Predictive validity of WXYfm and SAIN,LIM food nutrient profiling models in the Whitehall II cohort

Gabriel Masset

A thesis submitted for the degree of Doctor of Philosophy

University College London
Research Department of Epidemiology and Public Health

DECLARATION

I, Gabriel Masset, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

I was not involved in the development of the Whitehall II study and the design of the dietary assessment questionnaire. I did not collect the data (questionnaires, screening, and follow-up of events) used in this thesis.

ACKNOWLEDGMENTS

My first thanks go to my supervisors for their continuous support and their precious feedback over the three years of this project. I particularly appreciated the regular guidance of Dr Eric Brunner; the highly valuable but relaxed meetings in Oxford with Dr Mike Rayner and Dr Peter Scarborough; and the initial support given by Prof Gita Mishra which put me on the right tracks from the very first day.

My second thoughts are for the entire Whitehall II study team, which always gave me timely help to conduct this project. I am especially grateful to Dr Aida Sanchez for all data access and management issues and to the “study room” for the practical support and the advice regarding data collection. A special mention goes to all participants of the Whitehall II cohort, for their past and present contribution to scientific research.

A particular thanks to all my fellow PhDs, namely those from the very one “Room 350”. Sharing this experience with you all will remain unique.

I would further like to thank all the persons who helped me during this stay in the Department of Epidemiology and Public Health: the ever-smiling Sandy, the very precious Catherine, the graduate tutors Richard Marsh and Anne McNunn, and all other members of staff with whom regular chats livened the working days.

I would also like to acknowledge the British Heart Foundation, for the financial support and the very useful “fellows day”.

Last, this project could not have been possible without the perpetual support of my family and friends, merci!

Abstract

Background: Nutrient profiling (NP) aims to identify healthier food options according to the content of selected 'positive' nutrients e.g. fibre, protein, and 'negative' nutrients e.g. sodium, saturated fat. The British and French food safety agencies developed the WXYfm and SAIN,LIM models, respectively. Their predictive validity in relation to chronic disease has yet to be demonstrated.

Aim: To test the hypothesis that 'healthy' diets as defined by NP have predictive validity.

Methods: Between 1991-93, 7,251 participants of the Whitehall II study completed a 127-item food frequency questionnaire (FFQ). WXYfm and SAIN,LIM scores for each FFQ-item were used to derive energy-weighted aggregate diet scores (AS) for each participant and NP model. Validity was assessed against baseline factors including dietary quality indices. Prospective associations were examined with incident CHD, diabetes and cancer, and all-cause mortality (318, 754, 251, and 524 events, respectively—median follow-up time was approximately 17 years).

Results: AS were weakly associated with dietary quality indices. Cox modelling identified U-shaped associations (p quadratic trend <.05) between both AS and all outcomes except diabetes. Participants with middle AS had slightly reduced risk; SAIN,LIM estimates were significant for CHD and all-cause mortality. Dietary misreporting, particularly of energy-dense foods with high 'negative' nutrient content, was associated with BMI, hypertension and other risk factors, and explained much of the unexpected U-shaped AS-outcome associations. Alternative AS less sensitive to dietary misreporting confirmed the potential of NP as a public health tool. In particular, the WXYfm 'positive' nutrients predicted risk reduction for all outcomes.

Conclusions: Predictive validity of the NP approach was partly established. The prospective effects of AS on chronic disease outcomes were confounded by the association between vascular risk and energy misreporting. Further predictive validity studies of NP methods ideally require food-based dietary assessment (e.g. diet diaries, 24h recalls) with less reporting bias.

Table of contents

List of tables

List of figures

List of abbreviations

7DD	7-day (diet) diary
Afssa/Anses	French food safety agency (Agence française de sécurité sanitaire des aliments)
AHA	American Heart Association
AHEI	Alternative healthy eating index
ANOVA	Analysis of variance
AS	Aggregate score
BHF	British Heart Foundation
BMI	Body mass index
BMR	Basal metabolic rate
CHD	Coronary heart disease
CI	Confidence interval
CNC	Centre for Nutritional Epidemiology in Cancer Prevention and Survival
COMA	Committee on Medical Aspects of Food and Nutrition Policy
CVD	Cardiovascular disease
DQI	Diet Quality Index
DRI	Dietary Reference Intakes
ECG	Electro cardiogram
EE	Energy expenditure
EI	Energy intake
EPIC	European Prospective Investigation into Cancer and Nutrition
EU	European Union
EWS	Energy weighted score, WXYfm aggregate score
EWS+	Aggregate score for WXYfm with positive components only
FAO	Food and Agriculture Organisation
FDA	Food and Drug Administration
FFQ	Food-frequency questionnaire
FQS	Food Quality Score nutrient profiling model
FSA	Food Standards Agency
FSANZ	Food Standards Australia New Zealand
FVS	Food variety score
HDL	High-density lipoprotein
HEI	Healthy Eating Index
HER	High energy reporter
HR	Hazard ratio
ICD	International Classification of Diseases
LDL	Low-density lipoprotein
LER	Low energy reporter
MI	Myocardial infarction
MONICA	Multinational Monitoring of Trends and Determinants in Cardiovascular Disease
MRC	Medical Research Council
MUFA	Mono-unsaturated fatty acid
NFI	Nutritious Food Index nutrient profiling model
NHANES	National Health and Nutrition Examination Survey
NHS	National Health Service
NP	Nutrient profiling
NRF	Nutrient Rich Food nutrient profiling model
NSP	Non-starch polysaccharides
Ofcom	Office of Communications
OGTT	Oral glucose tolerance test
ONQI	Overall Nutritional Quality Index nutrient profiling model

OS	WXYfm overall score	SD	Standard deviation
PAL	Physical activity level	SEM	Structural equation modelling
PES	Percentage of energy score	SFA	Saturated fatty acid
PES(Q1)	SAIN,LIM aggregate score	SSCg3d	Nutrient profiling model developed for the British Food Standards Agency, earlier version of WXYfm
PUFA	Poly-unsaturated fatty acid	UCL	University College London
PWS	Portion weighted aggregate score	UK	United Kingdom
RACC	Reference Amounts Customarily Consumed	UN	United Nations
RDI	Recommended daily intake	US	United States of America
RFS	Recommended food score	WHO	World Health Organization
RRR	Ratio of restricted to recommended food components nutrient profiling model	WWS	Weight weighted aggregate score
SAIN,LIM	Nutrient profiling model developed for the French food safety agency	WXYfm	Nutrient profiling model developed for the British Food Standards Agency

Chapter 1: Introduction

Consumers from most developed countries are exposed to a vast array of dietary goals which range from nutrient recommendations (e.g. Dietary Reference Values in the UK (Department of Health, 1991)) to food-based guidelines (e.g. Dietary Guidelines for Americans in the US (US Department of Health and Human Services and US Department of Agriculture, 2005)). These recommendations were derived from a considerable body of evidence linking diet to health, and particularly non-communicable chronic disease, synthesised in several key reports (Department of Health, 1994; Department of Health, 1998; World Health Organization, 2003; World Health Organization, 2004; World Cancer Research Fund and American Institute for Cancer Research, 2007; Parkin *et al.*, 2011). They led to a commonly accepted definition of a healthy dietary pattern well illustrated by the British "Eatwell plate" and the US ChooseMyPlate.gov schemes (National Health Service, 2011; US Department of Agriculture, 2012), i.e. high intake of unrefined carbohydrates, fruit, and vegetables; moderate intake of dairy, meat, fish, and egg products; limited intake of added fats, and sweet or salty foods.

Despite such evidence and numerous government-endorsed programmes promoting healthier dietary choices (e.g. "5-a-day" in several European countries), most individuals fail to achieve healthy diets. In the UK, findings from the latest National Diet and Nutrition Survey indicated that adults' mean intake of fruit and vegetables, saturated fats, non-milk extrinsic sugars, and non-starch polysaccharides did not meet the recommendations (Department of Health & Food Standards Agency, 2011). Prevalence of diet-related risk factors is on the increase, with 2.6 million cases of diabetes diagnosed in 2009 and rising levels of obesity (Gonzalez *et al.*, 2009; Diabetes UK, 2010; National Obesity Observatory, 2011), and could hinder the downward trends in non-communicable diseases observed in the last decade (Office for National Statistics, 2011). In 2005, diet related illness was estimated to cost the NHS £6 billion per year (Rayner & Scarborough, 2005).

The general public does not appear to fully embrace public health messages, and alternative strategies are needed to help people make healthier dietary choices.

Consumers buy food instead of nutrients or food-groups; advice on individual foods could be very helpful to shift behaviours towards better options. A tool signposting the healthiest choices within a food basket or shelf could help practitioners in their day-to-day advice, and consumers in their weekly shopping (Muller & Ruffieux, 2011). Such a lever could also be used by regulators, manufacturers, and large retailers who have the power to shape a global food supply. Public health and food sector stakeholders could all benefit from such a tool, if objective and evidence-based.

Nutrient profiling could be this key "missing link" between nutrient recommendations and food-based guidelines (Darmon, 2009). This quantitative tool aims at "categorising foods according to their nutritional content" (Rayner *et al.*, 2004a) to derive an objective measure of the "healthiness" of an individual food. Nutrient profiling is based on the principle that some foods are more likely than other to contribute towards healthy dietary patterns. The hypothesis is that increased consumption of such "healthier foods" would, in turn, lead to reduced risk of chronic disease.

Nutrient profiling is currently being used for mandatory or voluntary food labelling. In Denmark, Norway, and Sweden, an official "Keyhole" logo appears on the packaging of healthier food options, as defined by a nutrient profiling model (Swedish National Food Administration, 2009). The Choices International program, developed mainly with funding from Unilever, allows foods from participating manufacturers to carry the "Choices logo" if within the appropriate nutrient content thresholds (Choices International Foundation, 2009). In the US, several charity logos have been recently developed (e.g. the American Heart Association Nutritional criteria for certified foods (2009)) alongside commercial and patented labels (e.g. NuVal, www.nuval.com). Nutrient profiling has also been used by governments to regulate health claims made on food. The US Food and Drug Administration (FDA) uses a simple model to allow access to such claims (Food and Drug Administration (FDA), 2008). A similar application of nutrient profiling was proposed by a recent EU regulation (n.1924/2006 (The European Parliament and the Council of the European Union, 2006)). The French food safety agency developed the SAIN,LIM nutrient profiling model for this purpose but no agreement could be reached between

member states (Agence française de sécurité sanitaire des aliments, 2008; Darmon *et al.*, 2009). The SAIN,LIM model was further proposed to be part of the French national diet and health program (Programme national nutrition santé) (Bourdillon *et al.*, 2010). In the UK, the WXYfm model developed for the Food Standards Agency (FSA) (Rayner *et al.*, 2005a) is currently being used by the regulator of broadcasting—Ofcom—to regulate advertising access during television programmes directed at children (Office of communications, 2007b) and product placement for all TV programmes produced under UK jurisdiction (Office of communications, 2011). Other potential applications include school vending (e.g. foods would need to pass a nutrient profiling model criterion to be sold on school sites (World Health Organization, 2006)) and fiscal and trade policies. For example, a soft-drink tax has been proposed by many scientists (Brownell *et al.*, 2009), and has been implemented in several US states. Such taxes need a nutrient profile model (if only rudimentary) to define which soft-drinks should be taxed (Jacobson & Brownell, 2000; Chriqui *et al.*, 2008; Sturm *et al.*, 2010). The WHO is currently working on a manual which aim is to set guiding principles to help its Member States in the implementation of nutrient profiling based policies (World Health Organization, 2011).

If nutrient profiling is to be used as a regulatory tool, it needs to be adapted to local cultures and production; the definition of a unique set of rules for a large region may prove difficult as illustrated in the EU. Further, to be accepted by all stakeholders, including public health practitioners and the food industry which often have diverging interests, it would need to be evidence-based, i.e. proven to improve the health status of individuals by promoting a population shift towards healthier eating. The validation of nutrient profiling has been investigated by several authors (Scarborough *et al.*, 2007b; Drewnowski & Fulgoni, 2008; Drewnowski *et al.*, 2008; World Health Organization, 2011). It is generally agreed that several steps common to the validation of any new scientific measurement need to be included (Cronbach & Meehl, 1955).

"Criterion" validity refers to the comparison of the new method with a known "gold-standard". There is no existing gold-standard measure of food quality, or food healthiness, and criterion validity *per se* cannot be tested. The comparison of nutrient profiling models to classification of foods obtained from "nutritional experts" has

been considered as criterion oriented validity (Scarborough *et al.*, 2007a). Both WXYfm and SAIN,LIM performed well against such classification but experts were shown to be culturally biased (Azais-Braesco *et al.*, 2006; Scarborough *et al.*, 2007a). The hypothesis beneath nutrient profiling is that diets containing higher amounts of healthy foods would in turn be healthier. "Construct" validity uses this hypothesis and relates to the ability of a nutrient profiling model to be associated with dietary goals (i.e. nutrient recommendations and food-based guidelines). Several measures exist to assess the integrated healthiness of a whole diet (e.g. Healthy Eating Index (McCullough *et al.*, 2000a; McCullough *et al.*, 2000b), data-driven dietary clusters (Martikainen *et al.*, 2003)). "Convergent" validity assesses the association between nutrient profiling and these measures. The WXYfm and SAIN,LIM models were both tested for construct and convergent validity and related well to dietary goals and measures of diet quality (Arambepola *et al.*, 2008; Maillot *et al.*, 2011). However, the dietary recommendations used in the construct and convergent validity testing are included in the nutrient profiling models' algorithms. Testing nutrient profiling models against objective measures of health status would avoid such loophole.

The "ideal" way of assessing nutrient profiling would be to demonstrate "predictive" validity (Drewnowski & Fulgoni, 2008), i.e. associations with prospective health outcomes (e.g. CVD or cancer). The most practical approach to test for predictive validity would be by means of prospective analysis of existing cohort study data. Intervention studies would be too long, expensive, and impractical. To date, only one commercial and patented nutrient profiling model was tested for predictive validity in two US cohorts. Individuals with diets containing higher amounts of healthy foods, as defined by the Overall Nutritional Quality Index (ONQI), had lower prospective risk of cardiovascular and total mortality (Chiuve *et al.*, 2011). For evidence-base policies to be effective, similar type of studies need to be carried out using data from other populations (and local data if possible) and using other nutrient profiling models—particularly those designed for regulatory purposes.

The Whitehall II cohort is well suited to test predictive validity within a British population. 10,308 well-characterised civil servants were recruited in 1985 and subsequently completed dietary assessment questionnaires in 1991; they have been

followed until 2010 for verified incident circulatory disease, diabetes, and cause of mortality (Marmot & Brunner, 2005).

The two nutrient profiling models mentioned above, WXYfm and SAIN,LIM, are both government-endorsed schemes developed for the respective food safety agencies. Their algorithms have been designed through intensive consultation processes and are freely available. Both schemes have been linked to healthier diets (Arambepola *et al.*, 2008; Darmon *et al.*, 2009; Maillot *et al.*, 2011) but associations with health outcomes have not yet been investigated.

This project aimed to assess the relationship between dietary quality derived from the WXYfm and SAIN,LIM nutrient profiling models and prospective health outcomes including coronary heart disease, diabetes, and cancer mortality within the Whitehall II cohort study. It was hypothesised that diets containing higher proportions of "healthier" foods, as defined by both models, would be predictive of improved health outcomes.

Individuals' diet can depend on a variety of factors and dietary assessment methods may be subject to bias. Hence, a further aim of this project was to analyse potential sources of confounding and bias which could have affected the observed associations for predictive validity. Such an investigation allowed adding an additional step in the development and validation process of nutrient profiling models: the identification of results-led models which would be able to predict adverse health outcomes with greater sensitivity and specificity.

Chapter 2: Literature review and background

This chapter (Chapter 2) reviews the published evidence on nutrient profiling (NP) validation in general. The goal of the project being to assess the NP concept and underlying hypothesis, i.e. diets containing more healthy foods promote better health status, focus is not put on the specific applications of individual NP models, e.g. food labelling designed to shift consumers' buying behaviours. This review is preceded by a short presentation of selected existing NP models and a detailed description of WXYfm and SAIN,LIM, the two models used in this project. The chapter then focuses on the evidence linking diet and health within the Whitehall II study, to assess the potential of the data with respect to NP validation.

2.1 Nutrient profiling schemes and their validation

The idea of assessing nutritional quality of individual foods was first introduced in the 1970s (Hansen, 1973; Sorenson *et al.*, 1976; Guthrie, 1977). It has attracted interest in the last few decades when attention was shifted from diets to foods. For example, there were moves in many countries to encourage consumers to rely on food labels to choose the healthier option: "there needs to be better, clearer information on nutrition labels connecting an individual food product to a consumer's overall diet. […] People shouldn't need a calculator or an advanced degree in math or nutrition to calculate what makes a healthy diet" (McClellan, 2003). There is a growing interest in NP illustrated by the publication in recent years of specific supplements on NP in three scientific journals (Eur J Nutr (2007) 46(S2), J Am Coll Nutr (2009) 28(4), Am J Clin Nutr (2010) 91(4)).

This section contains first a short description of NP parameters, i.e. the characteristics to be determined when designing a model, together with a few selected models and the details of the WXYfm and SAIN,LIM models. Literature was then searched for NP validation methods.

2.1.1 Nutrient profiling schemes characteristics

NP models are based on some specific parameters set when developing a model (Scarborough *et al.*, 2007c; Drewnowski *et al.*, 2008). Two broad approaches have been used to design food NP models:

- "across-the-board" schemes where all foods are scored/classified according to the same algorithm, to identify healthy foods in general;
- "category-specific" designs where specific algorithms are defined for a number of food groups, to identify healthier options within these categories.

In a recent study, Scarborough and colleagues investigated the pros and cons of each approach (Scarborough *et al.*, 2010). The "category-specific" method, with a limited number of categories, was considered more appropriate for promoting healthier diets.

Other features of nutrient profiling models include the following choices:

- The choice and number of nutrients. These can be positive (or valued) nutrients supposed to be beneficial and/or negative nutrients (to limit) which have been shown to be detrimental. Several studies have investigated the effect of different sets of nutrients with the same basic NP model, and conclusions supported the use of a limited number of nutrients (Agence française de sécurité sanitaire des aliments, 2008; Fulgoni *et al.*, 2009).
- The choice of recommended values for the selected nutrients. These are usually derived from national and international nutrient recommendation, and can be adapted to specific applications (e.g. school vending).
- "Reference amount" or "base" of the scheme (usually 100kcal, 100g or portion size), which indicates the amount of food on which the algorithm calculates the nutrient content. It has been argued that a 100kcal basis represents better the nutrient density of positive nutrients better, while a 100g basis is more appropriate for negative nutrients (Drewnowski *et al.*, 2009).
- The choice of an algorithm to combine the nutrient content information, and crucially the way of balancing positive and negative nutrients (e.g. sum or ratio) (Fulgoni *et al.*, 2009).
- The use of thresholds to separate "healthy" and "unhealthy" foods, allowing an easier interpretation of the foods rankings. Such thresholds have usually

been implemented in regulatory models and category-specific models to highlight the approved or healthy options.

Each feature needs careful consideration, and models under development usually undergo several steps of internal validity and/or peer-review before being published in their final version (Rayner *et al.*, 2004a; Drewnowski, 2005; Rayner *et al.*, 2005b; Rayner *et al.*, 2005c; Agence française de sécurité sanitaire des aliments, 2008).

Selected models are presented in tables 2.1 and 2.2. Some of these adopt an across-the board approach and some a category-specific approach. They include the WXYfm and SAIN,LIM models used in this project and presented in the following sections. These tables do not present an exhaustive list of existing NP models, but rather a selection of published models combining different aspects of the features presented above. An exhaustive review of existing nutrient profiling models was published by the British Food Standards Agency (Stockley *et al.*, 2007) and should be updated shortly.

Table 2.1: Summary of selected "across-the-board" food nutrient profiling models[#]

Name	Algorithm	Base	Valued Nutrients	Nutrients to limit	References
Nutritious Food Index (NFI) [a]	$NFI = \sum \left(w \cdot \%DV_{positive} + w \cdot \%DV_{negative} \right)$	Serving	Fibre, calcium, iron, zinc, magnesium, potassium, phosphorus, niacin, folate and vitamins A, C, B1 and B2.	Total fat, SFA, cholesterol, sodium.	(Gazibarich & Ricci, 1998)
Ratio of recommended to restricted food components (RRR)[b]	$RRR = \frac{\sum \left(Nutrient_{Good} / 6 \right)}{\sum \left(Nutrient_{restricted} / 5 \right)}$	Serving	Protein, fibre, calcium, iron and vitamins A and C.	Energy, SFA, total sugar, cholesterol, sodium.	(Scheidt & Daniel, 2004);
Food Quality Score 1, 2, and 3 (FQS 1,2,3)	$FQS_{1/2/3} = \frac{\sum_1^{n1/n2/n3} \%DV_{1/2/3} \Big/ n1/n2/n3}{\sum_1^5 \%DV \Big/ 5}$	2000kcal	$\mathbf{n_1}$: fibre, vitamins A, C, E, D, and B12, folate, calcium, magnesium, iron, potassium. $\mathbf{n_2}$: same, but category specific. $\mathbf{n_3}$: n1 + protein, phosphorous, zinc, copper, niacin, pantothenic acid, vitamins B1, B2, K and B6, manganese, selenium.	Denominator: energy, SFA, cholesterol, sodium, and energy from fats.	(Kennedy *et al.*, 2008)
Calories for Nutrient (CFN) [c]	$CFN = \frac{ED}{(\sum_1^{13} \%DV_{100g}) / 13}$	1000kcal	Protein, calcium, iron, zinc, magnesium, folate, niacin and vitamins A, C, B1, B2, B6 and B12.		(Zelman & Kennedy, 2005)
WXYfm	Section 2.1.2	100g	Protein, fibre, fruit/vegetable/nut content.	SFA, sodium, total sugars, energy.	(Rayner *et al.*, 2005a)
SSCg3d	Earlier version of WXYfm.	100g	n-3 fatty acids, fruit/vegetable content, calcium, iron.	SFA, sodium, added sugar, energy.	(Rayner *et al.*, 2005c)
FSANZ	Adapted from WXYfm model.	100g	Protein, fibre, fruit/vegetable/nut content.	SFA, sodium, total sugars, energy.	(Food Standards Australia New Zealand)
SAIN,LIM	Section 2.1.3	100kcal / 100g	5 from protein, fibre, calcium, iron, ALA, MUFA, vitamins C, D, and E	Sodium, SFA, added sugar.	(Darmon *et al.*, 2009)
Nutrient Rich Food (NRF9.3)[d]	$NRF9.3 = \frac{\sum_1^9 \%DV}{9} - LIM$	100kcal or RACC	Protein, fibre, calcium, iron, magnesium, potassium and vitamins A, C, E and B12.	SFA, added (or total) sugar, sodium.	(Fulgoni *et al.*, 2009)

[#]Abbreviations: $\%DV_i$, percent of daily value (recommended intake) for a nutrient in the reference amount or in amount i; SFA, saturated fatty acids; MUFA, Mono-unsaturated fatty acids; ALA, α-linolenic acid. [a] *w*, weight given to individual nutrients. [b] *Nutrient*: nutrient content per serving. [c] ED, energy density. [d] LIM, see SAIN,LIM model presentation in section 2.1.3; RACC, reference amount customarily consumed.

Table 2.2: Summary of selected "category-specific" food nutrient profiling models[#]

Name/Organisation	Number of categories	Reference amount	Valued nutrients	Nutrients to limit	Reference
Tripartite classification model [a]	14	100g	n-3 fatty acid, fibre, vitamin C, and folate.	SFA, sodium, sugar, and energy.	(Netherlands Nutrition Centre (NNC), 2005)
Food and Drug Administration	3	Serving	N/A	Total fat, SFA, sodium and cholesterol.	(Food and Drug Administration (FDA), 2008)
American Heart Association [b]	2	Serving	One from protein, fibre, iron, calcium and vitamins A and C.	Total fat, SFA, cholesterol, TFA, sodium.	(American Heart association (AHA), 2009)
Choices programme	28	Depends on category and nutrient.	Fibre (depending on category).	SFA, trans fatty acid, sodium, added sugars, energy.	(Choices International Foundation, 2009)
Australian Heart Foundation	>10 [c]	Serving or 100g	Both negative and positive nutrients.		(National Heart Foundation of Australia (AHF), 2009)
Nutrimap	7	100kcal	MUFA, PUFA, fibre, folic acid, vitamins D, C, and e, calcium, iron, and magnesium.	Total carbohydrates, sugars, total lipids, SFA, and sodium.	(Labouze *et al.*, 2007)
US National Heart Blood and Blood Institute – Go, Slow, and Whoa foods	8	N/A	N/A, Foods from each category are divided into one of the three healthiness groups.		(US Department of Health and Human Services - National Heart Lung and Blood Institute)
Keyhole logo	25	100g + %energy	Fibre (depending on category).	TFA (all products), total fat, SFA, total or refined sugars, and sodium (depending on the category).	(Swedish National Food Administration, 2009)
Overall Nutritional Quality Index (ONQI)	?	? Patented and undisclosed	Fibre, folate, vitamins A, C, D, E, B6, B12, potassium, calcium, zinc, n-3 fatty acids, total flavonoids, total carotenoids, magnesium, and iron..	SFA, TFA, sodium, added sugar, cholesterol. Further includes fat quality, protein quality, energy density and glycemic load as correcting factors.	(Katz *et al.*, 2010)

SFA, saturated fatty acid; MUFA, mono-unsaturated fatty acid; PUFA, poly-unsaturated fatty acid; TFA, trans fatty acid.
[#] These models being category specific, the algorithm varies from one food category to another, resulting in different reference amount and/or nutrients being used by the same scoring system. [a] 3 healthiness classes are defined for each food category. [b] 2 specific logos are defined for whole-grain and whole-oats products. [c] 10 main groups are defined, with some further sub-categories within these main groups.

2.1.2 WXYfm

The WXYfm model was developed for the Food Standards Agency (FSA) for the regulation of food advertising on television programs aimed at children (Rayner *et al.*, 2005a). The model first allocates points on the basis of the nutritional content per 100g of the food or drink. Foods and drinks are then classified into healthiness categories.

Step 1

"A" points are calculated as follows:

Total "A" points = (points for energy)
+ (points for saturated fats)
+ (points for total sugars)
+ (points for sodium)

A maximum of 10 points can be scored for each nutrient (table 2.3). Individual nutrient thresholds are derived from the Guideline Daily Amounts (Rayner *et al.*, 2004a; Rayner *et al.*, 2004b).

Table 2.3: Thresholds for points scored by each "A" nutrient of the WXYfm model

	Thresholds for individual points							
Nutrient(/100g)	**0**	**1**	**2**	**3**	**…**	**8**	**9**	**10**
Energy (kJ)	≤335	>335	>670	>1005	…	>2680	>3015	>3350
Saturated fat (g)	≤1.0	>1.0	>2.0	>3.0	…	>8.0	>9.0	>10.0
Total sugars (g)	≤4.5	>4.5	>9.0	>13.5	…	>36.0	>40.0	>45.0
Sodium (mg)	≤90	>90	>180	>270	…	>720	>810	>900

Thresholds are derived from Guideline Daily Amounts

Step 2

"C" points are calculated as follows:

Total "C" points = (points for protein)
+ (points for non-starch polysaccharide (NSP) fibre)
+ (points for fruit, vegetable and nuts)

A specific report on the definition and the calculation of fruit, vegetable and nuts content was published (Scarborough *et al.*, 2005). A maximum of 5 points can be scored for each nutrient/food component as indicated in table 2.4.

Table 2.4: Thresholds for points scored by each "C" nutrient of the WXYfm model

	Thresholds for individual points					
Nutrient (/100g)	**0**	**1**	**2**	**3**	**4**	**5**
Protein (g)	≤1.6	>1.6	>3.2	>4.8	>6.4	>8.0
NSP fibre* (g)	≤0.7	>0.7	>1.4	>2.1	>2.8	>3.5
Fruit, vegetable and nuts (g)	≤40	>40	>60	-	-	>80

Thresholds are derived from Guideline Daily Amounts
*NSP, Non-starch polysaccharide

Step 3

The overall score is calculated with the "A" and "C" total points:

Overall score = (total "A" points) – (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:

Overall score = (total "A" points)
– (fibre + fruit, vegetable and nuts points)

Step 4

The food or drink is then assigned into one of the healthiness categories (figure 2.1):

- A food is classified as "less healthy" when it scores 4 points or more.
- A food is classified as "healthier" when it scores 0 points or less.
- A drink is classified as "less healthy" when it scores 1 point or more.
- A drink is classified as "healthier" when it scores 0 points or less.

The FSA and Ofcom use only the "4 points" threshold for foods: a food can be advertised if it scores less than 4 points. There are therefore only two categories for both foods and drinks: those that can be advertised, and those that can't.

Since the WXYfm model was designed to control access to TV advertisement on programmes aimed at children, alcoholic drinks were not scored by the model.

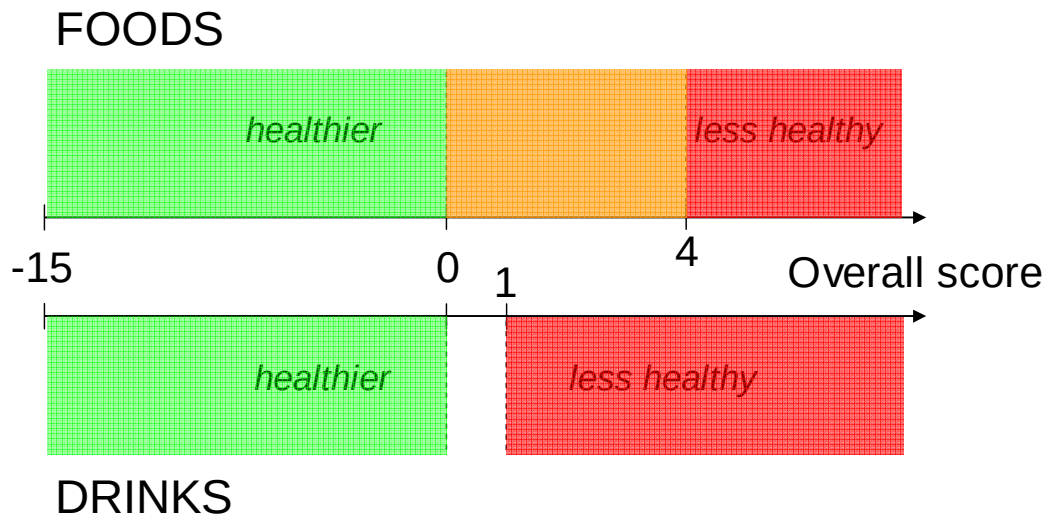


Figure 2.1: Classification into healthiness categories according to the WXYfm "overall score"

2.1.3 SAIN,LIM

The SAIN,LIM nutrient profiling model was proposed by the French food safety agency (Anses, formerly Afssa) (Agence française de sécurité sanitaire des aliments, 2008; Darmon *et al.*, 2009). This model is based on two previously published indicators: the Nutrient Density Score (NDS), based on qualifying nutrients (i.e. positive nutrients), and the LIM score, based on disqualifying nutrients (i.e. the nutrient to be limited) (Darmon *et al.*, 2005; Maillot *et al.*, 2007). Thresholds are defined for each of these sub-scores to define four healthiness categories or "quadrants".

Calculation of SAIN and LIM sub-scores

The SAIN score is an un-weighted arithmetic mean of the percentage adequacy for five positive nutrients. It is calculated for 100kcal of food, as follows:

$$\text{SAIN} = \frac{\sum_1^i ratio_i}{i} \times 100$$

$$\text{With } ratio_i = \left[\frac{nutrient_i}{RV_i} \right] \times \frac{100}{\varepsilon}$$

Where nutrient_i is the quantity (g, mg, or μg) of positive nutrient i in 100g of food, RV_i is the daily recommended value for nutrient i (table 2.5), and ε is the energy content of 100g of food (in kcal/100g).

The five basic nutrients included in the SAIN are protein, fibre, ascorbic acid, calcium, and iron. In addition to these five basic nutrients, optional nutrients are also used, which differ according to the lipid contents of individual foods. For foods providing less than 97% of their energy as lipids, vitamin D is used as an optional nutrient. This means that the vitamin D ratio is calculated for each food by using the $ratio_i$ algorithm and, when the vitamin D ratio is higher than the lowest ratio among the five basic ones, this lowest ratio is replaced by the vitamin D ratio in the SAIN algorithm. For foods providing more than 97% of their energy as lipids, four optional nutrients are used: vitamin D, vitamin E, α-linolenic acid, and mono-unsaturated fatty acids. The ratios calculated for these optional nutrients are compared with those obtained for the five basic nutrients, and up to two replacements are allowed between optional and basic nutrients in the SAIN algorithm.

Table 2.5: Recommended values (RV) and maximum recommended values (MRV) used to calculate each SAIN and LIM sub-scores, respectively

Sub-score	Nutrient	Value (RV or MRV)
SAIN	Protein (g)	65
	Fibre (g)	25
	Vitamin C (mg)	110
	Calcium (mg)	900
	Iron (mg)	12.5
	Vitamin D (μg)	5
	Vitamin E (mg)	12
	α-linolenic acid (g)	1.8
	Mono-unsaturated fatty acids (g)	44.4
LIM	Saturated fatty acids (g)	22
	Added sugars (g)[#]	50
	Sodium (mg)*	3,153

These values are based on French (Martin, 2001) and European (Eurodiet Core Report, 2000) nutritional recommendations. [#] If added sugars are not available, "free sugars", as defined by the WHO are used (World Health Organization, 2003). * Not including salt added at the table.

The LIM score is the mean percentage of the maximal recommended values for three nutrients: sodium, added sugars, and saturated fatty acids (SFA). The LIM score is calculated for 100g of food as follows:

$$\text{LIM} = \frac{\sum_1^3 ratio_j}{3}$$

$$\text{With } ratio_j = \left[\frac{nutrient_j}{MRV_j} \right] \times 100$$

Where $nutrient_j$ is the content (g, mg) of limited nutrient *j* in 100g of food, and MRV_j is the daily maximal recommended value for nutrient *j* (table 2.5). The LIM is multiplied by 2.5 for soft drinks.

Overall, the SAIN,LIM model is based on 8 basic nutrients (5 included in the SAIN plus 3 included in the LIM) plus 4 optional nutrients (in the SAIN only).

Threshold values for each sub-score

On the basis of a reference daily energy intake of 2000kcal, the optimum value for the SAIN is 100% for 2000kcal, which is equivalent to 5% for 100kcal of food. A SAIN value ≥5 therefore indicates a good nutrient density. The LIM is calculated for 100g and the reference value used to derive the threshold is based on food intake rather than on energy intake. Because the mean daily food intake (including solid foods only) observed in the French population was approximately 1330g/d (Volatier, 2000), the maximal value for the LIM score is 100% for 1330g, which is equivalent to 7.5% for 100g of food. As a result, a LIM value <7.5 indicates a low content of negative nutrients.

