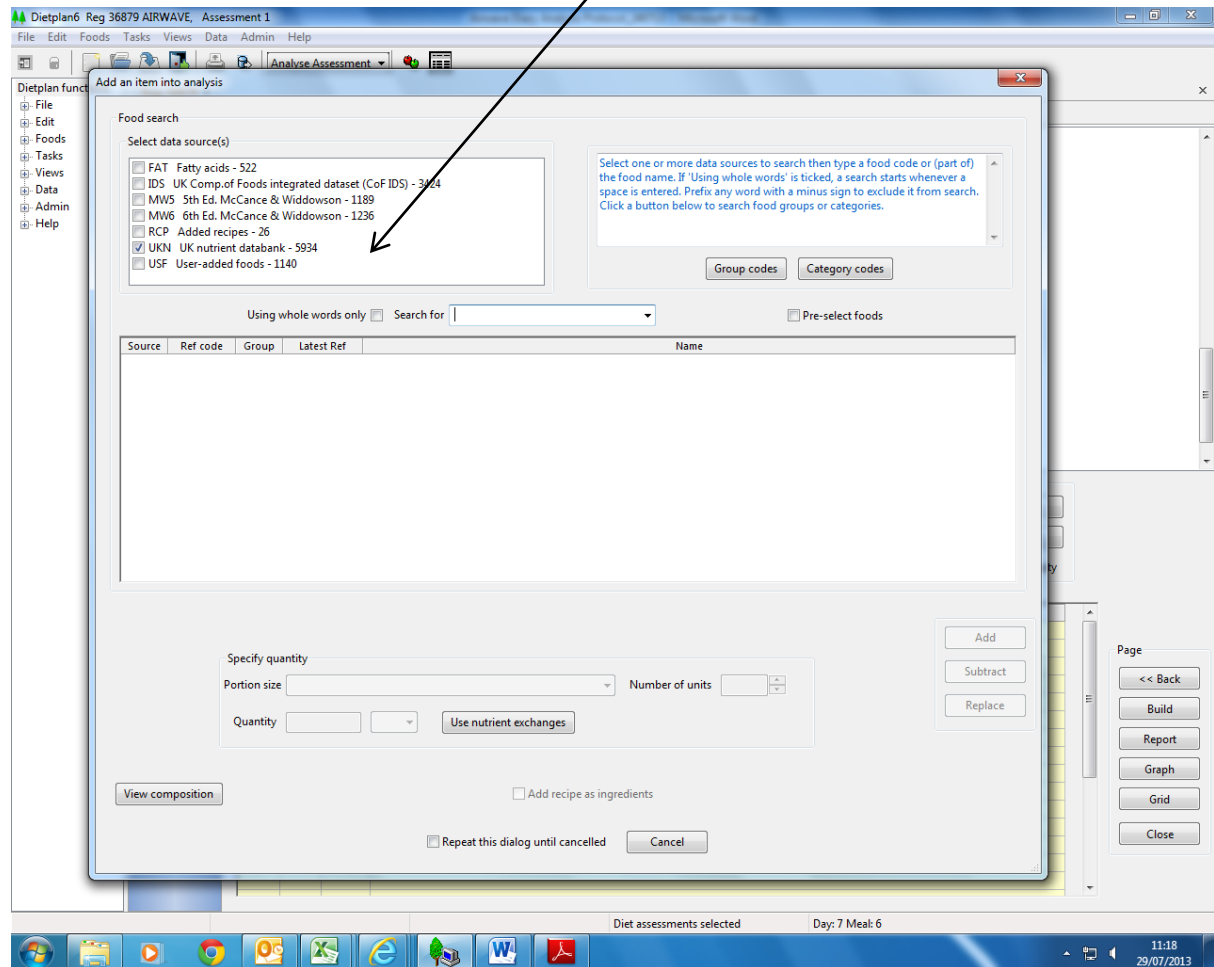
Enter meal times:
'no'

COMPOSITION SCREEN:

Database selection

Click on 'add item'

UKN to be the only database selected, unless otherwise indicated by the Airwave code book



Entering meals

The screenshot shows the Dietplan6 software interface. The 'Add Food' window is open, displaying a list of nutrients and their intake values. The 'Summary' tab is selected, showing a list of meals for Day 1 and Day 2. An arrow points from the text 'Enter each section of the daily diary under the meal numbers as follows:' to the meal list.

Nutrient	Unit	Intake	per 100g
Edible portion		1.00	1.00
Water	g	219.6	98.5
Total Nitrogen	g	0.12	0.06
Protein	g	1.0	0.4
Fat	g	0.4	0.2
Carbohydrate (mse)	g	1.1	0.5
Energy (kcal)	kcal	11	5
Energy (kJ)	kJ	49	22
Starch (mse)	g	Nil	Nil
Total sugars (mse)	g	1.1	0.5
NMES	g	---	---
Glucose	g	Nil	Nil
Fructose	g	Nil	Nil
Sucrose (mse)	g	Nil	Nil
Maltose (mse)	g	Nil	Nil
Lactose (mse)	g	1.1	0.5
NSP (Englyst)	g	Nil	Nil
Fibre (AOAC)	g	Nil	Nil
Saturated fats	g	0.2	0.1
Mono-unsat.fats	g	0.1	0+
Poly-unsat.fats	g	0+	0+
Trans.fats	g	0.02	0.01
Cholesterol	mg	1.4	0.6
Sodium (Na)	mg	10	4
Potassium (K)	mg	90	40

Source	RefNo	Weight	Food or ingredient
MW6	211	23.0	Semi-skimmed milk, average
MW6	1106	200.0	Tea, black, infusion, average

Enter each section of the daily diary under the meal numbers as follows:

Diet plan	Food diary section
Meal 1	before breakfast
Meal 2	breakfast
Meal 3	Mid-morning
Meal 4	lunch
Meal 5	tea
Meal 6	evening meal
Meal 7	later evening
Meal 8	'anything else'.

Start from day 1 meal 1 and enter the items as written in the diary

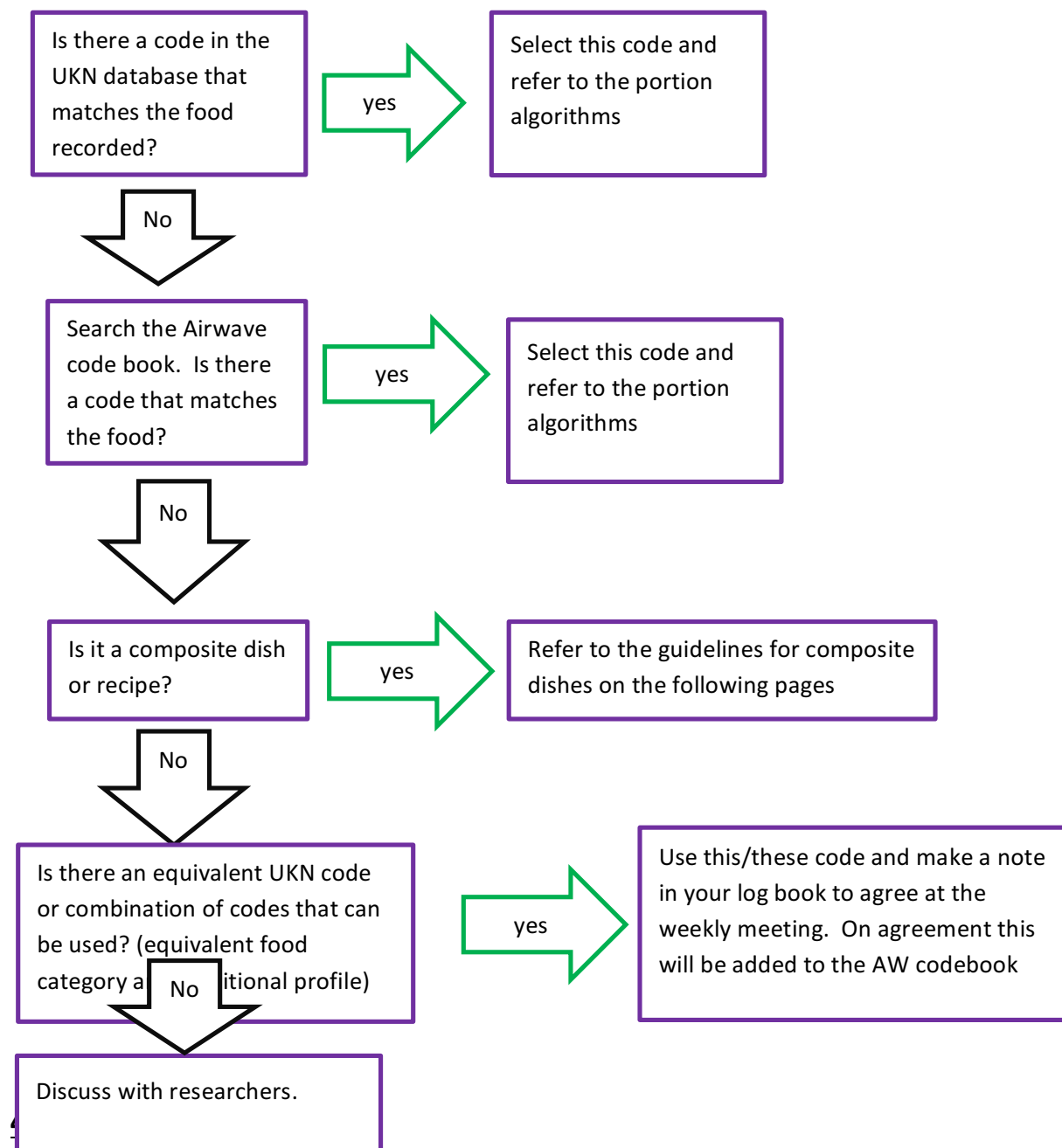
- ✓ Remember to save your work as you go to avoid any data loss.

4.0 Selecting food codes

4.1 General rules

- ✓ UKN is the default database. Only use RCP and USF databases if it's a food from the codebook.
- ✓ Never select a code for an item that states '(fats only)' at the end of the description.
- ✓ Items such as drinks and sandwiches should be broken into their constituent parts (see 3.2 Foods and meals not in the database)
- ✓ Items should be coded in the form in which they were eaten e.g. raw tomato / boiled potato where possible unless option not available.

Algorithm for code selection:



The following items should be broken into their constituent parts where possible:

- Hot beverages (except those made up with water)
- Sandwiches
- Home cooked meal when full recipe provided
- Homemade burgers

○ **Example 1:**

Instant coffee with semi-skimmed milk and white sugar

Enter as:

UKN 17-159 instant, made up with water

UKN 12-313 Semi-skimmed, pasteurised milk, average

UKN 17-063 sugar, white

Rules for entering composite dishes:

1.Canned ready meals: i.e. spaghetti, soup, beans etc.

Select UKN code which states 'canned –reheated'.

2.Dish stated as 'homemade' and where a recipe is provided:

Enter all ingredients as the amount consumed and in their correct state– i.e. cooked / raw directly into the assessment (i.e. do not set up a separate recipe code).