On the basis of its SAIN and LIM values and on the thresholds defined for each score, each food is classified into one of four possible SAIN,LIM quadrants as displayed in figure 2.2. Quadrant 1 includes foods with the most favourable nutrient profile (high nutrient density and low content of negative nutrients), whereas quadrant 4 includes foods with the least favourable nutrient profile (low nutrient density and high content of nutrients to limit). Foods from quadrants 2 and 3 are intermediate in terms of nutritional quality.

In regards to the European regulation n. 1924/2006 on claims made on food packaging, the Anses recommended that only foods within quadrants 1 and 2 should carry nutrient claims, and that only foods from quadrant 1 could carry health claims.

Similarly to WXYfm, alcoholic drinks were not scored by the SAIN,LIM model since the EU regulation n. 1924/2006 excluded this category from claim access.

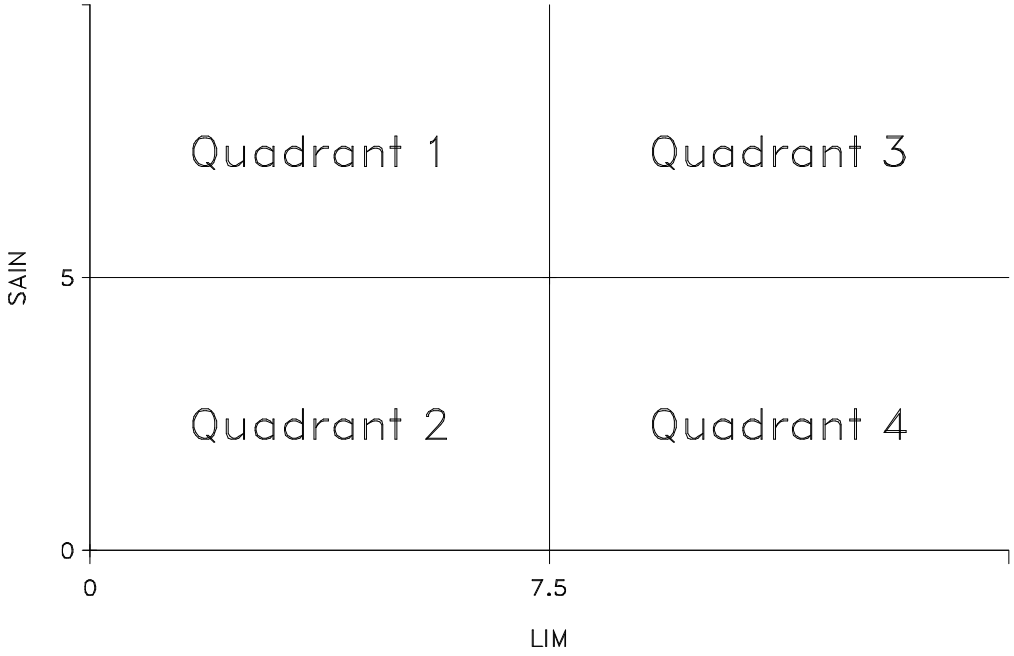


Figure 2.2: SAIN,LIM quadrants
Healthier foods are classified in quadrant 1; less healthy ones in quadrant 4.

2.1.4 Validation of nutrient profiling schemes

(i) Literature searches

As mentioned above, most NP models generally go through an internal validation process during their development. It was therefore decided that the most efficient approach to identify publications concerned with the validation of NP models was to search lists and references in known publications that already included some sort of validity testing. In addition, online searches were carried out in Pubmed and Google Scholar including the following search algorithm (all terms in title or abstract):

((nutrient OR nutritional OR food)
AND (profiles OR profile OR profiling OR profiler))
AND (validating OR validation OR valid OR validity)

Inclusion of the retrieved references was done on the basis of title, abstract, and main research objective. Studies which focused only on the application of NP models (e.g. assessing the impact of a health logo supported by a NP model) were excluded.

The search through known publications and related references identified 24 studies which aimed at validating or testing one or more NP model(s). The Pubmed search retrieved 193 articles, of which 15 were retained. The Google scholar search retrieved 17 articles, of which 5 were retained. All the retained publications had been previously identified.

(ii) Results, approaches to validate food nutrient profiling models

Several approaches aiming to validate one or more NP model have been published in recent years. These ranged from the comparison of rankings of a defined set of foods to mathematical programming and associations with prospective health outcomes within large scale longitudinal studies. In this section, five approaches are presented from the more simple (i.e. requiring the least material) to the more complex ones (i.e. requiring individual based data or more advance modelling):

a. Comparison of the rankings of foods from different NP models;
b. Comparison of the NP-derived rankings of foods with rankings from "nutrition experts";
c. Use of dietary survey data to compare NP with healthiness of diets and achievement of dietary goals, to test for construct and convergent validity;
d. Use of statistical modelling to design theoretical diets containing more or less healthy foods, to test for construct validity;
e. Predictive validity, i.e. prospective association with adverse health outcomes.

For each method, some publications compared several NP models to assess the most suited or most robust scheme; these were included in the respective sections. Most

NP models mentioned below were presented in the tables 2.1 and 2.2. Table 2.6 summarises the published approaches.

a. Comparison of foods rankings by several NP models

This first approach consists in the comparison of the rankings of a selected list of foods (usually chosen to represent well the intake of the target population) derived from two or more NP models. It has been commonly used in the development process of NP models, allowing scientists to assess their model (particularly different versions of the same model) with minimal effort and material.

Both the WXYfm and the SAIN,LIM were developed in several steps using such a validation method to improve the models (choice of nutrients, algorithm, reference amount, etc.) (Rayner *et al.*, 2005a; Rayner *et al.*, 2005c; Agence française de sécurité sanitaire des aliments, 2008). The final version of the WXYfm was further tested against the British Balance for Good Health (BGH, which preceded the Eatwell plate) and the results showed good agreement between the NP model and the BGH classification (i.e. healthier foods were more likely to belong to food groups which consumption was encouraged, and vice versa) (Arambepola *et al.*, 2008).

In addition, several studies compared rankings of foods from different NP models, either to compare existing models or to assess a new model. The Australia and New Zealand food safety agency developed a model derived from the WXYfm, and compared the rankings of foods from this new model with the rankings derived from the WXYfm and from an early version of the Choices program (Food Standards Australia New Zealand). In the US, Kennedy and colleagues (2008) proposed three models (Food Quality Scores 1, 2, and 3) to measure nutrient density based on the 2005 Dietary Guidelines for Americans. Kennedy et al. concluded that all three approaches ranked foods in a similar way and in agreement with the guidelines. Further, the proposed "Go, Slow, and Whoa" classification of foods by the US National Heart, Lung, and Blood Institute was tested against the nutrient rich food model (NRF9.3) and it was concluded that the proposed classification could be helpful in indentifying healthful food options (Drewnowski & Fulgoni, 2011). In a French dataset, the Nutrimap model proposed by Labouze and colleagues (2007) agreed well with the WXYfm model. The WXYfm and SAIN,LIM models were

included in an extensive study which compared existing NP models (Garsetti *et al.*, 2007). The researchers highlighted that while models agreed well on fruit and vegetables or sugars and oils, food groups unlikely to be the object of marketing claims, agreement was less good with composite and processed foods. The WXYfm and SAIN,LIM models were also included in an analysis of bakery products (Trichterborn *et al.*, 2011), and classified more foods as healthy than the Choices programme.

b. Comparison with rankings from nutrition experts

This second approach is very close to the first one, except that rankings derived from NP models are compared with rankings derived from "nutrition experts", who are hypothesised to give an external and true evaluation of foods healthiness. This method has also been used during the development of NP models since developers themselves assessed their model's rankings of foods. Yet, only a couple of systematic studies including the opinion of experts external to the development of the NP models under investigation have been published.

The WXYfm model, alongside the Nutrient Rich Food, the Calorie for Nutrient, and the Ratio of recommended to restricted food components models (table 2.1), was included in a French study which compared the rankings from these four models with the one obtained from 12 experts (Azais-Braesco *et al.*, 2006). Each expert had to classify into quintiles a list of 125 foods. The authors reported that the WXYfm model seemed to be the most consistent approach, with only a few "minor inconsistencies", e.g. fried onions classified better than currants.

The WXYfm model was also included in a British study involving 702 nutritional professional from the British Dietetic Association and the Nutrition Society (Scarborough *et al.*, 2007a; Scarborough *et al.*, 2007c). Each expert was e-mailed a random list of 40 foods out of a 120 food master list, and had to score each food on an absolute scale (6 categories from less healthy to more healthy). To assist with the categorisation, the energy (kcal), protein, carbohydrate, total sugar, fat, saturated fat, fibre, sodium, calcium and iron contents per 100g of foods were provided. Such "standard rankings" was then compared to rankings derived from the WXYfm, SSCg3d, Nutritious Food Index, Nutrient Rich Food, Ratio of recommended to

restricted food components, Dutch Tripartite, Australian Heart Foundation and American Heart Association models. The WXYfm and the SSCg3d were the most related to the "standard rankings".

The US Overall Nutritional Quality Index (ONQI) model was assessed by the experts from the committee involved in the model development (Katz *et al.*, 2010). It was shown that there was a good agreement overall and for specific food groups (except for fruit).

The use of the opinion of external experts could be considered as the closest approach to "criterion" validity (chapter 1) since it intends to be transparent and replicable. However, the standard rankings derived from nutritional experts could not be considered as a gold-standard. The main weakness was the cultural bias observed within the experts. For example, the results from the Scarborough et al. study showed that some words used in the descriptions of food overrode the experts' assessment of the nutritional composition of the food. For instance, "Take away stir fry vegetables" with a relatively low fat and saturated fat content and a relatively high fibre content was ranked as less healthy than dishes with a higher fat and saturated fat content.

c. Third approach: use of dietary survey, convergent validity

NP is based on the principle that healthier diets contain a higher proportion of healthier foods. This principle is used to test for convergent validity: NP models are assessed against healthiness of diets. Such an approach, which requires dietary survey data with associated nutrient content of foods, has been implemented by four research groups. In the first three examples, the NP scores of foods were aggregated at the participant level. Such aggregation is a necessary step for all studies linking food-based NP to characteristics of individuals (their whole diet in the case of convergent validity).

First, the approach was used during the development of the Nutrient Rich Food (NRF) model (Fulgoni *et al.*, 2009). Diets of participants from the US NHANES 1999-2002 surveys were ranked by all the versions of the NRF to be tested and by the Healthy Eating Index (HEI, see section 2.2.2) diet quality score. Each version of

the NRF was then regressed against the HEI. The NRF9.3 model (table 2.1) which accounted for most of the HEI variation was selected as the final version of the NP model (linear regression R^2 was 0.45).

Second, the HEI score was also used to assess the ONQI model using NHANES 2003-06 data (Katz *et al.*, 2010). The authors calculated an ONQI score for the reported total daily intake, and good agreement was obtained between quartiles of this ONQI score and the HEI (with around 4% of participants classified in opposite quartiles). The linear regression R^2 was lower than for the NRF9.3 (0.29, adjusted for age, sex, and ethnicity). Further, the ONQI score derived from the hypothetical Dietary Approaches to Stop Hypertension diet of the National Heart Lung and Blood Institute (2006) was significantly higher than the one derived from the average NHANES diet.

Third, the WXYfm model was tested against the Diet Quality Index (DQI (Patterson *et al.*, 1994)) among participants of the National Diet and Nutrition Survey of adults carried out in Great Britain in 2000–01 (Arambepola *et al.*, 2008). The energy intake from less healthy foods was closely related to the quartile classification of the DQI, whereas the trend was quite flat for healthier foods. Such result did indicate that the WXYfm model discriminated well dietary patterns, but raised concern on potential misreporting of intakes as a clear inverse trend appeared between the DQI and energy intake.

Fourth, the HEI was adapted to data from five EU national dietary surveys (Belgium, Denmark, France, Ireland, and Italy (Volatier *et al.*, 2007)). Participants classified in the first and fifth quintile of the adapted-HEI were defined as “healthy eating” and “not healthy eating”, respectively. In a second step, foods which consumption was statistically different between the two groups were identified as “indicator foods” of the healthy or the unhealthy patterns. In a final step, the indicator foods were scored by three NP models: WXYfm, Dutch Tripartite, and FDA requirements for health claims (Quinio *et al.*, 2007). Sensitivity and specificity between the “healthy eating” or “not healthy eating” indicator foods and the NP rankings were then assessed. Agreement was generally good, but some discrepancies were found, especially with

unhealthy foods consumed in conjunction with healthy foods and therefore classified in the "healthy eating" pattern, e.g. jam and butter associated with bread.

d. Fourth approach, modelling theoretical diets to test for construct validity

A French team used linear programming to test the SAIN,LIM model in two studies using the French national dietary survey (Volatier, 2000). First, a food database was used to create "healthy" (or "unhealthy") diets fulfilling a set of 40 nutrient recommendations (Darmon *et al.*, 2009). It was shown that healthy diets could not be reached by choosing only unhealthy foods, while the unhealthy diets could not be obtained with healthy foods only. Second, participants' reported consumption of foods was included, and each participant's diet was optimised to reach the full set of nutrient recommendations (Maillot *et al.*, 2011). The optimised diets contained more healthy foods and less unhealthy ones compared to the reported diets. In both studies, it was observed that unhealthy foods could be part of a diet fulfilling the whole set of recommendations, if outweighed by healthy foods.

Dutch researchers used a Monte Carlo simulation method within the Dutch national dietary survey to estimate the effect of introducing healthier options carrying the Choices programme logo (Roodenburg *et al.*, 2009). Theoretical diets were modelled by substituting reported foods with healthier options where possible. Favourable shifts were shown for most nutrients. The same approach was used in Greece, Spain, the USA, Israel, China and South Africa and findings were consistent (Roodenburg *et al.*, 2011). This approach was further extended to include the potential impact of highlighted dietary changes on blood cholesterol levels (Vyth *et al.*, 2011a). The modelling of blood lipids changes was done using existing meta-analysis results, and a slightly favourable change in the total cholesterol/HDL ratio was predicted.

e. Fifth approach, predictive validity

To date, one study assessed the relationship between a NP model and prospective health outcomes, using the US Nurses' Health Study and the Health Professionals Follow-up Study (Willett *et al.*, 1987; Colditz *et al.*, 1991). The ONQI NP model (table 2.2) was applied to the dietary questionnaires of both cohorts (Chiuve *et al.*, 2011). An aggregated diet score, ONQI-f, was calculated as the average ONQI scores weighted by portion consumption. Quintiles of ONQI-f were included in Cox

proportional hazards regressions. In both cohorts, risk reduction was observed for total chronic disease, CVD, diabetes, and all cause-mortality (p for linear trend ≤0.01); no association was found for cancer. The same analysis using an average ONQI weighted by energy intake did not yield significant results. The authors stressed the limitations associated with the study design, namely misreporting of dietary intakes, imprecision of food content from global questionnaire items (vs. specific branded foods or different types of preparation), and that ONQI was designed to score individual foods rather than diets. Further details are given in chapter 10.

Table 2.6 summarises the different approaches and the main findings.

Table 2.6: Summary of approaches used to validate nutrient profiling (NP) models

	(i)	(ii)	(iii)	(iv)	(v)
Design	Rankings of foods, comparison of models	Rankings against experts (=standard) ranking	Rankings against diet healthiness	Mathematical modelling of theoretical diets	Prospective associations with health outcomes
Type of validity		Criterion-oriented	Convergent	Construct	Predictive
Data needed	Food database with nutrient content	(i) + ranking from experts, obtained via internal or external survey	(i) + dietary survey	(i) (+ optional dietary survey)	Longitudinal data (cohort study) with dietary assessment at baseline
Pros	Easy to implement, requires little data	No dietary intake data needed, assess efficiency of model in specified region/cultural settings	Allows testing for convergent validity: is the model linked to healthier dietary patterns? Could be used for construct validity.	Allows linking NP to nutrient recommendations, construct validity. Can assess impact of food substitutions.	Allows assessing the link between NP and future health, i.e. testing the principle underlying the NP concept
Cons	No external validation, quite limited, except if compared with existing and validated models	Experts are biases and the standard rankings cannot be considered as gold standard. A systematic standard rankings needs recruitment of many nutritional experts and a rigorous design for rankings of foods.	Requires aggregating the NP scores at the dietary (participant) level to compare with a diet quality index. Or needs identification of 'indicator foods' associated with healthy/unhealthy patterns. Unhealthy foods can be integrated to healthy dietary patterns. Subject to reporting bias.	Needs some programming skills. Models very dependent on constraints and target values (optimised mathematical solutions). Theoretical diets can be too far from achievable diets.	Needs large longitudinal dataset, or could use nested case-control studies. Scoring of items from dietary questionnaire may not reflect true NP scores. Subject to reporting bias, and diet likely to change during follow-up.
Models included[1]	SSCg3d, WXYfm, SAIN,LIM, FSANZ, NRF, "Go, Slow, and Whoa", Nutrimap	WXYfm, NRF, CFN, RRR, SSCg3d, NFI, Dutch Tripartite, AHF, AHA, ONQI	NRF, ONQI, WXYfm, Dutch tripartite, FDA	SAIN,LIM, Choices	ONQI

[1]See tables 2.1 and 2.2 for models details.

2.1.5 Implications for research project

Many NP models have been published in the recent years. Only the basic characteristics of NP models, together with a few examples, were presented in this section since the aim of the project was not to discuss the features of all existing NP models.

The initial validation of a NP model usually involved assessing the rankings of foods derived from the NP model against rankings derived from pre-existing NP models or expert advice. Yet, relatively few studies have been formally carried out and published since this step often remained internal during the development process. Published results showed that most NP models agreed well with each other, especially for foods at the extremes of the healthiness classification (e.g. fruit and vegetables were almost always classified as healthy while sweets or salty snacks were considered unhealthy by most models). Some discrepancies did occur with composite or processed foods. Studies including rankings of foods derived from nutrition experts highlighted that such method could not be considered as a gold-standard since human perception is culturally biased.

The introduction of dietary survey data allowed assessing NP models against healthiness of dietary patterns, with the necessity of aggregating NP scores at the participant level. Such aggregated scores were associated with diet quality indices in the expected directions but the associations were relatively weak, highlighting that healthy dietary patterns were not exclusively composed of healthy foods. This was confirmed by a study carried out in five European countries which indicated that some specific unhealthy foods were often part of healthy dietary patterns, and were therefore considered as misclassified by the NP models.

The use of mathematical diet optimisation further confirmed this finding since diets fulfilling a full set of nutrient recommendations could contain some unhealthy foods, if outweighed by the healthy foods. Such technique was also used to demonstrate that a healthy diet could not be achieved by selecting unhealthy foods only.

The general consensus that NP could contribute towards healthier dietary intakes was further reinforced by a prospective study which showed that an aggregated diet score based on the ONQI model predicted reduced risk of all chronic disease except cancer.

However, all the validation methods described above—except predictive validity—suffered from a major loophole: nutrient recommendations included in the NP models were used in the validation process. The ONQI model used in the predictive validity investigation by Chiuve and colleagues is a patented model which is not publicly available (Chiuve *et al.*, 2011). The results obtained for ONQI therefore need to be confirmed using alternative NP models or datasets to conclude more generally on the predictive validity of NP.

WXYfm and SAIN,LIM models have been developed for national food safety agencies. Their algorithms are freely available and their development went through an open peer review process. Both models have been included in several validation studies using all the methods presented above but predictive validity.

Predictive validation of NP requires individual longitudinal data, with dietary assessment at baseline. The British Whitehall II study was initiated in 1985 and participants continue to be followed-up, the phase 11 clinical phase being currently underway. Dietary assessment was introduced in 1991, and this could act as baseline for a nutritional based prospective study. As mentioned by Chiuve and colleagues, such a design is not flawless and contains some intrinsic limitations. The next sections of this chapter therefore focus on the different aspects linking reported diets to health outcomes within the Whitehall II study, in order to assess whether the data could be used to test for predictive validity of the WXYfm and SAIN,LIM NP models.

2.2 Diet quality and health in the Whitehall II study

Several studies linking dietary intake to diverse health conditions used the Whitehall II study data. Results showing protective effects of healthy dietary patterns would confirm that the Whitehall II data suit the analysis of the predictive validity of NP.
Two main approaches have been used to determine dietary patterns of individuals (Kant, 2004; Waijers *et al.*, 2007):

- Data-driven methods using mainly factor or cluster analysis to define *a posteriori* patterns;
- Theoretically defined diet indexes or scores that assess compliance with *a priori* chosen criteria.

Both methods have been implemented in the Whitehall II data, and this section reviews the published evidence. Further to the literature held by the Whitehall II study team, the terms "diet", "dietary", and "food" were searched in conjunction with "Whitehall" and "Stress and health" (the alternative name for the Whitehall II study) in Pubmed and Google scholar.

2.2.1 Data-driven dietary patterns

With this first approach, statistical models are used to derive a specified number of dietary patterns within a given dataset (Blaikie, 2003). Principal component analysis and reduced rank regression derive dietary patterns, the "factors", which are defined by their relative association (the "factor loading") with several foods or food groups. Cluster analysis generates mutually exclusive groups of individuals based on the alikeness of their reported intake.

(i) Dietary clusters

This approach was first applied to the Whitehall II cohort to investigate socioeconomic differences in dietary patterns (Martikainen *et al.*, 2003). Six dietary clusters were derived. The "unhealthy" and "very unhealthy" clusters were associated with lower employment grade in both men and women. In contrast, the

"French" dietary cluster was associated with higher employment grade. These dietary intake differences accounted for about 25—50 per cent of the grade differences in HDL cholesterol and serum triglyceride levels. Together with other behavioural risk factors (mainly smoking), dietary clusters accounted for a third of the socioeconomic gradient in CHD incidence (Marmot *et al.*, 2008).

Dietary clusters were further associated with prospective risk of diabetes and fatal and non-fatal CHD (Brunner *et al.*, 2008). In this study, four clusters were derived and participants in the "healthy" cluster were at lower risk compared to participants in the "unhealthy" cluster, even after adjustment for socioeconomic position.

(ii) Principal component analysis

The relationship between depression and diet was investigated within the Whitehall II participants using principal component analysis (Akbaraly *et al.*, 2009a). Two dietary patterns were extracted: "whole foods" and "processed foods". Higher consumption of the "whole foods" pattern was associated with lower odds of depression, while a high consumption of the "processed foods" pattern was associated with increased odds. These associations were robust to adjustment.

Similar patterns were derived in a study investigating the association between diet and cognitive function (Akbaraly *et al.*, 2009b). The "whole foods" pattern was linked to lower cognitive function deficit while the "processed foods" one was associated with an increased deficit. Both associations were attenuated by education attainment.

(iii) Reduced rank regression

The use of reduced rank regression to derive dietary patterns using intermediate dependent variables was first presented by Hoffmann and colleagues (Hoffmann *et al.*, 2004). The method was applied twice to the Whitehall II food frequency questionnaire.

In the first investigation, reduced-rank regression was used to determine a dietary pattern associated with insulin resistance (McNaughton *et al.*, 2008). Such a pattern (characterised by high consumption of low-calorie/diet soft drinks, onions, sugar-sweetened beverages, burgers and sausages, crisps and other snacks, and white bread; and low consumption of medium-/high-fibre breakfast cereals, jam, French dressing/vinaigrette, and wholemeal bread) was associated with an increased risk of type 2 diabetes.

Serum total and HDL cholesterol, and triglyceride levels were used as dependent variables to derive a relatively similar dietary pattern in the second study (McNaughton *et al.*, 2009). This pattern was associated with an increased risk of CHD robust to adjustment.

2.2.2 Predefined dietary score, the alternative healthy eating index

The Healthy Eating Index (HEI), based on the Dietary Guidelines for Americans and the Food Guide Pyramid, was originally designed to score the diets of the US NHANES dietary surveys participants (Kennedy *et al.*, 1995). The HEI yielded only small associations with major chronic disease (McCullough *et al.*, 2000a; McCullough *et al.*, 2000b). The alternative healthy eating index (AHEI) was an attempt to improve the original score, and it was shown to reduce risk of major chronic disease in the Nurses' Health Study and the Health Professionals Follow-up Study (McCullough *et al.*, 2002). The AHEI is presented in more details in chapter 4.

The AHEI was applied to the Whitehall II food frequency questionnaire in three separate studies. Adherence to the AHEI was associated with a reversion of the metabolic syndrome status (Akbaraly *et al.*, 2010) and with a steeper decline in serum LDL cholesterol over 10 years of follow-up (Bouillon *et al.*, 2011). Further, high AHEI scores were shown to be protective against all-cause mortality, and especially CVD mortality, in survival analyses using data over 18 year of follow-up (Akbaraly *et al.*, 2011).

2.2.3 Food or nutrient specific analyses

The Whitehall II dietary data were also used to study the impact on health status of specific foods and/or nutrients.

Combined consumption of tea and coffee was modestly associated with a reduced incidence of type 2 diabetes; the effect was not significant for separate consumption of tea or coffee (Hamer *et al.*, 2008). Diabetes was also linked to overall dietary glycemic index and glycemic load (Mosdol *et al.*, 2007). The former was not associated with incident risk while an inverse association was observed for glycemic load, which did not follow the hypothesised harmful effect.

Plasma phospholipids were associated with CHD in a nested case-control study conducted on men (Clarke *et al.*, 2009): saturated fatty acids were shown to double the risk while poly-unsaturated fats halved it. These associations were highly attenuated by adjustment for serum lipids (HDL and LDL cholesterol) and inflammatory markers.

2.2.4 Whitehall II and predictive validity of nutrient profiling

The Whitehall II dietary data were consistently associated with health outcomes in the expected directions, except in one study focusing on glycemic index and glycemic load. Data-driven and a priori defined measures of dietary quality yielded relatively similar findings, which showed the robustness of the data. The Whitehall II study data therefore appeared to suit the scope of predictive validation of NP models.

In addition, contemporary risk factors could be used to deepen the investigation of NP validity. Cross-sectional associations between a NP model and risk factors available in the Whitehall II data could be tested (i.e. testing for concurrent validity).

Most studies presented in this section (2.2) were based on the assumption that reported intakes were markers of true intake. Validation of reported dietary intake is

essential for our project since any bias concerning the specific level of intakes could alter the prospective association between NP scores and health outcomes.

2.3 Validation of the Whitehall II food frequency questionnaire, energy misreporting

Food frequency questionnaires (FFQ) were widely adopted to assess dietary intakes in epidemiological studies (see chapter 4 for a detailed description of the Whitehall II FFQ). Their validation has been an important step in the development of nutritional epidemiology. Appendix 1 describes the validation process and the existing dietary assessment methods. This section first presents the results of the validation study conducted on the Whitehall II FFQ. It then focuses on misreporting of dietary intakes, which is the main information bias linked to the FFQ tool.

2.3.1 Validity of the Whitehall II FFQ

The Whitehall II FFQ was assessed against a 7-day diet diary (7DD) and several biomarkers in a sub-sample (n=860) of the study population (Brunner *et al.*, 2001). It was concluded that the FFQ performed well, especially against the available biomarkers (serum cholesteryl ester fatty acids, plasma α-tocopherol and β-carotene). Compared to the 7DD, the FFQ tended to over-estimate intake of plant-derived micronutrients, and to underestimate fat intake. Reported mean energy intake from the two dietary assessment methods was similar in men, and some 10% higher according to the FFQ in women. Approximately 34% of participants with a completed FFQ were identified as low-energy reporters. This last observation confirmed previous findings made in large scale nutritional studies. Issues regarding total energy intake have been long debated (Prentice *et al.*, 1986; Pryer *et al.*, 1997; Poppitt *et al.*, 1998; Pomerleau *et al.*, 1999). The next section focuses on the detection of energy misreporters and on the usual methods used to correct for this reporting error.

2.3.2 Detecting energy misreporting

Energy intake is highly regulated and its expenditure can be assessed quite easily, using direct or indirect methods. It is possible to estimate the correspondence between reported energy intake (EI) and the measured or calculated energy expenditure (EE).

The Goldberg cut-offs technique lies in the fundamental equation EI = EE (at constant body weight) to detect high and low energy reporters (Goldberg *et al.*, 1991). Total EE depends on the physical activity level (PAL) and on the basal metabolic rate (BMR). The BMR is a measure of EE in a complete rest status; it can be estimated using age, sex, and body weight (FAO/WHO/UNU, 1985; Schofield *et al.*, 1985; Department of Health, 1991). EE is given by the following equation:

$$EE = BMR \cdot PAL$$

This first equation can be rewritten as follows:

$$EI / BMR = PAL$$

Thus, the reported energy intake of an individual and its respective calculated BMR can be compared to the expected PAL for that individual. Measurement errors occur in all the elements of this equation. The Goldberg method defines the confidence limits for the different terms in the equation in order to determine whether the mean reported energy intake is plausible (details given in chapter 8).

Since direct measures of energy expenditure were not available in the Whitehall II data, the Goldberg cut-offs technique was implemented in a sub-sample of the study population (n=865) with both FFQ and 7DD data (Stallone *et al.*, 1997). Using a single PAL category for all individuals, 33.3% of participants were defined as low energy reporters. The use of a single PAL category to detect low energy reporters was a limitation of the Stallone study because individuals with high energy expenditures may have not been detected as under-reporters.

Among Whitehall II participants, low energy misreporting was further linked to higher BMI and lower employment grade (Brunner, 1997), which was in line with previous observations.

2.3.3 Tackling dietary misreporting

Two main approaches have been used to account for energy misreporting:

- excluding low and high energy reporters from the analysis;
- adjusting for total energy intake.

The first approach has been used in a few studies (Pryer *et al.*, 1995; McNulty *et al.*, 1996; Shortt *et al.*, 1997; Drummond *et al.*, 1998), but it introduces bias of an unknown size into the data and may exclude the most interesting individuals. Table 2.7 summarises the four models that have been proposed to adjust for total energy intake in the association between disease and nutrient intakes (Willett & Stampfer, 1998).

Table 2.7: Disease risk models for adjusting for total energy intake in epidemiological studies

Method	Model
Residual	Disease = Nutrient residual
Standard Multivariate	Disease = Nutrient + ε_{tot}
Energy partition	Disease = $\varepsilon_{nutrient}$ + ε_{other}
Multivariate density	Disease = (Nutrient/ ε_{tot}) + ε_{tot}

"Nutrient residual"' is the residual from the regression of a specific nutrient on total calories. "Nutrient" is the absolute intake. ε_{tot} is the total energy intake. $\varepsilon_{nutrient}$ and ε_{other} are the energy intake from the specific nutrient and from the other sources, respectively.

Stallone and colleagues tested both the exclusion of low energy reporters and the adjustment for total energy intake within the Whitehall II data (Stallone *et al.*, 1997). It was concluded that the latter approach was preferable. Exclusion of under-reporters was not recommended by the authors since low energy reporting was strongly associated with employment grade.

In addition, energy misreporting is usually associated with differential misreporting of foods, i.e. some foods tend to be systematically over-reported (e.g. fruit and vegetables) and other systematically under-reported (e.g. snacks and sweets),

resulting in apparent healthier diets in low energy reporters (Livingstone & Black, 2003). To address this issue, Rosner et al. (1989) introduced regression calibration which uses diet diary reported intake to correct the epidemiological association (odds ratio or hazard ratio) for a single nutrient or food. The method has been further developed to obtain corrected estimates for all the food items of a FFQ; details are given in chapter 8 (Rosner & Gore, 2001). Also, biomarkers have been used to correct reported FFQ intakes (Kaaks *et al.*, 1994; Kaaks, 1997; Rosner *et al.*, 2008). Limited biomarker data were available within the Whitehall II study, and it was not possible to apply such methods.

2.4 Conclusion

This chapter first highlighted the main gap in literature to be addressed: predictive validity of publicly available and government-endorsed NP models. In order to link a NP model to global dietary intakes (i.e. convergent validity) or to prospective health outcomes (i.e. predictive validity), researchers had to aggregate the NP scores for individual foods to produce an aggregate score that indexed the nutritional quality of the diet. Hence, such an "aggregate score" has to be designed for both the WXYfm and SAIN,LIM models prior to assess the predictive validity of the two NP models.

The Whitehall II data were used in several publications showing a link between dietary patterns and health status. Such data were therefore considered to be appropriate to test for predictive validity of the WXYfm and SAIN,LIM models.

However, self-reported dietary assessment tools are subject to bias, namely misreporting of intakes. A relevant proportion of Whitehall II participants were shown to misreport their intake, which will need to be taken into account in the subsequent analyses.

Chapter 3: Aim and objectives

As highlighted in the previous chapter, the body of evidence surrounding nutrient profiling (NP) models validation is increasing. Still, there is a lack of studies assessing the underlying hypothesis of the NP concept, i.e. that diets containing more healthy foods would be protective against adverse health outcomes. Such investigation has only been carried out on a patented and not publicly available NP model, and this gap in the literature needs to be addressed.

The main aim of this PhD thesis was to assess the predictive validity of two government-endorsed NP models, the British WXYfm and the French SAIN,LIM, using the Whitehall II cohort study data. It was hypothesised that diets containing higher proportions of healthier foods, as defined by both models, would be protective against incident coronary events, diabetes, cancer mortality, and all-cause mortality.

Five specific objectives were defined to achieve the project aim (figure 3.1):

1. To derive NP aggregate scores summarizing the foods' NP scores at the diet level. The so-defined aggregate scores will be used as main exposures in subsequent analyses.
2. To assess the construct, convergent, and concurrent validity of aggregate scores, by testing their respective associations with dietary intake (nutrients, foods, and food groups), data-driven dietary clusters, the Alternative Healthy Eating Index, and risk factors; and to identify potential confounders of the association between aggregate scores and prospective health status.
3. To build a survival analysis model using Cox proportional hazard regressions, with NP-derived aggregate scores as main exposure and including potential confounding factors identified in objective 2, to assess the predictive validity of the two NP models (figure 3.2).
4. To interpret the observed results, i.e. to examine in detail the role of bias and confounding on the prospective associations between aggregate scores and adverse health outcomes.
5. To derive alternative NP models and/or aggregate scores based on results of objectives 3 and 4.