3.Dish stated as 'homemade' and complex* recipe provided: i.e. ingredients are not available in their correct state in the database AND the recipe provided is not comparable to those in the UKN database. Check the standard recipes used if listed in McCance(2).

*e.g. recipe with full quantities provided, but difficult to estimate actual proportion eaten.

In this instance a recipe should be set up using the 'Recipe analysis' option:

The code for the recipe should be the participant code prefixed by AW.

If multiple recipes are required for one participant then add 'a', 'b', 'c'... etc.. at the end of the code.

e.g. AW44399a

The recipe name should state the description of the dish and the participant code

e.g. Fruit cake – recipe for AW44399

Please ensure that the correct 'group code' is assigned to the recipe.

Please note these recipes should not be included in the Airwave codebook as they are unique to each participant.

4.Dish stated as 'homemade ' and no recipe provided:

Select UKN code which states a 'homemade' version of this dish, if available.

5.Ready/Microwave meal (including ready to eat pizza, pies) with brand name provided:

Check the retailer / manufacturer website and select the closest equivalent code from diet plan that reflects the food eaten and the macronutrient profile.

NB: this may involve selecting 2 or more codes to combine to provide a similar food/nutrient profile.

See below for how to match foods not on the database.

6.Fast food burgers:

Enter as the complete food (e.g. Big Mac etc.)

If there is not a code in UKN database please check the excel Airwave code book. If there is no code in the code book then a recipe will be required (see recipe set up protocol).

7.Retail sandwiches:

Do not use sandwiches in dietplan, match the sandwich to those online. Sandwiches can be found on Tesco and Waitrose websites. Use Waitrose for M&S and Waitrose sandwiches only (ie. Premium/luxury type sandwiches) and Tesco's for all other brands if brand not stated use Tesco as default.

Foods and meals not in the database

- Sandwiches
- Weight watchers/ reduced fat, calorie, sugar etc. meals
- Foods from cafés and fast foods e.g. costa, burger king, KFC etc.

How to code

- Find the nutritional information:
 1. Check Airwave codebook for item
 2. Check Portion Information folder: \\wmdi-nutritech\Dietary assessment protocol\Files for shared access\Portion information\retail & fast food portion info
 3. Check retailers website
 4. Search online
- Calorie match (see screen shot below)
 1. Click 'use nutrient exchanges' button
 2. Choose kcal,
 3. Type in the amount of calories you want to match in 'exchange value'
 4. Click 'view composition'
 5. If the protein, carbohydrates, sugars, fat are with 10% this amount can be used

If outside of 10%

- Check what nutrient is outside of the 10% range, e.g. protein is too low
- Add foods using the **ingredients list** (only include foods if they are on the ingredient list)
 1. If more protein is required add more of the protein source
 2. If carbohydrates are required add add more of the carbohydrate source
 - Be careful as some e.g. pasta, bread also contain protein
 - If sugar is needed add sugar.
 - There are different types of sugar. Some products have ingredients which are the same e.g. inverted sugar, white sugar and honey, glucose fructose syrup so only code white sugar. **Be mindful of adding sugar as it will increase carbohydrates.**
- If fat is needed add oil or fat (from the source of fat in the product), check saturated fat.
 - One type of fat may not be suitable, both palm oil and rapeseed oil may be required.

It is possible to calculate the portion size by using the 'nutrient exchange' function in Dietplan6

Some foods e.g. Burger King burger's, pasta salads, Subway, sandwiches must be coded separately

- Code the main ingredients using the ingredient list
 1. Calorie match the ingredients using their primary nutrient
 - E.g. a ham salad sandwich
 - Calorie match bread to carbohydrates
 - Calorie match ham to protein
 - Add green salad portion as for homemade sandwich
 2. Compare fats, carbohydrates, sugar, protein for the meal to nutritional information
 3. Adjust weights of ingredients so nutrients are with 10%

4. Some extra ingredients from the ingredients list may need to be added e.g. oil or sugar

Tips

- Weights do not always match exactly this does not matter if the nutrients does
- For the first few 'real' diaries have a go and ask for them to be checked
- Sandwiches – use Tesco's for all sandwiches and Waitrose for Waitrose and M&S and premium/luxury sandwiches
- You may need to use more than one type of the same food e.g. sponge cake and fatless sponge

8.Homemade sandwiches:

Unless otherwise stated:

2x medium slices of bread (as per general questions response or default code)

Fat spread (if stated or stated in general questions) – NB quantity for each slice

Filling: unless otherwise stated use 'average' amounts for cheese / meats / fish / egg.

Unless otherwise stated 20g lettuce / 34g tomato / 23g cucumber

'Salad' – default portion 20g of 'green salad' (UKN 15-380) other editions to sandwiches use ~15g.

9. Mixed Salads

If there is not enough information provided to be able to enter the weight of each vegetable included use 'green salad' UKN 15-380, see code book for default portion sizes.

4.3 Cooking methods:

Not all the different cooking options for each food type are available in Dietplan, therefore a comparable cooking method may need to be selected if the one recorded in the food diary is not available.

Vegetables

- 'steamed' or 'microwaved' - enter as boiled 'unsalted water' if 'steamed' not available
 - 'stir fried' - if fried option not available enter as 'raw' (select type of oil as specified in the general questions as the 'average amount', refer to the codebook for the default code)
 - 'stir fried' in cook in sauce - enter as boiled 'unsalted water' if 'steamed' not available
 - 'roasted' - select 'baked' or grilled if available + oil (select type of oil as specified in the general questions as the 'average amount', refer to the codebook for the default code)
 - 'casseroled' or 'stewed' - select 'boiled'
- 'boiled' = default cooking method unless vegetables are eaten as part of a salad

Meat / fish

- 'steamed' - enter as 'casseroled' or 'stewed' if 'steamed' not available
- 'fried' - select 'grilled' + oil if fried option not available (select type of oil as specified in the general questions as the 'average amount', refer to the codebook for the default code)
- 'stir fried in cook in sauce' - enter as 'casseroled' or 'stewed'
- 'roast' - enter as 'grilled' if 'roast' option not available
- 'grilled' - enter as 'roast' if 'grilled' option not available

Some fish e.g. tuna steak do not have a cooked code so the raw must be used.

Default cooking methods if not specified:

Potato, root vegetables, peas - 'boiled' unsalted water

Meat and fish – grilled

4.4 Weight changes on cooking

- If the raw weight of a food is given the cooked weight needs to be calculated and entered into Dietplan.
- If the weight is written in the 'amount' column we assume this is a cooked weight unless it is part of recipe

Weight Gain

- Foods which gain weight upon cooking (they expand when cooked)
 - E.g. pasta, rice, couscous, lentils
- Find the water gain in McCance and Widdowson's The Composition of Foods
- The number given is the percentage of water gained, this means the amount of water added. Therefore if it gives +144 water gain the total is x2.44

▪ E.g. Dried pasta boiled +123%
200g dry pasta, water gain 123%
 200×2.23
= 446g boiled pasta

Or (an alternative method)
 $200g + 123\%$

▪ E.g. Dried rice boiled +172%
241g dry rice, water gain 172%
 241×2.72
= 655.52g boiled rice

Or (if using calculator)
 $241 + 172\%$

Water loss

- Foods which lose weight upon cooking
 - E.g. meat, fish
- Find the water loss in McCance and Widdowson's The Composition of Foods
- The number given is the % of water lost during cooking
 - E.g. Raw stewing beef -36%

450g stewing beef, water loss -36%
 $450g \times 0.36 = 162g \text{ lost}$
 $= 450g \text{ raw stewing beef} - 162g \text{ loss}$
= 288g eaten

Or if using a calculator $450 - 36\% = 288g$

- E.g. Raw chicken casserole
 - 175g stewing beef, water loss -25%
 - $175\text{g} \times 0.25 = 43.75\text{g lost}$
 - $= 175\text{g raw chicken} - 43.75\text{g loss}$
 - $= \underline{131.25\text{g eaten}}$

Or if using a calculator $175\text{g} - 25\% = 131.25\text{g}$

Tips

- Code foods in the form they are eaten e.g. we eat chicken cooked so it has to be coded cooked
- Some foods e.g. tuna steak only have raw code in Dietplan so the raw code has to be used
- Some grains e.g. couscous, bulgur wheat, quinoa only have the raw code in Dietplan. The quantity in the Airwaves food portion is cooked weight so:
 - Code 50% of cooked weight and 50% water for couscous and bulgur wheat
 - Code 40% cooked weight and 60% water for quinoa
- Water loss does not have to be calculated for vegetables in a recipe

All sausages to be entered as cooked weights. Unless specified otherwise e.g. brand and weight as the following weights and NOT those stated in the UKN dietplan6 dataset.

Default sausage weights (each):

	raw	cooked
Thick	47g	35g
Cocktail	12g	9g
Thin	25g	19g
Premium (e.g. 'best' / 'finest' / taste the difference')	65g	49g

(Based on average supermarket weight per sausage recorded in Foodbase data)

Default cooked weight is not stated: 34g (average thick, thin and premium)

Conversion of liquids to grams:

- Liquids have a specific gravity so 100ml is not always 100g
- Conversion factors must be used as weight must be coded
- Used for milk, cream, ice cream, egg, oils, fizzy drinks, 100% fruit juices, alcoholic drinks
- Do not use for water, squash, tea, coffee

How to calculate the conversion factor

- All conversion factors are after the contents page in the Food Portion Sizes book (orange book)
- Multiple value in 'specific gravity' column by the ml given

E.g. 200ml of whole milk

Conversion factor 1.031

$$200\text{ml} \times 1.031 \\ = \underline{206.2\text{g milk}}$$

E.g. 330ml can cola

Conversion factor 1.04

$$330\text{ml} \times 1.04 \\ = \underline{343.2\text{g cola}}$$

Tips

- Some fizzy drinks e.g. cola in the book say not for low calorie but this factor can be used
- If the respondent has used a small amount e.g. 3tbs of milk the conversion factor does not need to be used
- The weights of milk for tea and coffee in section 1.2 **do not** need a conversion factor
- If the amount has been given in ounces, litres or pints convert to ml first then use conversion factor (values are at the top of the same page as the conversion factors)
- Weights of glasses and cups are given in ml in this document
- Some alcoholic drinks have pints and half pints in the drop down, these already have the conversion factor so it does not need to be calculated again
- Be careful of yogurt as the respondent sometimes puts 125ml or 500ml, they probably mean 1 pot which is 125g or 500g so the conversion factor does not need to be used
- If a volume of food is given check the food density pdf in \\WMDI-NUTRITECH\Dietary Analysis\Airwave\Files for shared access\Portion information

Spoon volumes not listed in the FSA book:

Serving spoon = 40ml. MAFF Food Atlas (5)

Ladle = 120ml (mean volume of the two serving spoons featured in MAFF Food Atlas (5))

If a participant enters 'spoon' but does not specify what size, then apply appropriate default sizes:

Meat in sauce / vegetables – tablespoons

Vinegar / herbs / spices / sugar - teaspoons

Cup: for tea / coffee and other hot drinks refer to portion algorithm.

If a 'cup' is used to measure liquids other than hot drinks take the default volume to be 250ml (UK cooks measure) and apply specific gravities as appropriate. See food density pdf [\\wmdi-nutrittech\Dietary assessment protocol files for shared access \Portion information\food density

Other arbitrary measures:

'a handful' of Fruit and vegetables code 100g

'a handful' of nuts and dried fruits code 30g

'a splash of oil' code 5g

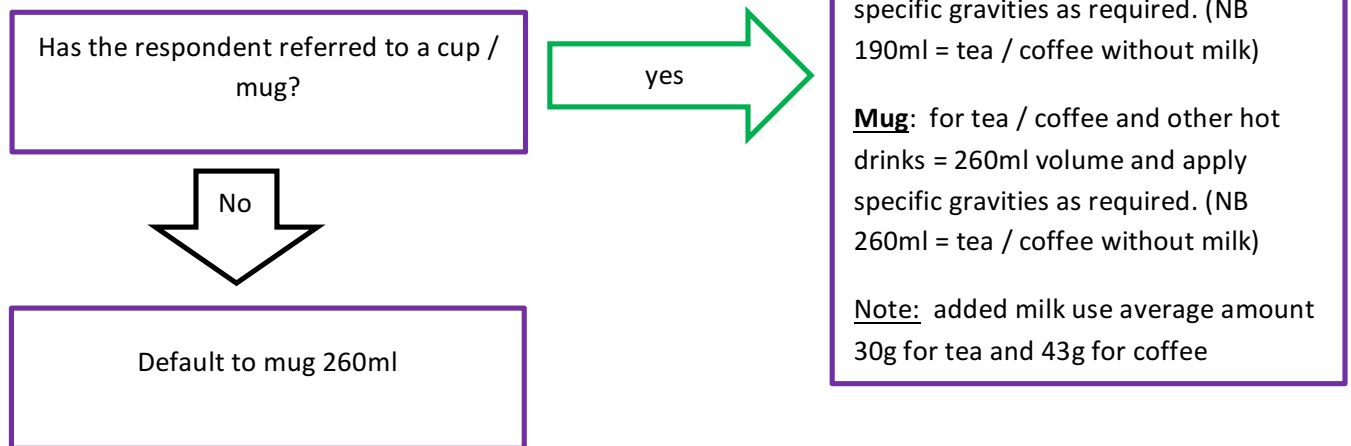
'a drizzle of salad dressing' = 10g (25th percentile for salad dressing see reference Foodbase)

Imperial measures: Conversion of imperial measures to grams are in the FSA Food Portion Sizes book (3) (page viii).

Note that 1cl = 10ml, so they might say 75cl of wine which is 750ml which is a standard bottle.

5.0 Portion size algorithms

Algorithm for portion size estimation – hot drinks



NB:

Adding milk to hot drinks

Tea / coffee entered as the cup / mug volume (i.e. 190 / 260ml) then add the milk.

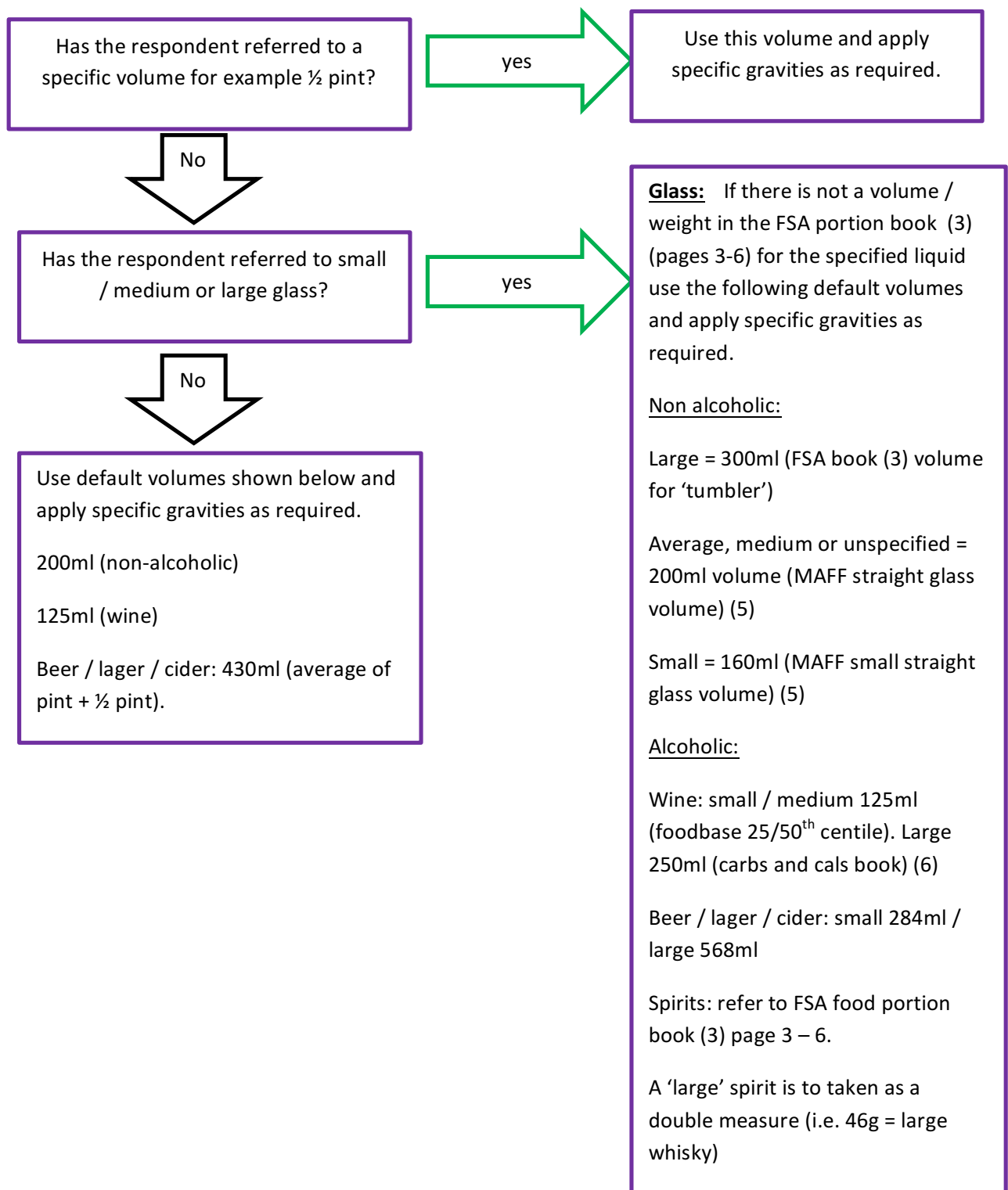
e.g. mug of tea with average milk

Tea infusion average 260g + semi skimmed milk pasteurised average 30g

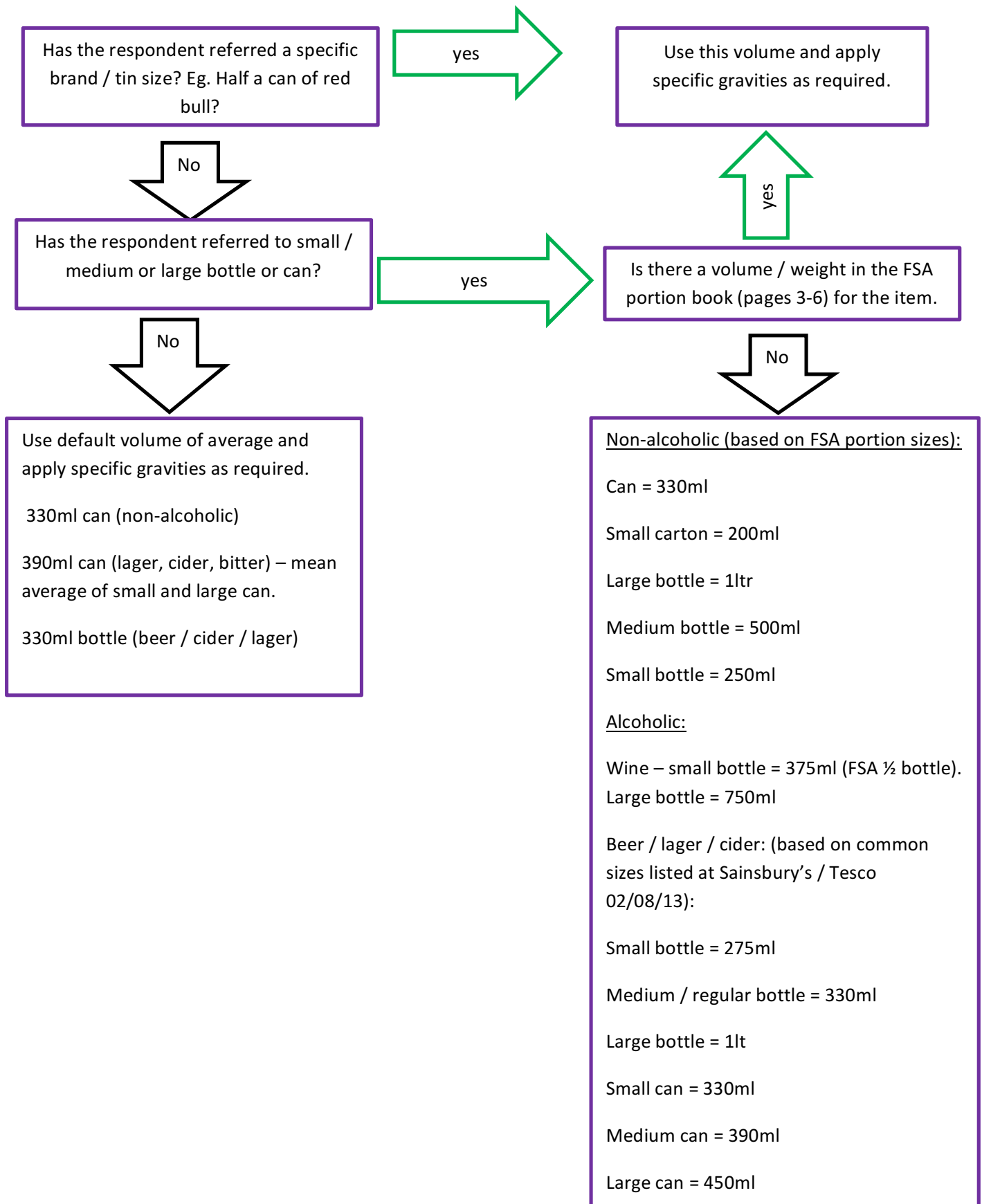
Sugar in drinks

If a participant records '1 sugar' – enter 1 level teaspoon (=4g sugar).