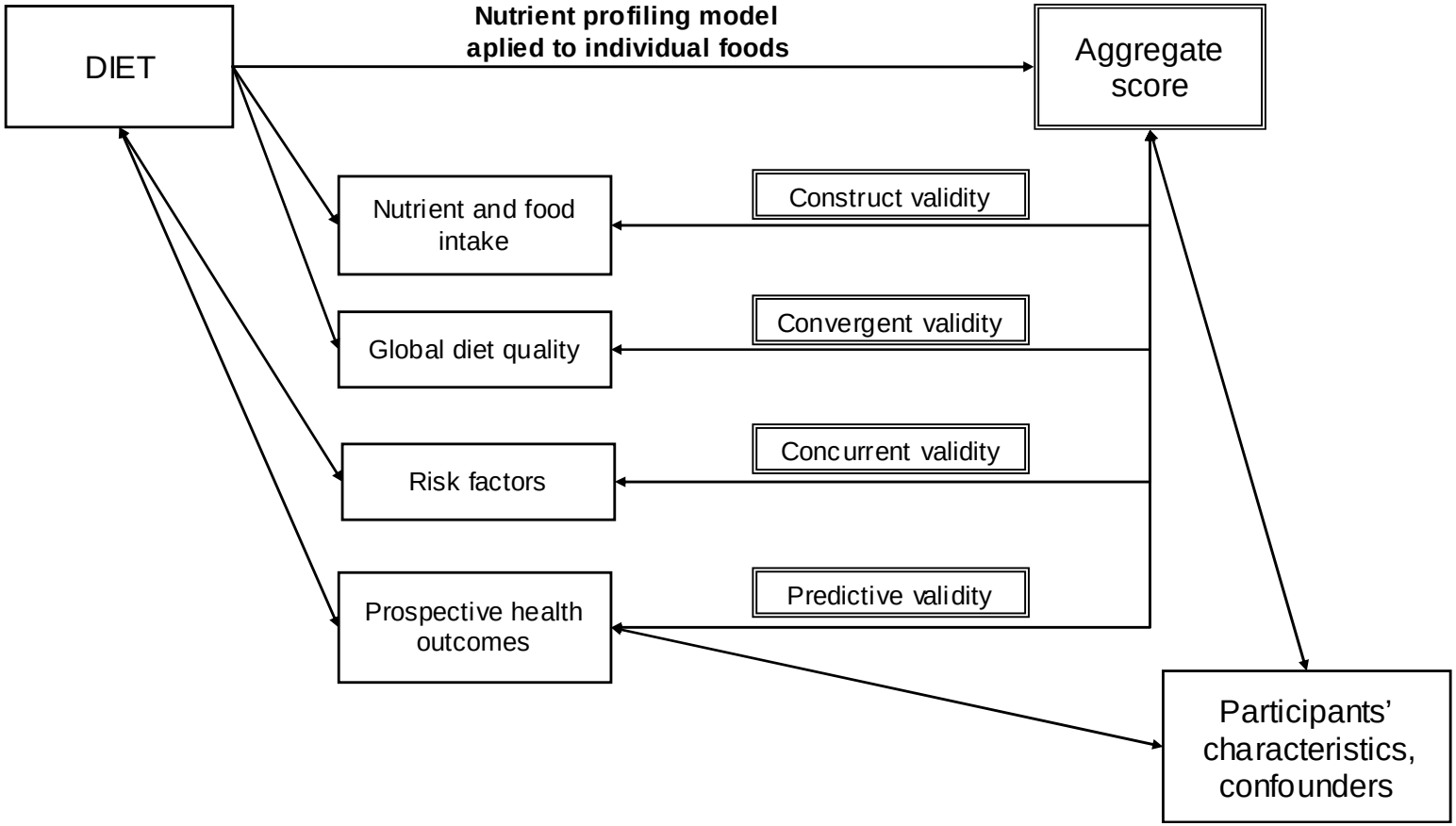


Figure 3.1: Theoretical framework for objectives 1, 2, and 3 (double-framed boxes)

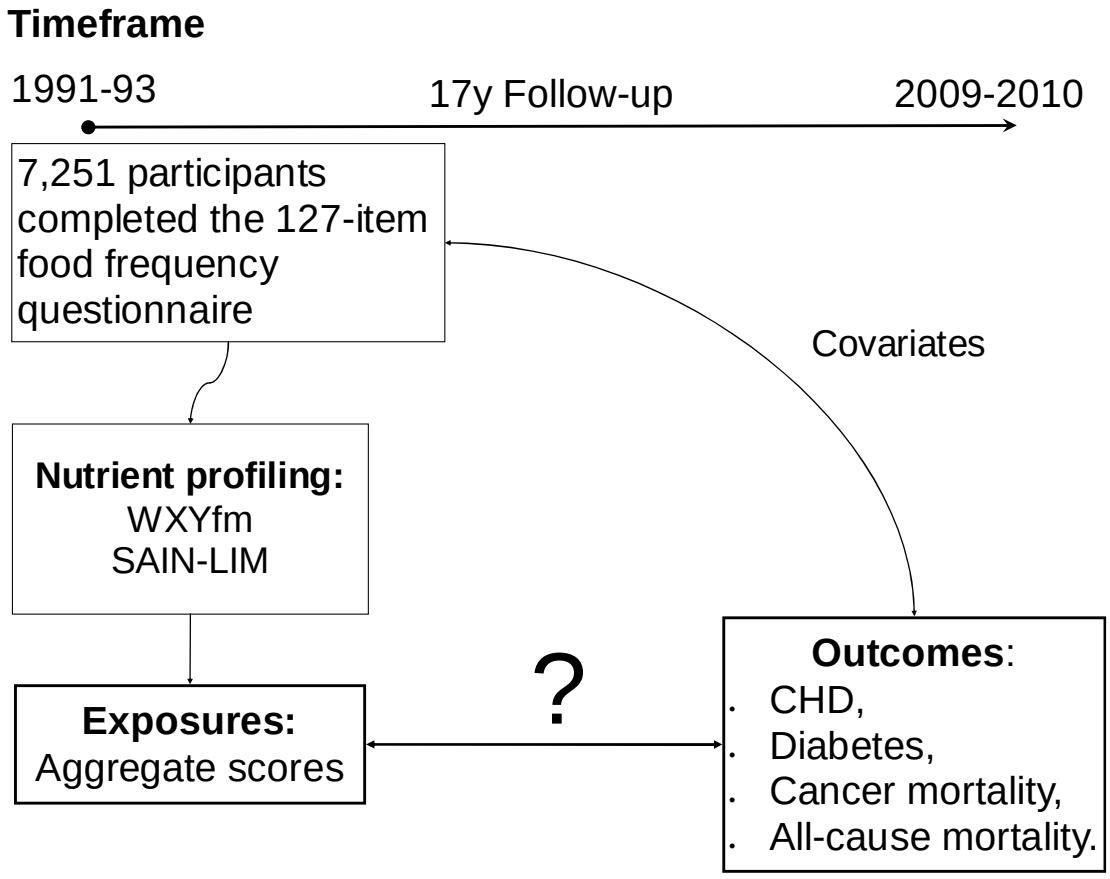


Figure 3.2: Framework for the predictive validity model (objective 3)

Chapter 4: Material and methods, implementation of nutrient profiling models to the Whitehall II data and design of aggregate scores

This chapter describes in details all the tools, data, and statistical methods that were used consistently throughout the thesis. Special focus is put on the Whitehall II data and the design of the aggregate scores.

WXYfm and SAIN,LIM nutrient profiling (NP) models were presented in details in chapter 2. All the main analyses were carried out on the British WXYfm, which was the first model to be applied to the Whitehall II data. The second model, SAIN,LIM, was used as a comparison tool to assess the extent to which observed results depended on the WXYfm algorithm. Its implementation is presented in chapter 7.

4.1 The Whitehall II study

The Whitehall II cohort was set up in 1985 following the results of the first Whitehall study which highlighted the social gradient in cardiovascular disease, using civil service employment grade as indicator for socio-economic position. The gradient was robust to traditional risk factors (e.g. smoking, physical activity), sparking demand for further research (Marmot *et al.*, 1978; Rose & Marmot, 1981).
This section presents briefly the whole Whitehall II population, the dietary assessment tools, the outcome assessment variables, and all other variables included in the analyses.

4.1.1 The cohort

The target population of the Whitehall II study was all civil servants aged 35-55 years working in the London offices of 20 Whitehall departments between 1985 and 1988. A response rate of 73% led to a final sample of 3,413 women and 6,895 men (Marmot & Brunner, 2005). The whole cohort has been invited to a research clinic at

5-year intervals, and a postal questionnaire was sent to participants between clinic phases (table 4.1). The last clinical phase, phase 9, was ended in December 2009, and data for CHD and diabetes events were available until this date. Follow-up of mortality through the NHS registry provided date and cause of death (99.9 % of participants flagged) until January 2010. Phase 10 was the last completed phase (January to March 2011). It was modified compared to previous questionnaire phases since it used a sub-sample of the study population to assess new measurements for the future phases, subsequent to the ageing of the Whitehall II participants. Phase 11 started in January 2012. In this project data from phase 3 was used, together with outcome data from phase 3 until the end of follow-up.

The Whitehall II study has been funded by the Medical Research Council, the British Heart Foundation, the National Heart Lung and Blood Institute (USA), and the National Institute on Ageing (USA). Ethical approval for the study was obtained from the Joint University College London/University College Hospital Committees on the Ethics of Human Research. All participants gave written informed consent for their participation at each phase of the study.

Table 4.1: Completed phases of the Whitehall II study

Phase	Dates	Type	n	Dietary assessment
1	1985-88	Screening/Questionnaire	10,308	
2	1989-90	Questionnaire	8,133	
3	1991-93	Screening/Questionnaire	8,637	FFQ + 7DD
4	1995-96	Questionnaire	8,629	
5	1997-99	Screening/Questionnaire	7,830	FFQ
6	2001	Questionnaire	7,344	
7	2003-04	Screening/Questionnaire	6,967	FFQ
8	2006	Questionnaire	7,180	
9	2007-09	Screening/Questionnaire	6,762	FFQ
10*	2011	Screening/Questionnaire	277	

FFQ, food frequency questionnaire; 7DD, 7 day diet diary.

* Phase 10 was run on a sub-sample of the study population to test new measurements for phase 11.

4.1.2 Dietary assessment measures

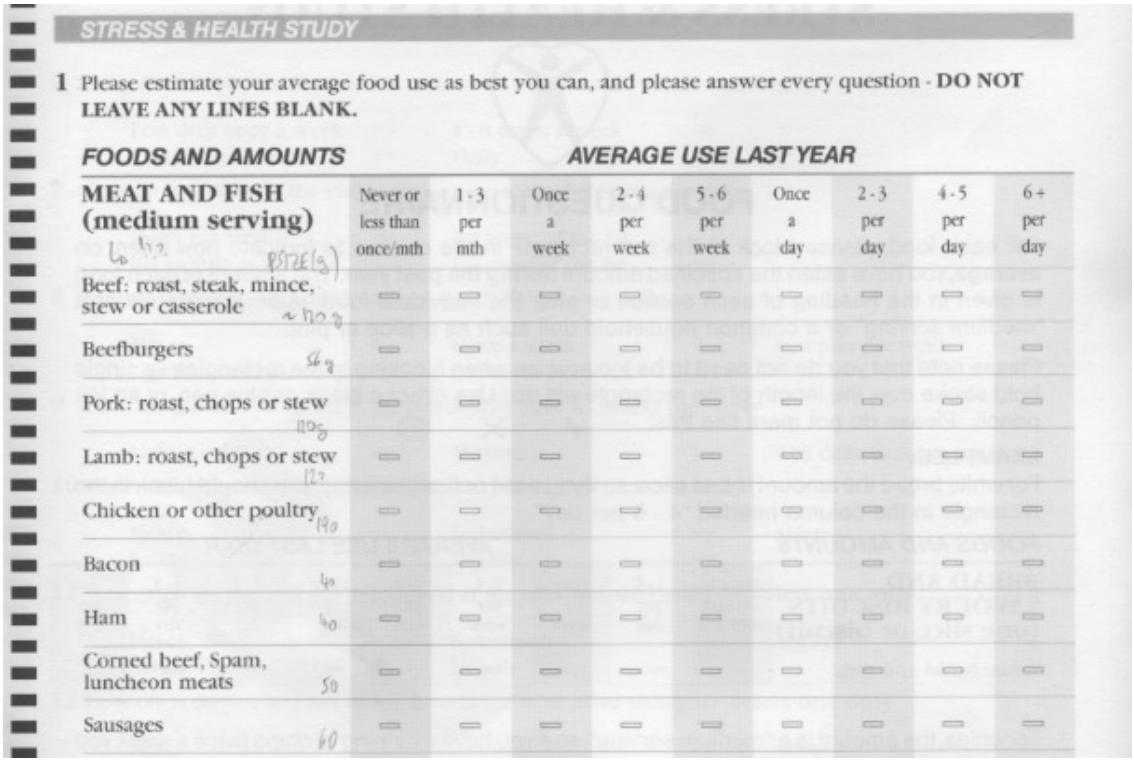
Detailed dietary assessments were introduced at phase 3 (1991-94). As a result, the baseline population for this project was the remaining participants at phase 3. Dietary intakes were reported in a 127-item anglicized version of the Willett food frequency questionnaire (FFQ) (Willett *et al.*, 1985; Willett, 1998) and a 7-day diet diary (7DD) as used in the UK arm of the EPIC study (Bingham *et al.*, 1994). The FFQ and 7DD were completed by 8,225 and 6,726 respondents, respectively. 1,350 7DD have been coded by the Whitehall II study team and further coding has been done through collaboration with the MRC Centre for Nutritional Epidemiology in Cambridge (see chapter 8 for further details on 7DD data). FFQs were also used in phases 5, 7 and 9, in a slightly altered version. Only the phase 3 FFQ was used in the present project.

For all items in the administered FFQ, participants were asked to report their frequency of eating a common unit or portion size in nine predefined categories ranging from "never or less than once per month" to "6+ per day" (figure 4.1). The FFQ also contained a series of supplemental questions about representativeness of listed items, use of added fat in cooking, use of salt, consumption of meat fat and regular food supplement intake. Reported intakes were then converted into grams/day using standard portion sizes. Energy and nutrient content of the reported diets were derived based on the 4th and 5th editions of *McCance & Widdowson's The composition of Foods* and its supplementary tables, and added food composition records (Paul & Southgate, 1978; Holland *et al.*, 1988; Holland *et al.*, 1989; Holland *et al.*, 1991a; Holland *et al.*, 1991b; Holland *et al.*, 1992a; Holland *et al.*, 1992b; Holland *et al.*, 1993; Chan *et al.*, 1994; Chan *et al.*, 1995).

A validation study of the FFQ was done on a sub-sample of participants with both the FFQ and the diet diary coded (Brunner *et al.*, 2001). It was concluded that both dietary assessment methods performed relatively well, correlations between methods being higher when energy-adjusted. More details are given in chapters 2 and 8.

Missing values were a concern as some participants did not report consumption of all the items. All the following analyses only included participants with less than 10

missing items, for which the consumption was assumed to be null (Rosner & Gore, 2001).



STRESS & HEALTH STUDY

1 Please estimate your average food use as best you can, and please answer every question - **DO NOT LEAVE ANY LINES BLANK.**

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
MEAT AND FISH (medium serving)	Never or less than once/mth	1-3 per mth	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Beef: roast, steak, mince, stew or casserole									
Beefburgers									
Pork: roast, chops or stew									
Lamb: roast, chops or stew									
Chicken or other poultry									
Bacon									
Ham									
Corned beef, Spam, luncheon meats									
Sausages									

Figure 4.1: Phase 3 food frequency questionnaire, extract from the first page.

4.1.3 Outcomes

(i) Mortality follow-up

The Whitehall II study is linked to the National Health Service (NHS) death and electronic patient records using the NHS identification number assigned to all British citizens. A total of 10,297 participants (99.9%) were successfully traced and have been followed through these registers. Mortality data (median follow-up 17.7 years, range [0.08-18.4]), including the cause of death, were available through the NHS Central Registry until 31 January 2010. Death certificates were coded using the 9th or 10th revision of the International Classification of Disease (ICD). In addition to all-cause mortality, we analyzed mortality from cancer, except non-melanoma skin cancer (ICD-9 140-209 except 173, and ICD-10 C00-C97 except C44).

A total of 915 incident deaths were recorded within the 171,267 person-years of follow-up (mean (SD) was 16.8 (2.67) years per person). Of these, 419 were attributable to cancer, 259 to CVD, and 143 to CHD.

(ii) Fatal CHD and non-fatal myocardial infarction

Deaths were classified as coronary if ICD-9 codes 410-414 or ICD-10 codes I20-I25 were present on the death certificate. Potential cases of non-fatal myocardial infarction (MI) up to December 30 2009, for those who attended phase 9, have been ascertained by questionnaire items on chest pain (Rose, 1982), doctors' diagnoses, and hospitalizations (NHS Hospital Episode Statistics database). 12-lead electrocardiograms (ECG) were performed at phases 3, 5, 7 and 9 and assigned Minnesota codes (Macfarlane *et al.*, 1990). Details of physician diagnoses and investigation results were sought from clinical records for all potential cases. Based on all available data from questionnaires, ECGs, and cardiac enzymes, definite non-fatal MI was defined using the MONICA criteria (Tunstall-Pedoe *et al.*, 1994). MI was defined as positive if a questionnaire or clinical record of diagnosed MI was obtained in the presence of an ischemic ECG, and defined as negative when self reported only. Classification of MI was carried out blind to other study data by two trained coders, with adjudication by a third in the rare event of disagreement.

416 incident fatal CHD and non-fatal MI were identified in the 140,641 py of follow-up (mean (SD) was 14.5 (5.24) years per person).

(iii) Diabetes

Incident cases of diabetes have been identified by self-report of doctor's diagnosis, diabetic medication and 2-hour 75g oral glucose tolerance test (OGTT) at phases 3, 5, 7 and 9, according to the 1999 WHO classification (World Health Organization, 1999). Incident diabetes was dated at the day of the clinic visit for those first identified through OGTT. For those identified by self-report, the midpoint between the first instance of self-reported diabetes and the previous phase was used. For those who had not developed diabetes up to phase 9, follow up was censored on December 30 2009 (phase 9 closing date).

A total of 927 incident cases of diabetes were identified with a mean (SD) follow-up of 13.9 (4.27) years per person (total was 114,209 py).

4.1.4 Covariates

The link between diet and health can be confounded by external factors which need to be taken into account in the predictive validity analysis. The potential confounding factors were selected on the basis that they were associated with both the health outcomes and dietary intake.

(i) Socio-demographic variables

Age (date of birth), sex, and ethnicity (white, south Asian, black, other) were obtained at recruitment (phase 1). Phase 3 questionnaire included questions on marital status (married or cohabiting, single, divorced, widowed) and occupational position based on current (or last for retired participants) British civil service employment grade defined on the basis of salary and grouped into six categories ranging from senior administrators to clerical and office support staff. The occupational position was used a proxy for socio-economic position

(ii) Health behaviours

Smoking status was assessed by questionnaire at phases 1, 2, and 3 which allowed deriving three categories: current, former, and never smoker. The questionnaire also included hours and frequency of participation in vigorous (e.g. running, hard swimming, playing squash), moderately (e.g. dancing, cycling, leisurely swimming) and mildly energetic physical activity. Metabolic equivalents (METs) associated to each level of activity were used to categorise participants in three groups of activity (High/Intermediate/Low, more details are given in chapter 8). Total energy and alcohol intake were estimated from the FFQ.

(iii) Health status at baseline (phase 3)

Prevalence of coronary heart disease (CHD) was based on clinically verified events and included non fatal myocardial infarction and definite angina (see previous section). Diabetes was diagnosed using the WHO definition (see previous section). The presence of any other "longstanding illness" was self-reported.

Blood pressure was measured twice in the sitting position after 5 minutes rest using Hawksley random-zero sphygmomanometer (Hawksley, Lancing, Sussex, United Kingdom). Hypertension was defined as systolic or diastolic blood pressure ≥ 140 or ≥ 90mmHg, respectively, or by the use of hypertensive drugs (Akbaraly *et al.*, 2011).

Serum cholesterol was determined by the cholesterol oxidate peroxidase colorimetric method (BCLkit; Boehringer, Mannheim, Germany). HDL cholesterol was determined after dextran sulphate-magnesium chloride precipitation of non-HDL cholesterol (Brunner *et al.*, 1993). Serum triglycerides were measured in a centrifugal analyser by enzymic colorimetric methods. Dyslipidaemia was defined as LDL cholesterol ≥ 4.1mmol/L or by the use of lipid-lowering drugs (Akbaraly *et al.*, 2011).

Height was measured to the closest mm using a stadiometer with the participant standing completely erect with the head in the Frankfort plane. Weight was measured twice with all items of clothing removed except underwear. A Soehnle (Backnang, Germany) scale was used to read weight to the nearest 0.1kg. If the reading alternated between two measures (0.1kg apart with the participant standing still) the higher reading was recorded. Body mass index (BMI) was derived as weight (kg) divided by height-squared (m^2). Four categories were defined as follow: underweight, BMI <20kg/m²; normal (reference group), BMI ≥20-<25kg/m^2; overweight, BMI ≥25-<30 kg/m^2; obese, BMI ≥30kg/m²) (Akbaraly *et al.*, 2011).

Fibrinogen, factor VII, Von Willebrand's factor, interleukin 6 and C-reactive protein were measured at phase 3 from serum stored at −70°C (Brunner *et al.*, 1997; Kumari *et al.*, 2000; Nabi *et al.*, 2008; Elovainio *et al.*, 2010).

4.2 Deriving a participant's aggregate score

In order to analyse the prospective associations between NP and adverse health outcomes, it was necessary to translate the NP scores for FFQ-items into one or more variables that characterised a participant. The resulting variable(s), the "aggregate score(s)", represented the exposure in the subsequent analyses. The goal was to rank participants according to their relative consumption of FFQ-items as scored by the NP models, which was the first research objective (chapter 3). The selection of an appropriate aggregate score was done using the WXYfm model. The choice of an aggregate score for the SAIN,LIM model is presented in chapter 7. Both NP models did not score alcoholic drinks. As a result, the aggregate scores did not include alcohol FFQ-items in their algorithms.

4.2.1 Classification of phase 3 FFQ-items by the WXYfm nutrient profiling model

The WXYfm model was applied to the phase 3 FFQ using the corresponding nutrient composition dataset (section 4.1). Data on Englyst fibre content being missing, thresholds for the Association of Official Analytical Chemists (AOAC) fibre definition were used (Cho S *et al.*, 2007). Foods and drinks were then classified in the healthiness categories (table 4.2). The complete classification of FFQ-items is shown in appendix 2.

Table 4.2: Classification of phase 3 FFQ-items in the WXYfm healthiness categories

n per WXYfm category	"healthier"	"intermediate"	"less healthy"
Foods (n=105)	51	8	46
Drinks (n=17)	10	N/A	7

Excludes alcohol items

4.2.2 Classification of participants: the aggregate scores

Two approaches for deriving an aggregate score for each participant were developed. The first approach used the WXYfm overall score. The second one used the healthiness categories.

(i) Weight-weighted, energy-weighted and portion-weighted arithmetic mean scores

Three weighted arithmetic means were computed on the WXYfm overall score (OS) (ranging from -13 to 28 in the phase 3 questionnaire) of each item (except alcohol):

- weight-weighted aggregate score $WWS = \frac{\sum_{i=1}^{122} OS_i \cdot w_i}{\sum_{i=1}^{122} w_i}$;
- energy-weighted aggregate score $EWS = \frac{\sum_{i=1}^{122} OS_i \cdot \varepsilon_i}{\sum_{i=1}^{122} \varepsilon_i}$;
- portion-weighted aggregate score $PWS = \frac{\sum_{i=1}^{122} OS_i \cdot p_i}{\sum_{i=1}^{122} p_i}$.

With OS_i the WXYfm overall score of item i, w_i the weight of intake of item i, ε_i the energy intake from item i and p_i the portion intake from item i. All these intakes are expressed per day.

(ii) Percentage of energy intake from WXYfm healthiness categories

In a previous study aiming to rank individuals based on their dietary quality, Arambepola and colleagues (2008) calculated the energy intake from each WXYfm category, merging the healthier and intermediate categories for foods. This resulted in two categories for all food and drinks: those who could be advertised (according to the Food Standards Agency and Ofcom), i.e. “healthier” and “intermediate” ones;

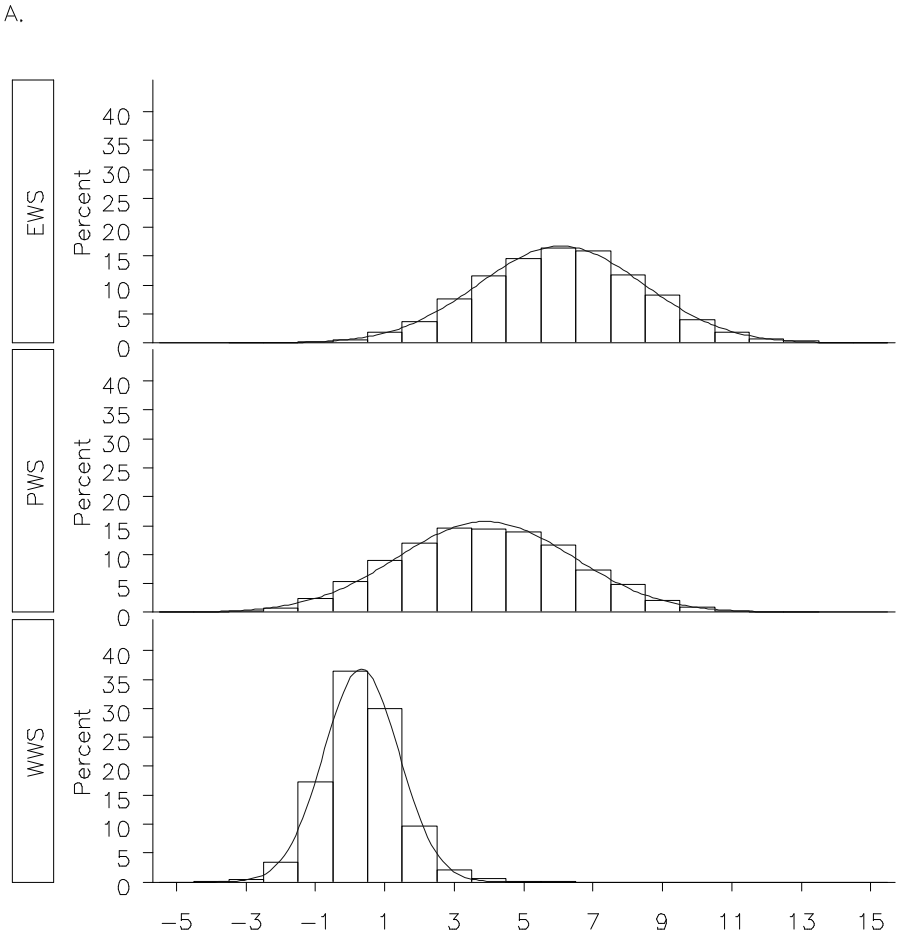
and those would could not, i.e. “less healthy” foods and drinks. An aggregate score was derived from this approach:

$$\text{Percentage of Energy score} \quad \text{PES} = \frac{(\varepsilon_{\text{WXYfm}})_{\text{adv}}}{\varepsilon_{\text{total}}}$$

With $(\varepsilon_{\text{WXYfm}})_{\text{adv}}$ the energy intake from the “healthier” and “intermediate” foods and drinks and $\varepsilon_{\text{total}}$ the total energy intake (excluding alcohol).

Distribution of all the aggregate scores is shown in figure 4.2. The energy-weighted EWS yielded unhealthier scores compared to the WWS and PWS. This was certainly linked to the fact that most energy dense foods had unhealthier WXYfm scores, energy content being included in the NP model as a negative component. The score range was much narrower with the WWS than with the EWS and PWS. The PES distribution showed that on average, participants had around 55% of their energy intake from healthier and intermediate foods and drinks.

The rankings derived from all the aggregate scores described above were compared in order to select the most appropriate aggregate score for subsequent analyses.



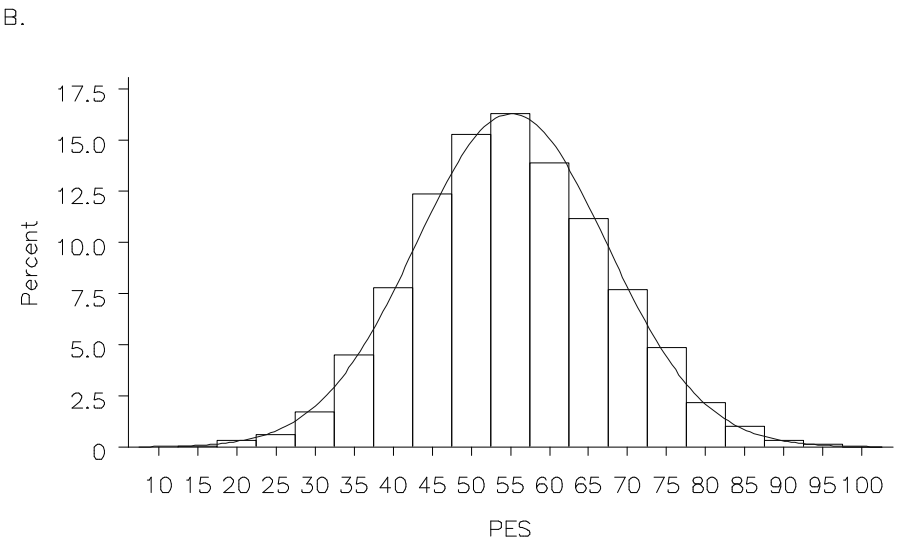


Figure 4.2: Distribution of WXYfm aggregate scores (n=8,253)
A. EWS, PWS, WWS. B. PES

4.2.3 Rankings of participants and selection of an aggregate score for further analyses

(i) Rankings obtained with all aggregate scores

Spearman's rank correlation coefficients were computed on the raw scores. All aggregate scores were highly correlated with each other (table 4.3). The highest correlation was between EWS and PES, the two scores weighted by energy intake.

Table 4.3: Spearman's correlations between aggregate scores (n=8,253)

	WWS	EWS	PWS	PES
WWS	1			
EWS	0.79	1		
PWS	0.77	0.79	1	
PES	0.81	0.82	0.68	1

(ii) Rationale of the two approaches

The first approach (WWS, EWS and PWS) did not rely on the arbitrary thresholds of the healthiness categories and could be used to assess the WXYfm overall score. Each weighting scale had its own drawbacks. The weighting by energy from each food may have over-represented the energy dense foods, as illustrated in figure 4.2, while weighting on weight intake may have put more emphasis on drinks. The weighting by portion seemed to be a good alternative yet the same weight was given to foods consumed in very different amounts.

The second approach, PES score, relied on the healthiness categories and was an appropriate candidate to validate the respective thresholds.

(iii) Selection of one aggregate score

The WXYfm ranking of foods primarily depends on the overall score. Validating this scale was considered as the main priority to assess the model. The thresholds

presently used by Ofcom to determine the foods and drinks which can be advertised were determined arbitrarily and would also need to be validated.

All the aggregate scores described above did yield similar rankings of participants. The selection of two aggregate scores was not thought judicious as subsequent results might be very similar and would further depend on the divergent aggregate scores algorithms rather than on dietary intakes and the WXYfm.

Not all existing NP models rely on healthiness categories to separate foods. Chiuve and colleagues, in their analysis of the ONQI NP model applied an aggregate score, ONQI-f, similar to the PWS (Chiuve *et al.*, 2011). Portion sizes are commonly defined in the US for nutritional labelling purposes, which is not the case in Europe. The authors assessed an alternative to the ONQI-f, weighted by energy, but did not present the results. Energy was retained by Fulgoni and colleagues to validate their own NP scheme (Fulgoni *et al.*, 2009), and by Arambepola *et al*. in their study of the WXYfm model (Arambepola *et al.*, 2008). Weighting by weight of intake led to a narrow distribution of participants (figure 4.2) and therefore less inter-individual variance. As a result, the EWS aggregate score was retained to validate the WXYfm NP model in further analyses.

4.3 Statistical methods and other analysis tools

For all the statistical methods implemented in this project, significance was calculated at the 5% level (α=0.05). Quartiles of aggregate scores were used in most cases. Most statistical analyses and data handling was done using the SAS 9.1 package, special mention was made otherwise.

4.3.1 Cross-sectional associations

Spearman's rank correlation was used for comparison of two continuous variables (Kirkwood & Sterne, 2003). It is a non-parametric measure calculated in a similar way to Pearson's correlation using the ranks instead of the actual values. The

correlation coefficients range from -1 to 1 with 0 indicating no correlation and 1 (or -1) similar (or inverse) rankings.

Cohen's κ-statistic was used to measure agreement between categorical variables. It is thought to be more robust than percentage agreement as it takes into account agreement occurring by chance. Since all comparisons were made on ordered variables (e.g. aggregate score quartiles), the weighted statistic which penalises greater disagreement was used (Cohen, 1968; Ludbrook, 2002). Its value can range from -1 to 1, but usually ranges from 0 (agreement no better than chance) to 1 (perfect agreement).

To measure heterogeneity across groups of participants (e.g. quartiles of aggregate scores), one-way analysis of variance (ANOVA) was used for continuous variables. Linear trends were assessed by including the ordered categories in regression models. χ^2 tests were used for categorical variables. For variables with only two levels, and for which trends could be estimated, the Cochran-Armitage test was used (SAS Institute Inc.). To assess the mean difference between two continuous variables measured in the same set of individuals, paired t-tests were used.

4.3.2 Survival analyses

The main objective of this project was to assess the predictive validity of the WXYfm and SAIN,LIM models. This was done using survival analysis where occurrence of the outcome and survival time (i.e. the time between the baseline and the event occurrence) are taken into account. Similarly to cross-sectional associations, heterogeneity across groups as well as trends and quantified associations were estimated. The goal was to compare the hazard *h(t)*, which is the probability that the event occur at time *t*, between groups.

(i) Log-rank test

Heterogeneity of survival across groups or strata of individuals was assessed using Log-rank (or Mantel-Cox χ^2) test (Mantel, 1966). It is based on the comparisons in

each stratum of the number of *observed* events with the *expected* number of events if there was no difference between groups (Kirkwood & Sterne, 2003). This test was mainly used to identify potential confounding factors.

(ii) Cox proportional hazards regression

To quantify the prospective associations between the aggregate scores and health outcomes, i.e. the ratio of hazards between two groups, Cox proportional hazards regression models were implemented (Cox, 1972; Cox & Oakes, 1984). The mathematical form of the model is:

$$h(t) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n)$$

Where *h(t)* is the hazard (risk) at time *t*, $h_0(t)$ is the baseline hazard (i.e. the hazard for an individual in whom all exposure variables = 0) at time *t*, and x_1 to x_n are the n exposure variables.

When there is a single binary exposure variable (x_1=1 or x_1=0), the hazard ratio (HR) at time *t* is given by:

$$\mathrm{HR}(t) = \frac{h_0(t)\exp(\beta_1)}{h_0(t)} = \exp(\beta_1)$$

The regression coefficient is therefore the log(hazard ratio), and the model assumes proportional hazards (see below).

These models were implemented as follows:

- Follow-up time in years was used for the outcome variables; all prevalent cases were excluded from the analysis. Date of diagnosis was used as event date for CHD, cancer mortality and all-cause mortality. Incident diabetes could not be dated exactly and was therefore interval-censored (see section 4.2). The approximation of the likelihood function developed by Breslow (1974) was used in case of tied events (i.e. occurring at the same time point).
- Quartiles of aggregate scores were used as main exposure, with the least healthy quartile being the reference group in all analyses. Linear trend was assessed by including the raw aggregate score, or its quartiles, as a continuous variable. Quadratic trends were estimated by including both the

continuous aggregate score and (aggregate score)-squared in the models (Greenland, 1995).

- Covariates were included either as continuous or categorical (dummy variables and a reference group). Interaction between exposure and covariates was tested by including interaction terms in the models.

(iii) Proportional hazards assumption

Both methods presented above assume "proportional hazards", i.e. the ratio of hazards between two groups (0 and 1) is constant over time:

$$\frac{h_1(t)}{h_0(t)} = \text{constant}$$

Several methods have been described to test the proportional hazards assumption. Scaled Schoenfeld residuals, defined for each failing individual and for all covariates included in the model, were used in this project (Schoenfeld, 1982). For the covariate X, the Schoenfeld residual of failing participant i at time *t* is as follows:

$$\text{Residual } (X,i,t) = X_i(t) - \text{expected value of } X(t)$$

Where the expected value of X at time *t* is the mean of X weighted by the likelihood of failure for each individual in the risk set at time *t*. If the Schoenfeld residuals are shown to be associated with time, then the proportional hazard assumption is violated for covariate X. The Schoenfeld residuals were tested using the stphtest command of Stata 11, which used the approximation developed by Grambsch and Therneau (1994).

4.3.3 The Alternative Healthy Eating Index

This diet quality index was used to test for convergent validity of the WXYfm and SAIN,LIM NP models. It was developed by McCullough et al. (McCullough *et al.*, 2002) to assess the Dietary Guidelines for Americans in the Nurses' Health Study and the Health Professionals Follow-up Study (see chapter 2 for more details).

Within the Whitehall II data, the Alternative Healthy Eating Index (AHEI) was scored on the basis of the intake levels of 9 components (Akbaraly *et al.*, 2011). The

original components of the index include vegetables, fruit, nuts and soy, the ratio of white (seafood and poultry) to red meat, cereal fibre, trans fat, the ratio of poly-unsaturated fatty acids to saturated fatty acids (PUFA/SFA), long-term multivitamin use (<5 or ≥5y) and alcohol consumption. As cereal fibre was not available in our nutrient dataset we adapted the score by replacing it with total fibre. Each component had the potential to contribute 0 to 10 points to the total score, with the exception of multivitamin use, which contributed either 2.5 or 7.5 points (table 4.4). All the component scores were summed to obtain a total AHEI score ranging from 2.5 to 87.5; higher scores corresponded to a healthier diet.

Table 4.4: Construction of the Alternative Healthy Eating Index (AHEI)

Components		**Criteria for min. scores**	**Criteria for max. scores**	**Possible score range**	**AHEI scores in the participants *** M ±SD
Vegetable (serving /day)		0	5	0-10	5.6 (2.9)
Fruit (serving /day)		0	4	0-10	5.9 (3.1)
Nuts and Soy (serving /day)		0	1	0-10	3.2 (3.0)
Ratio of white to red meat		0	4	0-10	5.1 (2.8)
Total Fibre (% of energy)		0	24	0-10	7.6 (3.0)
Trans Fat (% of energy)		≥4	≤0.5	0-10	8.4 (2.7)
Ratio of PUFA to SFA		≤0.1	≥1	0-10	5.2 (2.7)
Duration of multivitamin Use		<5 year	≥5 year	2.5-7.5	4.2 (2.4)
Alcohol serving/day	Men	0 or >3.5	1.5-2.5	0-10	4.7 (3.7) [$]
Alcohol serving/day	Women	0 or >2.5	0.5-1.5	0-10	
Total Score				2.5-87.5	50.0 (12.0)

PUFA, Poly-unsaturated fatty acids; SAF, saturated fatty acids.
*Each AHEI component contributed from 0 to 10 points to the total AHEI score, except the multivitamin component which was dichotomous and contributing either 2.5 points (for non-use) or 7.5 points (for use) A score of 10 indicates that the recommendations were fully met, whereas a score of 0 represents the least healthy dietary behaviour. Intermediate intakes were scored proportionately between 0 and 10.
[$] Mean score for men and women combined.

4.3.4 Analysis strategy, selection bias

(i) Complete-cases analysis

To analyse independently the effect of aggregate scores on health outcomes, the survival analyses models were adjusted for various confounding factors. Since several models with different levels of adjustment were implemented, it was necessary to keep the same sample for comparison purposes and to assess the effect of the confounders. Complete-cases analysis (i.e. including only participants with the full-set of information) was chosen, as illustrated in figure 4.3. Such conservative strategy was likely to introduce some selection bias.

(ii) Selection bias

To analyse the potential selection bias which occurred by opting for complete-cases analysis, several baseline dietary and non-dietary factors were tabulated according to the inclusion in the main analysis (table 4.5).

There was no difference in energy intake and for most macronutrients, only carbohydrates and protein intake were different between included and excluded participants. Also there was no difference in fruit and vegetable intake. Yet, more participants included in the main analysis were classified in the healthy and Mediterranean dietary clusters previously proven to reduce chronic disease incidence (chapters 2 and 5).

Individuals with no missing information were less likely to smoke, to be obese, to have hypertension, to have reported a longstanding illness or a “fair” or “poor” health over the past year. They had lower levels of several inflammatory markers and HDL cholesterol, though there was no difference in terms of dyslipidaemia status. There was no difference between the two groups in terms of ethnicity and physical activity. Overall, results from table 4.5 did indicate that the sample not included in the analysis had a poorer health status than participants with no missing information.

Such observations followed previous findings and indicated that a selection did occur, with only the fittest participants included in the main analysis sample.

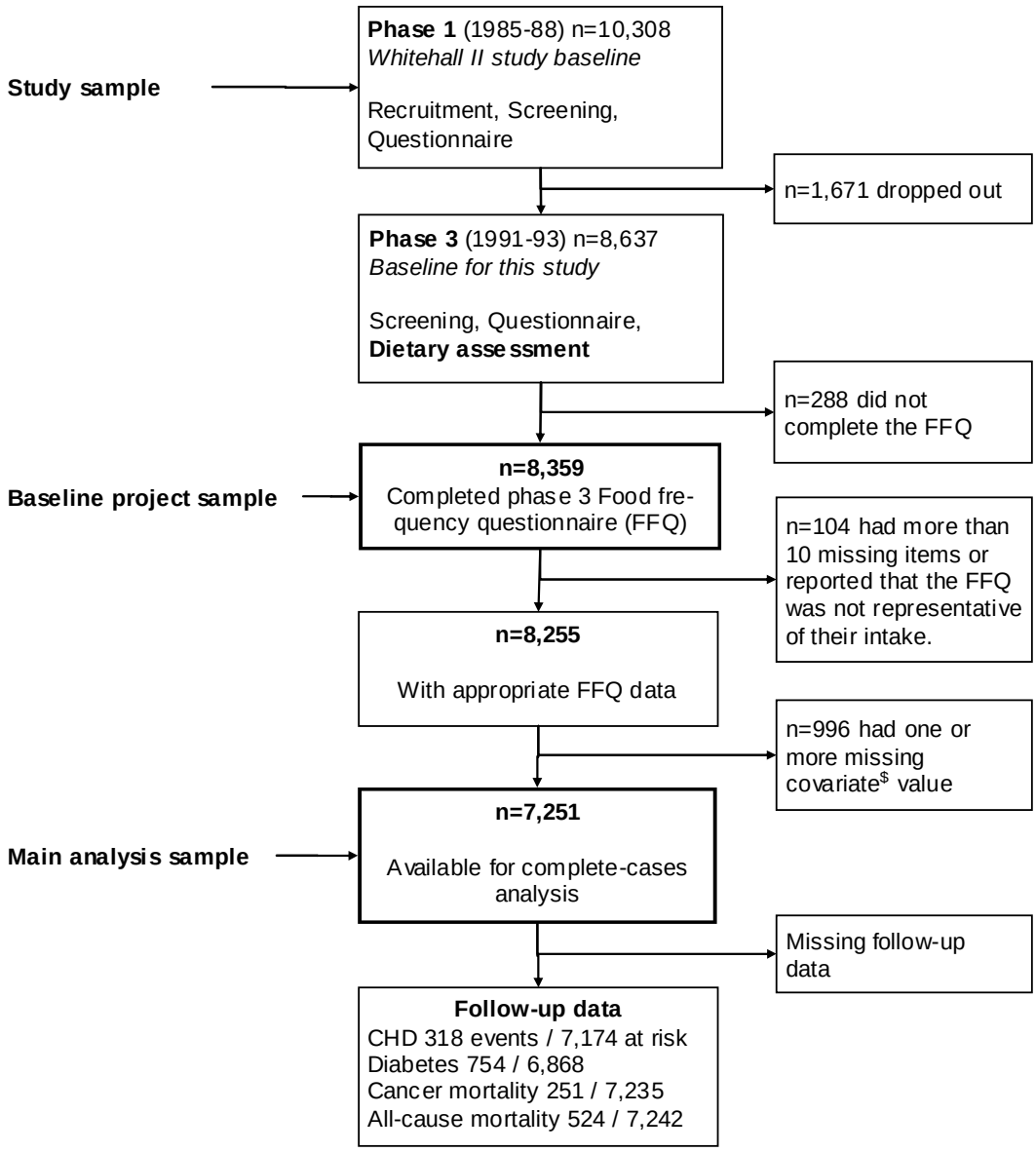


Figure 4.3: Strategy for selection of main analysis sample
$ Covariates included: age, sex, and ethnicity, marital status, employment grade, smoking status, physical activity level, energy and alcohol intake, BMI, hypertension and dyslipidaemia status, and prevalence of longstanding illness.

Table 4.5: Baseline characteristics of participants according to their inclusion in the complete-cases analyses

Variable (mean and SEM, or %, as indicated	Excluded n ≤ 1,108		Included n=7,251		
	Mean	SEM	Mean	SEM	p[$]
Energy intake (kcal/d)	2,248	24.4	2,242	8.07	0.807
Total fat (%en)	33.3	0.21	32.9	0.07	0.070
SFA (%en)	13.6	0.13	13.3	0.04	0.089
MUFA (%en)	10.1	0.07	9.93	0.02	0.024
PUFA (%en)	6.33	0.07	6.37	0.03	0.615
Total carbohydrates (%en)	48.0	0.23	48.9	0.08	<.001
Protein (%en)	18.0	0.12	17.6	0.04	0.001
Alcohol (%en)	4.02	0.17	3.89	0.06	0.432
Fruit and vegetables (g/d)	500	9.24	514	3.40	0.187
% Healthy and med. clusters	48.3		52.3		0.023
Age (y)	50.2	0.19	49.6	0.07	0.001
Ethnicity (% white)	91.8		91.3		0.617
Grade (% high)	14.8		17.6		
Grade (% intermediate)	63.1		66.4		<.001
Grade (% low)	22.1		16.1		
% never smoker	44.5		51.3		
% ex-smoker	35.3		35.1		<.001
% current smoker	20.3		13.6		
% inactive	67.6		66.0		0.350
BMI (kg/m^2)	26.3	0.16	25.2	0.04	<.001
% underweight	3.13		4.41		
% normal weight	34.0		48.9		<.001
% overweight	43.8		37.6		
% obese	19.0		9.06		
Systolic blood pressure (mmHg)	122	0.54	121	0.16	0.001
% Hypertension	27.2		20.3		<.001
Cholesterol - HDL (mmol/L)	1.32	0.02	1.44	0.00	<.001
Cholesterol - LDL (mmol/L)	4.30	0.05	4.40	0.01	0.045
Triglycerides (mmol/L)	2.72	0.11	1.38	0.01	<.001
% Dyslipidaemia	59.0		59.3		0.890
% longstanding illness	40.5		33.5		<.001
% fair or poor self reported health	17.1		10.7		<.001
Interleukin-6	2.14	0.09	1.93	0.03	0.036
C-reactive protein	2.69	0.23	1.87	0.05	<.001
Fibrinogen	2.52	0.02	2.41	0.01	<.001

SEM, standard error of the mean; %en, percent of energy intake; SFA, saturated fatty acids; MUFA, monounsaturated FA; PUFA, polyunsaturated FA; Med. Mediterranean. [$] ANOVA for continuous variables, χ^2 for categorical ones.

4.4 Conclusion

The selection of the EWS aggregate score was the first research objective of the project. The next step is to ensure that the tool designed to assess the predictive validity of the WXYfm model is adequate. This is done in the next chapter using cross-sectional data to assess the association between the EWS and dietary intake. Associations with non-dietary factors were also investigated since such variables could act as confounders in the relationship between the EWS and prospective health outcomes.

Chapter 5: Construct, convergent, and concurrent validity of the EWS aggregate score

In this chapter, cross-sectional data are used to assess whether the EWS aggregate score is appropriate to test for predictive validity of WXYfm.

First, the EWS rankings of participants were associated with intakes of food groups, FFQ-items, and nutrients. Construct validity would be confirmed if participants classified as healthier by the EWS would obtain more favourable intakes profiles. In particular, the EWS would need to discriminate participants at the FFQ-item level to reflect the food-based nutrient profiling (NP) concept.

Second, convergent validity was assessed by linking EWS to measures of global dietary quality previously shown to be protective against adverse health outcomes. Positive associations were expected between EWS and these measures.

Last, associations between EWS and non-dietary factors such as demographic characteristics and health behaviours were investigated to identify potential confounders for the predictive validity models. Risk factors of chronic disease are markers of contemporary health status of participants. Concurrent validity of EWS would be obtained if individuals classified as healthier would display better risk profiles.

Most results were presented by quartile of EWS, with the first quartile containing participants with the less healthy scores, i.e. containing the least amount of healthy foods.

5.1 Dietary intakes and EWS, construct validity

5.1.1 Food group intakes across EWS quartiles

Intakes of most food groups were significantly associated with EWS in both men and women, with some strong trends across quartiles as highlighted in table 5.1. Fish, dairy products, breakfast cereals, and fruit, vegetables and nuts were positively associated with EWS quartiles. On the contrary, the trend was towards reduced intake of bread, snacks and sweets, prepared meals, sauces and spreads, and drinks in the fourth (healthiest) quartile. For meat, no clear association appeared, the participants in the first quartile having the lowest intake.

Trends displayed in table 5.1 were in line with expectations and similar in both sexes: food groups containing healthier foods were eaten in higher quantities in the healthiest quartiles; the contrary was true for food groups with more unhealthy foods. The steep increase in fruit and vegetable intake was confirmed by plasma β-carotene levels which were positively associated to the EWS classification (table 5.2).

Yet, the increased dairy products intake among participants in the fourth quartile of EWS and the pattern of meat intake yielded the question of increased saturated fat intake in these healthiest individuals. Since these two food groups contain foods of diverse nutrient composition, special focus was put on intakes of meat and dairy products.

Table 5.1: Crude food group mean intakes across EWS quartile (4: healthier)

	Men (n=5,627)					Women (n=2,522)				
Group (g/d)	**1**	**2**	**3**	**4**	**p***	**1**	**2**	**3**	**4**	**p***
Meat products and eggs	140	147	148	141	0.008	121	127	131	127	0.108
Fish and shellfish	29.3	32.9	35.4	42.0	<.001	31.0	35.1	37.6	46.7	<.001
Bread and crackers	107	107	106	98.5	0.002	87.1	85.0	81.7	75.5	0.005
Breakfast cereals	36.2	41.1	44.4	43.0	<.001	36.4	37.5	41.5	43.7	0.069
Potatoes, rice and pasta	182	201	205	201	<.001	157	184	181	180	<.001
Dairy products	376	428	512	659	<.001	408	502	659	828	<.001
Meals[#]	25.9	27.0	25.4	19.5	<.001	22.0	23.2	20.0	16.3	<.001
Fat spreads	30.0	21.0	16.5	10.8	<.001	24.6	17.4	13.6	8.7	<.001
Snacks and sweets	152	116	87.4	52.0	<.001	121	80.8	62.2	37.5	<.001
Sauces and other spreads	49.5	45.8	40.7	30.1	<.001	39.0	36.0	30.3	24.8	<.001
Drinks[§]	733	723	723	685	0.007	720	720	732	692	0.343
Fruit and nuts	181	217	241	311	<.001	237	284	317	429	<.001
Vegetables	212	235	250	275	<.001	223	251	272	332	<.001

[#] Meals included quiche, pizza and lasagne. [§]Excluded alcohol and milks. *Heterogeneity ANOVA across quartiles.

Table 5.2: Mean levels of plasma β-carotene across EWS quartiles (4: healthier)

	Men (n=3,975)					Women (n=1,873)				
Quartiles	**1**	**2**	**3**	**4**	**p***	**1**	**2**	**3**	**4**	**p***
β-carotene (mol/L)	0.82	0.85	0.86	0.97	<.001	0.93	1.12	1.01	1.11	0.029

* Heterogeneity ANOVA across quartiles.

5.1.2 Meats and dairy products intakes across EWS quartiles

Similarly to table 5.1, intakes of almost all items in table 5.3 were significantly associated with the EWS rankings. Consumption trends were clearly dependent on the WXYfm overall score of each item. Leanest products with low overall score (i.e. “healthier”) were consumed in greater amount within the fourth quartile, whereas foods with high overall score (i.e. “less healthy”) had their intake highest in the first quartile. This indicated that the EWS algorithm did discriminate participants according to their consumption of individual FFQ-items, and their respective WXYfm score.

This discrimination was well illustrated by the intake of saturated fat from meat and dairy products, lowest in the healthiest quartile (table 5.3). Inverse trends were also observed for sugar and sodium intake from meat products. They were not for dairy products, which suggested that the positive components outweighed sodium and sugar for this group.

5.1.3 Nutrient intakes across EWS quartiles

Consumption trends observed for meat, dairy and other food groups were reflected in the nutrient intakes (table 5.4). Lower intake of all fats including cholesterol and unsaturated fatty acids was observed in the fourth quartile. Energy intake was strongly and inversely associated with EWS. Despite such trend, crude intakes of most micro-nutrients were higher in participants classified in the healthier quartiles. The few exceptions were iron, and vitamins A, D, and E which displayed inverse or non-significant associations. These were reversed to positive when intakes were energy-adjusted (not shown). Intake of the negative sodium nutrient was lowest in the fourth quartile.

Overall, the above results did indicate that the EWS construct was valid. It discriminated participants at the FFQ-item level and therefore appeared to be appropriate to assess the predictive validity of WXYfm.

Table 5.3: Crude meat and dairy products mean intakes across EWS quartiles (4: healthier)

FFQ-item (g/d)	WXYfm score[#]	Men (n=5,627) 1	2	3	4	p*	Women (n=2,522) 1	2	3	4	p*
Beef	1	22.0	24.6	25.2	24.9	<.001	20.1	21.4	20.9	19.5	0.494
Beefburgers	19	1.93	1.80	1.72	1.12	<.001	0.91	1.10	0.78	0.51	<.001
Pork	0	9.33	10.9	10.9	9.67	<.001	8.61	9.45	9.30	8.71	0.565
Lamb	1	9.80	11.0	11.2	11.4	0.003	11.2	11.9	11.7	9.63	0.012
Chicken	-1	33.3	39.8	45.5	51.0	<.001	36.0	41.0	47.6	57.2	<.001
Bacon	21	5.81	4.69	4.22	3.20	<.001	3.91	3.36	3.24	2.30	<.001
Ham	12	6.51	6.40	6.12	5.08	<.001	5.11	4.60	4.99	4.17	0.042
Corn beef, spam, luncheon meats	16	3.72	3.41	3.08	2.26	<.001	2.74	1.91	1.94	1.37	<.001
Sausages	20	5.77	5.03	4.49	3.18	<.001	3.52	3.09	2.65	1.69	<.001
Pies	14	12.8	10.6	8.29	5.02	<.001	6.08	4.94	4.43	2.27	<.001
Liver products	-3	1.61	1.73	1.75	1.38	0.002	1.52	1.46	1.39	1.40	0.804
Meat soup	3	11.4	11.5	11.2	11.5	0.986	7.43	8.94	9.66	9.46	0.181
Eggs	14	16.5	15.5	13.8	11.2	<.001	14.3	13.8	12.7	8.99	<.001
SFA from meat (g)		6.05	5.99	5.73	5.05	<.001	4.80	4.76	4.70	4.09	<.001
Sodium from meat (mg)		547	507	470	390	<.001	397	373	371	301	<.001
Sugars from meat (g)		0.52	0.47	0.42	0.32	<.001	0.32	0.30	0.29	0.22	<.001

(Continued)

Table 5.3 (continued)

FFQ-item (g/d)	WXYfm score[#]	Men (n=5,627)					Women (n=2,522)				
		1	2	3	4	p*	1	2	3	4	p*
Whole milk	2	122	123	116	102	0.095	117	124	125	77.7	0.024
Semi skimmed milk	0	145	190	239	302	<.001	158	215	324	346	<.001
Skimmed milk	-1	46.7	50.4	84.7	188	<.001	63.1	76.8	124	277	<.001
Channel Island milk	2	1.95	2.11	4.13	0.62	0.207	1.08	4.25	0.75	6.28	0.473
Sterilised milk	0	3.77	1.35	7.63	14.0	0.002	1.88	8.22	7.22	36.9	<.001
Dried milk	20[$]	1.01	0.78	0.78	0.79	0.095	1.26	1.14	1.34	1.37	0.688
Soya milk	-1	0.90	1.69	2.75	2.28	0.516	2.14	3.21	0.76	3.31	0.534
Coffee whitener	20	1.96	1.57	1.17	0.60	<.001	1.37	1.07	0.74	0.54	<.001
Single cream	12	1.61	1.52	1.24	0.73	<.001	1.84	1.52	1.03	0.49	<.001
Double cream	15	1.43	1.35	1.06	0.57	<.001	1.99	1.36	1.04	0.48	<.001
Yoghurt	-1	26.6	35.7	37.1	40.0	<.001	39.0	46.9	55.0	68.0	<.001
Cheese	22	21.6	17.8	15.0	10.4	<.001	18.5	15.5	12.5	8.07	<.001
Cottage cheese, fromage frais	2	1.77	2.26	2.52	2.88	<.001	3.59	3.79	4.95	5.83	<.001
SFA from dairy (g)		11.0	10.7	10.6	9.74	<.001	10.6	10.9	11.3	10.5	0.535
Sodium from dairy (mg)		359	366	396	451	<.001	366	401	475	546	<.001
Sugars from dairy (g)		18.7	21.6	25.9	33.7	<.001	20.8	25.6	34.0	43.0	<.001

SFA, saturated fatty acid. [#] see chapter 2. [$] This score is calculated on 100g of powder, not as consumed dry milk.
* Heterogeneity ANOVA across quartiles.

Table 5.4: Crude energy and nutrient mean intakes across EWS quartiles (4: healthier)

	Men (n=5,627)					Women (n=2,522)				
Nutrient (unit/d)	**1**	**2**	**3**	**4**	**p***	**1**	**2**	**3**	**4**	**p***
Energy (kcal)	2,540	2,389	2,275	2,097	<.001	2,185	2,067	2,021	1,960	<.001
Total fat (%en)	37.6	34.5	32.1	28.2	<.001	37.6	34.2	31.6	27.2	<.001
SFA (%en)	15.8	14.0	12.8	10.9	<.001	16.2	14.1	12.7	10.6	<.001
MUFA (%en)	11.1	10.5	9.92	8.68	<.001	11.0	10.3	9.58	8.03	<.001
PUFA (%en)	7.17	6.76	6.28	5.64	<.001	6.73	6.50	5.98	5.33	<.001
Total carbohydrates (%en)	46.8	48.2	48.9	50.6	<.001	47.0	48.5	49.6	52.2	<.001
Protein (%en)	14.7	16.4	17.7	19.7	<.001	16.2	18.0	19.5	21.7	<.001
Alcohol (%en)	4.11	4.29	4.70	5.00	<.001	2.46	2.57	2.63	2.42	0.742
Sodium (mg)	3,192	3,011	2,875	2,596	<.001	2,745	2,565	2,546	2,441	<.001
Potassium (mg)	3,830	4,043	4,205	4,453	<.001	3,688	4,003	4,282	4,783	<.001
Calcium (mg)	1,147	1,189	1,257	1,385	<.001	1,121	1,215	1,394	1,598	<.001
Magnesium (mg)	361	379	392	408	<.001	331	353	373	408	<.001
Phosphorus (mg)	1,625	1,706	1,774	1,868	<.001	1,549	1,656	1,812	1,996	<.001
Iron (mg)	13.0	13.3	13.3	12.9	0.002	11.8	12.0	12.1	12.4	0.067

(Continued)

Table 5.4 (continued)

	Men (n=5,627)					Women (n=2,522)				
Nutrient (unit/d)	**1**	**2**	**3**	**4**	**p***	**1**	**2**	**3**	**4**	**p***
Vitamin A (µgRE)	1,318	1,307	1,277	1,142	<.001	1,246	1,223	1,211	1,260	0.646
Vitamin D (µg)	4.68	4.63	4.67	4.82	0.445	4.40	4.47	4.62	5.00	0.009
Thiamin (mg)	1.90	2.02	2.04	2.03	<.001	1.72	1.83	1.90	2.00	<.001
Riboflavin (mg)	2.23	2.39	2.52	2.67	<.001	2.15	2.34	2.61	2.89	<.001
Niacin (mgNE)	23.4	24.9	25.3	25.0	<.001	21.6	22.8	23.3	24.0	<.001
Vitamin C (mg)	117	134	142	165	<.001	136	163	173	221	<.001
Vitamin E (mg)	5.55	5.64	5.57	5.29	<.001	5.49	5.49	5.50	5.70	0.297
Vitamin B6 (mg)	2.33	2.55	2.63	2.72	<.001	2.22	2.45	2.57	2.77	<.001
Vitamin B12 (µg)	6.67	7.25	7.74	8.24	<.001	6.62	7.17	7.90	8.86	<.001
Total folic acid (µg)	320	344	350	360	<.001	305	333	345	381	<.001
Panthothenic acid (µg)	5.57	6.00	6.32	6.77	<.001	5.36	5.90	6.48	7.25	<.001
Biotin (µg)	42.5	45.0	46.4	48.3	<.001	38.5	41.9	44.4	47.3	<.001
Cholesterol (mg)	264	259	252	232	<.001	246	239	239	221	<.001
Fibre (g)	24.6	25.9	26.3	26.6	<.001	23.0	24.3	24.8	27.0	<.001

%en, percent of energy intake; SFA, Saturated fatty acids; MUFA, Mono-unsaturated fatty acid; PUFA, Poly-unsaturated fatty acid; RE, retinol equivalent; NE, niacin equivalent.

* Heterogeneity ANOVA across quartiles.

5.2 Convergent validity against dietary quality indices

Two existing measures of dietary pattern quality previously associated with reduced risk of prospective health outcomes within the Whitehall II population (the Alternative Healthy Eating Index (AHEI) and data-driven clusters, chapters 2 and 4) were used to assess the convergent validity of the EWS aggregate score.

5.2.1 The Alternative Healthy Eating Index

This index has been recently applied to the Whitehall II data and was shown to be predictive of lower cardiovascular and all-cause mortality (Akbaraly *et al.*, 2011).

Rank correlations between the EWS and the AHEI were similar for men and women: -0.260 and -0.263, respectively. Such values were in the expected direction (the WXYfm overall score being on an inverse scale, smaller values of EWS represented healthier foods/diets) but were relatively low. Linear regressions yielded R^2 values below 0.10 (not shown), which was lower than previous results linking NP models to the Healthy Eating Index (chapter 2). These weak associations were well illustrated in table 5.5: significant but weak positive gradients were observed between the EWS quartiles and the AHEI. Quartiles cross-tabulations further confirmed the poor agreement as measured by the κ-statistic between the two scores (table 5.5).

Table 5.5: EWS and AHEI quartiles cross-tabulation (4: healthier)

	EWS quartiles							
	Men (n=5,626)				**Women (n=2,518)**			
	1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Mean AHEI*	36.9	40.1	41.7	44.7	39.8	43.3	44.5	48.8
AHEI quartiles[#]								
1	479	344	291	195	225	155	141	98
2	422	373	353	308	154	155	146	118
3	286	389	372	376	160	182	176	145
4	218	301	391	528	91	138	167	267

AHEI: Alternative Healthy Eating Index.
*p<.001 in both sexes for heterogeneity ANOVA across quartiles.
[#]Weighted κ-statistic were 0.159 in men and 0.167 in women.

5.2.2 Dietary clusters

The cluster approach was applied to the Whitehall II data to define four mutually exclusive groups in which participants had similar dietary characteristics: unhealthy, sweet, Mediterranean and healthy (Brunner *et al.*, 2008). Compared with the unhealthy pattern (white bread, processed meat, chips, and whole milk), the healthy pattern (fruit, vegetables, whole-meal bread, low-fat dairy, and little alcohol) was found to reduce incident risk of CHD and diabetes.

Cross-tabulations of the clusters and the EWS quartiles showed some agreement between the two methods (table 5.6). Both the sweet and unhealthy clusters were less represented in the third and fourth quartiles of EWS, with a stronger trend for the smaller sweet cluster. In contrast, a strong positive gradient appeared for the healthy cluster. For the Mediterranean cluster, there was a positive trend in men, and an inverse quadratic association in women. Yet, some disagreement was observed: 32.2% and 20.8% of men and women respectively classified in the EWS fourth quartile were in the unhealthy cluster; and 22.2% and 28.2% were respectively classified in the healthy cluster despite being in the EWS first quartile.

Analysis of variance confirmed the overall trends (table 5.6). According to the EWS, participants classified in the sweet cluster had the highest intake of less healthy foods. Individuals from the healthy cluster did have the healthiest aggregate scores.

Associations between the EWS and the two measures of dietary quality were in the expected directions, which confirmed the convergent validity of the aggregate score. The relationships were relatively weak, illustrating that the EWS ranked participants differently compare to measures based on total intake and global dietary patterns.

Table 5.6: Cross-tabulation of clusters and EWS quartiles (4: healthier) in men and women

Cluster (n, column %)	Men (n=5,303)					Women (n=2,301)				
	Quartiles				Mean	Quartiles				Mean
	1	2	3	4	EWS*	1	2	3	4	EWS*
Unhealthy	531	457	475	420	6.61	253	205	172	114	6.16
	40.0	34.1	35.7	32.2		43.5	35.2	29.3	20.8	
Sweet	353	304	177	53	7.56	68	45	30	6	6.88
	26.6	22.7	13.3	4.1		11.7	7.7	5.1	1.1	
Mediterranean	148	233	287	258	6.03	97	148	112	49	5.85
	11.2	17.4	21.6	19.8		16.7	25.4	19.1	8.9	
Healthy	295	347	392	573	5.71	164	185	274	379	4.53
	22.2	25.9	29.5	43.9		28.2	31.7	46.6	69.2	

* $p<.001$ in both sexes for heterogeneity ANOVA across clusters.

5.3 Non-dietary characteristics

5.3.1 Socio-demographic characteristics

In both men and women, age, marital status, ethnicity, and employment grade were significantly associated with the EWS quartiles (table 5.7). Age was positively associated with the quartiles, which was coherent with previous observations. White ethnicity was inversely associated with the aggregate score quartiles, suggesting that ethnic minorities had healthier diets. No clear trend could be observed for marital status in men, whereas an inverse gradient appeared in women, indicating that married women had healthier dietary patterns. There was a positive association between low employment grade and EWS; and the fourth quartile contained fewer high graded participants. This pattern between employment grade and dietary intakes had been observed among Whitehall II participants when using the cluster distribution (Brunner *et al.*, 2008), but not with the AHEI diet index (Akbaraly *et al.*, 2011).

5.3.2 Health behaviours and self-perception of health

Table 5.7 reveals that participants in the fourth quartile of EWS were less likely to be current smokers, the difference was not significant in women. Such improved health behaviour was confirmed by answers relating to the importance of health and the self-control over one's health. Indeed, participants in the healthier quartiles were more likely to find health "extremely important" and to agree strongly that keeping healthy depended on them.

There was no association between physical activity and the EWS rankings. This result was counter intuitive and against previous observations made in Whitehall II. It confirmed that the EWS did classify individuals differently compared to data-driven clusters and the AHEI diet index.

Table 5.7: Socio-demographic characteristics and health behaviours across EWS quartiles (4: healthier)

Variable (mean or %)	Men					Women				
	1	2	3	4	p*	1	2	3	4	p*
Age (y)	48.5	49.1	49.6	50.2	<0.001	49.7	50.2	50.8	51.0	0.001
% living alone[#]	18.9	15.2	16.9	18.4	0.045	42.7	33.3	37.9	32.5	<0.001
Ethnicity (% white)	97.4	96.7	93.9	86.3	<0.001	94.7	92.0	84.5	74.5	<0.001
Grade (% high)	20.5	23.3	23.9	21.9		6.24	6.60	7.20	3.57	
Grade (% intermediate)	73.9	71.9	70.0	68.9	<0.001	61.0	56.5	50.6	50.4	<0.001
Grade (% low)	5.65	4.81	6.11	9.27		32.8	36.9	42.2	46.0	
% never smoker	49.0	48.1	49.1	47.8		51.9	54.6	56.5	60.5	
% ex-smoker	36.0	41.0	37.7	40.2	0.016	29.2	27.2	26.6	24.8	0.125
% current smoker	14.9	10.9	13.2	11.9		18.9	18.2	16.8	14.7	
METs [§]	3.92	3.95	4.01	3.87	0.699	3.31	3.35	3.22	3.28	0.899
% inactive	63.1	63.5	60.8	64.7	0.180	73.7	72.7	73.7	73.2	0.978
% Agree strongly "Keeping healthy depends on me"	28.1	31.6	34.3	40.9	<.001	32.4	34.0	39.7	42.2	<.001
% "Health is extremely important"	41.5	39.7	42.0	48.2	0.001	43.9	48.0	50.9	54.1	0.042

[#] Never married/cohabiting, divorced, or widowed. [§] Metabolic equivalents. * Heterogeneity ANOVA or χ^2

5.3.3 Concurrent validity against baseline risk factors and inflammatory markers

BMI and systolic blood pressure were higher in men and women in the fourth quartile of EWS, i.e. participants classified as healthiest by the aggregate score had higher levels of obesity and hypertension (table 5.8). The trend was similar for blood lipids in men, with higher levels of total cholesterol, LDL cholesterol, and triglycerides among participants in the fourth quartile of EWS. Associations were not significant in women. The trends observed on inflammatory markers followed similar patterns. In men and women, levels of fibrinogen and Von Willebrand's factor were significantly higher in the fourth quartile. In women, C-reactive protein levels were positively associated with the EWS quartiles (table 5.8). For interleukin-6, a U-shaped association was observed in men only. There was an increased prevalence of longstanding illnesses in participants classified in the third and fourth quartiles of EWS.

Despite being associated with healthier dietary intake patterns and more favourable health behaviours, the EWS aggregate score was positively associated with a less favourable risk profile. Therefore, the concurrent validity of the EWS could not be established.