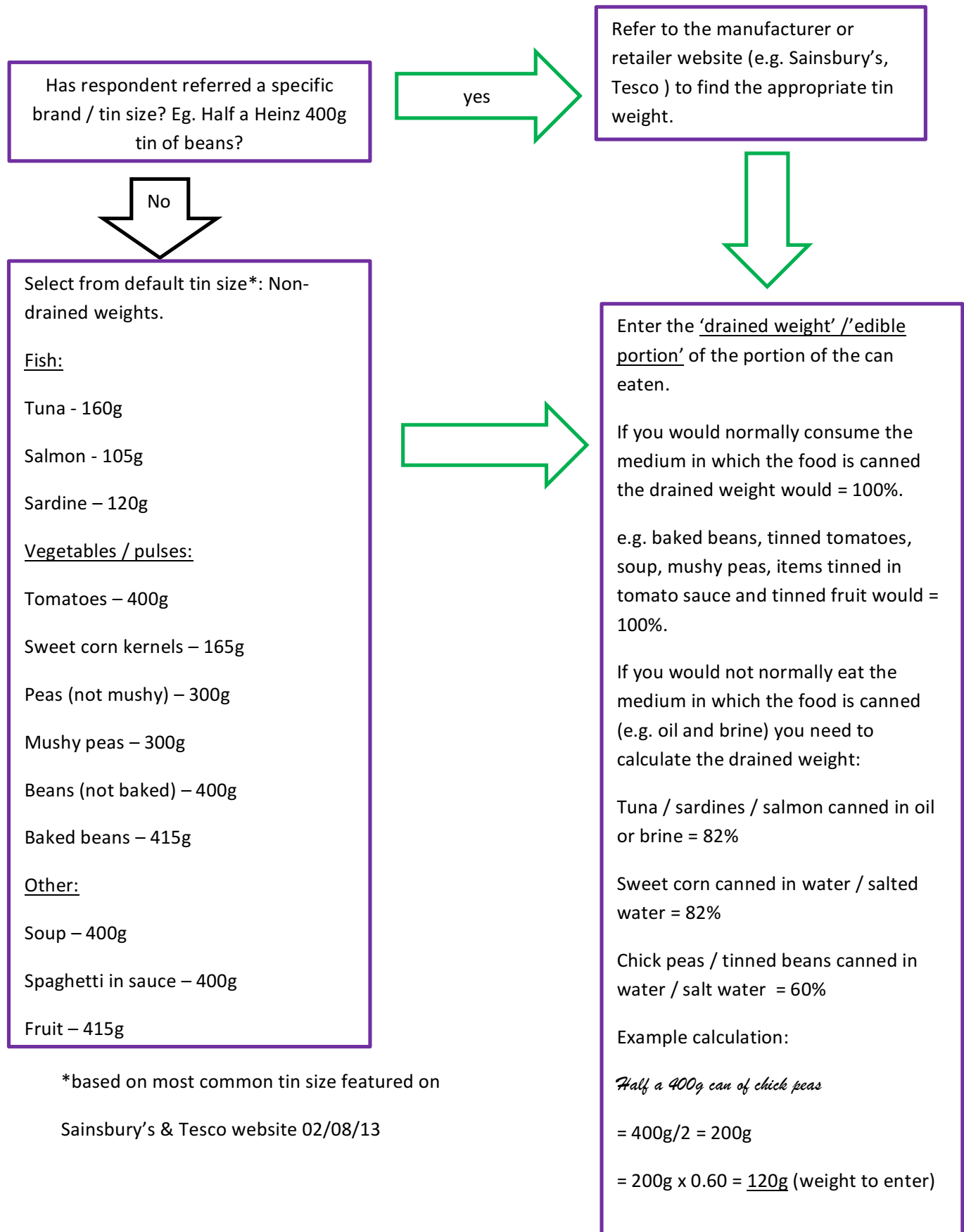
Algorithm for portion size estimation – cold / ambient / alcoholic drinks described by a glass.



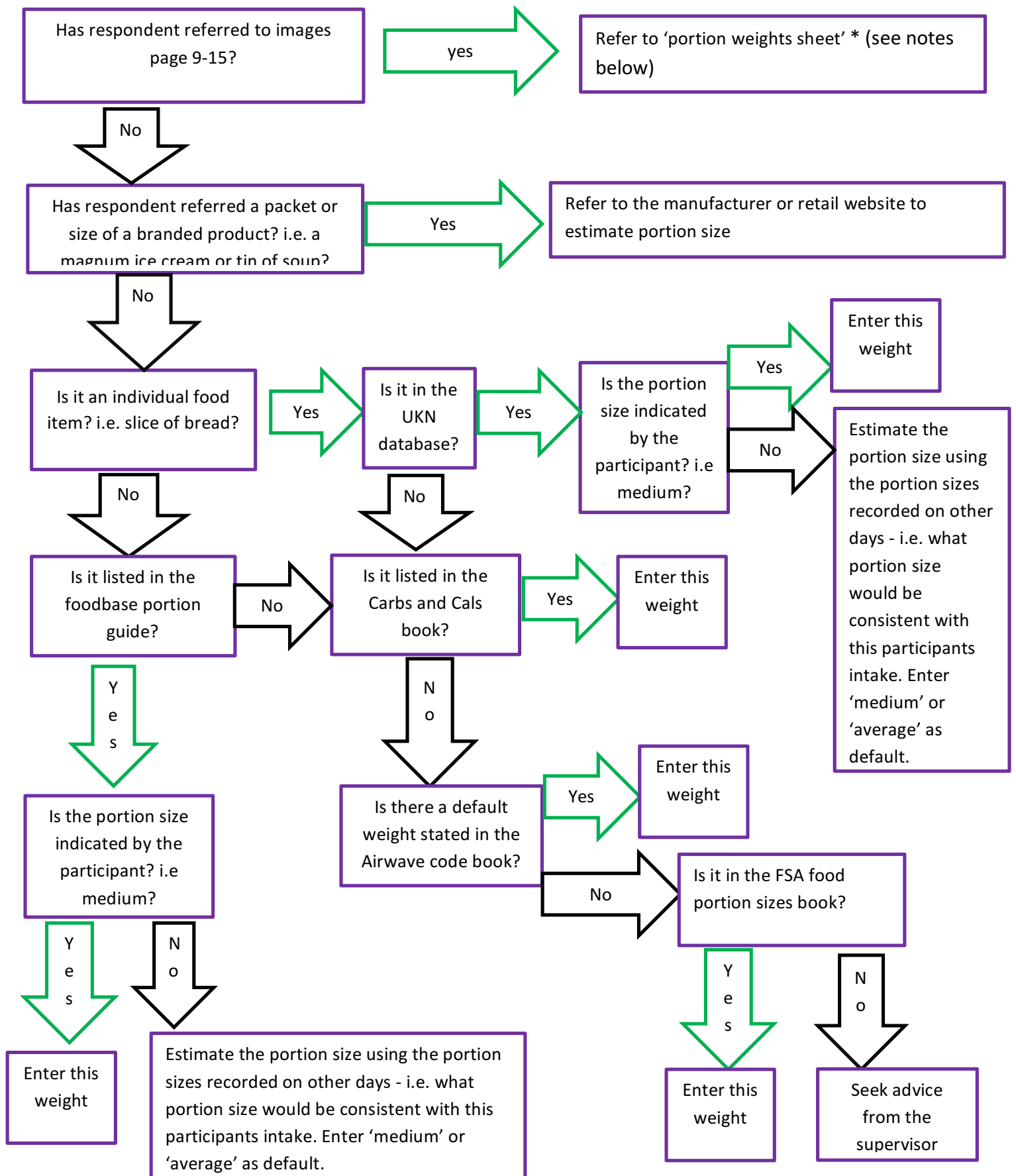
Algorithm for portion size estimation – cold / ambient / alcoholic drinks described by a bottle or can.



Algorithm for portion size estimation – food: canned and tinned, when recorded as a proportion of a can / tin.



Algorithm for portion size estimation FOOD (NOT canned / tinned):



*Notes to the Airwave food portions

- Check that you select the portion size relative to the food type. For example picture set 14 related to salad and boiled leaf vegetables,

e.g. photo 14a = 51g for boiled spinach

= 15g for lettuce

- If a participant has only referred to one portion size against 2 foods that are from different food groups:

Eg. Chicken in curry sauce & boiled rice 8b

Select portion 8b for the rice (131g) and select the 'b' size portion for the chicken curry, in this example it would be 19b (115g).

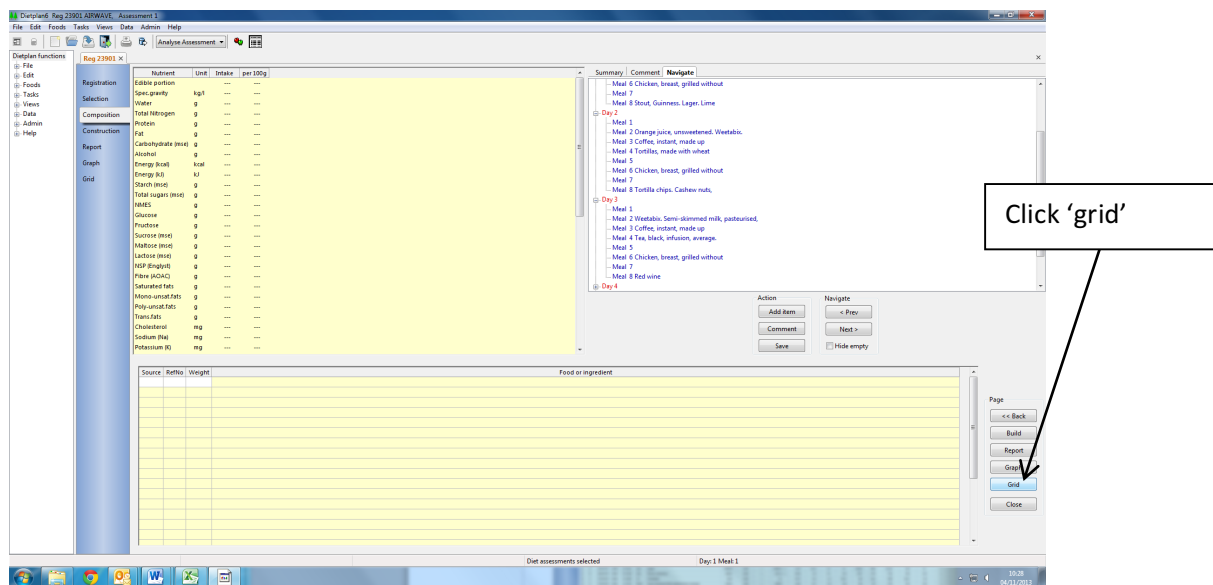
- If a participant has only referred to one portion size against 2 foods that are from the same food group:

Eg. Carrots, broccoli, green beans 13c

Select 13c = 107g divide by 3 = 36g per type of vegetable

6.0 Error checking

- Save the assessment
- Check the following against the diary you have inputted:
 - ✓ Registration code is the same as the code on the front of the diary you have inputted
 - ✓ Meal occasions have been entered in the correct meal number (1-8)
 - ✓ Codes represent the foods in the diary
 - ✓ Portion sizes are entered correctly
 - ✓ Any changes are saved
- Select 'grid'



When grid view opens you can double check foods and quantities:

Check the date has been registered by Dietplan and is correct

Note the total at the bottom of the 'Energy kcal' column - this can be used to double check the mean kcal intake per day. Range should be approx. 1500 - 3000kcal

'back' takes to the composition screen to allow any changes to be made.