Table 5.8: Risk factors and inflammatory markers levels across EWS quartiles (4: healthier)

	Men					Women				
Variable (Mean or %)	**1**	**2**	**3**	**4**	**p***	**1**	**2**	**3**	**4**	**p***
BMI (kg/m^2)	24.8	25.0	25.1	25.6	<.001	25.1	25.4	25.9	26.4	<.001
% underweight	3.85	3.57	4.17	2.64		9.00	7.24	3.67	4.46	
% normal weight	53.4	50.7	48.6	43.8	<.001	47.4	47.1	45.2	42.0	<.001
% overweight	36.3	40.2	39.5	44.8		31.6	30.8	35.5	33.8	
% obese	6.45	5.51	7.67	8.73		12.1	14.8	15.7	19.7	
Systolic blood pressure (mmHg)	121	121	122	123	<.001	116	117	119	119	<.001
% Hypertension[#]	18.9	19.1	23.0	27.2	<.001	15.6	14.1	20.1	22.2	0.001
Cholesterol - Total (mmol/L)	6.47	6.40	6.49	6.57	0.002	6.54	6.54	6.48	6.54	0.805
Cholesterol - LDL (mmol/L)	4.45	4.39	4.45	4.50	0.044	4.31	4.31	4.24	4.31	0.687
Cholesterol - HDL (mmol/L)	1.32	1.32	1.32	1.32	0.950	1.68	1.70	1.68	1.66	0.390
Triglycerides (mmol/L)	1.54	1.54	1.64	1.70	0.001	1.21	1.17	1.23	1.25	0.346
% Dyslipidaemia	63.3	59.1	62.0	63.7	0.066	53.6	54.2	53.8	52.9	0.976
% longstanding illness	32.7	31.2	36.1	35.4	0.023	32.9	33.1	35.4	38.1	0.183
Fibrinogen (g/L)	2.31	2.31	2.34	2.40	<.001	2.52	2.59	2.61	2.65	0.006
Von Willebrand's factor (IU/dl)	105	107	106	110	0.007	106	108	110	115	0.002
C-reactive protein (mg/L)	1.80	1.82	1.62	1.89	0.472	1.92	2.28	2.57	2.24	0.073
Interleukin-6 (ng/L)	1.91	1.71	1.74	1.89	0.037	2.24	2.16	2.38	2.27	0.540

[#] Hypertension was defined as systolic or diastolic blood pressure ≥ 140 or ≥ 90 mmHg, respectively, or by the use of hypertensive drugs.

* Heterogeneity ANOVA or χ^2

5.4 Discussion

The EWS aggregate score was significantly associated with improved dietary intakes. Most micro-nutrients crude intakes were positively associated with the EWS, despite a strong inverse association between the EWS and energy intake. Participants were classified based on their relative intake of FFQ-items, which related well to the NP concept and confirmed the construct validity of the aggregate score.

Convergent validity was also confirmed: both the dietary clusters and the AHEI diet index were significantly associated with the EWS in the expected direction. The relationships were weaker than in previous investigations (Fulgoni *et al.*, 2009; Katz *et al.*, 2010) and some disagreement appeared. These disagreements indicated that the EWS was not a simple copy of existing diet quality indices. Consistent with the NP concept, the EWS did classify participants according to their consumption of individual FFQ-items, unlike the AHEI which used total intake; and independently of their global dietary pattern, unlike the dietary clusters.

Associations with non-dietary factors revealed that several characteristics including employment grade could act as confounder in the survival analyses models. The risk factors and inflammatory markers levels indicated that individuals in the healthiest quartiles of EWS had riskier profiles. This was especially true for obesity levels and hypertension. The concurrent validity of the aggregate score was not confirmed. These surprising results could be linked to the higher prevalence of longstanding illnesses in the participants classified as healthiest by the EWS. Such existing condition may have encouraged individuals to have healthier dietary patterns and to be more health conscious, as illustrated by the trend with former smokers (table 5.7).

In summary, the EWS aggregate score translated well the NP concept at diet level. It was considered as adequate to assess the predictive validity of the WXYfm model. The survival analysis implemented in the next chapter need to account for the potential confounders highlighted in this chapter. In particular, the EWS was

associated with energy intake, employment grade, and BMI which are all markers of dietary misreporting (chapter 2). Such possibility is taken into account in the models of the next chapter, and is further discussed in chapter 8.

Chapter 6: Survival analysis, predictive validity of WXYfm

In this chapter, the first results for the main objective of the project, i.e. testing for predictive validity of the WXYfm nutrient profiling (NP) model, are reported. The initial hypothesis was that diets containing higher proportions of healthy foods would be protective against prospective chronic disease and mortality. The EWS aggregate score was shown to translate well the WXYfm model at diet level, and it was expected to observe risk reduction of incident events in participants with a healthier aggregate score. The inclusion of potential confounding factors in the Cox proportional hazards models allowed assessing the independent effect of EWS on fatal and non-fatal CHD, diabetes, cancer mortality, and all-cause mortality. A brief methods section giving the exact specifications of the models is included first. The discussion section highlights possible explanations for the observed results.

6.1 Methods, Cox regressions

Survival analyses were run by fitting Cox proportional hazard regressions using follow-up time in years as time variable. Individuals with prevalent cases of disease at baseline were excluded from the models (chapter 4). Participants were classified in quartiles of EWS, the first and least healthy one served as reference group. Linear trend was assessed by including the quartiles as continuous variable in two ways, including directly the EWS in the models did not change the results. Tests for quadratic trends were done by including EWS and EWS-square in the regressions (Greenland, 1995). Inclusion of covariates followed the assessment of their relationship with health outcomes using Log-rank tests (not shown).

All associations presented in chapter 5 were displayed in sex-specific quartiles. As most trends were similar for both men and women, it was chosen to combine sexes in the Cox regression models allowing for stronger estimates to be obtained. Women had healthier diets according to EWS (Cochran-Armitage trend test p-value was

below 0.001); Cox models were adjusted accordingly. Analyses by sex were conducted and conclusions were not modified (not shown).

Base models were adjusted for age, sex, and ethnicity (white, Asian, black, other). A second model was further adjusted for marital status (married/cohabiting, single, divorced or widowed), employment grade (civil service scale), smoking status (never, ex, current), physical activity level (low, intermediate, high), and energy and alcohol intake. A last model included BMI categories (underweight, normal weight, overweight, obese), hypertension and dyslipidaemia status, and prevalence of longstanding illness.

The proportional hazard assumption was tested using scaled Schoenfeld residuals. It was not met for sex with diabetes as outcome, for longstanding illness and dyslipidaemia with all-cause mortality as outcome, and for BMI categories with CHD. Analyses were stratified in such cases as interaction tests were not significant (not shown).

All analyses were carried out on a complete-case dataset (n=7,251, not including outcome variables) to allow comparison between different levels of adjustment.

6.2 Results, unexpected U-shaped associations

The first observation was that the expected risk reduction was not obtained for any outcome, with no significant estimate for individual quartiles (table 6.1). Linear risk increase of CHD was observed with models 1 and 2. All other significant values were obtained for the quadratic trend tests, suggesting U-shaped associations between the EWS aggregate score and prospective health outcomes.

More specifically, the U-shape was observed for all outcomes in model 1, and was strongest for CHD and all-cause mortality. A significant linear risk increase was also observed for CHD. The associations between EWS and cancer mortality and diabetes were very weak.

The socio-demographic and health behaviour covariates included in model 2 did attenuate the U-shapes, quadratic trend tests not being significant for diabetes, but conclusions were similar to model 1.

Table 6.1: Cox regression estimates for EWS quartiles (4: healthier)

Outcome (cases / n)	Quartile /trend	Model 1			Model 2			Model 3		
		HR	**95 % CI**		**HR**	**95 % CI**		**HR**	**95 % CI**	
CHD (318 / 7,174)	1	Ref			Ref			Ref[#]		
	2	0.78	0.56	1.10	0.81	0.58	1.14	0.82	0.58	1.15
	3	1.06	0.77	1.45	1.09	0.79	1.49	1.03	0.75	1.41
	4	1.31	0.96	1.79	1.34	0.97	1.84	1.22	0.89	1.69
	Linear	**1.12**	**1.01**	**1.25**	**1.13**	**1.01**	**1.25**	1.09	0.98	1.21
	p quadratic	**<.001**			**0.001**			**0.003**		
Diabetes (754 / 6,868)	1	Ref*			Ref*			Ref*		
	2	0.95	0.77	1.17	0.99	0.80	1.22	1.00	0.81	1.24
	3	0.89	0.72	1.10	0.93	0.75	1.15	0.89	0.72	1.10
	4	1.08	0.88	1.33	1.12	0.90	1.38	1.04	0.84	1.28
	Linear	1.02	0.95	1.09	1.03	0.96	1.10	1.00	0.93	1.07
	p quadratic	**0.042**			0.085			0.401		
Cancer mortality (251 / 7,235)	1	Ref			Ref			Ref		
	2	0.94	0.66	1.35	0.94	0.65	1.35	0.94	0.65	1.35
	3	1.01	0.71	1.43	0.97	0.68	1.38	0.95	0.66	1.36
	4	0.95	0.66	1.36	0.90	0.62	1.30	0.87	0.60	1.26
	Linear	0.99	0.88	1.11	0.97	0.86	1.09	0.96	0.85	1.08
	p quadratic	**0.032**			**0.041**			0.057		
All-cause mortality (524 / 7,242)	1	Ref			Ref			Ref[§]		
	2	0.85	0.66	1.09	0.85	0.66	1.10	0.85	0.66	1.09
	3	0.89	0.69	1.14	0.89	0.69	1.14	0.86	0.67	1.11
	4	1.04	0.81	1.33	1.02	0.79	1.30	0.97	0.76	1.25
	Linear	1.02	0.94	1.10	1.01	0.93	1.10	0.99	0.92	1.08
	p quadratic	**<.001**			**0.002**			**0.004**		

Model 1 adjusted for age, sex, and ethnicity. Model 2 adjusted for marital status, employment grade, smoking status, physical activity level, and energy and alcohol intake. Model 3 adjusted for BMI, hypertension and dyslipidaemia status, and prevalence of longstanding illness.
HR, hazard ratio; CI, confidence interval. [#] Stratified for BMI categories. * Stratified for sex.
[§] Stratified for longstanding illness and dyslipidaemia.

The further adjustment in model 3 did have some effect on individuals in the fourth quartile of EWS which obtained lower hazard ratio estimates. It resulted in non-significant risk increase for CHD, which suggested that the trends in models 1 and 2 were partly due to confounding by the risk factors. The quadratic associations remained significant for CHD and all-cause mortality. There was a slight suggestion of an inverse linear trend for cancer mortality.

The U-shapes did indicate that EWS aggregate score was associated with health outcomes, although not in the expected direction. The strongest risk reductions were observed for CHD and all-cause mortality, in the middle quartiles of EWS. These specific estimates, especially for the second quartile, were less susceptible to further adjustment than the fourth quartile estimates, suggesting that the risk reduction was effectively due to EWS. Yet, confidence intervals were wide and no strong conclusions could be drawn from these individual results.

6.3 Discussion

Predictive validity of the EWS aggregate score was not demonstrated by the results of table 6.1. Risk reduction was hypothesised for participants in the healthiest quartile of EWS, but quadratic trends were obtained for all outcomes, suggesting that these were not chance findings. Participants in the middle (e.g. second and third) quartiles of EWS were at lower risk than the reference and least healthy individuals; though the estimates were not significant. The unexpected results came with participants in the fourth quartile, i.e. which reported diets had highest content of healthy foods. They were at equal, or higher, risk of incident chronic disease than individuals from the first quartile.

The above results must be taken with caution as they could not make abstraction of several limitations (see chapter 10 for details). Most of these limitations were common with the investigation of the ONQI NP model by Chiuve et al. (2011). The ONQI-f aggregate score was shown to be associated with linear risk reduction of all chronic disease but cancer. This suggested that NP models should be associated with a protective effect even though they are not designed to score whole dietary patterns.

Several factors could have explained these diverging results: (i) the NP models, (ii) the cohort studies, (iii) the aggregation methods, (iv) dietary misreporting, and (v) diet variety.

The first explanatory factor is the NP models used in both investigations. The WXYfm model is a relatively simple across-the-board NP algorithm which incorporates seven nutrients in its calculation. The ONQI model is known to contain 30 nutrients and adjusting factors (Katz *et al.*, 2010). The incorporation of such extra-nutrients could have explained the protective associations. Further details of this patented model cannot be disclosed. The ONQI algorithm could not be applied to the Whitehall II dataset since some nutrient content information was missing (chapters 2 and 10).

Second, the WXYfm and ONQI models were implemented on different cohort studies. The combined analysis of the US-based Nurses' Health Study and the Health Professionals' Follow-up Study yielded more than 100,000 participants and over 30,000 events, which largely exceeds the sample used in the present project. However, it was not possible to access the US-based data, and no cross-comparison could be made. The Whitehall II FFQ was an anglicised version of the US questionnaires (chapter 4); it might have been more adapted to the US population.

Third, the EWS aggregate score was weighted by energy and the ONQI-f by portion size. Analyses run with the PWS and WWS aggregate scores (respectively weighted by portion size and weight of intake, see chapter 4) yielded similar results for WXYfm (not shown). Chiuve and colleagues reported that an ONQI-f score weighted on energy was implemented but did not yield significant results.

Fourth, chapter 5 highlighted systematic relationships between EWS and energy intake, BMI, and employment grade; all markers of misreporting of intakes (Brunner, 1997). Such dietary misreporting could have led to a number of participants with an observed EWS ranking not reflecting their true intake. The models implemented in the present chapter were adjusted for total energy intake, employment grade, and BMI. Models excluding participants at the extreme of the energy intake distribution showed results similar to table 6.1 (results not shown).

Yet, residual confounding linked to dietary misreporting—and its associations with risk factors such as BMI and hypertension (Macdiarmid & Blundell, 1998)—could have occurred.

Last, NP models are designed to score individual foods, not diets. In line with this concept, the EWS did not incorporate any information on the dietary patterns of individuals apart from total energy intake. Other characteristics of diets, in particular diet variety, may have been related to EWS in a way that would confound its association with health outcomes. One could, for example, have obtained a very healthy EWS score by having a diet restricted to only a few very healthy foods. Such dietary patterns may be detrimental (Michels & Wolk, 2002; Savy *et al.*, 2005) and could have explained the U-shaped associations.

The detailed investigation of these explanatory factors shaped the reminder of this project as the focus shifted towards the understanding of the observed prospective associations. In order to verify whether the present results were due to WXYfm alone, an alternative NP model, SAIN,LIM (see chapter 2), was applied to the Whitehall II data. This across-the-board model was developed by the French food safety agency and has been shown to relate well to nutrient recommendations in the French population (Maillot *et al.*, 2011). Its algorithm differs from WXYfm in several aspects (chapter 2), which makes this model a good comparison tool. To analyse the impact of misreporting, more refined techniques were implemented for the detection of global energy reporters and for correcting intake at the FFQ-item level. Further, diet variety was related to the EWS aggregate score and to prospective health outcomes to assess its potential as a confounding factor. To better understand the relationship between the WXYfm-derived EWS and prospective health status in the Whitehall II population, WXYfm components (i.e. the nutrients incorporated in its algorithm) were investigated separately to assess: (i) whether they predicted prospective chronic disease as hypothesised, i.e. intake of positive nutrients would be protective, and intake of negative nutrients would predict increased risk; and (ii) whether the EWS aggregate score was evenly correlated to all its components.

6.4 Conclusion, explaining the U-shapes

The initial hypothesis was not verified since no significant risk reduction of any outcome was observed, and the predictive validity of WXYfm could not be confirmed. The results from table 6.1 were highly unexpected, especially given the prior report from Chiuve et al. Yet, such inconclusive results did not rule out the existence of an association between WXYfm and prospective health outcomes. Explaining the observed U-shapes became a priority, and was used as a thread for the reminder of the project.

Chapter 7: Validity of the SAIN,LIM model

The previous chapter did not fully confirm the predictive validity of the WXYfm model. The observed associations may have been due to the NP model itself, and it was necessary to assess an alternative model. In this chapter, the SAIN,LIM model was tested for construct, convergent, concurrent, and predictive validity; in a similar way as WXYfm was in chapters 5 and 6, with the same hypotheses and research questions.

Similarly to the WXYfm model, the SAIN,LIM is an "across-the-board" algorithm using thresholds to define healthiness categories (detailed description in chapter 2). Its specificity lies in the separation of the positive and negative scales: the SAIN and LIM sub-scores, respectively. The algorithm is bi-dimensional and two thresholds are used to define four healthiness categories or "quadrants" (figure 2.2). The selection of an aggregate score was the first main objective of this chapter and needed to take into account this duality of the SAIN,LIM model. The following analyses kept the structure of the two previous chapters.

7.1 Classification of FFQ-items according to the SAIN,LIM nutrient profiling model

Summary statistics displayed in table 7.1 highlighted that the SAIN,LIM score could not be calculated for the "energy-free" foods (i.e. tea, coffee, and diet fizzy drinks) because the SAIN sub-score is calculated per 100kcal. As expected, the two subscales of the SAIN,LIM model were negatively correlated (r=-0.58). A composite index, (LIM minus SAIN), was created to compare SAIN,LIM to WXYfm. A very high correlation (0.9) was achieved between WXYfm and this composite scale, which indicated a high level of agreement between the two NP models. WXYfm and (LIM minus SAIN) were more associated with the LIM than with the SAIN, suggesting that both WXYfm and SAIN,LIM models depended more on the negative nutrients. Associations obtained with the composite (LIM minus SAIN) score were not further presented as they were very similar to the WXYfm results.

Table 7.1: Summary statistics and rank correlations between WXYfm and SAIN,LIM

	WXYfm	SAIN	LIM	LIM minus SAIN
Summary statistics				
n	130	126	126	126
Mean	4.38	11.7	14.4	2.67
Standard deviation	10.4	17.8	17.8	29.0
Correlations				
WXYfm	1			
SAIN	-0.58	1		
LIM	0.90	-0.58	1	
LIM minus SAIN	0.90	-0.73	0.95	1

The (LIM minus SAIN) score was created for comparison with the WXYfm algorithm

Cross-tabulation results shown in table 7.2 confirmed the correlation coefficients: agreement between the two NP models was high but not perfect. Channel Island, whole, and dry milks were classified as "less healthy" by WXYfm whereas they were in the best quadrant of the SAIN,LIM model, calcium being a positive nutrient in the French algorithm. On the other end, coleslaw, tinned fruit, vegetable soup, and white bread were in the fourth (and worst) quadrant of SAIN,LIM despite being "healthier" according to WXYfm. Foods like milks and white bread were widely consumed and classification differences may have consequences on the aggregate scores rankings of participants. Appendix 2 includes classification of all FFQ-items with both NP models.

Table 7.2: WXYfm categories vs. SAIN,LIM quadrants, number of items

	SAIN,LIM quadrants				
WXYfm categories	1	2	3	4	**Total**
Healthier	44	9	3	4	60
Intermediate	2	1	6	2	11
Less healthy	3	2	17	33	55
Total	49	12	26	39	126

Merging the 2nd and 3rd quadrants of SAIN,LIM, agreement using the weighted κ-statistic was 0.60

7.2 SAIN,LIM aggregate score: PES(Q1)

The aggregate scores algorithms developed for the WXYfm model (chapter 4) were applied to the SAIN,LIM model. Since the SAIN and LIM sub-scores ranked items

on separate scales, the use of an energy-weighted score similar to EWS (which always referred to WXYfm) was only feasible separately, and not for the whole SAIN,LIM model. Two scores, EW(SAIN) and EW(LIM), were created.

In order to combine the two sub-scores, the use of the "quadrants" was necessary. The French food safety agency originally recommended that only foods from the first and healthiest quadrant gained access to health claims (Agence française de sécurité sanitaire des aliments, 2008). As a result, the PES(Q1) aggregate score, which calculated the percentage of energy from foods of the first quadrant, was derived.

In line with the results on food classification, the EW(LIM) was the aggregate score most correlated to EWS (table 7.3). The PES(Q1) was well correlated to all other scores including EWS. The EW(SAIN) was the score least correlated to EWS.

Table 7.3: Rank correlation between EWS and SAIN,LIM-derived aggregate scores (n=8,149)

	PES(Q1)	EW(SAIN)	EW(LIM)	EWS
PES(Q1)	1			
EW(SAIN)	0.63	1		
EW(LIM)	-0.67	-0.66	1	
EWS	-0.70	-0.62	0.91	1

EWS refers to the WXYfm aggregate score

The PES(Q1) aggregate score was the only algorithm to combine the SAIN and LIM scales in one score. It was retained for further analyses. Using a similar protocol to the previous chapters, participants were classified into quartiles of PES(Q1).

7.3 Dietary intakes and PES(Q1), construct validity

7.3.1 Food group intakes across PES(Q1) quartiles

Associations between food group intakes and PES(Q1) were very similar to those observed in chapter 5 (table 7.4). Most associations were highly significant, except

for breakfast cereals and drinks, and revealed a generally healthier pattern of participants in the fourth quartile of PES(Q1).

Some differences with the EWS were noticed. The trend for potatoes, rice and pasta intake was reversed and became negative. The bread, meat, and dairy products trends were stronger with PES(Q1), confirming the classification differences highlighted in section 7.1. This was probably due to the inclusion of calcium and iron in the SAIN algorithm. On the other hand, the trends were weaker for snacks and sweets, and for fat spreads. This last observation may have been due to the inclusion of energy as a negative component in the WXYfm model.

7.3.2 Meat and dairy products intakes across PES(Q1) quartiles

Similarly to table 7.4, levels of intake and trends displayed in table 7.5 followed the results obtained with the EWS aggregate score, particularly for the meat group. For the dairy products, a few differences were observed, notably for whole milk which was positively associated with PES(Q1), in line with its classification in the first quadrant. This resulted in higher intakes of saturated fats from dairy products in the fourth quartile of PES(Q1). Despite this specific trend, certainly due to the inclusion of calcium in the SAIN algorithm, the intake profile of participants in the fourth quartile was generally better. Table 7.5 confirmed that PES(Q1), like EWS, did discriminate participants at the FFQ-item level.

Table 7.4: Crude food group mean intakes across PES(Q1) quartiles (4: healthier)

	Men (n=5,627)					Women (n=2,522)				
Group (g/d)	**1**	**2**	**3**	**4**	**p***	**1**	**2**	**3**	**4**	**p***
Meat products and eggs	136	145	146	148	<.001	112	127	129	137	<.001
Fish and shellfish	30.3	33.6	35.8	39.9	<.001	30.8	36.8	40.1	42.7	<.001
Bread and crackers	135	116	95.1	73.0	<.001	110	91.1	71.0	56.9	<.001
Breakfast cereals	38.2	41.8	42.4	42.2	0.027	37.9	41.5	40.1	39.6	0.729
Potatoes, rice and pasta	208	205	196	181	<.001	194	179	173	156	<.001
Dairy products	298	379	462	835	<.001	309	380	508	1,198	<.001
Meals[#]	27.6	26.9	23.9	19.2	<.001	24.6	22.4	20.1	14.4	<.001
Fat spreads	27.1	21.1	17.2	13.0	<.001	21.9	16.9	13.6	11.9	<.001
Snacks and sweets	141	115	89.9	62.1	<.001	110	81.4	60.9	49.5	<.001
Sauces and other spreads	51.9	45.7	38.8	29.7	<.001	40.5	36.3	28.3	25.1	<.001
Drinks[§]	701	730	719	713	0.282	706	712	706	740	0.416
Fruit and nuts	165	218	261	306	<.001	217	313	351	385	<.001
Vegetables	215	239	250	267	<.001	227	263	288	298	<.001

[#] Meals included quiche, pizza and lasagne. [§]Excluded alcohol and milks. * Heterogeneity ANOVA across quartiles

Table 7.5: Crude meat and dairy products mean intake across PES(Q1) quartiles

	Men (n=5,627)					Women (n=2,522)				
FFQ-item (g/d)	**1**	**2**	**3**	**4**	**p***	**1**	**2**	**3**	**4**	**p***
Beef	22.5	23.1	25.4	25.6	<.001	18.5	20.5	20.7	22.2	0.056
Beefburgers	2.07	1.83	1.47	1.20	<.001	0.98	0.78	0.74	0.80	0.354
Pork	9.48	10.4	10.7	10.2	0.031	8.19	9.05	8.45	10.4	0.012
Lamb	10.6	10.6	10.9	11.2	0.500	10.8	11.0	10.9	11.6	0.771
Chicken	28.5	40.4	45.9	54.9	<.001	30.1	43.5	49.6	58.5	<.001
Bacon	5.45	4.86	4.14	3.47	<.001	3.63	3.55	3.09	2.54	<.001
Ham	6.19	6.83	5.85	5.24	<.001	4.94	4.97	4.43	4.53	0.323
Corn beef, spam, luncheon meats	3.66	3.43	3.04	2.33	<.001	2.82	1.99	1.66	1.49	<.001
Sausages	5.91	4.98	4.10	3.49	<.001	3.43	2.81	2.51	2.21	<.001
Pies	13.0	10.6	7.74	5.30	<.001	6.36	4.50	3.88	2.99	<.001
Liver products	1.67	1.71	1.55	1.53	0.268	1.36	1.45	1.55	1.41	0.636
Meat soup	10.2	11.6	11.8	12.0	0.119	7.68	9.26	9.61	8.93	0.334
Eggs	16.7	15.0	13.2	12.1	<.001	13.9	14.0	12.4	9.45	<.001
SFA from meat (g)	6.16	5.86	5.53	5.28	<.001	4.63	4.69	4.54	4.50	0.579
Sodium from meat (mg)	529	513	458	414	<.001	382	380	353	328	<.001
Sugars from meat (g)	0.51	0.47	0.40	0.34	<.001	0.32	0.29	0.27	0.24	<.001

(Continued)

Table 7.5 (continued)

	Men (n=5,627)					Women (n=2,522)				
FFQ-item (g/d)	**1**	**2**	**3**	**4**	**p***	**1**	**2**	**3**	**4**	**p***
Whole milk	82.6	96.4	122	163	<.001	74.0	83.0	94.7	191	<.001
Semi skimmed milk	123	165	201	387	<.001	120	152	202	570	<.001
Skimmed milk	40.4	55.3	70.4	203	<.001	54.9	67.2	124	295	<.001
Channel Island milk	0.71	1.19	1.83	5.10	0.036	0.42	0.83	2.62	8.47	0.176
Sterilised milk	1.61	2.26	4.88	18.0	<.001	2.19	2.60	4.52	44.8	<.001
Dried milk	0.87	0.86	0.78	0.84	0.840	1.06	1.23	1.40	1.42	0.267
Soya milk	2.18	0.68	1.71	3.06	0.313	2.30	1.19	2.14	3.78	0.620
Coffee whitener	1.78	1.58	1.17	0.77	<.001	1.20	0.83	0.91	0.79	0.052
Single cream	1.51	1.42	1.28	0.90	<.001	1.92	1.26	1.02	0.69	<.001
Double cream	1.40	1.22	1.04	0.75	<.001	1.96	1.24	1.02	0.64	<.001
Yoghurt	20.1	34.3	37.7	47.3	<.001	32.9	49.8	57.9	68.3	<.001
Cheese	19.7	17.0	15.9	12.2	<.001	16.4	15.1	13.4	9.71	<.001
Cottage cheese, fromage frais	2.16	2.26	2.29	2.72	0.076	3.48	4.46	5.10	5.12	0.017
SFA from dairy (g)	9.05	9.51	10.3	13.1	<.001	8.47	8.76	9.53	16.6	<.001
Sodium from dairy (mg)	303	334	373	562	<.001	297	335	400	757	<.001
Sugars from dairy (g)	14.6	19.3	23.3	42.6	<.001	15.8	19.9	26.7	61.1	<.001

SFA, Saturated fatty acids. * Heterogeneity ANOVA across quartiles.

7.3.3 Nutrient intakes across PES(Q1) quartiles

The analysis of crude nutrient intakes (table 7.6) further revealed some similarities with the EWS aggregate score. Energy intake was negatively associated with PES(Q1), as were all fat categories. Protein and most micro-nutrients were positively linked to the aggregate score.

Contrary to the EWS, the energy from carbohydrates was not related to PES(Q1), neither was the intake of thiamine and vitamin A in men. Some significant associations were observed with the PES(Q1) which were absent with the EWS: vitamin D in men, and vitamin A and alcohol in women. The most surprising associations were obtained with cholesterol and fibre which displayed positive and negative trends, respectively. This did not follow initial expectations and was the second result, together with the whole milk intake, which suggested that the PES(Q1) aggregate score might not discriminate healthy dietary patterns as well as EWS.

The differences between the EWS and PES(Q1) aggregate scores did not change the global conclusion that participants in the fourth quartile of PES(Q1) had improved nutrient intakes, which was in line with observations from tables 7.4 and 7.5. The construct of PES(Q1) was deemed valid and the next step was to assess its convergent validity. The PES(Q1) aggregate score did classify participants with respect to their FFQ-item intake, which made it adequate to test for predictive validity of the SAIN,LIM model.

Table 7.6: Crude energy and nutrient mean intakes across PES(Q1) quartiles (4: healthier)

	Men (n=5,627)					Women (n=2,522)				
Nutrient (unit/d)	**1**	**2**	**3**	**4**	**p***	**1**	**2**	**3**	**4**	**p***
Energy (kcal)	2,504	2,369	2,244	2,184	<.001	2,136	2,029	1,935	2,134	<.001
Total fat (%en)	35.8	33.8	32.5	30.3	<.001	35.6	33.3	31.6	30.1	<.001
SFA (%en)	14.4	13.5	13.1	12.4	<.001	14.7	13.3	12.7	12.8	<.001
MUFA (%en)	10.8	10.3	9.90	9.17	<.001	10.6	10.0	9.42	8.84	<.001
PUFA (%en)	7.24	6.76	6.29	5.58	<.001	6.80	6.47	6.01	5.26	<.001
Total carbohydrates (%en)	48.7	48.8	48.6	48.4	0.307	49.3	49.7	49.5	48.9	0.191
Protein (%en)	14.8	16.3	17.5	19.9	<.001	15.9	18.0	19.5	22.1	<.001
Alcohol (%en)	4.21	4.38	4.78	4.72	0.007	2.70	2.43	2.79	2.16	0.012
Sodium (mg)	3,248	3,032	2,776	2,618	<.001	2,779	2,588	2,404	2,525	<.001
Potassium (mg)	3,786	3,990	4,115	4,641	<.001	3,610	3,890	4,118	5,139	<.001
Calcium (mg)	1,055	1,133	1,194	1,595	<.001	1,005	1,090	1,223	2,012	<.001
Magnesium (mg)	370	378	382	410	<.001	334	351	355	424	<.001
Phosphorus (mg)	1,587	1,667	1,712	2,007	<.001	1,471	1,585	1,671	2,286	<.001
Iron (mg)	13.5	13.4	13.1	12.5	<.001	12.2	12.4	12.0	11.8	0.055

(Continued)

Table 7.6 (continued)

	Men (n=5,627)					Women (n=2,522)				
Nutrient (unit/d)	**1**	**2**	**3**	**4**	**p***	**1**	**2**	**3**	**4**	**p***
Vitamin A (μgRE)	1,275	1,279	1,228	1,261	0.245	1,152	1,188	1,228	1,372	<.001
Vitamin D (μg)	4.48	4.64	4.70	4.99	<.001	4.09	4.58	4.82	4.99	<.001
Thiamin (mg)	2.01	2.03	1.97	1.98	0.062	1.82	1.84	1.81	2.00	<.001
Riboflavin (mg)	2.13	2.30	2.40	2.99	<.001	1.98	2.15	2.35	3.50	<.001
Niacin (mgNE)	24.2	25.0	24.8	24.5	0.029	22.1	23.2	23.0	23.4	0.015
Vitamin C (mg)	111	132	148	167	<.001	131	166	186	209	<.001
Vitamin E (mg)	5.51	5.65	5.48	5.40	0.022	5.33	5.61	5.56	5.67	0.071
Vitamin B6 (mg)	2.35	2.51	2.59	2.78	<.001	2.22	2.40	2.50	2.88	<.001
Vitamin B12 (μg)	6.37	7.03	7.40	9.11	<.001	5.95	6.76	7.61	10.2	<.001
Total folic acid (μg)	332	342	344	357	<.001	312	331	344	377	<.001
Panthothenic acid (μg)	5.41	5.85	6.11	7.29	<.001	5.05	5.60	6.05	8.29	<.001
Biotin (μg)	42.8	43.9	45.1	50.5	<.001	37.4	40.2	41.4	53.1	<.001
Cholesterol (mg)	249	250	247	260	0.011	224	230	229	262	<.001
Fibre (g)	26.6	26.4	25.8	24.6	<.001	24.5	25.6	24.8	24.1	0.074

%en: percent of energy intake; SFA, saturated fatty acid; MUFA: Mono-unsaturated fatty acid; PUFA: Poly-unsaturated fatty acid; RE, retinol equivalent; NE, niacin equivalent. * Heterogeneity ANOVA across quartiles.

7.4 Convergent validity of PES(Q1)

Intake levels from tables 7.4 to 7.6 did suggest that the PES(Q1) construct was performing almost as well as EWS, which was weakly associated with the alternative healthy eating index (AHEI) and the dietary clusters (chapter 5). Convergent validity of PES(Q1) was tested against the same measures.

The PES(Q1) aggregate score was poorly correlated to the AHEI, with values of 0.181 and 0.121 for men and women, respectively. This was well illustrated by table 7.7 in which positive and significant—but quite weak—gradients were observed across quartiles of PES(Q1). Quartiles cross-tabulations further strengthened this observation, with a number of participants classified in opposite quartiles (table 7.7), which resulted in poor agreement between the AHEI and PES(Q1). All figures indicated that the PES(Q1) was less related to AHEI than the EWS.

Table 7.7: PES(Q1) and AHEI quartiles cross-tabulation (4: healthier)

	PES(Q1) quartiles							
	Men (n=5,626)				Women (n=2,518)			
	1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Mean AHEI[#]	37.6	40.7	41.9	43.0	40.6	45.3	45.7	44.9
AHEI quartiles[§]								
1	440	325	286	258	213	127	125	154
2	424	378	325	329	146	149	140	138
3	322	358	396	347	174	175	167	147
4	219	346	400	473	97	178	199	189

AHEI, alternative healthy eating index. [#]p<.001 in both sexes for heterogeneity ANOVA across quartiles. [§]Weighted κ-statistic: 0.118 in men, 0.077 in women.

The conclusions were similar for dietary clusters: weak associations in the expected directions, with a number of misclassified individuals (table 7.8). Mean PES(Q1) values by dietary cluster were analogous to EWS, with lower scores obtained in the sweet cluster, and highest ones in the healthy cluster (table 7.8).

Table 7.8: Dietary clusters and PES(Q1) quartiles (4: healthier) cross-tabulation

	Men (n=5,303)					Women (2,301)				
	Quartiles				Mean	Quartiles				Mean
Cluster	1	2	3	4	PES(Q1)*	1	2	3	4	PES(Q1)*
Unhealthy	545	466	460	412	28.5	262	172	167	143	34.2
Sweet	330	265	205	87	25.3	73	37	20	19	30.3
Med.	155	249	280	242	30.2	102	147	113	44	33.9
Healthy	298	354	394	561	32.3	149	229	275	349	40.3

Med. Mediterranean. * p<.001 in both sexes for heterogeneity ANOVA across clusters

The trends displayed in tables 7.7 and 7.8 remained very close to those observed with the EWS aggregate score, and convergent validity of the PES(Q1) was confirmed. The weakest associations obtained between PES(Q1) and AHEI concurred with the slight differences observed for construct validity between the two aggregate scores.

7.5 Non-dietary characteristics and PES(Q1)

7.5.1 Socio-demographic characteristics

Associations between PES(Q1) and age, ethnicity, and employment grade were almost equal to those obtained for the EWS (table 7.9). This was equally true for marital status, but the association was not significant in women.

7.5.2 Health behaviours and self-perception of health

Table 7.9 estimates were also very close to the trends observed with the EWS aggregate score. The health consciousness of participants in the healthier quartiles of the PES(Q1) aggregate score was well illustrated by the two questions relating to health perception. Current smoking was lower in healthier men, but the association was not significant. In women, a U-shaped association was observed for current smoking, the difference between quartiles being marginally significant. As for EWS, physical activity was not related to PES(Q1).

7.5.3 Concurrent validity against risk factors and inflammatory markers

Consequent to observations made above, it was no surprise that most trends displayed in table 7.10 were alike those in chapter 5. Participants in the healthier quartiles of the PES(Q1) aggregate score were more likely to be obese or overweight, and to suffer from hypertension and dyslipidaemia. In both sexes, there was a positive but non-significant association with prevalence of longstanding illnesses. Associations with inflammatory markers were slightly attenuated compared to the EWS.

In summary, the cross-sectional associations obtained for the SAIN,LIM derived PES(Q1) were similar those previously observed with the EWS aggregate score. Construct and convergent validity were confirmed, but associations were weaker than for EWS. Predictive validity was tested next, to assess whether these small differences were reflected in the prospective associations.

Table 7.9: Socio-demographic characteristics and health behaviours across PES(Q1) quartiles (4: healthier)

	Men					Women				
Variable (mean or %)	**1**	**2**	**3**	**4**	**p***	**1**	**2**	**3**	**4**	**p***
Age (y)	48.5	49.0	49.7	50.2	<.001	49.6	50.0	50.7	51.4	<.001
% living alone[§]	18.6	15.9	17.9	16.9	0.237	43.2	35.5	32.8	35.0	0.001
Ethnicity (% white)	96.6	95.5	94.6	87.7	<.001	95.4	90.4	83.5	76.4	<.001
Grade (% high)	20.1	22.1	25.6	21.8		6.27	8.47	5.93	2.92	
Grade (% intermediate)	73.4	71.9	69.7	69.6	<.001	60.0	57.3	53.5	47.6	<.001
Grade (% low)	6.50	6.03	4.67	8.62		33.8	34.2	40.5	49.5	
% never smoker	47.9	48.8	48.5	48.9		51.9	55.2	60.5	55.8	
% ex-smoker	36.8	38.7	40.0	39.4	0.067	28.2	28.0	25.8	25.8	0.050
% current smoker	15.2	12.5	11.5	11.7		19.9	16.8	13.7	18.4	
METs[#]	3.93	3.92	3.97	3.94	0.975	3.37	3.37	3.36	3.06	0.206
% inactive	62.9	63.9	62.2	63.0	0.830	72.4	73.5	71.8	75.6	0.447
% Agree strongly "Keeping healthy depends on me"	27.1	32.8	34.6	40.4	<.001	32.1	36.3	40.7	39.3	<.001
% "Health is extremely important"	39.0	40.7	42.6	49.0	<.001	45.7	44.2	51.8	55.1	0.003

[§] Never married/cohabiting, divorced, or widowed . [#] Metabolic equivalents. * Heterogeneity ANOVA or χ^2

Table 7.10: Risk factors and inflammatory markers levels across PES(Q1) quartiles (4: healthier)

	Men					Women				
Variable (Mean or %)	**1**	**2**	**3**	**4**	**p***	**1**	**2**	**3**	**4**	**p***
BMI (kg/m^2)	24.8	25.0	25.2	25.6	<.001	25.2	25.2	25.9	26.5	<.001
% underweight	4.50	3.43	3.35	2.94		9.31	5.88	4.55	4.61	
% normal weight	53.1	52.1	49.3	42.0	<.001	45.9	50.9	47.0	37.9	<.001
% overweight	35.9	37.4	40.3	47.2		32.3	28.9	31.6	38.9	
% obese	6.49	7.01	7.00	7.84		12.5	14.3	16.8	18.6	
Systolic blood pressure (mmHg)	121	122	122	123	0.003	117	116	119	119	<.001
% Hypertension[#]	19.2	22.1	21.9	25.1	0.004	15.5	13.9	20.2	22.4	<.001
Cholesterol - Total (mmol/L)	6.41	6.43	6.50	6.58	<.001	6.49	6.49	6.50	6.62	0.190
Cholesterol - LDL (mmol/L)	4.39	4.41	4.46	4.52	0.006	4.28	4.25	4.26	4.38	0.157
Cholesterol - HDL (mmol/L)	1.33	1.32	1.32	1.32	0.748	1.69	1.69	1.69	1.67	0.854
Triglycerides (mmol/L)	1.53	1.59	1.63	1.67	0.018	1.18	1.23	1.22	1.25	0.462
% Dyslipidaemia	59.5	60.8	63.0	64.9	0.024	54.3	52.3	50.4	57.4	0.100
% longstanding illness	31.8	33.3	35.8	34.6	0.136	32.6	35.2	35.1	36.5	0.541
Fibrinogen (g/L)	2.32	2.33	2.33	2.39	0.007	2.58	2.56	2.58	2.65	0.087
Von Willebrand's factor (IU/dl)	105	109	106	109	0.055	106	108	110	115	0.003
C-reactive protein (mg/L)	1.94	1.66	1.67	1.84	0.300	2.12	2.12	2.43	2.34	0.488
Interleukin-6 (ng/L)	1.79	1.74	1.81	1.90	0.296	2.21	2.26	2.19	2.38	0.617

[#] Hypertension was defined as systolic or diastolic blood pressure ≥ 140 or ≥ 90 mmHg, respectively, or by the use of hypertensive drugs.

* Heterogeneity ANOVA or χ^2

7.6 Predictive validity of PES(Q1), similar U-shapes

The Cox regression models implemented for the PES(Q1) aggregate score kept the same specifications as those for the EWS aggregate score. Prospective associations between PES(Q1) quartiles and CHD, diabetes, cancer mortality, and all-cause mortality were investigated with three levels of adjustment (see chapter 6 for more details). PES(Q1) was associated with sex (Cochran-Armitage trend test p-value was below 0.001) and Cox models were adjusted accordingly. The reference quartile remained the first, and least healthy, one.

Table 7.11 presents hazard ratio estimates for all models. Similarly to EWS, some significant quadratic trends were observed while the null-hypothesis could not be rejected for the linear trend tests. Yet, table 7.11 suggested that the PES(Q1) was more strongly related to prospective incidence of health outcomes than the EWS. Significant associations showed risk reduction and were relatively robust to adjustment.

Borderline significant linear risk reduction of cancer mortality, and all-cause mortality to a lesser extent, was observed in the fully-adjusted model. Despite the non-significance, it was suggested that participants in the fourth quartile were less at risk than the reference group. Also, there was a significant risk reduction of all-cause mortality robust to adjustment within participants in the second quartile of PES(Q1).

No linear risk increase of CHD was observed. The quadratic trend was highly significant and the third quartile was associated with a significant risk reduction in models 1 and 3.

There was a suggestion of a linear risk increase for diabetes in models 1 and 2; the trend was attenuated and non-significant in model 3.

An interesting aspect of the three models, already observed with the EWS, was that the adjustment of model 3 tended to bring down the estimates, especially the fourth

quartile ones. This led to a slight attenuation of the quadratic trends, which remained significant for all outcomes except diabetes.

Table 7.11: Cox regression estimates for PES(Q1) quartiles (4: healthier)

Outcome	Quartile /trend	Model 1 HR	95 % CI		Model 2 HR	95 % CI		Model 3 HR	95 % CI	
CHD (318 / 7,174)	1	Ref			Ref			Ref[#]		
	2	0.80	0.58	1.09	0.83	0.60	1.13	0.79	0.58	1.09
	3	**0.71**	**0.51**	**0.98**	0.74	0.53	1.04	**0.71**	**0.51**	**0.99**
	4	1.23	0.91	1.67	1.27	0.94	1.73	1.21	0.89	1.64
	Linear	1.06	0.95	1.17	1.07	0.96	1.19	1.05	0.95	1.17
	p quadratic	**0.002**			**0.010**			**0.010**		
Diabetes (754 / 6,868)	1	Ref*			Ref*			Ref*		
	2	0.88	0.71	1.10	0.92	0.74	1.14	0.90	0.72	1.11
	3	1.01	0.82	1.25	1.07	0.87	1.32	1.02	0.82	1.25
	4	1.11	0.90	1.38	1.15	0.93	1.42	1.06	0.85	1.31
	Linear	1.05	0.98	1.12	1.06	0.99	1.14	1.03	0.96	1.10
	p quadratic	0.249			0.704			0.801		
Cancer mortality (251 / 7,235)	1	Ref			Ref			Ref		
	2	0.80	0.56	1.13	0.78	0.55	1.11	0.79	0.55	1.11
	3	0.76	0.53	1.08	0.74	0.52	1.06	0.73	0.51	1.05
	4	0.73	0.51	1.06	0.71	0.49	1.03	0.69	0.48	1.01
	Linear	0.90	0.80	1.02	0.90	0.79	1.01	0.89	0.79	1.00
	p quadratic	**0.029**			**0.022**			**0.027**		
All-cause mortality (524 / 7,242)	1	Ref			Ref			Ref[§]		
	2	**0.71**	**0.56**	**0.92**	**0.73**	**0.57**	**0.93**	**0.72**	**0.56**	**0.92**
	3	0.86	0.68	1.09	0.88	0.69	1.12	0.87	0.68	1.10
	4	0.80	0.62	1.03	0.81	0.63	1.05	0.79	0.61	1.02
	Linear	0.95	0.88	1.03	0.96	0.88	1.04	0.95	0.87	1.03
	p quadratic	**0.011**			**0.029**			**0.035**		

Model 1 adjusted for age, sex, and ethnicity. Model 2 adjusted for marital status, employment grade, smoking status, physical activity level, and energy and alcohol intake. Model 3 adjusted for BMI, hypertension and dyslipidaemia status, and prevalence of longstanding illness
HR, hazard ratio; CI, confidence interval. * Stratified for sex. [#] Stratified for BMI categories.
[§] Stratified for longstanding illness and dyslipidaemia.

7.7 Discussion

The results from this chapter largely followed ones from chapters 5 and 6. The PES(Q1) construct was shown to be valid. Associations with existing measures of dietary quality were significant and in the expected direction, but weaker than for the EWS aggregate score. Concurrent validity could not be confirmed, with a riskier profile for participants classified as healthiest by the PES(Q1). Similarly to the EWS aggregate score, survival analyses estimates displayed strong quadratic trends and no significant linear risk reduction. Based on the PES(Q1) aggregate score results, predictive validity of the SAIN,LIM model could not be established despite significant risk reduction of CHD and all-cause mortality observed for participants in the middle quartiles of the PES(Q1).

The comparison of results obtained with EWS and PES(Q1) were subject to a methodology limitation. The aggregation algorithms used for the two NP models were not exactly similar because of the dual nature of the SAIN,LIM model. Unlike EWS, the PES(Q1) aggregate score did not take into account the exact SAIN and LIM scores obtained by each FFQ-item. In section 7.1, a composite NP score, the "LIM minus SAIN", was created for comparison purposes. An aggregate score, EW(LIM-SAIN), was derived from this composite NP model. Results were very similar to the EWS aggregate score (not shown) and confirmed the observations made with PES(Q1). The EW(LIM-SAIN) aggregate score was not retained for further analyses.

The implementation of the SAIN,LIM model was done to assess whether the predictive validity results obtained for WXYfm, which did not follow the hypothesis, could be explained by the NP model itself. The very similar conclusions drawn with SAIN,LIM, via the PES(Q1) aggregate score, suggested that the above assumption could be refuted. However, the WXYfm and SAIN,LIM models ranked foods relatively similarly and were both across-the-board algorithms. Therefore, both models could be similarly flawed in ways that would entail the U-shaped associations. It was not possible to apply another NP model within the timeframe of the project. As highlighted in chapter 6, other factors could have explained the predictive validity results. The following chapter focuses on these factors.

Chapter 8: Explaining the quadratic trends

The predictive validity hypothesis that diets containing higher content of healthy foods would be protective against prospective chronic disease was not fully confirmed by the two previous chapters. Several factors could have impacted and biased the estimates obtained for WXYfm and SAIN,LIM, compromising the predictive validity results.

The goal of this chapter was to explore in detail the potential impact of such factors to obtain a better understanding of the U-shaped associations. Three factors were investigated: (i) misreporting of dietary intakes; (ii) dietary variety; and (iii) nutrient profiling (NP) models' components and respective aggregate scores. The analysis of these three factors was done in three separate sections. The hypotheses, specific methods, and results were all presented in the respective sections. A global discussion gathered conclusions for all three factors.

Given the similarity of the results obtained with the two NP models, only the results for WXYfm and EWS are presented in the main text. Results for the SAIN,LIM model and the PES(Q1) aggregate score are reported in appendix 3.

8.1 Dietary misreporting in the Whitehall II cohort and implication on health outcomes

Dietary misreporting was previously shown to be relatively common among Whitehall II participants and to be related to some risk factors (chapter 2). Chapters 5 and 7 suggested that energy misreporting was systematically associated with both aggregate scores. Therefore, it may have confounded the prospective associations obtained in chapters 6 and 7. The aim of this section was to apply more refined methods to detect and measure dietary misreporting in order to verify whether misreporting could have entailed the U-shaped associations.

First, energy misreporting which relates to global food intake (i.e. some participants reporting less—or more—foods than actually consumed) was assumed to be associated with the aggregate scores rankings. The Goldberg cut-off method was used to determine whether an individual was a low, acceptable, or high energy reporter. Sensitivity analyses excluding misreporters (i.e. low and high energy reporters) were run to assess the effect of energy misreporting on the predictive validity results.

Second, the direct comparison of FFQ and 7-day diet diary (7DD) intakes allowed to verify the hypothesis that foods commonly considered as less healthy are under-reported with the FFQ, while the healthy foods are over-reported. If confirmed, such differential misreporting of FFQ-items may have further influenced the aggregate scores. As a result, the regression calibration technique was applied using the 7DD data as reference measure to obtain predicted, or corrected, intakes for each FFQ-item. Corrected aggregate scores were derived from these predicted intakes and included in Cox models. It was assumed that the corrected aggregate scores would reflect unhealthier diets as put forward by the differential misreporting of foods hypothesis. The corrected aggregate scores would result in new rankings of individuals, and potentially very different prospective associations with incident health outcomes.

The Goldberg and regression calibration methods are presented first as they were specifically used for misreporting analyses. All the results are presented for the WXYfm-derived EWS aggregate score only. Appendix 3 includes the results for the PES(Q1) aggregate score.

8.1.1 Methods, Goldberg cut-off and regression calibration

(i) Dietary data

Detailed presentation of the Whitehall II data was done in chapter 4. This section includes a brief summary of dietary assessment tools used for the Goldberg cut-offs and regression calibration models.

a. Food frequency questionnaire

FFQ intakes were available for 8,225 participants. The main issue regarding FFQ data was missing values. All the following analyses only included participants with less than 10 missing items. For the regression calibration model which assumes normality of distributions, all missing values were set to 0.001 and intakes were log-transformed.

b. 7-day diet diary

At phase 3, participants were given a 7-day diary (7DD) at the clinic, with the instruction to complete it at home and send it back with the provided envelope. 6,726 diaries were received. To date, 1,350 diaries have been coded by the Whitehall II study team.

In order to match the 7DD data with FFQ-items, foods reported in the diet diaries were regrouped into items corresponding to the FFQ ones. Hence, some foods recorded in the 7DD were not used in the analyses as there was no FFQ equivalent (e.g. condiments and spices). For coffee, tea, drinking chocolate and Horlicks, some rescaling was needed in order to fit the FFQ data. Mean daily intake of each "7DD-item" was obtained by dividing total intake by number of days recorded. Participants with less than 5 completed days were excluded from the analyses. For foods with no reported consumption, intakes were set to 0.001. Intakes were log-transformed for the regression calibration model.

(ii) Detection of energy misreporters using the Goldberg cut-off method

This technique is based on the fundamental equation that energy expenditure equals energy intake when body weight is constant (chapter 2). It uses calculated basal metabolic rate (BMR) available with anthropometric measurements of Whitehall II participants and estimated physical activity derived from the general questionnaire. The reported energy intake (EI_{rep}) of a participant is validated by defining an acceptable range for the EI_{rep}/BMR ratio (Goldberg *et al.*, 1991) given by the equation below:

$$PAL \times \exp\left(u_{min} \times \frac{F/100}{\sqrt{n}}\right) < EI_{rep}/BMR < PAL \times \exp\left(u_{max} \times \frac{F/100}{\sqrt{n}}\right)$$

With PAL the physical activity level category of the individual; u_{min}=-1.96 and u_{max}=1.96 for the 95% confidence limit of a normal distribution; n the number of subjects, here n=1 as the technique was applied individually. F is the factor that accounts for the individual variation in intake, BMR and energy requirements, and is given by:

$$F = \sqrt{\frac{CV_{wEI}^2}{d} + CV_{wB}^2 + CV_{tP}^2}$$

Where CV_{wEI} is the within-subject coefficient of variation in energy intake, d is the number of days of diet assessment, CV_{wB} is the coefficient of variation of estimated versus measured BMR, and CV_{tP} is the individual's day-to-day variation in PAL (physical activity is assumed to vary on a daily basis).

The Goldberg technique has been previously implemented in the Whitehall II data using a single PAL category for the definition of acceptable reporters, which might have led to some misclassification (chapter 2). Black investigated the validity of the Goldberg cut-offs at the individual level (Black, 2000b). She concluded that misclassification of low and high energy reporters was minimised when dividing the population into three categories of PAL. This approach was retained for the present analysis, and the terms in the Goldberg equation were derived as follows:

- EI_{rep} was obtained from the phase 3 FFQ.
- The BMR was calculated using 1991 Committee on Medical Aspects of Food Policy (COMA) equations (Department of Health, 1991).

- Classification of individuals into three physical activity level (PAL) categories was done using the 1991 COMA recommendations. Since Whitehall II participants were all working in civil service offices, the occupational category "light" was used. Non-occupational physical activity was derived from the phase 3 questionnaire (participants reported their average hours per week of mild, moderate and vigorous activities).
- The parameters included in the F factor have been estimated in several studies (FAO/WHO/UNU, 1985; Black, 2000a). The FFQ assessing usual dietary intake, the CV_{wEI}/d term was set to 0. CV_{wB} and CV_{tP} were set to 8.5% and 15%, respectively, using Black's recommendations.

The Goldberg cut-offs were computed for these values of CV_{wB}, CV_{tP} and for each PAL category depending on the participant's reported energy expenditure. The respective cut-off values are given in table 8.1.

Table 8.1: PAL categories and associated Goldberg cut-offs for use at phase 3

PAL category (value)	**EI_{rep}/BMR cut-off values**				
		Lower		**Upper**	
Mild (1.4)	Low energy reporter	0.999	Acceptable reporter	1.96	High energy reporter
Moderate (1.5)		1.07		2.10	
Heavy (1.6)		1.14		2.24	

EI_{rep}, reported energy intake; BMR, basal metabolic rate; PAL, physical activity level (target value for the EI_{rep}/BMR ratio).

(iii) Predicting individual FFQ-items true intake, regression calibration

A case study conducted on fruit and vegetables indicated that regression calibration was the most appropriate method when only one alternative to the FFQ was available (appendix 7). 7DD data were used as the alternative measure, and the same algorithm was applied to all FFQ-items.

a. Terminology and notations

Let T_{ij} denote the true intake of food i in participant j, which cannot be observed. The food record (7-day diary), R_{ij} , and FFQ, Q_{ij} , are two surrogate measures of T_{ij} and are measured with some error. The 7DD data is considered as the closest measure to true food intake and follows a simple random error model:

$$R_{ij} = T_{ij} + \varepsilon_{Rij} \quad [1]$$

Where the errors are independent of T_{ij} and of each other and are normally distributed with a mean of zero (i.e. $\varepsilon_{Rij} \sim N(0, \sigma^2)$).

Food frequency questionnaires are likely to be biased measures of true intake. Therefore, FFQ measures are assumed to follow the linear model defined below:

$$Q_{ij} = \alpha_{Qi} + \beta_{Qi}T_{ij} + \varepsilon_{Qij} \quad [2]$$

Where ε_{Qij} have the same properties as above; α_{Qi} is the systematic bias and β_{Qi} is the scaling bias of the FFQ, for food i.

In order to estimate the systematic and scaling bias parameters, the FFQ variable of interest must be regressed on another reference measure following model [1].

b. Regression calibration, Rosner & Gore method

Regression calibration uses 7DD as this reference measure, true intake is represented by the diet diary reported value and equation [2] becomes:

$$Q_{ij} = \alpha_{Qi} + \beta_{Qi}R_{ij} + \varepsilon_{Qij} \quad [3]$$

Under the strong assumption that random errors of both methods are not correlated, i.e. errors in the FFQ and 7DD are independent:

$$\begin{cases} \mathrm{cov}(\varepsilon_{Rij}, T_{ij}) = 0 \\ \mathrm{cov}(\varepsilon_{Rij}, \varepsilon_{Qij}) = 0 \end{cases} \quad [4]$$

Our goal was to predict the diet diary value (representing true intake) for all participants, including those not in the validation sub-sample, using regression calibration estimates from the validation sub-sample. This was achieved by implementing a linear model between FFQ and 7DD reported values in the validation sub-sample.

Our model followed the approach developed by Rosner and Gore. It included non-dietary covariates (Z_{ij} and the associated γ_i regression parameters) associated to both true and reported intake (sex, age, BMI, physical activity, employment grade, ethnicity, and smoking status) as well as all FFQ-items (Q_{ij}) as potential predictors of the diet diary value (Rosner & Gore, 2001). This followed observations that some FFQ-items were more associated to a diet diary intake than the respective FFQ-item (e.g. FFQ hamburger was a better predictor of diet diary chips than FFQ chips). The model was, for food i and participant j:

$$R_{ij} = \lambda_{0i} + \Sigma\, \lambda_{REGi} Q_{ij} + \Sigma\, \gamma_i Z_{ij} + \varepsilon_{ij} \quad [5]$$

A stepwise selection of FFQ intake variables was implemented to retain only FFQ-items (Q_{ij}) which contributed significantly to the model ($p<0.01$), with all non-dietary covariates (Z_{ij}) forced in the model. Once parameter estimates were obtained using the general least squares method, the predicted true (diet diary) intakes ($\hat{R}_{ij}$) could be calculated in the whole population:

$$\hat{R}_{ij} = \hat{\lambda}_{01i} + \Sigma \hat{\lambda}_{REGi} Q_{ij} + \Sigma \hat{\gamma}_i Z_{ij} \quad [6]$$

As all models were linear, variables were log-transformed to reach distributions closer to the normal one.

8.1.2 Results

(i) Goldberg cut-off, energy misreporting

a. Distribution of energy misreporters and association with dietary patterns

The implementation of the Goldberg cut-off technique identified 5,884 out of 8,033 (73.3%) participants who reported intake within the "acceptable" range (table 8.2), in line with previous observations (chapter 2). Compared to women, more men were low energy reporters, yet more were in the acceptable range. The almost 30% of participants which misreported their intake may have had a strong influence on the relationship between aggregate scores and health outcomes.

Table 8.2: Energy misreporting among Whitehall II participants (column %)

Reporting level	Men (n=5,561)	Women (n=2,472)
Under	21.5	14.3
Acceptable	73.6	72.5
Over	4.9	13.2

Table 8.3 highlighted that energy reporting levels were highly related to intake of most food groups. Generally, there was a positive association between higher energy reporting and energy dense groups like sweets and snacks or spreads. The trend was negative for food groups with low energy density, namely drinks, and fruit and vegetables; and for meat, fish, and potatoes, rice and pasta. This was consistent with previous observations in which low energy reporters tended to over-report foods commonly considered as healthy, and to under-report foods considered as less healthy (Macdiarmid & Blundell, 1998; Livingstone & Black, 2003).

The associations observed in table 8.3 were translated at the nutrient level, as shown in table 8.4. Overall, high energy reporters retrieved more energy from fats except cholesterol, and less from protein and alcohol, in line with previous findings (Livingstone & Black, 2003). The association with carbohydrates was weak. Intake of micro-nutrients was generally higher among low energy reporters, confirming the trends observed for fruit and vegetables, meat, and fish. The very strong positive gradient observed for energy intake confirmed that classification of participants into energy reporting levels depended mainly on reported food intake.

Low energy reporters appeared to have healthier diets (tables 8.3 and 8.4), which resulted in a systematic inverse association between energy reporting level and the EWS aggregate score (table 8.5). The reported intakes of participants classified in the fourth quartile of EWS were therefore more likely not to represent their true intake and energy misreporting would have led to misclassification of participants, in line with the expectations.

Results were similar for the PES(Q1) aggregate score, and confirmed the hypothesis that energy misreporting could have resulted in misclassification of participants (appendix 3).

Table 8.3: Food group intake by reporting level

Food group (g/2,000kcal)	Men				Women			
	Under	**Acceptable**	**Over**	**p***	**Under**	**Acceptable**	**Over**	**p***
n	1,195	4,093	273		354	1,791	327	
Meat products and eggs	142	123	107	<.001	144	128	102	<.001
Fish and shellfish	37.3	29.4	26.0	<.001	42.5	37.6	31.8	<.001
Bread and crackers	75.2	93.7	83.0	<.001	76.5	82.4	75.4	0.029
Breakfast cereals	37.6	36.0	33.1	0.167	30.2	39.6	38.0	0.001
Potatoes, rice and pasta	194	169	154	<.001	197	175	155	<.001
Dairy products	409	410	534	<.001	453	498	842	<.001
Meals[#]	19.7	21.3	21.1	0.068	19.7	21.0	16.1	0.001
Fat spreads	13.4	17.3	17.0	<.001	13.5	15.7	15.8	0.004
Snacks and sweets	67.5	88.0	107	<.001	61.2	70.8	84.1	<.001
Sauces and other spreads	29.1	36.2	38.9	<.001	26.9	32.2	32.4	0.001
Drinks[§]	751	617	502	<.001	921	731	584	<.001
Fruits and nuts	253	201	181	<.001	365	319	265	<.001
Vegetables	264	206	176	<.001	356	272	205	<.001

[#] Meals included quiche, pizza and lasagne. [§]Excluded alcohol and milks. *Heterogeneity ANOVA across reporting levels.

Table 8.4: Nutrient densities by reporting level

Nutrient (unit/2,000kcal)[#]	Men				Women			
	Under	Acceptable	Over	p*	Under	Acceptable	Over	p*
Energy (crude kcal)	1,547	2,446	3,905	<.001	1,192	1,999	3,317	<.001
Total fat (%en)	31.4	33.4	35.8	<.001	30.9	32.6	34.5	<.001
SFA (%en)	12.4	13.5	15.2	<.001	12.2	13.3	15.2	<.001
MUFA (%en)	9.71	10.1	10.7	<.001	9.29	9.75	9.96	<.001
PUFA (%en)	6.15	6.55	6.62	<.001	5.98	6.19	6.01	0.116
Total carbohydrates (%en)	48.2	48.8	48.2	0.011	49.7	49.4	49.1	0.569
Protein (%en)	18.2	16.9	16.3	<.001	19.7	18.7	18.4	<.001
Alcohol (%en)	5.59	4.32	3.00	<.001	3.17	2.62	1.33	<.001
Sodium (mg)	2,447	2,516	2,521	<.001	2,413	2,511	2,484	0.002
Potassium (mg)	3,942	3,525	3,389	<.001	4,422	4,065	4,012	<.001
Calcium (mg)	1,061	1,055	1,182	<.001	1,161	1,212	1,545	<.001
Magnesium (mg)	354	331	314	<.001	373	357	347	<.001
Phosphorus (mg)	1,551	1,490	1,520	<.001	1,673	1,665	1,808	<.001
Iron (mg)	11.9	11.4	10.3	<.001	12.7	12.2	10.5	<.001

(Continued)

Table 8.4 (continued)

Nutrient (unit/2,000kcal)[#]	Men				Women			
	Under	Acceptable	Over	p*	Under	Acceptable	Over	p*
Vitamin A (μgRE)	1,095	1,078	1,105	0.497	1,356	1,188	1,157	<.001
Vitamin D (μg)	4.29	4.05	3.65	0.001	4.84	4.62	3.98	0.001
Thiamin (mg)	1.80	1.72	1.63	<.001	1.88	1.83	1.75	<.001
Riboflavin (mg)	2.15	2.09	2.19	0.003	2.26	2.33	2.71	<.001
Niacin (mgNE)	23.2	21.2	19.0	<.001	24.9	23.0	19.4	<.001
Vitamin C (mg)	141	119	110	<.001	204	173	146	<.001
Vitamin E (mg)	5.07	4.72	4.57	<.001	5.96	5.53	4.80	<.001
Vitamin B6 (mg)	2.43	2.19	2.03	<.001	2.57	2.46	2.33	<.001
Vitamin B12 (μg)	6.81	6.34	6.55	<.001	7.67	7.21	7.74	0.009
Total folic acid (μg)	325	295	271	<.001	365	337	308	<.001
Panthothenic acid (μg)	5.64	5.26	5.28	<.001	6.13	5.95	6.36	<.001
Biotin (μg)	41.9	38.9	39.0	<.001	43.1	41.4	42.2	0.021
Cholesterol (mg)	231	214	213	<.001	246	230	220	<.001
Fibre (g)	23.4	22.4	20.3	<.001	26.1	25.1	21.2	<.001

%en, percent of energy intake; SFA, saturated fatty acid; MUFA, Mono-unsaturated fatty acid; PUFA, Poly-unsaturated fatty acid; RE, retinol equivalent; NE, niacin equivalent.

[#] Except energy intake. *Heterogeneity ANOVA across reporting levels

Table 8.5: Energy misreporting across EWS quartiles (4: healthier)

	Men				Women			
Column %	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
% Under	12.6	16.0	22.7	34.8	8.70	11.7	16.1	21.1
% Acceptable	79.5	79.4	73.2	62.3	75.0	76.9	70.3	67.3
% Over	7.87	4.60	4.17	2.96	16.3	11.5	13.6	11.5

χ^2 $p<0.001$ for both sexes

b. Reporting level and non-dietary characteristics of participants

Acceptable energy reporters were more likely to be never smokers, of white ethnicity, and of high employment grade (table 8.6). In line with previous observations (Brunner, 1997; Stallone *et al.*, 1997), there was a strong inverse association between energy reporting level and high BMI, with most obese and overweight participants being under-reporters. There was an inverse association with physical activity, significant in women only. Associations with blood pressure and blood lipids were significant in men and suggested a better profile for acceptable and over-reporters. Low energy reporters had therefore a less favourable risk profile which was consistent with previous findings (Macdiarmid & Blundell, 1998).

Low energy reporters being more likely to be *mis*classified in the EWS fourth quartile (table 8.5), the association between energy misreporting and higher prevalence of vascular risk would have confounded the prospective associations between EWS and chronic disease. In line with the concurrent validity results in chapter 5, the increased risk factors levels among under-reporters classified in the fourth quartile of EWS would have entailed the U-shaped associations. Conclusions were similar for the PES(Q1) aggregate score (appendix 3) and sensitivity analyses were run to test such assumption.

Table 8.6: Non-dietary characteristics of participants by energy reporting level

	Men				Women			
Mean or %	**Under**	**Acceptable**	**Over**	**p***	**Under**	**Acceptable**	**Over**	**p***
Age (y)	49.5	49.3	49.1	0.513	50.4	50.2	51.3	0.020
% living alone	21.3	16.1	17.7	<.001	36.2	36.1	39.8	0.438
Ethnicity (% white)	87.3	95.5	93.0	<.001	80.0	88.2	86.7	<.001
Grade (% high)	17.3	24.2	18.8		4.49	6.65	3.06	
Grade (% intermediate)	73.4	70.5	70.8	<.001	50.6	57.6	44.0	<.001
Grade (% low)	9.29	5.28	10.3		44.9	35.8	52.9	
% never smoker	42.7	50.3	48.1		51.1	57.2	53.9	
% ex-smoker	41.0	37.9	41.0	<.001	27.9	27.2	24.5	0.015
% current smoker	16.2	11.8	10.8		21.0	15.6	21.7	
BMI (kg/m^2)	26.4	24.8	23.7	<.001	27.4	25.5	24.8	<.001
% underweight	2.02	3.68	7.98		3.31	5.92	10.0	
% normal weight	35.9	52.2	63.9	<.001	35.2	47.5	46.0	<.001
% overweight	47.8	38.8	25.5		35.2	32.5	34.6	
% obese	14.22	5.27	2.66		26.2	14.0	9.39	
% inactive	62.2	62.5	66.7	0.359	68.8	72.8	79.2	0.008
Mets[§]	3.85	3.95	3.72	0.299	3.24	3.34	2.86	0.030

(Continued)

Table 8.6 (continued)

	Men				Women			
Mean or %	**Under**	**Acceptable**	**Over**	**p***	**Under**	**Acceptable**	**Over**	**p***
% Hypertension[#]	26.4	20.9	21.2	<.001	20.1	17.7	16.5	0.467
Systolic BP (mmHg)	123	122	121	0.002	118	118	117	0.853
% Dyslipidaemia	65.0	61.7	54.5	0.005	50.9	53.0	58.2	0.152
Cholesterol - Total (mmol/L)	6.58	6.46	6.27	<.001	6.46	6.52	6.63	0.194
Cholesterol - LDL (mmol/L)	4.51	4.44	4.27	0.001	4.22	4.28	4.41	0.086
Cholesterol - HDL (mmol/L)	1.28	1.33	1.38	<.001	1.66	1.69	1.67	0.493
Triglycerides (mmol/L)	1.76	1.56	1.42	<.001	1.32	1.20	1.21	0.029
% longstanding illness	31.8	34.3	33.7	0.267	36.0	34.5	36.1	0.765

[#] Hypertension was defined as systolic or diastolic blood pressure ≥ 140 or ≥ 90 mmHg, respectively, or by the use of hypertensive drugs.
[§] Metabolic equivalents. * Heterogeneity ANOVA or χ^2 test.

c. Reporting levels and prospective health outcomes, sensitivity analysis

Sensitivity analyses were conducted to assess the effect of misreporting on the aggregate score—health outcome relationship. Cox regression models similar to chapters 6 and 7 'model 3' were run excluding energy misreporters. Table 8.7 contains the parameter estimates for the EWS aggregate score.