7.0 Log the diary as complete

- Open the file via shared drive: [\\wmdi-nutritech\Dietary assessment protocol files for shared access / logging of food diaries in-out/020414_MASTER_FD_LOG].
- Search for the registration code and input the participant date of birth, gender and the date the diary was entered into Dietplan and your initials in the designated columns (highlighted in orange).
- Place in the 'completed' diaries box and mark off on the sheet attached to the box.

References:

1. Food Standards Agency. Foodbase 2005 [30.08.13]. Available from: http://www.foodbase.org.uk/results.php?f_report_id=82.
2. Food Standards Agency. McCance and Widdowson's: The composition of foods. 6th Summary ed. Cambridge: Royal Society of Chemistry: 2002
3. Food Standards Agency. Food Portion Sizes. 3rd ed. London: TSO: 2010.
4. Charrondiere R, Haytowitz D, Stadlmayr B. FAO/ FOODS Density Database Version 2.0. FAO: Rome, Italy, 2012
5. Nelson M, Atkinson M, Meyer J: *A Photographic Atlas of Food Portion Sizes*. London: MAFF Publications: 1997
6. Cheyette C, Balolia Y. Carbs and cals and protein and fats 1st ed. United Kingdom: Chello Publishing: 2010.

Appendix 1

Practice some matching, weight changes on cooking and conversion factors

Try coding these foods (create new analyses using registration your initials foods eg KLfoods)

- Sainsbury's chicken and sweetcorn sandwich
- Weight watchers lasagne
- Costa lemon muffin
- Supermarket bought chicken pasta salad
- Burger King Double Whopper
- Dolmio Light sauce

Try calculating the weight change of these foods

- 500g raw mince beef
- 100g raw back bacon
- 88g raw pasta
- Risotto recipe had third 224g rice, 550ml stock, 84g onions, 150g mushrooms
- Lasagne had a quarter 130g vegetables, 300g raw minced beef, 200g tomatoes, 450g water, garlic clove, herbs, 200g raw lasagne, 400g cheese sauce

Try calculating the weight in grams of these foods

- 200ml of semi-skimmed milk
- 3 pints of lager
- 250ml ice cream
- 1 glass red wine
- 250ml Tesco's chocolate milk shake

Appendix 4: Systematic literature review databases

PubMed

PubMed systematic literature search terms (table 13) include GWAS "genome-wide" OR "genome AND identification" OR "genome AND association" together with search terms for phenotypic trait "HbA1c OR glycated haemoglobin A1c". Papers were reviewed based on inclusion/exclusion criteria (table 14). The criteria are applied to ensure only variants with highest and consistent effect size are included, strengthening the statistical power for this thesis analysis.

DIAGRAM

The DIAGRAM consortium is a group of researchers sharing reviews and large-scale studies to characterise the genetic basis of T2D focusing on samples of European descent. The database publications were systematically reviewed as per set criteria in table 13 and 14 (link: www.diagram-consortium.org).

MAGIC

MAGIC consortium represents a collaborative database of multiple GWAS identifying loci associated with glycaemic and metabolic traits. MAGIC investigators studies fasting glucose, fasting insulin, 2 hour oral glucose test and HbA1c, as well as perform meta-analysis of more sophisticated measures of insulin secretion and sensitivity. MAGIC has identified dozens of loci influencing these traits. The database publications were systematically reviewed as per set criteria in table 13 and 14 (link: www.magicinvestigators.org).

National Human Genome Institute Research GWAS database

NHGIR is a US based organization which catalog published GWAS. Its curators conducts weekly PubMed systematic literature searches and peer reviews for GWAS based on following search terms "genome-wide" OR "genome AND identification" OR "genome AND association". Papers are eligible for the NHGIR database if they meet the following criteria:

Inclusion criteria:

1. Minimum 100,000 SNPs in the initial stage before quality control filters are applied.

2. Statistical significance SNP-trait/phenotype p-value $< 10^{-5}$ in the overall (initial GWAS and replication) population.

Exclusion criteria:

1. The study was published in a language other than English.
2. SNPs assayed were limited to those in candidate genes.
3. Samples were assayed to measure somatic variation (e.g. in tumor samples).
4. The study does not include any new GWAS data.

Additional search terms were applied to find GWAS specific to HbA1c trait and T2D phenotype (table 13). These GWAS were selected and reviewed based the set inclusion/exclusion criteria (table 14). Furthermore papers with no study replications were also excluded, to avoid any potential false positive results. Information on the reviewed papers are gathered and assess on following: study date: disease/trait information: initial sample size: replication sample size: platform (manufacturer): number of SNPs passing quality control metrics (use of multiple platforms without imputation, the total number of imputed SNPs, or "pooled" to denote studies of pooled DNA). For each identified SNP following information was reviewed: chromosomal region: gene: rs-number and risk allele: risk allele frequency in controls: p-value and any relevant text: OR (or % variance explained, SD, or unit difference for quantitative traits), 95% CI and any relevant text.

Full methodology for NHGIR systematic literature review:

Weekly PubMed searches are done using the terms "genome-wide" OR "genome AND identification" OR "genome AND association", with limits on the current year and human status.

Studies and associations are eligible for inclusion in the NHGRI GWAS catalog if they meet the following criteria:

1. Inclusion of at least 100,000 SNPs in the initial stage, before quality control filters are applied.

2. Statistical significance (SNP-trait p-value $< 1.0 \times 10^{-5}$) in the overall (initial GWAS + replication) population.
 - a. If a study does not report a combined p-value, the p-value and effect size from the largest sample size will be reported as long as the initial and replication samples each show an association of $p < 1.0 \times 10^{-5}$.
 - b. If a study does not include a replication stage, significant SNPs from the discovery stage will be reported.
 - c. SNP-trait associations that are described as previously known at the time of publication and are statistically significant in the GWAS sample, but are not attempted for replication, are reported.

Studies and associations are excluded if:

1. The study was published in a language other than English.
2. SNPs assayed were limited to those in candidate genes.
3. Samples were assayed to measure somatic variation (e.g., in tumor samples).
4. The study does not include any new GWAS data.

Information on the following study-level fields is extracted: author (last name of first author): study date (online publication date, if available): PubMed URL: publication title: disease/trait information: initial sample size (summing across multiple Stage 1 populations, if applicable): replication sample size (summing across multiple populations, if applicable): platform (manufacturer): number of SNPs passing quality control metrics (using "up to [maximum number of SNPs]" if multiple platforms are used without imputation, the total number of imputed SNPs, or "pooled" to denote studies of pooled DNA, as applicable): whether the study was one of copy number variants (initially excluded: additional studies to be added).

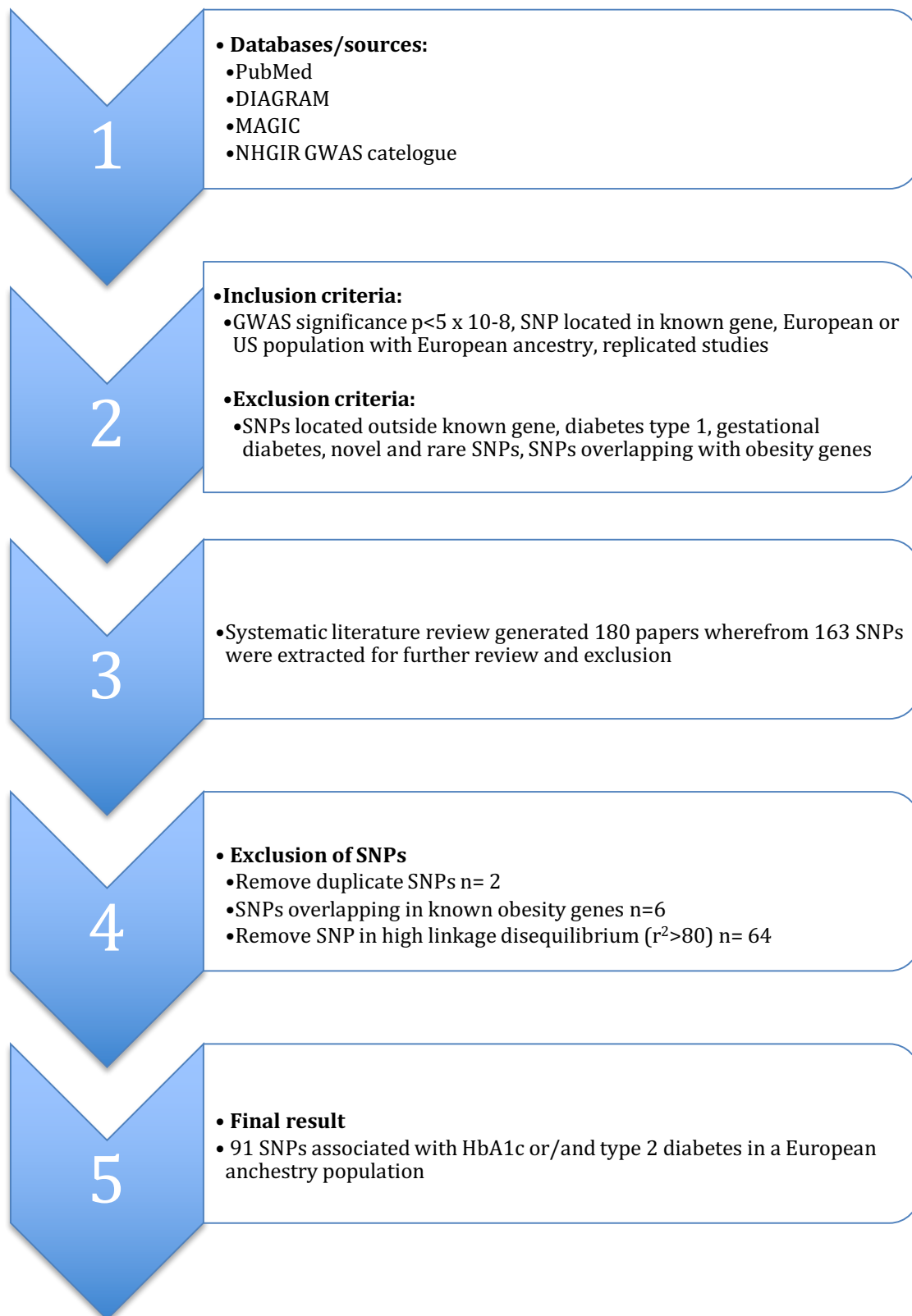
For each identified SNP, we extract: chromosomal region (from UCSC Genome Browser): gene (as reported): rs number and risk allele (as reported): risk allele frequency in controls (if not available among all controls, among the control group with the largest sample size): p-value and any relevant text (e.g., subgroups where applicable): OR (or % variance explained, SD increment, or unit difference for quantitative traits), 95% CI and any relevant text (e.g.,

subgroups). If the p-value, OR, and 95% CI fields are not available for the combined population, we extract estimates from the population group with the largest sample size.