Compared to the original model, the exclusion of misreporters led to attenuated and non-significant U-shapes for CHD and all-cause mortality. The linear trends remained non-significant but with lower hazard ratio estimates. For diabetes, a significant quadratic trend was observed. A linear risk reduction was suggested for cancer mortality. For all outcomes, hazard ratio estimates of the fourth quartile and the linear trend were lower when including acceptable reporters only. This strongly suggested that the higher proportion of low energy reporters misclassified in the fourth quartile of EWS confounded the prospective associations obtained in chapter 6. The quadratic trend would have been explained by the higher risk factor levels among low-energy reporters.

However, the exclusion of energy misreporters led to a smaller sample size, slightly wider confidence intervals, and non-significant estimates. The absence of linear and protective trends could be further explained by the more homogeneous sample obtained once energy misreporters were excluded. It was therefore not possible to conclude more categorically on the effect of energy misreporting. A method using the full sample should allow obtaining more robust estimates.

For the PES(Q1) aggregate score, most quadratic trends were also attenuated when excluding energy misreporters (appendix 3). Estimates changes were smaller, suggesting that the aggregate score was less sensitive to misreporting.

Table 8.7: Hazard ratio estimates for sensitivity analyses excluding energy misreporters, EWS quartiles (4: healthier)

Outcome, cases/total (numbers for acceptable reporters only)		Model 3 (chapter 6)			Model 3, acceptable reporters only		
		HR	95	% CI	HR	95	% CI
CHD, 318 / 7,174 (220 / 5,263)	1	Ref			Ref[#]		
	2	0.82	0.58	1.15	0.86	0.59	1.27
	3	1.03	0.75	1.41	1.15	0.80	1.67
	4	1.22	0.89	1.69	1.18	0.80	1.75
	Linear	1.09	0.98	1.21	1.08	0.95	1.23
	p quadratic trend	**0.003**			0.061		
Diabetes, 754 / 6,868 (511 / 5,060)	1	Ref*			Ref*		
	2	1.00	0.81	1.24	0.88	0.69	1.12
	3	0.89	0.72	1.10	0.83	0.65	1.07
	4	1.04	0.84	1.28	1.01	0.78	1.31
	Linear	1.00	0.93	1.07	0.99	0.91	1.08
	p quadratic trend	0.402			**0.048**		
Cancer mortality, 251 / 7,235 (185 / 5,309)	1	Ref			Ref		
	2	0.94	0.65	1.35	0.98	0.66	1.47
	3	0.95	0.66	1.36	0.95	0.63	1.42
	4	0.87	0.60	1.26	0.75	0.48	1.19
	Linear	0.96	0.85	1.08	0.92	0.80	1.06
	p quadratic trend	0.057			0.697		
All-cause mortality, 524 / 7,242 (372 / 5,312)	1	Ref			Ref[§]		
	2	0.85	0.66	1.09	0.88	0.66	1.17
	3	0.86	0.67	1.11	0.84	0.63	1.13
	4	0.97	0.76	1.25	0.92	0.68	1.24
	Linear	0.99	0.92	1.08	0.97	0.88	1.07
	p quadratic trend	**0.004**			0.137		

Model 3 adjusted for age, sex, ethnicity, marital status, employment grade, smoking status, physical activity level, energy and alcohol intake, BMI, hypertension and dyslipidaemia status, and prevalence of longstanding illness.

HR, hazard ratio; CI, confidence interval. * Stratified for sex. [#] Stratified for BMI categories.
[§] Stratified for longstanding illness and dyslipidaemia.

(ii) Differential misreporting of FFQ-items, comparison with diet diary data

The systematic association observed between energy reporting and the aggregate scores rankings would have been due to differential misreporting of FFQ-items. The Whitehall II FFQ was shown to over-estimate plant based micro-nutrients compared to the 7-day diet diary (Brunner *et al.*, 2001), but the analysis was not conducted for specific food items. Diet diaries represent better true levels of intake (Willett, 1998). Therefore, the 7-day diet diary (7DD) reported intakes were compared to the FFQ intakes to identify which food groups and FFQ-items were under or over reported in the FFQ. Paired t-tests were conducted for all food groups and for all FFQ-items and their 7DD equivalent.

Table 8.8 revealed that fruit and vegetable intake as reported in the FFQ was indeed higher than the 7DD reported intake. The inverse was true for the snacks and sweets groups. Dairy products and starchy foods appeared to be over-reported in the FFQ, while drinks were under-reported. The differences observed with dairy products and drinks could be due to some misclassification of milk drunk with tea or coffee. Small but significant differences were observed in all other food groups except fish products and spreads.

The differential intake between the two methods was at the food item level, as illustrated in appendix 4 which contains the t-tests results for all items. In some food groups, the difference between FFQ and 7DD data was related to the WXYfm score of the items (e.g. bacon and sausages were under-reported in the FFQ, while chicken was over-reported).

The absolute difference in reported intakes between the FFQ and 7DD tools may not be the best way to assess the relationship between the two methods since they measure different aspects of dietary intake. Analysing linear associations or rankings between the two methods may therefore be more relevant. For this purpose, the regression calibration method was applied in an attempt to correct FFQ reported intakes, i.e. to predict the participant's true intake of each FFQ-item.

Table 8.8: Mean difference between FFQ and 7DD reported intakes (n=1,349)

	Difference FFQ – 7DD in g/d		
Food group	**Mean**	**(95% CL)**	**p***
Meat products and eggs	-7.10	(-11.2; -3.01)	0.001
Fish and shellfish	-1.55	(-3.34; 0.25)	0.091
Bread and crackers	-8.62	(-12.1; -5.17)	<.001
Breakfast cereals	8.66	(6.64; 10.7)	<.001
Potatoes, rice and pasta	40.8	(36.0; 45.6)	<.001
Dairy products	277	(247; 306)	<.001
Meals[2]	7.08	(5.46; 8.71)	<.001
Fat spreads	0.49	(-0.35; 1.33)	0.254
Snacks and sweets	-17.2	(-20.7; -13.8)	<.001
Sauces and other spreads	12.5	(10.8; 14.2)	<.001
Drinks	-139	(-167; -111)	<.001
Fruits and nuts	130	(121; 139)	<.001
Vegetables	90.9	(84.1; 97.7)	<.001

*Paired t-test. [2]Meals include quiche, pizza and lasagne. CL, confidence limit.

(iii) Predicted true intakes of FFQ-items, survival analysis including corrected aggregate scores

a. Predicted true intakes

The regression calibration model [5] of section 8.1.1 was implemented to each FFQ-item, using the 7DD as reference measure. A "predicted intakes" dataset was created for the whole study sample by using the regression calibration parameter estimates. Individual estimates for all FFQ-items are given in appendix 4. The validity of FFQ reported intakes vs. 7DD ones was highest for foods consumed on a regular basis and in easily identified quantities (e.g. beverages, breakfast cereals, and some spreads); it was lowest for meat, fish , eggs, and vegetables. This was consistent with previous observations (Rosner & Gore, 2001) and showed that the absolute difference of reported intakes observed in table 8.8 between FFQ and 7DD data was not necessarily a sign of non-validity.

b. Corrected aggregate score

A corrected EWS (EWS_C) aggregate score was derived from the predicted intakes dataset, using the EWS algorithm from chapter 4. The EWS_C yielded unhealthier values compared to the original EWS (table 8.9). This followed the original hypothesis of fruit and vegetable over-reporting and snacks under-reporting. The

rank correlation between the original and corrected aggregate score was quite low (0.25). The dispersion of observations in figure 8.1 illustrated well the low correlation coefficient and indicated that the EWS_C would derive a very different rankings of participants compared to the EWS. The corrected PES(Q1) aggregate score yielded similar results (appendix 3).

Table 8.9: Summary statistics for original EWS and regression calibration derived EWS_C

Variable	n	Mean	Std Dev	Minimum	Median	Maximum
EWS	7,463	6.09	2.38	-2.50	6.11	16.8
EWS_C	7,463	9.00*	4.98	-2.99	8.86	21.6

* Significantly different from original score (paired t-test p<0.001)

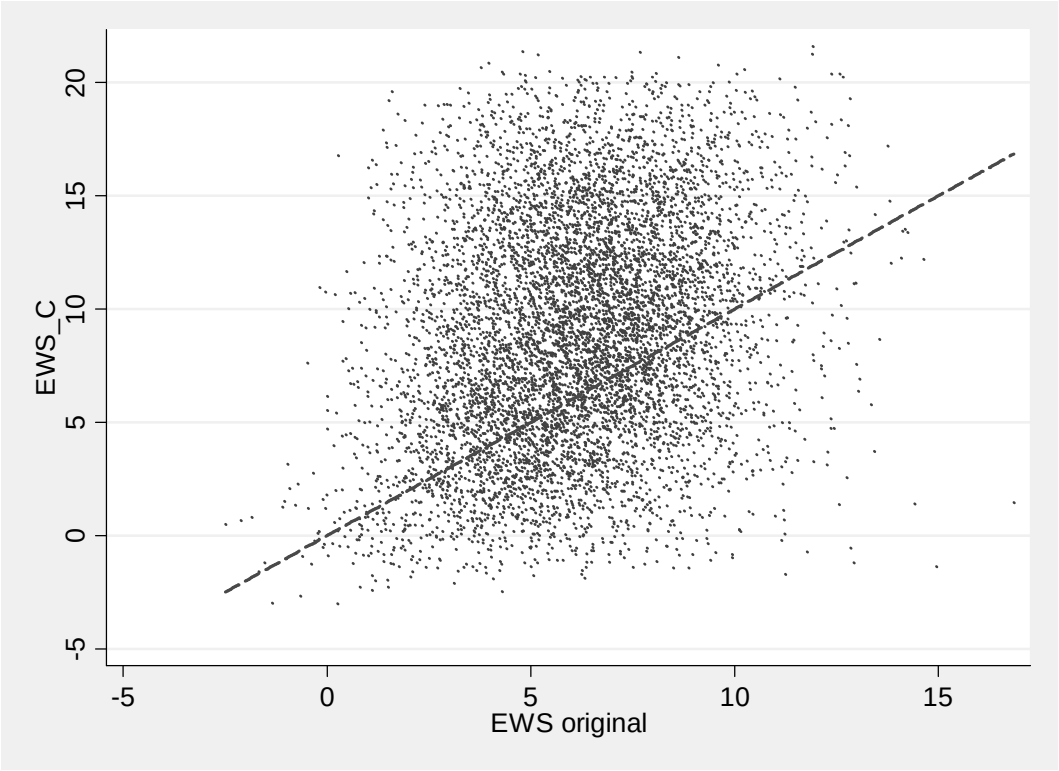


Figure 8.1: Regression calibration derived EWS_C vs. original EWS
The dashed line represents the y=x function

c. Proportional hazards regressions

The corrected EWS_C aggregate score clearly yielded new rankings of participants, with very different estimates compared to the EWS (table 8.10). None of the quadratic trend tests was significant but the expected risk reduction was not

obtained. Instead, most individual quartile estimates indicated higher risk compared to the original aggregate score. This was particularly noticed for cancer, with a 62% risk increase for participants in the EWS_C fourth quartile, which resulted in a significant positive linear trend. A linear risk increase was also suggested for CHD, but the estimates were not significant. No specific association could be observed for all-cause mortality and diabetes.

The results for the corrected PES(Q1) were similar, with all quadratic trends not significant, and most point estimates suggesting increased risk or no association (appendix 3).

In summary, the corrected FFQ-items were associated with less healthier diets as measured by the aggregate scores, in line with the differential misreporting of foods assumption. The differences observed between the original and corrected aggregate scores confirmed that differential misreporting occurred and had an impact on the aggregate scores rankings. However, the survival analysis models yielded highly unexpected results and no specific conclusion could be drawn from table 8.10 estimates. The regression calibration method involved several limitations, including some key assumptions which could have led to the unexpected results.

Table 8.10: Cox regression estimates across EWS and EWS_C quartiles (4: healthier)

Outcome	Quartile/ trend	Original EWS			EWS_C		
		HR	95	% CI	HR	95	% CI
CHD	1	Ref			Ref[#]		
(318 / 7,174)	2	0.82	0.58	1.15	1.09	0.78	1.52
	3	1.03	0.75	1.41	1.20	0.86	1.67
	4	1.22	0.89	1.69	1.26	0.91	1.75
	Linear	1.09	0.98	1.21	1.08	0.97	1.20
	p quadratic	**0.003**			0.956		
Diabetes	1	Ref*			Ref*		
(754 / 6,868)	2	1.00	0.81	1.24	**1.29**	**1.06**	**1.58**
	3	0.89	0.72	1.10	0.98	0.80	1.22
	4	1.04	0.84	1.28	0.97	0.78	1.20
	Linear	1.00	0.93	1.07	0.96	0.90	1.03
	p quadratic	0.402			0.105		
Cancer	1	Ref			Ref		
mortality	2	0.94	0.65	1.35	1.26	0.85	1.85
(251 / 7,235)	3	0.95	0.66	1.36	1.41	0.96	2.06
	4	0.87	0.60	1.26	**1.62**	**1.11**	**2.36**
	Linear	0.96	0.85	1.08	**1.17**	**1.04**	**1.31**
	p quadratic	0.057			0.902		
All-cause	1	Ref			Ref[§]		
mortality	2	0.85	0.66	1.09	0.90	0.70	1.15
(524 / 7,242)	3	0.86	0.67	1.11	0.94	0.73	1.21
	4	0.97	0.76	1.25	1.06	0.83	1.35
	Linear	0.99	0.92	1.08	1.02	0.94	1.11
	p quadratic	**0.004**			0.071		

Models adjusted for age, sex, ethnicity, marital status, employment grade, smoking status, physical activity level, energy and alcohol intake, BMI, hypertension and dyslipidaemia status, and prevalence of longstanding illness * Stratified for sex. [#] Stratified for BMI categories. [§] Stratified for longstanding illness and dyslipidaemia. HR, hazard ratio; CI, confidence interval.

8.1.3 Misreporting and quadratic trends, limitations

First, the regression calibration model assumed that diet diary data were an unbiased estimate of true intake. Diet diaries are also prone to reporting error and it was not possible to verify whether the associations obtained with the corrected aggregate scores were closer to the true epidemiological associations. Also, the regression calibration models assumed that errors in the FFQ and the 7DD were independent,

which was likely to be flawed. Yet, this may not have altered the estimates too much (Spiegelman *et al.*, 1997; Rosner & Gore, 2001).

Second, the FFQ and 7DD tools do not measure the same aspect of food intake. For foods rarely consumed, the 7DD data contained more non-consumers than consumers and normal distributions were not obtained with the log-transformation. As an example, strawberries are likely not to be reported in a 7DD if the diary is completed in winter. Regression calibration estimates were therefore particularly affected for such foods as illustrated in appendix 4. Statistical models dealing with such issue were recently developed for 24-hour recall data (Tooze *et al.*, 2006; Kipnis *et al.*, 2009; Zhang *et al.*, 2011). Implementation of these techniques would have required considerable adaptation of the regression calibration models and was beyond the scope of this project.

Last, the corrected versions of EWS and PES(Q1) were linear combinations of FFQ-items predicted intakes, all obtained by the regression calibration model. The errors for each item intake estimate were therefore added up, and the resulting corrected aggregate score may not be very meaningful. This could have explained the relative independence observed between the original and corrected aggregate scores (figure 8.1 and appendix 3), and the surprising survival analysis results.

The identification of energy misreporters with the Goldberg method also included some limitations which was the reason to apply the regression calibration models. The exclusion of energy misreporters led to a loss of statistical power and wider confidence intervals. The estimates from table 8.6 were subject to a selection bias since misreporting was shown to be associated with several non-dietary covariates and risk factors. The definition of the Goldberg thresholds was based on self-report and estimated measures, all prone to some error (FAO/WHO/UNU, 1985; Black, 2000a; Black, 2000b). The method was not initially designed to be applied individually and the use of three PAL categories was done to limit the number of misclassified individuals (Black, 2000b). Last, the method assumed constant body weight for participants across time. As a result, some participants may have been misclassified if low (or high) reported energy intake was associated with an effective weight decrease (or increase).

8.1.4 Conclusion

Energy under-reporting was associated with improved aggregate scores rankings and less favourable risk factors profiles. Cox regressions excluding energy misreporters yielded attenuated U-shapes and lower hazard ratio estimates for the linear trend tests compared to the original models (chapters 6 and 7). The observed risk reduction was not significant, but the sensitivity analysis results confirmed that the association between low energy reporting and higher prevalence of vascular risk was likely to have confounded the predictive validity results. The absence of significant linear associations may have been due to the sample restriction or to selection bias, with acceptable energy reporters being a relatively homogeneous group.

The association between energy reporting and aggregate scores rankings was linked to differential misreporting of FFQ-items. The comparison with 7DD data confirmed that participants tended to over-report healthy foods and under-report less healthy foods when using the FFQ. Differential misreporting of foods had an impact on the prospective associations between aggregate scores and health outcomes. However, the regression calibration method did not yield consistent results and it was not possible to conclude precisely on the influence of differential misreporting of FFQ-items.

The two aggregate scores retained in the present analysis were weighted by energy intake. This certainly increased their sensitivity to energy misreporting and to differential under-reporting of the energy dense foods. In chapter 4, aggregating algorithm using different weighting scales were proposed. Compared to the EWS and PES(Q1), these alternative aggregate scores—which also relied on the exact reported amounts of each FFQ-item—yielded similar rankings and comparable survival analyses results (not shown). Aggregating algorithms relying less on the exact reported intake may be better suit the analysis of predictive validity of NP models using FFQ data.

The use of the Goldberg cut-off and regression calibration techniques allowed understanding better the impact of dietary misreporting on the predictive validity results. Yet, both methods were associated with strong limitations, and were

therefore not used in the subsequent analyses. Standard methods, such as adjusting for total energy intake and BMI were retained to account for energy misreporting.

8.2 Diet variety and aggregate scores

The WXYfm and SAIN,LIM NP models are "across-the-board" algorithms, i.e. the same algorithm is applied to all foods regardless of the food category. Across-the-board models are designed to identify healthier foods *per se* rather than healthier versions of foods within food groups. As a result, both WXYfm and SAIN,LIM categorise all foods from some food groups as healthy while other food groups have all their items classified as unhealthy. The aggregate scores EWS and PES(Q1) used in the previous chapters did not take diet variety into account, and one could have obtained a very high (or low) ranking by having a very restricted diet, which may be detrimental to health (Michels & Wolk, 2002; Savy *et al.*, 2005). On the other hand, an individual eating a more varied diet may have obtained an average ranking despite having more balanced intakes.

Hence, it was assumed that participants in the extreme quartiles of both EWS and PES(Q1) had a lower dietary variety than those in the middle quartiles; such association would have explained part of the U-shapes. To assess such an assumption, dietary variety was first associated with the aggregate scores and with prospective health outcomes. Variety was then included as a confounding variable in chapters 6 and 7 Cox models. Dietary variety might not have had the same effect in participants having less healthy or healthier diets, as defined by the aggregate scores. The role of diet variety as an effect modifier was investigated by including interaction terms in the Cox regressions.

8.2.1 The food variety score

The food variety score (FVS), or diet variety score, was used to capture diet variety within the Whitehall II FFQ. It was the number of FFQ-items reported to be consumed more than once a week (Drewnowski *et al.*, 1996; Drewnowski *et al.*,

1997; Hatloy *et al.*, 1998). Within the 7,251 participants of the complete-case analyses (chapters 6 and 7), the FVS ranged from 0 to 99, with a mean of 43.1.

8.2.2 Food variety score and aggregate scores

A slight quadratic association was observed between the EWS aggregate score and the food variety score (table 8.11). This first result followed the initial assumption and was confirmed by regression models using squared aggregate scores ($p<.001$ for EWS^2). Since the relationship was highly significant, the FVS was considered as a potential confounder, and associations with prospective health outcomes were investigated. Results were similar for the PES(Q1) aggregate score (appendix 3).

Table 8.11: Mean food variety score across EWS quartiles (4: healthier)

	EWS				
	1	**2**	**3**	**4**	**p***
FVS	42.9	44.8	43.7	41.0	<.001

FVS, food variety score.
* Heterogeneity ANOVA across quartiles

8.2.3 Food variety score and health outcomes

The association between the FVS and prospective health outcomes was first analysed using log-rank tests for heterogeneity across quartiles of FVS. The tests highlighted that the FVS was associated with CHD and all-cause mortality (not shown). These results were confirmed by Cox regression models which suggested that variety had a protective effect on CHD, cancer and all-cause mortality (table 8.12). These trends were robust to adjustment and confirmed the role of diet variety in preventing prospective chronic disease. As a result, the FVS was included in the survival analysis models of chapters 6 and 7.

Table 8.12: Cox regression estimates across the food variety score quartiles

Outcome	Quartile /trend	Model 1 HR	95 % CI		Model 2 HR	95 % CI		Model 3 HR	95 % CI	
CHD (318 / 7,174)	1	Ref			Ref			Ref[#]		
	2	**0.72**	**0.53**	**0.97**	0.75	0.55	1.02	0.73	0.53	1.00
	3	0.87	0.64	1.18	0.93	0.67	1.29	0.86	0.62	1.20
	4	**0.59**	**0.43**	**0.81**	**0.64**	**0.44**	**0.93**	**0.59**	**0.40**	**0.86**
	Linear	**0.87**	**0.79**	**0.96**	0.90	0.80	1.01	**0.87**	**0.77**	**0.98**
Diabetes (754 / 6,868)	1	Ref*			Ref*			Ref*		
	2	0.98	0.80	1.20	1.00	0.81	1.23	1.01	0.82	1.24
	3	0.92	0.74	1.14	0.97	0.77	1.22	0.92	0.74	1.16
	4	1.03	0.84	1.26	1.07	0.85	1.36	1.01	0.80	1.29
	Linear	1.01	0.94	1.07	1.02	0.95	1.10	1.00	0.92	1.07
Cancer mortality (251 / 7,235)	1	Ref			Ref			Ref		
	2	0.73	0.52	1.02	0.75	0.53	1.06	0.75	0.53	1.06
	3	**0.65**	**0.45**	**0.93**	0.69	0.47	1.01	0.68	0.46	1.00
	4	**0.68**	**0.48**	**0.95**	0.74	0.49	1.12	0.73	0.49	1.10
	Linear	**0.88**	**0.78**	**0.98**	0.90	0.79	1.03	0.90	0.79	1.03
All-cause mortality (524 / 7,242)	1	Ref			Ref			Ref[§]		
	2	**0.75**	**0.60**	**0.95**	0.80	0.63	1.02	0.80	0.63	1.02
	3	**0.75**	**0.59**	**0.96**	0.83	0.64	1.08	0.81	0.62	1.06
	4	**0.68**	**0.53**	**0.86**	**0.74**	**0.56**	**0.99**	**0.74**	**0.55**	**0.99**
	Linear	**0.89**	**0.82**	**0.96**	0.92	0.84	1.01	0.91	0.83	1.00

Model 1 adjusted for age, sex, and ethnicity. Model 2 adjusted for marital status, employment grade, smoking status, physical activity level, and energy and alcohol intake. Model 3 adjusted for BMI, hypertension and dyslipidaemia status, and prevalence of longstanding illness
HR, hazard ratio; CI, confidence interval. * Stratified for sex. [#] Stratified for BMI categories.
[§] Stratified for longstanding illness and dyslipidaemia.

8.2.4 Diet variety, confounder or effect modifier

(i) Variety as a confounding factor

Cox regression models from chapters 6 and 7 were run and further included the food variety score as a covariate. The hazard ratio estimates obtained for all the outcomes were very similar to the original results for both EWS and PES(Q1) (appendix 5.1). Quadratic trends were very slightly attenuated and remained significant (or borderline significant for all-cause mortality). This was not in line with the

expectation that adjusting for the lower diet variety in the healthiest individuals would result in attenuated U-shapes. The difference in diet variety between the aggregate score quartiles was small (table 8.11), and this would explain the very small changes compared to the original Cox regressions.

Above models assessed diet variety as a confounding factor assuming that the effect would be constant across aggregate score values. This assumption might have been flawed and models including interaction terms were run to assess whether diet variety was an effect modifier.

(ii) Interaction between aggregate scores and the food variety score

Interaction terms (θ) were introduced in the Cox models between each aggregate score quartile and the food variety score (aggregate score quartile*FVS). As an example for EWS, the model was specified as follows, with the first quartile of EWS (EWS_Q1) as the reference group:

$$\text{Outcome (event/t)} = h_o(t)\exp(\alpha + \beta_2 \cdot \text{EWS\_Q2} + \beta_3 \cdot \text{EWS\_Q3} + \beta_4 \cdot \text{EWS\_Q4} + \theta_2 \cdot \text{EWS\_Q2*FVS} + \theta_3 \cdot \text{EWS\_Q3*FVS} + \theta_4 \cdot \text{EWS\_Q4*FVS} + \gamma \cdot \text{FVS} + \text{covariates (age, sex, ethnicity)} + \varepsilon)$$

For EWS, interaction terms were found significant for cancer and all-cause mortality, and borderline significant for CHD (not shown). For PES(Q1), the interaction was significant with cancer mortality, and almost significant for CHD (not shown). To interpret these results, Cox models were stratified by FVS tertiles. Figure 8.2 displays hazard ratio estimates for the EWS aggregate score and model 1 (adjusted for age, sex and ethnicity). In these models, the reference group was the EWS first quartile (least healthy) combined with the second tertile of FVS. The PES(Q1) estimates are displayed in appendix 3.3.

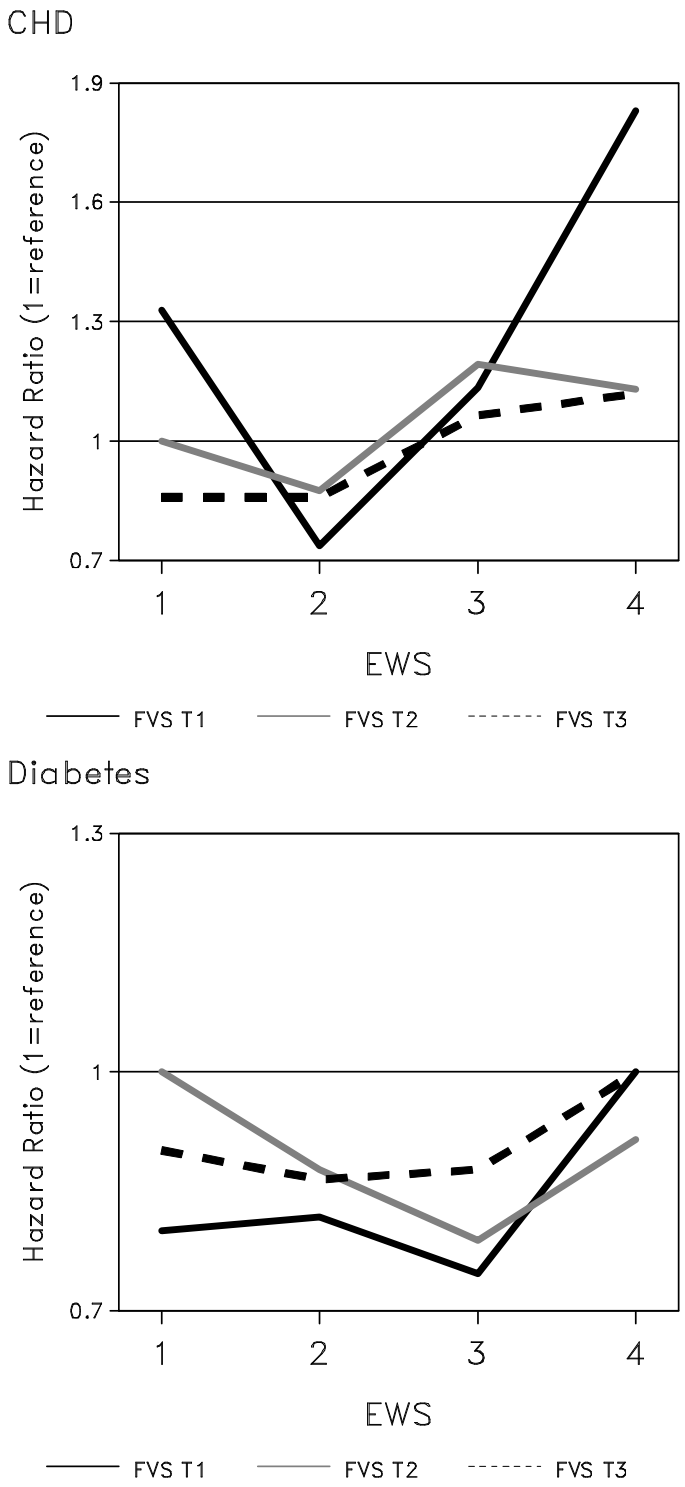


Figure 8.2: Hazard ratio estimates across EWS quartiles (4: healthier), stratified by FVS tertiles
The reference group was the 1st quartile of EWS with the 2nd tertile of FVS. Models were adjusted for age, sex, and ethnicity

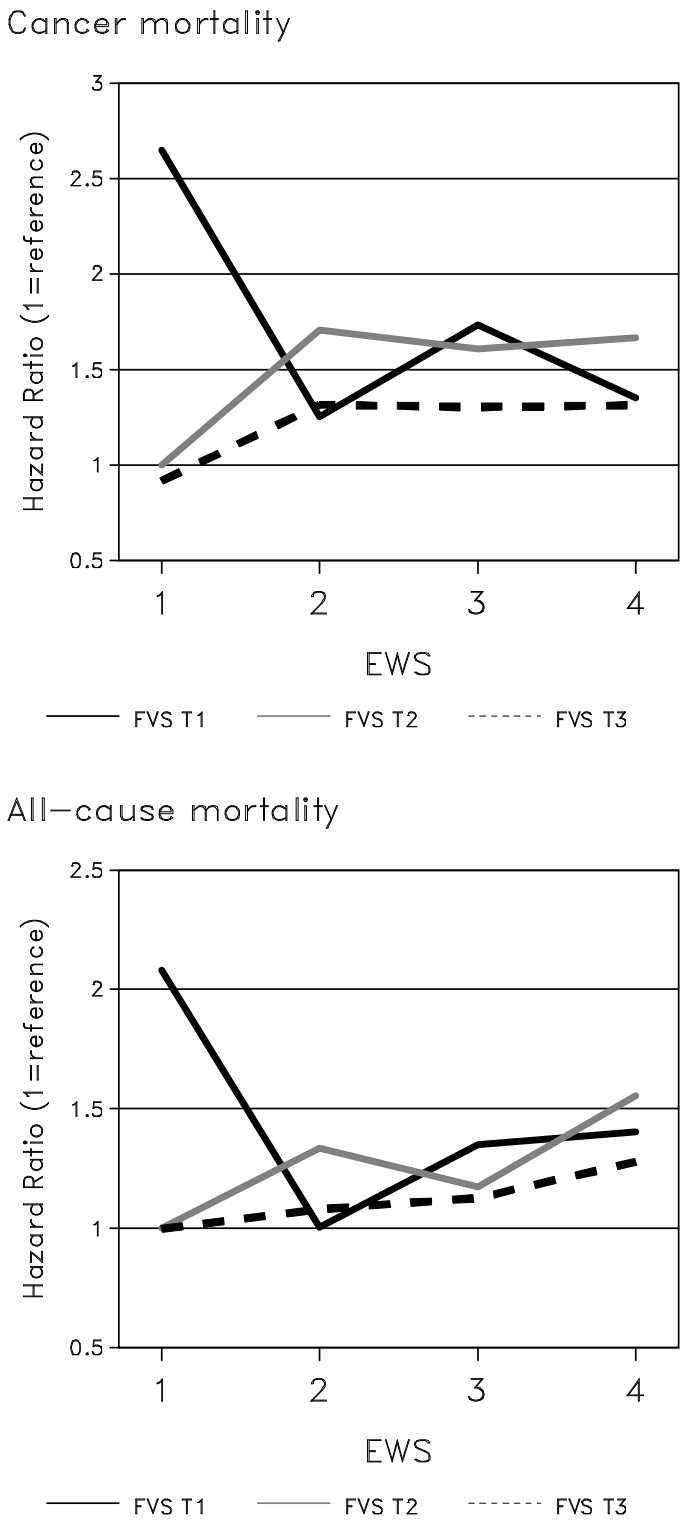


Figure 8.2 (continued): Hazard ratio estimates across EWS quartiles (4: healthier), stratified by FVS tertiles

The reference group was the 1st quartile of EWS with the 2nd tertile of FVS. Models were adjusted for age, sex, and ethnicity.

For CHD, cancer and all-cause mortality, the interaction was well illustrated by the crossing trends between the FVS tertiles (figure 8.2). Prospective associations between EWS and the outcomes were mainly observed in participants with low diet variety, the associations were much weaker for participants in FVS second and third tertiles.

More precisely, the EWS was inversely associated with cancer and all-cause mortality for those participants with low diet variety (FVS 1st tertile): significant higher risks were observed for individuals in the EWS first quartile and trends flattened in the 2nd, 3rd and 4th quartiles. The prospective associations were slightly positive for participants in the 2nd and 3rd tertile of FVS, risk levels were alike. For CHD, a strong J-shape was observed within the FVS 1st tertile. Similarly to the mortality outcomes, trends were slightly positive for participants with mid to high diet variety. For diabetes, trends were similar across FVS tertiles, further confirming the non-significant interaction tests.

The associations displayed in figure 8.2 were well linked to the results of chapter 6 and uncovered details on the origins of the quadratic trends. The most visible example was obtained with the CHD outcome. The strong J-shape obtained with participants in the FVS first tertile, more likely to be classified in the EWS extreme quartiles (table 8.10), related well with table 6.2 estimates (model 1). The protective effect of EWS on cancer and all-cause mortality observed within participants classified in the FVS first tertile would have similarly explained the trends of table 6.2. The U-shapes would have been due to the positive trends observed for participants in the FVS second and third tertiles, together with the flattening of the trend for participants in the FVS first tertile. For diabetes, there was no interaction with diet variety and the trends for all FVS tertiles suggested a U-shape.

The results for the PES(Q1) aggregate score were alike, with a protective effect on cancer and all-cause mortality among participants in the FVS first tertile, and no specific trend for the FVS second and third tertiles. For CHD, a similar strong J-shape was observed whereas the associations were weak for diabetes (appendix 3.3).

8.2.5 Limitations

The index chosen to measure diet variety, the food variety score, may not be the most adequate tool. It measures quite well absolute variety (e.g. eating different types of chocolate bars) but does not give information on food group variety (i.e. eating foods from all food groups). We conducted further analyses using the diet diversity score, a measure of food group variety (Drewnowski *et al.*, 1996; Dubois *et al.*, 2000; Savy *et al.*, 2005). Conclusions were similar: participants with a lower diet diversity score were more at risk and were more sensitive to the aggregate score classification (results not shown).

In addition, the FVS depended on the grouping of foods in the FFQ. As an example, Whitehall II participants could report in much more details their fruit and vegetable intake (34 items), as opposed to meat or fish (16 items). The results obtained with the FVS were linked to the FFQ used in this study and may not have reflected the true variety of participants' diets.