In extracting information, we follow these additional guidelines: Missing or not applicable fields are denoted as follows: allele not reported: NS, not significant (no associations at $p < 1.0 \times 10^{-5}$ identified): NR, not reported: Where multiple genetic models are available, effect sizes (OR's or beta-coefficients) are prioritized as follows: 1) genotypic model, per-allele estimate: 2) genotypic model, heterozygote estimate, 3) allelic model, allelic estimate.

Focusing on risk alleles, we invert ORs < 1 and their associated confidence intervals, and report the opposite allele if available. If 95% CIs are not published, we estimate these using standard errors where available. If more than one SNP within a gene met the above criteria, we report one SNP unless there was evidence for an independent association. Associations attributed to a combination of one or more genetic variants are denoted as such in the rs number column (e.g., "rs1015362-G + rs4911414-T," "3-SNP haplotype 1"). If available, rs numbers for SNPs comprising the haplotype are indexed so that they are searchable using the SNP search features. Genes attributed to a SNP are extracted verbatim from the published report: "intergenic" and "NR" (not reported) were used to denote a location which was not attributed to a particular gene (if it appeared that gene information was sought) or an absence of reporting on location information, respectively. The term "pending" is used to identify an eligible GWAS for which SNP information has not yet been extracted: studies of CNVs, which are known to be incompletely ascertained, are also noted as pending.

Appendix 5: PRISMA flowchart of systematic literature review



Appendix 6: List of genetic markers found from systematic literature review

Reference	SNP rs ID	Gene	Risk allele	OR or beta	P-value	SE or [95% CI] [0.03-0.05] %	Sample
HbA1c NHGIR	rs1046896	FN3KRP	T	0.04	2E-26	increase	up to 46,368 European ancestry individuals
HbA1c NHGIR	rs13266634	SLC30A8	A	0.02	5E-8	[NR] % decrease	14,618 European ancestry females
		RPS3AP42 -				[0.02-0.04] %	
HbA1c NHGIR	rs1387153	MTNR1B	T	0.03	4E-11	increase	up to 46,368 European ancestry individuals
HbA1c NHGIR	rs552976	G6PC2	A	0.02	5E-10	[NR] % increase	14,618 European ancestry females
						[0.08-0.10] %	
HbA1c NHGIR	rs16926246	HK1	C	0.09	3E-54	increase	up to 46,368 European ancestry individuals
						[0.03-0.05] %	
HbA1c NHGIR	rs1799884	GCK	T	0.04	1E-20	increase	up to 46,368 European ancestry individuals
						[0.05-0.07] %	
HbA1c NHGIR	rs1800562	HFE	G	0.06	3E-20	increase	up to 46,368 European ancestry individuals
						[0.01-0.03] %	
HbA1c NHGIR	rs2779116	SPTA1	T	0.02	3E-9	increase	up to 46,368 European ancestry individuals
						[0.02-0.04] %	
HbA1c NHGIR	rs4737009	ANK1	A	0.03	6E-12	increase	up to 46,368 European ancestry individuals
						[0.04-0.06] %	
HbA1c NHGIR	rs1402837	ABCB11	G	0.05	8E-18	increase	up to 46,368 European ancestry individuals
						[0.024-0.040]	
HbA1c NHGIR	rs560887	G6PC2	T	0.03	1E-17	unit decrease	4,088 European ancestry non-diabetic elderly individuals
						[0.04-0.08] %	
HbA1c NHGIR	rs6474359	ANK1	T	0.06	1E-8	increase	up to 46,368 European ancestry individuals
HbA1c NHGIR	rs7072268	HK1	A	0.05	2E-25	[NR] % increase	14,618 European ancestry females
HbA1c NHGIR	rs730497	GCK	A	0.03	6E-12	[NR] % increase	14,618 European ancestry females
						[0.02-0.04] %	
HbA1c NHGIR	rs7998202	ATP11AUN	G	0.03	5E-9	increase	up to 46,368 European ancestry individuals
						[0.02-0.04] %	
HbA1c NHGIR	rs855791	TMPRSS6	A	0.03	3E-14	increase	up to 46,368 European ancestry individuals

						[0.07-0.11]		
						mmol/L		
T2D NHGIR	rs10423928	GIPR	A	0.09	3E-12	increase		15,234 individuals
T2D NHGIR	rs10440833	CDKAL1	A	1.25	2E-22	[1.20-1.31]		8,130 European ancestry cases, 38,987 European ancestry controls
T2D NHGIR	rs10501320	MADD	G	NR	1E-88	NR		10,701 European ancestry individuals
		UBA52P6 -						12,171 European ancestry cases, 56,862 European ancestry controls
T2D NHGIR	rs10811661	DMRTA1	T	1.18	1E-27	[1.13-1.24]		1,804 Mexican ancestry cases, 779 Mexican ancestry controls
T2D NHGIR	rs10838687	MADD	T	NR	7E-12	NR		10,701 European ancestry individuals
T2D NHGIR	rs10923931	NOTCH2	T	1.13	4E-8	[1.08-1.17]		4,549 European ancestry cases, 5,579 European ancestry controls
T2D NHGIR	rs10946398	CDKAL1	C	1.16	1E-8	[1.10-1.22]		1,924 European ancestry cases, 2,938 European ancestry controls
		UBA52P6 -						
T2D NHGIR	rs10965250	DMRTA1	G	1.2	1E-10	[1.13-1.27]		8,130 European ancestry cases, 38,987 European ancestry controls
								12,171 European ancestry cases, 56,862 European ancestry controls
		HHEX -						6,952 East Asian ancestry cases, 11,865 East Asian ancestry controls
T2D NHGIR	rs1111875	EXOC6	C	1.15	3E-19	[1.11-1.19]		5,561 South Asian ancestry cases, 14,458 South Asian ancestry controls
								1,804 Mexican ancestry cases, 779 Mexican ancestry controls
								12,171 European ancestry cases, 56,862 European ancestry controls
								6,952 East Asian ancestry cases, 11,865 East Asian ancestry controls
		CDC123 -						5,561 South Asian ancestry cases, 14,458 South Asian ancestry controls
T2D NHGIR	rs11257655	CAMK1D	T	1.06	3E-9	[1.01-1.11]		1,804 Mexican ancestry cases, 779 Mexican ancestry controls
T2D NHGIR	rs11558471	SLC30A8	A	NR	3E-18	NR		10,701 European ancestry individuals
T2D NHGIR	rs11603334	ARAP1	A	NR	3E-102	NR		10,701 European ancestry individuals
		ZFAND6 -						
T2D NHGIR	rs11634397	FAH	G	1.06	2E-9	[1.04-1.08]		8,130 European ancestry cases, 38,987 European ancestry controls
T2D NHGIR	rs11642841	FTO	A	1.13	3E-8	[1.08-1.18]		8,130 European ancestry cases, 38,987 European ancestry controls
								12,171 European ancestry cases, 56,862 European ancestry controls
								6,952 East Asian ancestry cases, 11,865 East Asian ancestry controls
								5,561 South Asian ancestry cases, 14,458 South Asian ancestry controls
T2D NHGIR	rs11717195	ADCY5	T	1.09	2E-8	[1.05-1.14]		1,804 Mexican ancestry cases, 779 Mexican ancestry controls
								12,171 European ancestry cases, 56,862 European ancestry controls
								6,952 East Asian ancestry cases, 11,865 East Asian ancestry controls
T2D NHGIR	rs12571751	ZMIZ1	A	1.09	2E-10	[1.06-1.13]		5,561 South Asian ancestry cases, 14,458 South Asian ancestry controls

								1,804 Mexican ancestry cases, 779 Mexican ancestry controls
						[0.05-0.09] mmol/L increase		
T2D NHGIR	rs1260326	GCKR	T	0.07	3E-10			15,234 individuals
		CDC123 -						
T2D NHGIR	rs12779790	CAMK1D	G	1.11	1E-10	[1.07-1.14]		4,549 European ancestry cases, 5,579 European ancestry controls 12,171 European ancestry cases, 56,862 European ancestry controls 6,952 East Asian ancestry cases, 11,865 East Asian ancestry controls 5,561 South Asian ancestry cases, 14,458 South Asian ancestry controls
		RPS3AP49 -						
T2D NHGIR	rs12970134	MC4R	A	1.08	3E-8	[1.03-1.12]		1,804 Mexican ancestry cases, 779 Mexican ancestry controls
		KRT18P24 -						
T2D NHGIR	rs13292136	CHCHD2P9	C	1.11	3E-8	[1.07-1.15]		8,130 European ancestry cases, 38,987 European ancestry controls
T2D NHGIR	rs1470579	IGF2BP2	C	1.14	2E-9	[1.09-1.19]		8,130 European ancestry cases, 38,987 European ancestry controls
T2D NHGIR	rs1531343	RPSAP52	C	1.1	4E-9	[1.07-1.14]		8,130 European ancestry cases, 38,987 European ancestry controls
		RPL29P30 -						
T2D NHGIR	rs1549318	LARP6	T	NR	2E-10	NR		10,701 European ancestry individuals
T2D NHGIR	rs1552224	ARAP1	A	1.14	1E-22	[1.11-1.17]		8,130 European ancestry cases, 38,987 European ancestry controls 12,171 European ancestry cases, 56,862 European ancestry controls
T2D NHGIR	rs163184	KCNQ1	G	1.09	2E-14	[1.04-1.13]		
								12,171 European ancestry cases, 56,862 European ancestry controls
T2D NHGIR	rs17106184	FAF1	G	1.1	4E-9	[1.07-1.14]		
								12,171 European ancestry cases, 56,862 European ancestry controls
T2D NHGIR	rs1727313	MPHOSPH9	C	1.06	1E-8	[1.04-1.08]		
		KRT18P24 -						12,171 European ancestry cases, 56,862 European ancestry controls
T2D NHGIR	rs17791513	CHCHD2P9	A	1.21	3E-8	[1.13-1.31]		
T2D NHGIR	rs1801214	WFS1	T	1.13	3E-8	[1.08-1.18]		8,130 European ancestry cases, 38,987 European ancestry controls 12,171 European ancestry cases, 56,862 European ancestry controls
T2D NHGIR	rs1801282	PPARG	C	1.16	6E-10	[1.10-1.23]		
		KCNQ1:KC						
T2D NHGIR	rs231362	NQ1OT1	G	1.08	3E-13	[1.06-1.10]		8,130 European ancestry cases, 38,987 European ancestry controls
		RNA5SP94 -						
T2D NHGIR	rs243021	MIR4432	A	1.08	3E-15	[1.06-1.10]		8,130 European ancestry cases, 38,987 European ancestry controls
T2D NHGIR	rs2877716	ADCY5	C	0.09	7E-16	[0.07-0.11]		15,234 individuals

						mmol/L increase	
T2D NHGIR	rs2943640	NYAP2 - MIR5702	C	1.09	7E-9	[1.05-1.13]	12,171 European ancestry cases, 56,862 European ancestry controls
T2D NHGIR	rs2943641	NYAP2 - MIR5702	C	1.19	9E-12	[1.13-1.25]	679 French ancestry cases, 697 French ancestry controls
T2D NHGIR	rs3132524	POU5F1	G	1.07	4E-9	[1.04-1.09]	12,171 European ancestry cases, 56,862 European ancestry controls
T2D NHGIR	rs3802177	SLC30A8	G	1.16	2E-18	[1.11-1.22]	12,171 European ancestry cases, 56,862 European ancestry controls
T2D NHGIR	rs4402960	IGF2BP2	T	1.17	2E-9	[1.11-1.23]	1,464 European ancestry cases, 1,467 European ancestry controls
T2D NHGIR	rs4430796	HNF1B	G	1.13	9E-10	[1.07-1.09]	12,171 European ancestry cases, 56,862 European ancestry controls
T2D NHGIR	rs4457053	ZBED3-AS1	G	1.08	3E-12	[1.06-1.11]	8,130 European ancestry cases, 38,987 European ancestry controls
T2D NHGIR	rs4458523	WFS1	G	1.09	2E-9	[1.06-1.13]	12,171 European ancestry cases, 56,862 European ancestry controls
T2D NHGIR	rs4502156	NPM1P47 - C2CD4B	T	NR	4E-20	NR	10,701 European ancestry individuals
T2D NHGIR	rs4506565	TCF7L2	T	1.36	5E-12	[1.20-1.54]	1,924 European ancestry cases, 2,938 European ancestry controls
T2D NHGIR	rs4607103	ADAMTS9- AS2	C	1.09	1E-8	[1.06-1.12]	4,549 European ancestry cases, 5,579 European ancestry controls
T2D NHGIR	rs4689388	WFS1	T	1.16	1E-8	[1.10-1.21]	679 French ancestry cases, 697 French ancestry controls
T2D NHGIR	rs4712523	CDKAL1	G	1.2	2E-12	[1.14-1.26]	679 French ancestry cases, 697 French ancestry controls
T2D NHGIR	rs4790333	SGSM2	T	NR	3E-9	NR	10,701 European ancestry individuals
T2D NHGIR	rs4812829	HNF4A	A	1.07	5E-8	[1.01-1.12]	12,171 European ancestry cases, 56,862 European ancestry controls
T2D NHGIR	rs5015480	HHEX - EXOC6	C	1.18	1E-15	[1.13-1.23]	8,130 European ancestry cases, 38,987 European ancestry controls
T2D NHGIR	rs5215	KCNJ11	C	1.08	3E-11	[1.04-1.12]	12,171 European ancestry cases, 56,862 European ancestry controls
T2D NHGIR	rs5219	KCNJ11	T	1.14	7E-11	[1.10-1.19]	1,161 European ancestry cases, 1,174 European ancestry controls
T2D NHGIR	rs5945326	KRT18P48 - DUSP9	A	1.27	3E-10	[1.18-1.37]	8,130 European ancestry cases, 38,987 European ancestry controls
T2D NHGIR	rs6235	PCSK1:CTD- 2337A12.1	G	NR	1E-26	NR	10,701 European ancestry individuals
T2D NHGIR	rs6808574	BCL6 - LPP- AS2	C	1.07	6E-9	[1.04-1.09]	12,171 European ancestry cases, 56,862 European ancestry controls
T2D NHGIR	rs6813195	RPS3AP18 - RPS14P6	C	1.08	4E-14	[1.06-1.10]	12,171 European ancestry cases, 56,862 European ancestry controls

		SLC35D3 -					
T2D NHGIR	rs6937795	RPL35AP3	A	1.04	7E-6	[1.02-1.06]	12,171 European ancestry cases, 56,862 European ancestry control
T2D NHGIR	rs702634	ARL15	A	1.06	7E-9	[1.04-1.09]	12,171 European ancestry cases, 56,862 European ancestry control
T2D NHGIR	rs7178572	HMG20A	G	1.08	2E-11	[1.04-1.13]	12,171 European ancestry cases, 56,862 European ancestry control
T2D NHGIR	rs7578326	LOC646736	A	1.11	5E-20	[1.08-1.13]	8,130 European ancestry cases, 38,987 European ancestry controls
T2D NHGIR	rs7578597	THADA	T	1.15	1E-9	[1.10-1.20]	4,549 European ancestry cases, 5,579 European ancestry controls
T2D NHGIR	rs7612463	UBE2E2	C	1.1	7E-9	[1.04-1.16]	12,171 European ancestry cases, 56,862 European ancestry control
T2D NHGIR	rs7754840	CDKAL1	C	1.12	4E-11	[1.08-1.16]	1,464 European ancestry cases, 1,467 European ancestry controls
T2D NHGIR	rs7756992	CDKAL1	G	1.2	2E-26	[1.16-1.25]	12,171 European ancestry cases, 56,862 European ancestry control
T2D NHGIR	rs7766070	CDKAL1	G	1.2	8E-9	[1.13-1.27]	1,399 European ancestry cases, 5,275 European ancestry controls
							2,112 lean type 2 diabetes cases, 4,123 obese type 2 diabetes case:
T2D NHGIR	rs7795991	CDKAL1	A	1.26	7E-10	[1.17-1.35]	54,412 controls
		RBMX2P4 -					12,171 European ancestry cases, 56,862 European ancestry control
T2D NHGIR	rs7901695	RPL26P21	G	1.05	7E-7	[1.03-1.07]	
T2D NHGIR	rs7903146	TCF7L2	T	1.38	2E-31	[1.31-1.46]	1,464 European ancestry cases, 1,467 European ancestry controls
T2D NHGIR	rs7957197	TCF7L2	T	NR	2E-20	NR	10,701 European ancestry individuals
T2D NHGIR	rs7961581	OASL	T	1.07	2E-8	[1.05-1.10]	8,130 European ancestry cases, 38,987 European ancestry controls
		TSPAN8 -					
T2D NHGIR	rs8042680	LGR5	C	1.09	1E-9	[1.06-1.12]	4,549 European ancestry cases, 5,579 European ancestry controls
T2D NHGIR	rs8050136	FTO	A	1.17	1E-12	[1.12-1.22]	1,161 European ancestry cases, 1,174 European ancestry controls
T2D NHGIR	rs8090011	FTO	A	1.23	7E-14	[1.18-1.32]	1,924 European ancestry cases, 2,938 European ancestry controls
							2,112 lean type 2 diabetes cases, 4,123 obese type 2 diabetes case:
T2D NHGIR	rs849134	LAMA1	G	1.13	8E-9	[1.09-1.18]	54,412 controls
T2D NHGIR	rs849135	JAZF1	A	1.13	3E-9	[1.09-1.18]	8,130 European ancestry cases, 38,987 European ancestry controls
T2D NHGIR	rs864745	JAZF1	G	1.12	2E-9	[1.08-1.17]	12,171 European ancestry cases, 56,862 European ancestry control
T2D NHGIR	rs896854	JAZF1	T	1.1	5E-14	[1.07-1.13]	4,549 European ancestry cases, 5,579 European ancestry controls
		TP53INP1:L					
		OC1019270					
T2D NHGIR	rs9502570	O2	T	1.06	1E-9	[1.04-1.09]	8,130 European ancestry cases, 38,987 European ancestry controls
		RREB1 -					
T2D NHGIR	rs972283	SSR1	A	1.06	1E-9	[1.04-1.08]	12,171 European ancestry cases, 56,862 European ancestry control
		KLF14 -					
T2D NHGIR	rs9936385	MIR29A	G	1.07	2E-10	[1.05-1.10]	8,130 European ancestry cases, 38,987 European ancestry controls

T2D NHGIR	rs9939609	FTO	C	1.13	1E-12	[1.09-1.18]	12,171 European ancestry cases, 56,862 European ancestry control
-----------	-----------	-----	---	------	-------	-------------	--

MAGIC

Manning et al 2012

Fasting insulin	rs7607980	COBLL1-GRB14	T
	rs2943634	IRSI	C
	rs4841132	PPP1R3B	A
	rs4691380	PDGFC	C
Fasting glucose		UHUHRF1B	
	rs4646949	P1	T
	rs2785980	LYPLAL1	T
	rs6048205	FOXA2	A
	rs1371614	DPYSL5	T
	rs13179048	PCSK1	C

Dupuis et al 2010

Fasting glucose	rs2293941	PDX1	A
HbA1c	rs1483121	OR4S1	G
T2D	rs10830963	MTNR1B	G
	rs4607517	GCK	A

Fasting insulin etc	rs2191349	DGKB	
	rs11708067	TMEM195	T
BMI adjust - no effect		ADCY5	A

Fasting insulin	rs7944584	MADD	A			
	rs10885122	ADRA2A	G			
	rs174550	FADSI	T	0.017	8.3*10-9	0.003
	rs11605924	CRY2	A			
	rs11920090	SLC2A2	T	0.02	3.3*10-6	0.004
	rs7034200	GLIS3	A	0.018	1.2*10-9	0.003
Fasting insulin	rs340874	PROX1	C	0.013	6.6*10-6	0.003

		C2CD48 -				
	rs11071657	FAM148B	A			
	rs780094	GCKR	C			
	rs35767	IGF1	G			
Scott et al						
2012						
Fasting	rs4869272	PCSK1	T			
glucose	rs11619319	PDX1	G			
	rs983309	PPP1R3B	T			
	rs6943153	GRB10	T	0.0154	1.6 * 10 ⁻¹²	0.002
	rs6113722	FOXA2	G			
	rs16913693	IKBKAP	T	0.0434	3.5 *10 ⁻¹¹	0.007
	rs3829109	LOC728489	G	0.0172	1.1*10 ⁻¹⁰	0.003
	rs3783347	WARS	G	0.0168	1.3*10 ⁻¹⁰	0.003
	rs2302593	GIPR	C			
	rs9368222	CDKAL1	A			
	rs10747083	P2RX2	A			
	rs6072275	TOP1	A	0.0159	1.7*10 ⁻⁸	0.003
	rs7651090	IGF2BP2	G			
	rs576674	KL	G	0.0167	2.3*10 ⁻⁸	0.003
	rs11715915	AMT	C	0.012	4.9*10 ⁻⁸	0.002
BMI adjusted	rs17762454	RREB1	T			
	rs7708285	ZBED3	G			
	rs2657879	GLS2	G	0.0157	3.9*10 ⁻⁸	0.003
	rs1421085	FTO	C			
	rs9884482	TET2	C			
	rs10195252	GRB14	T			
	rs1167800	H1P1	A			
	rs2820436	LYPLAL1	C			
	rs2745353	RSPO3	T	0.0143	5.5*10 ⁻⁹	0.002
	rs731839	PEPD	G	0.0145	1.7*10 ⁻⁸	0.003
	rs4865796	ARL15	A			

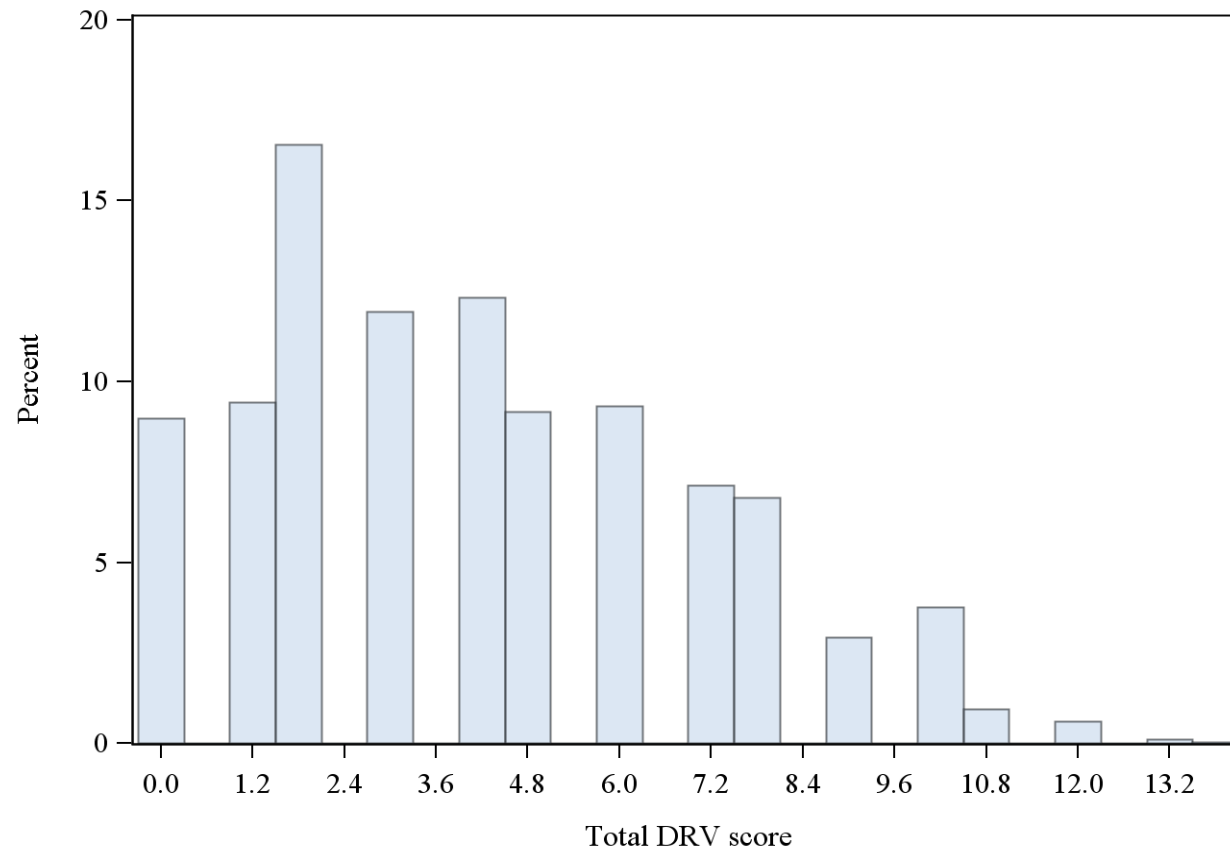
Fasting insulin	rs2972143	IRS1	G			
BMI adjusted	rs1530559	YSK4	A	0.0145	3.4*10-8	0.003
	rs2943645	IRS1	T			
	rs2126259	PPP1R3B	T			
	rs17036328	PPARG	T			
	rs974801	TET2	G			
		ANKRD55				
	rs459193	MAP3K1	G			
	rs6822892	PDGFC	A			
	rs4846565	LYPLAL1	G			
	rs3822072	FAM13A1	A	0.0116	1.8*10-8	0.002
	rs6912327	UHRF1BP1	T			
2hr glucose	rs6975024	GCK	C			
	rs11782386	PPP1R3B	C			
	rs1019503	ERAP2	A			

Appendix 7: Final list of 87 genetic markers included in genetic risk score

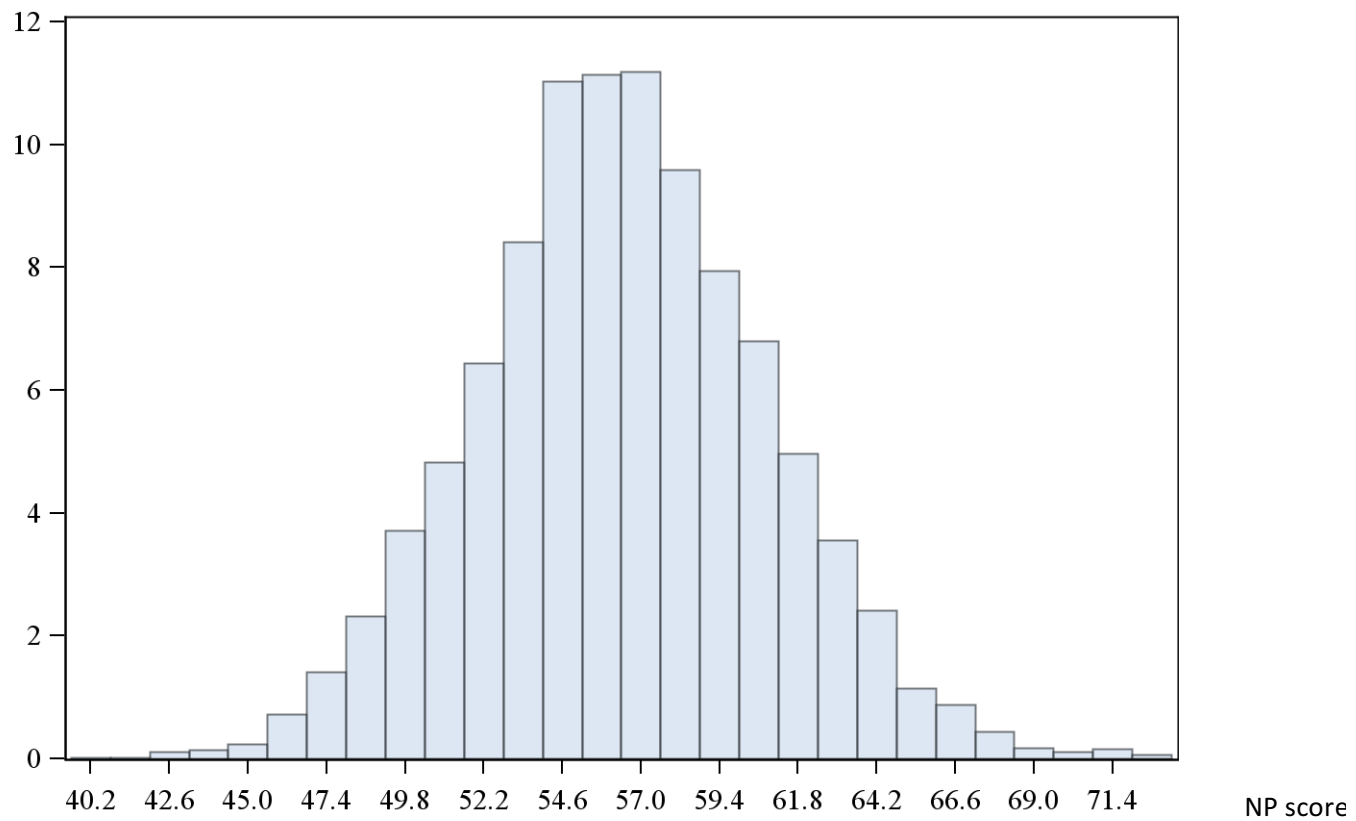
Genetic marker (SNP rs-ID)	Chromosome position	Risk allele
rs17791513	9:81905590	A
rs7034200	9:4289050	A
rs10965250	9:22133284	G
rs3829109	9:139256766	G
rs16913693	9:111680359	T
rs11782386	8:9201787	C
rs2126259	8:9185146	T
rs4841132	8:9183596	A
rs983309	8:9177732	T
rs6474359	8:41549194	T
rs13266634	8:118184783	T
rs1167800	7:75176196	A
rs6943153	7:50791579	T
rs730497	7:44223721	A
rs849134	7:28196222	G
rs864745	7:28180556	C
rs1074708	7:159102498	T
rs972283	7:130466854	A
rs9502570	6:7258617	T
rs6912327	6:34764922	T
rs3132524	6:31136714	C
rs1800562	6:26093141	G
rs4712523	6:20657564	G
rs6937795	6:137291281	A
rs2745353	6:127452935	T
rs6235	5:95728898	G
rs4457053	5:76424949	G
rs702634	5:53271420	A
rs3822072	4:89741269	A
rs4458523	4:6289986	G
rs4691380	4:157720124	C
rs6813195	4:153520475	C
rs9884482	4:106081636	C
rs4607103	3:64711904	C
rs7612463	3:23336450	C
rs6808574	3:187740523	C
rs4402960	3:185511687	T
rs11920090	3:170717521	T
rs1801282	3:12393125	C
rs11717195	3:123082398	T
rs855791	22:37462936	A
rs4812829	20:42989267	A
rs6072275	20:39743905	A
rs6113722	20:22557099	G

rs243021	2:60584819	A
rs7578597	2:43732823	T
rs1260326	2:27730940	T
rs7578326	2:227020653	A
rs560887	2:169763148	T
rs10195252	2:165513091	T
rs1530559	2:135755629	A
rs10423928	19:46182304	A
rs731839	19:33899065	G
rs12970134	18:57884750	A
rs1046896	17:80685533	T
rs4430796	17:36098040	G
rs4790333	17:2262703	T
rs9936385	16:53819169	C
rs8042680	15:91521337	C
rs11634397	15:80432222	G
rs7178572	15:77747190	G
rs1549318	15:71109147	T
rs4502156	15:62383155	T
rs3783347	14:100839261	G
rs576674	13:33554302	G
rs2293941	13:28491198	A
rs7998202	13:113331868	G
rs1531343	12:66174894	C
rs2657879	12:56865338	G
rs1727313	12:123640853	C
rs1387153	11:92673828	T
rs174550	11:61571478	T
rs10838687	11:47312892	T
rs163184	11:2847069	G
rs231362	11:2691471	G
rs5215	11:17408630	C
rs1111875	10:94462882	C
rs12571751	10:80942631	A
rs7072268	10:71099913	T
rs16926246	10:71093392	C
rs11257655	10:12307894	T
rs4506565	10:114756041	T
rs7901695	10:114754088	C
rs17106184	1:50909985	G
rs2785980	1:219700519	T
rs340874	1:214159256	C
rs10923931	1:120517959	T

Appendix 8: Histogram of the DRV score



Appendix 9: Histogram of the nutrient profile score



Appendix 10: Histogram of the genetic risk score