8.2.6 Conclusion

The EWS and PES(Q1) aggregate scores were not designed to take into account diet variety, but their algorithm, combined with the across-the-board nature of the WXYfm and SAIN,LIM models, led to systematic associations between variety and both aggregate scores. Diet variety played a crucial role in predicting prospective health outcomes (table 8.12). It revealed to be an effect modifier since associations between aggregate scores and prospective chronic disease only appeared in participants with a low diet variety. As a result, an alternative aggregation method taking more into account diet variety may reveal a better predictor of health outcomes. Though, the inclusion of variety in an aggregation algorithm would go against the food-based NP concept which aim is to determine healthiness of individual foods based exclusively on nutrient content.

8.3 Nutrient profiling components analysis

The two previous sections revealed that both dietary misreporting and diet variety impacted on the predictive validity results. The aggregating algorithms used for the WXYfm and SAIN,LIM NP models relied on the exact reported intake and were weighted by energy intake. Both EWS and PES(Q1) may therefore be too sensitive to dietary misreporting. The aggregate scores were designed not to take into account diet variety but were actually associated to it. As a result, alternative aggregating algorithms less related (or related in a different way) to both energy misreporting and diet variety could be designed.

To effectively derive new aggregating algorithms, it is necessary to understand better the link between NP models components (i.e. the nutrients included in the models and the way they are computed), the aggregate scores, and health outcomes. Two research questions arose for this section: (i) did all components of the NP models predict health outcomes as hypothesised, i.e. were negative nutrients associated with increased risk, and positive nutrients with reduced incidence; and (ii) were aggregate scores equally correlated to all components included in the NP models, or were they driven by just a few components. The related hypotheses were that the U-shaped associations observed in chapters 6 and 7 might be due to the fact that some components failed to predict outcomes in the expected direction, or that aggregate scores were correlated to a few components only. To conduct such analyses, "component scores" were created for each component of the NP models. These "component scores" were first included in Cox regression models to assess their relationship with the outcomes of interest. Their association with the aggregate scores was then assessed.

8.3.1 Methods, component scores

To assess the crude effect of each component, energy residuals were estimated, i.e. crude intake of each component was regressed against energy intake and residuals were retained for inclusion in the survival analysis models.

To analyse the effect of the NP and aggregation algorithms on each component, "component scores" were derived in a similar way to the EWS aggregate score, i.e. energy-weighted means. For WXYfm, the allocated points (0 to 10 for the four negative components, and 0 to 5 for the three positive ones) were used; for the SAIN,LIM model, the content/recommendation ratios were used (see chapter 2 for the NP models algorithms).

Survival analysis models were run individually for each energy residual and component score. The Z-scores (i.e. standardised variable ~ N(0,1)) of the residuals or the component scores were included in the Cox regression models. Hazard ratios were estimated for an increase of one standard deviation, reflecting the linear trend. The Cox regression models were adjusted for age, sex, and ethnicity following the specifications of chapters 6 and 7 "model 1".

Rank correlations were computed between the component scores, their respective nutrient crude intake, and the NP aggregate scores in order to assess the component scores' relationship with the respective component intake and with the aggregate score.

8.3.2 Survival analysis, energy residuals and component scores

(i) Nutrient energy residuals

Hazard ratio estimates for one standard deviation increase of the nutrient energy residuals Z-scores are displayed in figure 8.3. No significant associations were observed for the negative nutrients. Saturated fats were positively associated with all outcomes, in line with the expectations. Similar results were obtained for sodium, except for a weak inverse association with CHD. This was not expected but the hazard ratio estimate was close to 1. More surprisingly, sugar intake was suggested to reduce risk of cancer and all-cause mortality; associations were positive for CHD and null for diabetes.

Each positive component, except vitamin C, displayed at least one significant association. Fruit, vegetable, and nuts (FVN) were protective against cancer and all-cause mortality. Dietary fibres were inversely associated with diabetes and all-cause mortality. Iron was protective against CHD, diabetes, and all-cause mortality. These results followed the hypotheses and similar trends were suggested for the other outcomes (except a null-association between FVN and diabetes). On the contrary, significant positive associations were obtained between protein intake and diabetes, and between calcium and CHD. These unexpected associations could be explained by the high protein and/or calcium content in some otherwise less healthy foods (e.g. red meat and full-fat dairy products rich in saturated fat, luncheon meats and savoury pies high in sodium).

Overall, most estimates obtained for the nutrient residuals were in the expected directions. Unexpected associations were weak and not significant, except for the protein and calcium components of the SAIN,LIM model.

(ii) WXYfm Component scores

Hazard ratio and 95% confidence intervals estimates for Z-scores of all WXYfm component scores are displayed in figure 8.4.

The trends previously observed in figure 8.3 for the fibre and FVN positive components were confirmed and strengthened. Inverse associations were also obtained for the protein component, with a significant risk reduction of all-cause mortality. These results suggested that taken together, the three positive components of the WXYfm model might yield significant inverse associations for all outcomes. It followed the hypothesis beneath the NP model and indicated that the WXYfm and/or the component scores algorithms did strengthen the protective effect of positive components.

The component scores algorithms also altered the negative components estimates, but the expected positive associations were not obtained since all estimates indicated either null or protective effects. In particular, CHD risk incidence was significantly reduced with all components but sugar. Risk reduction was also observed for

diabetes and all-cause mortality, the hazard ratios of both outcomes were significant for the sugar component. Associations for cancer mortality were non-significant.

Overall, the estimates from figure 8.4 suggested that the positive components of the WXYfm NP model predicted slightly lower risk of prospective health outcomes, in accordance with the hypothesis. On the contrary, the negative components did not predict increased risk as expected. The observed inverse associations could have entailed the U-shapes since lower negative component scores for participants classified in the fourth quartile of EWS would not have led to reduced risk.

Results for the SAIN,LIM negative components were similar, especially for sodium (appendix 3.4). In line with figure 8.3, the protein and calcium positive components were associated with increased incidence of CHD and diabetes, which could have further explained the J-shaped associations observed in table 7.11. The unexpected results for calcium could be linked to the over-reporting of dairy products (table 8.8).

Section 8.1 showed that energy-dense foods' intake tended to be under-reported. These energy-dense foods are usually classified as less healthy by the WXYfm and SAIN,LIM NP models due to their higher content of negative components. The unexpected results obtained with the negative component scores may therefore be related to dietary misreporting, and more specifically to the association between energy under-reporting and vascular risk.

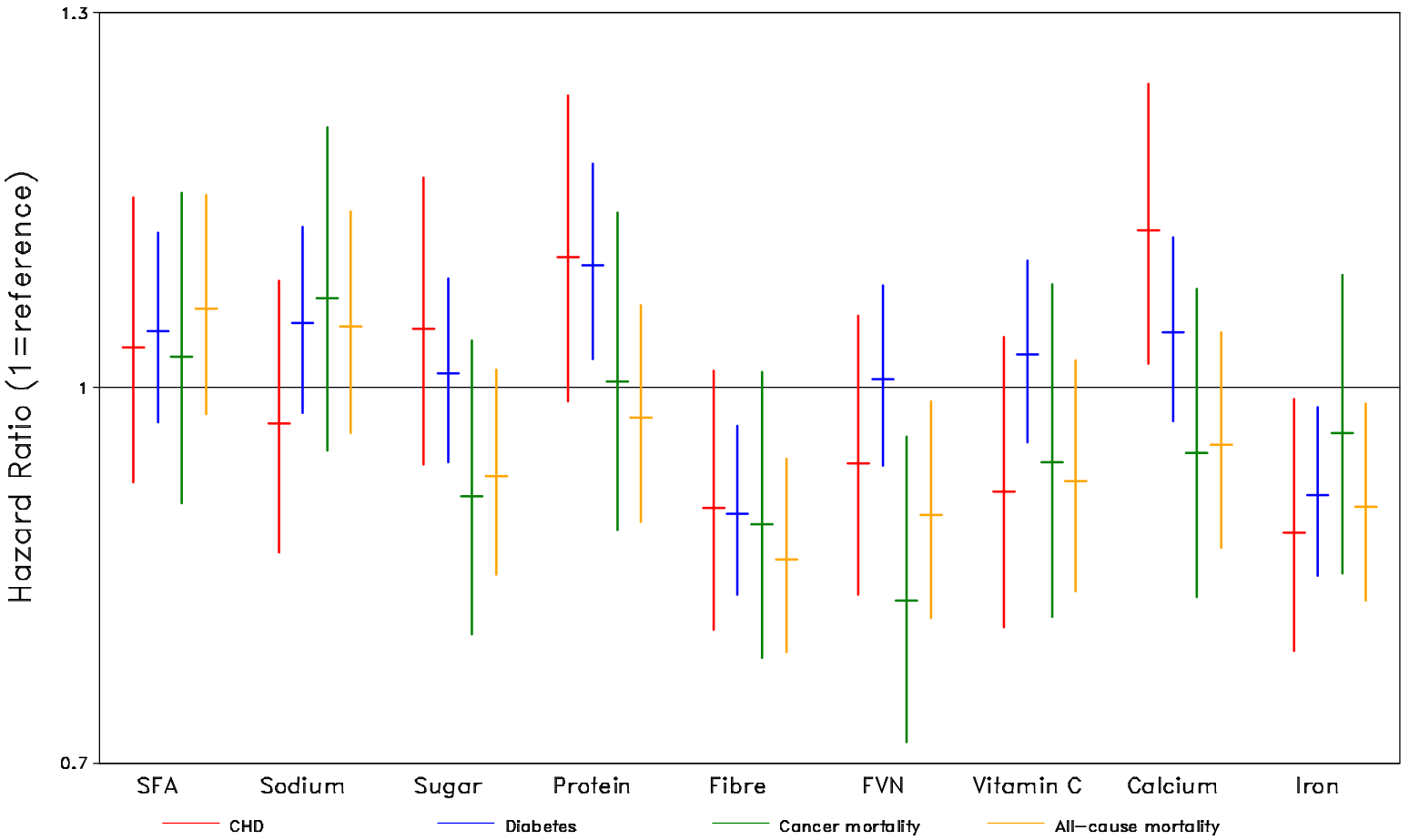


Figure 8.3: Hazard ratio estimates and 95% CI for energy residuals Z-scores (n=7,251)
Models adjusted for age, sex, and ethnicity; diabetes models stratified for sex.
SFA, Saturated fatty acid; FVN, fruit, vegetable, and nuts content.

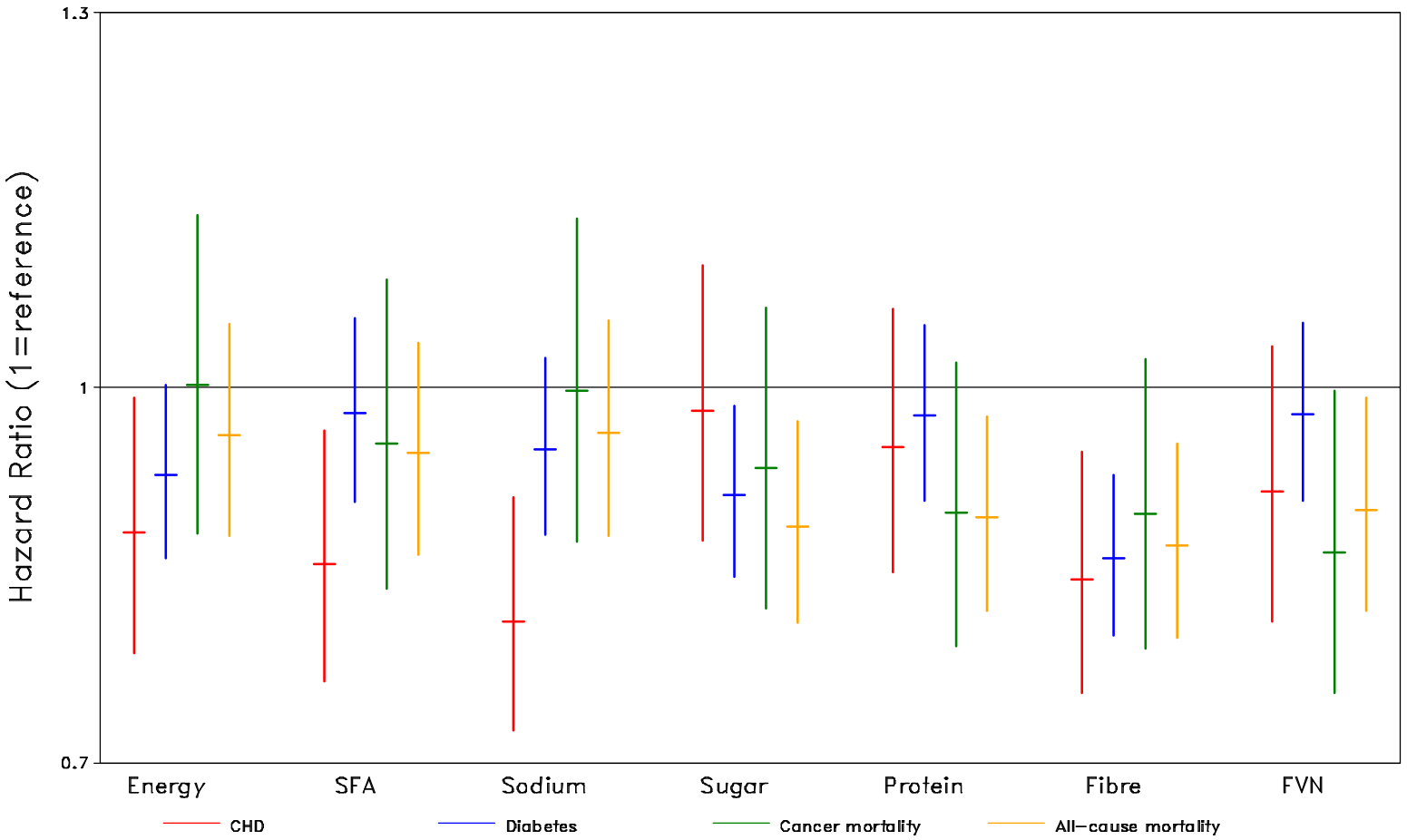


Figure 8.4: Hazard ratio estimates and 95% CI for WXYfm "component scores" Z-scores (n=7,251)
Models adjusted for age, sex, and ethnicity; diabetes models stratified for sex.
SFA, Saturated fatty acid; FVN, fruit, vegetable, and nuts content.

8.3.3 Component scores and energy reporting

Table 8.13 confirmed that the negative component scores of the WXYfm model were positively associated with energy reporting. As a result, these component scores were significantly lower among low energy reporters. On the other hand, the protein and fibre positive components were weakly correlated to reported energy intake, and higher values of the component scores were observed among acceptable reporters. The fruit, vegetable, and nut component was inversely correlated to reported energy intake, further confirming that energy under-reporting was associated with higher intake of fruit and vegetables.

Results were similar for the SAIN,LIM components (appendix 3.4), suggesting that the survival analysis results for the negative components were indeed confounded by energy misreporting and its association with vascular risk.

Table 8.13: WXYfm component scores and energy reporting

Component score	Rank correlation with reported energy intake	Mean component score by reporting level			
		Low	Acceptable	High	p*
Energy	0.23	2.25	2.50	2.43	<.001
Saturated fat	0.25	2.59	2.91	3.03	<.001
Sugar	0.19	1.35	1.54	1.67	<.001
Sodium	0.20	2.40	2.68	2.54	<.001
Protein	0.01	2.52	2.59	2.50	<.001
Fibre	0.06	1.26	1.35	1.25	<.001
Fruit, vegetable and nut	-0.22	0.77	0.67	0.63	<.001

*Heterogeneity ANOVA across reporting levels.

8.3.4 Component scores and EWS aggregate score

Component scores were all positively related to their respective nutrient intake, the correlations were low for energy and protein and highest for fruit, vegetable and nuts (table 8.13). The crude intake of nutrients was therefore not equally reflected by the pointing system of the WXYfm NP model and the component scores. This was in line with figure 8.4 which showed different results from figure 8.3.

Rank correlations between component scores and the EWS aggregate score were higher for the negative components (energy, saturated fats, sugar, and sodium). They were null for the protein and fibre components and relatively high for the fruit, vegetable, and nuts component. This indicated that the EWS relied mainly on the negative components and on fruit, vegetable, and nuts content; both the fibre and protein components having very little influence on the aggregate score. This was in accordance with the WXYfm algorithm which gave more weight to the negative nutrients. More importantly, it could have led to the U-shapes obtained in chapter 6 since participants in the healthiest quartile of EWS did not necessarily have increased intake of the protein and fibre components (shown to be protective in figure 8.4). It further strengthened the observations made previously with negative components: EWS was highly correlated to these components which did not predict increased risk as expected.

The PES(Q1) aggregate score was evenly correlated to all its components (appendix 3.4). This was probably due to the SAIN,LIM algorithm which used similar ratio scales for both positive and negative components; and it could have entailed the stronger risk reduction observed in chapter 7 with PES(Q1) compared to EWS (the fibre, vitamin C and iron components were protective against prospective health outcomes).

Table 8.14: Rank correlations between component scores, respective nutrient intake, and the EWS aggregate score (n=7,251)

	Energy	**SFA**	**Sugar**	**Sodium**	**Protein**	**Fibre**	**FVN content**
Nutrient intake	0.23	0.58	0.42	0.46	0.22	0.56	0.71
EWS	0.85	0.90	0.61	0.49	0.00	0.02	-0.45

SFA, saturated fatty acid; FVN, fruit, vegetable, and nuts.

8.3.5 Limitations

First, residual confounding could have occurred as the survival analysis models were only adjusted for age, sex, and ethnicity. Similar analyses were run using models 2 and 3 of chapters 6 and 7 and similar conclusions were drawn (not shown).

Second, the use of FFQ reported intakes might have been a stronger limitation in this section since the content of some nutrients could vary hugely within the FFQ-item categories, e.g. sugar and saturated fat content in the "yoghurt" item. Also, the estimation of sodium intake was most probably not reliable as it did not take into account salt added at the table.

Third, energy reporting was shown to be associated with under-reporting of unhealthy foods high in the negative components. The hazard ratios estimates differences between figures 8.3 and 8.4 highlighted a strong effect of the aggregation algorithm. The Cox models estimates were therefore likely to be confounded by differential under-reporting of the unhealthy foods, which would have particularly affected the negative components estimates.

8.3.6 Conclusion

The component analysis gave a considerable insight on the potential origin of the U-shaped associations obtained in chapters 6 and 7. First, the negative component scores did not predict increased risk as originally expected. Therefore, the reduced dietary content of negative components in participants classified as healthiest by the aggregate scores was not associated with reduced risk. Second, the WXYfm and EWS algorithms emphasised these negative components. The combined effects of these two factors could have, alone, entailed the quadratic trends for EWS.

The SAIN,LIM model was associated more evenly with all its components, which was reflected in the associations between component scores and PES(Q1) (appendix 3). This would have explained the stronger risk reductions observed in chapter 7 compared to the EWS which was poorly related to the protein and fibre

components—both protective for most outcomes. Yet, some positive components of PES(Q1) were associated with increased incidence of CHD and diabetes. This would have led to the J-shaped associations observed for these two outcomes in table 7.11.

The unexpected results obtained for the negative component scores, which appeared to have entailed the U-shapes, would have been explained by the confounding effect of the association between low energy reporting and vascular risk (table 8.13). The difference between the estimates of figures 8.3 and 8.4 indicated that the results were linked to the component scores algorithm (i.e. the EWS aggregate core algorithm) and not to the crude intake of the negative nutrients. This further showed the influence of the aggregate scores algorithm on the survival analysis estimates and confirmed that energy-weighted aggregation algorithms may not be the best solution when using FFQ data prone to misreporting.

8.4 Discussion

This chapter was introduced to better understand the quadratic trends observed in chapters 6 and 7. The sensitivity of the aggregate scores to energy misreporting—associated with vascular risk status—appeared as the main potential explanatory factor for the U-shaped associations.

Energy reporting was inversely associated with diet healthiness as measured by the aggregate scores leading to a higher proportion of low energy reporters in the healthiest quartiles of EWS and PES(Q1). Low energy reporting was also linked to low employment grade and higher BMI (associated with higher energy needs), which led to less favourable levels of vascular risk factors among low energy reporters. Energy misreporting would have therefore confounded the associations between the aggregate scores and prospective health outcomes. This was confirmed by sensitivity analyses excluding energy misreporters which yielded attenuated quadratic trends. Lower hazard ratio estimates were specifically obtained for the fourth quartile of EWS and for the linear trend tests. The PES(Q1) aggregate score was slightly less affected by energy misreporting, but conclusions were similar.

The use of the 7DD data confirmed that the systematic associations observed between the aggregate scores and energy reporting was due to differential misreporting of FFQ-items. Participants tended to over-report foods considered as healthy (e.g. fruit and vegetables) while under-reporting the energy dense unhealthy foods (e.g. snacks and sweets, some meat products). Regression calibration was applied in an attempt to correct the reported intakes of each FFQ-item. The corrected aggregate scores derived from the corrected FFQ-items intakes indicated less healthy diets, in line with the expectations. Cox models including the corrected aggregate scores yielded surprising results: most hazard ratios estimates indicated higher risk when compared to the original aggregate scores. The regression calibration model was subject to a range of assumptions likely to be flawed, for instance normal distribution of intakes for all FFQ and 7DD items, which could have entailed the unexpected results observed with the corrected aggregate scores.

In addition, diet variety was associated with the EWS and PES(Q1) aggregate scores in the expected way (i.e. participants in the least healthy and healthiest quartiles had a slightly lower variety) and diet variety itself was protective against chronic disease. Yet, diet variety did not confound the prospective associations observed in chapters 6 and 7. Instead, it acted as an effect modifier: stratified Cox regressions showed that associations between aggregate scores and prospective chronic disease were only observed in participants with a low diet variety. The interaction between aggregate scores and diet variety highlighted the difficulty of aggregating the food-based NP concept at the diet level without taking into account other characteristics of dietary patterns.

Last, components of the WXYfm and SAIN,LIM NP models were associated separately with prospective health outcomes. Cox models including energy adjusted intakes (residual method) yielded weak estimates in the expected directions, except for the sugar and protein components. When using component scores, derived in a similar way to the aggregate scores, the prospective associations between individual components and health outcomes were changed. The protective effect of the WXYfm positive components was confirmed and strengthened, including for proteins. On the other hand, some negative components were found to be inversely associated with prospective risk, deviating from the hypothesis beneath the NP

models. Similar results were obtained for SAIN,LIM, though the positive protein and calcium component scores predicted increased prospective risk. Therefore, the component scores (similar to the aggregate scores algorithms) yielded associations not reflecting the crude intake of the NP components, highlighting the influence of the aggregating algorithm on the prospective associations. The unexpected results obtained for the negative component scores—explained by the association between negative component scores and energy under-reporting—could have entailed the U-shaped associations since participants in the healthiest quartiles of the aggregate scores would not have benefited from lower negative component scores.

The WXYfm NP model puts more weight on the negative components. As a result, the derived EWS aggregate score depended mainly on the negative components. This would have reinforced the unexpected effect on prospective risk of the negative components and would have resulted in the quadratic trends observed in chapter 6. The SAIN,LIM model did not emphasise the negative nutrients and the derived PES(Q1) aggregate score was associated more evenly with all its components. This would have explained the stronger risk reduction observed in chapter 7 for the middle quartiles (some positive components were protective against chronic disease). The quadratic trends would have been due to the unexpected effects of the protein, calcium, and negative components.

8.5 Conclusion

The results from this chapter confirmed that all three factors had some influence on the U-shapes initially observed between the aggregate scores and prospective health outcomes. The WXYfm and SAIN,LIM NP models rely on the energy density of foods since it is highly correlated to the content of negative nutrients and it is inversely linked to nutrient density: energy dense foods have low NP scores. The aggregate scores were both weighted by energy intake and were therefore particularly sensitive to the exact reported intake of the less healthy energy dense foods. Section 8.1 showed that energy misreporting was linked to differential misreporting of the energy dense foods. Low energy reporters therefore obtained higher aggregate scores rankings and were more likely to obtain low negative

component scores (section 8.3). In parallel, low energy reporting was associated with higher levels of several risk factors. As a result, it appeared that the strongest reason for the U-shaped associations was energy under-reporting which over-influenced the aggregate scores rankings and confounded the prospective associations.

The impact of the EWS and PES(Q1) aggregate scores algorithms, well illustrated in sections 8.2 and 8.3, suggests that their algorithms may have not been the most adequate given the Whitehall II data limitations. An aggregating algorithm relying less on the exact reported amounts, particularly of energy-dense foods, may better suit FFQ data. It was difficult to quantify the impact of diet variety on the quadratic trends, but it was suggested that diet scores linked to diet variety may be able to better predict prospective chronic disease risk. These conclusions were used in the next chapter to design alternatives to the EWS and PES(Q1) aggregate scores with the aim to verify the impact of dietary misreporting and diet variety on the U-shaped associations.

Chapter 9: Results-driven aggregate score models

The analyses in chapter 8 identified three potential explanatory factors for the quadratic associations initially observed between aggregate scores and prospective health outcomes. The goal of this chapter was to modify the aggregate scores and/or the nutrient profiling (NP) models algorithms to verify the results from the previous chapter and to obtain better predictors of health outcomes.

First, there was a strong suggestion that the U-shaped associations were due to the association between low-energy reporting and increased vascular risk. Diet variety was further shown to be intrinsically linked to future health status. It was therefore assumed that an aggregation method which would be less sensitive to absolute reported intake and depend more on diet variety may capture better the healthiness of individuals' diets. Second, differential under-reporting of less healthy foods was shown to particularly affect the negative components of the NP models. As a result, an alternative algorithm which would put less weight on the negative nutrients might be less sensitive to such misreporting.

A brief methods section presented two new aggregation techniques based on the above assumptions. Cox regressions were run to analyse the predicting ability of these new algorithms. Similarly to the previous chapter, the results for the SAIN,LIM model were presented in appendix 6.

9.1 Methods

9.1.1 The "Recommended Food Score" aggregate score

The first new aggregating method was based on the recommended food score (RFS) developed by Kant et al. (2000) which had been linked to prospective health outcomes in various studies (Kant *et al.*, 2000; Michels & Wolk, 2002; Kant *et al.*, 2004; Kaluza *et al.*, 2009). The RFS was simply the number of healthy foods and

drinks (i.e. WXYfm overall score under 4 for foods and under 0 for drinks; quadrant 1 for SAIN,LIM) reported to be consumed more than once a week. It was therefore very similar to the food variety score (FVS) introduced in chapter 8 but did not take into account the less healthy foods. The threshold used for the WXYfm model was the one recommended by the Food Standards Agency and used by Ofcom to regulate TV advertising (see chapter 2 for more details). The RFS(WXYfm) acronym was used for the WXYfm derived RFS. For the SAIN,LIM model, the RFS(SAIN,LIM) which counted foods in the first quadrant was used (results presented in appendix 6.1).

9.1.2 The EWS+ aggregate score

The second aggregation method relied on an altered NP algorithm: it did not use the negative components. The EWS+ aggregate score thus followed a similar algorithm to the EWS, but depended exclusively on the positive components points of the WXYfm model (appendix 6.3 presents the classification of FFQ-items using the WXYfm positive components only). A similar score was applied to the SAIN,LIM model, EW(SAIN) which used the SAIN sub-score only (appendix 6.1).

9.1.3 Survival analysis models

Quartiles of the two new aggregate scores were used to fit Cox proportional hazards regressions. Two models were used: chapters' 6 and 7 model 1 (adjusted for age, sex and ethnicity) and model 3 (fully adjusted). The reference group remained the first quartile for all aggregate scores.

9.2 Results

Within the complete-cases analysis sample (n=7,251), the RFS(WXYfm) had a mean value of 26.8 and ranged from 0 to 56. It was slightly correlated to the EWS (rank correlation = -0.21). Similarly, the EWS+ was moderately correlated to the EWS (r=-0.22) highlighting that the EWS depended more on the negative components.

Table 9.1 presents Cox regressions estimates for the RFS(WXYfm) aggregate score. Significant reduced incidence of all-cause mortality was obtained for all quartiles in model 1, and remained significant for the second quartile in model 3. Some significant risk reduction of cancer mortality was also observed for the third quartile in model 1. Quadratic trend tests were significant for both cancer and all-cause mortality outcomes. For CHD, linear risk reduction was slightly suggested, while no clear trend could be highlighted for diabetes. Results for the SAIN,LIM derived RFS(SAIN,LIM) score were very similar (appendix 6.2), which was explained by the high correlation between RFS(WXYfm) and RFS(SAIN,LIM) (r=0.94).

Estimates for the EWS+ aggregate score are presented in table 9.2. The results contrasted clearly with the original EWS since inverse and significant linear trends appeared for all outcomes in model 1. Such risk reduction confirmed the original hypothesis that diets containing more nutrient dense foods would be protective against prospective health outcomes. When adjusting the model for the full range of confounding variables most of the trends were attenuated, except for the CHD outcome. Risk reduction of diabetes, cancer and all-cause mortality was still suggested, with close to significant linear trends for diabetes and all-cause mortality. Quadratic trend tests were significant for all-cause mortality in model 3, but it did not appear clearly in the individual quartile estimates. Protective trends were not obtained with the EW(SAIN) aggregate score (appendix 6.2), which concurred with the results from appendix 3.4 (individual components). It suggested that the choice of positive components and pointing system of the WXYfm model might be more adequate for the Whitehall II data.

Table 9.1: Cox regression estimates across quartiles of the RFS(WXYfm) aggregate score (4: healthier)

Outcome (cases / n)	Quartile/ trend	Model 1			Model 3		
		HR	95	% CI	HR	95	% CI
CHD (318 / 7,174)	1	Ref			Ref[#]		
	2	0.90	0.65	1.23	0.91	0.66	1.25
	3	0.79	0.58	1.07	0.88	0.63	1.22
	4	0.83	0.61	1.14	0.92	0.65	1.31
	Linear	0.93	0.84	1.03	0.97	0.87	1.09
	p quadratic	0.410			0.457		
Diabetes (754 / 6,868)	1	Ref*			Ref*		
	2	0.89	0.72	1.10	0.90	0.73	1.12
	3	0.83	0.67	1.02	0.89	0.72	1.11
	4	1.06	0.87	1.29	1.06	0.85	1.33
	Linear	1.02	0.95	1.08	1.02	0.95	1.10
	p quadratic	0.704			0.752		
Cancer mortality (251 / 7,235)	1	Ref			Ref		
	2	0.76	0.53	1.09	0.79	0.55	1.13
	3	**0.67**	**0.47**	**0.95**	0.72	0.50	1.04
	4	0.77	0.55	1.08	0.83	0.57	1.22
	Linear	0.91	0.81	1.02	0.94	0.83	1.06
	p quadratic	**0.004**			**0.006**		
All-cause mortality (524 / 7,242)	1	Ref			Ref[§]		
	2	**0.71**	**0.55**	**0.92**	**0.75**	**0.58**	**0.97**
	3	**0.72**	**0.57**	**0.91**	0.80	0.62	1.02
	4	**0.77**	**0.61**	**0.97**	0.86	0.66	1.12
	Linear	0.92	0.85	1.00	0.96	0.88	1.05
	p quadratic	**0.001**			**0.002**		

Model 1 adjusted for age, sex, and ethnicity. Model 3 further adjusted for marital status, employment grade, smoking status, physical activity level, and energy and alcohol intake, BMI, hypertension and dyslipidaemia status, and prevalence of longstanding illness.
[#] Stratified for BMI categories * Stratified for sex. [§] Stratified for longstanding illness and dyslipidaemia. HR, hazard ratio; CI, confidence interval.

Table 9.2: Cox regression estimates across quartiles of the EWS+ aggregate score (4: healthier)

Outcome (cases / n)	Quartile/ trend	Model 1			Model 3		
		HR	**95**	**% CI**	**HR**	**95**	**% CI**
CHD (318 / 7,174)	1	Ref			Ref[#]		
	2	**0.67**	**0.49**	**0.90**	**0.71**	**0.52**	**0.96**
	3	**0.66**	**0.48**	**0.89**	**0.70**	**0.51**	**0.96**
	4	**0.64**	**0.47**	**0.87**	**0.66**	**0.48**	**0.90**
	Linear	**0.86**	**0.78**	**0.95**	**0.88**	**0.79**	**0.97**
	p quadratic	0.333			0.462		
Diabetes (754 / 6,868)	1	Ref*			Ref*		
	2	**0.80**	**0.66**	**0.98**	0.84	0.68	1.03
	3	**0.81**	**0.67**	**0.99**	0.89	0.72	1.09
	4	**0.79**	**0.65**	**0.96**	0.84	0.68	1.03
	Linear	**0.93**	**0.87**	**0.99**	0.95	0.89	1.02
	p quadratic	0.440			0.819		
Cancer mortality (251 / 7,235)	1	Ref			Ref		
	2	0.79	0.56	1.10	0.91	0.64	1.28
	3	0.78	0.56	1.09	0.94	0.66	1.34
	4	**0.64**	**0.45**	**0.92**	0.78	0.53	1.13
	Linear	**0.87**	**0.78**	**0.98**	0.93	0.83	1.05
	p quadratic	**0.024**			0.073		
All-cause mortality (524 / 7,242)	1	Ref			Ref[§]		
	2	**0.75**	**0.59**	**0.95**	0.87	0.68	1.10
	3	**0.77**	**0.61**	**0.97**	0.91	0.71	1.17
	4	**0.70**	**0.55**	**0.89**	0.81	0.63	1.05
	Linear	**0.90**	**0.83**	**0.97**	0.94	0.87	1.02
	p quadratic	**<.001**			**0.006**		

Model 1 adjusted for age, sex, and ethnicity. Model 3 further adjusted for marital status, employment grade, smoking status, physical activity level, and energy and alcohol intake, BMI, hypertension and dyslipidaemia status, and prevalence of longstanding illness.
[#] Stratified for BMI categories * Stratified for sex. [§] Stratified for longstanding illness and dyslipidaemia. HR, hazard ratio; CI, confidence interval.

9.3 Discussion

The two new aggregating algorithms were applied to verify in practical terms the results from chapter 8 which identified three potential explanatory factors for the U-shaped associations obtained in chapters 6 and 7.

First, the RFS aggregate score was very similar to the Food Variety Score used in chapter 8 but it only included healthy foods. It did not rely on the exact reported amount and the energy density of each item. It was therefore less sensitive to dietary misreporting than the EWS and PES(Q1) aggregate scores. Compared to EWS and PES(Q1), the estimates for both RFS(WXYfm) and RFS(SAIN,LIM) were more in accordance with the original hypothesis: attenuated quadratic trends and significant risk reduction for some individual quartiles. These results confirmed that the original aggregate scores developed in chapter 4 were too sensitive to dietary misreporting and that diet variety was intrinsically linked to prospective chronic disease risk. The RFS may therefore be a more appropriate aggregating algorithm for FFQ data. However, the risk reduction observed in table 9.1 and appendix 6.2 was attenuated by further adjustment, and the quadratic trend tests remained significant for cancer and all-cause mortality. The RFS algorithm did not account for the effect of differential under-reporting of unhealthy foods on the negative components (chapter 8), which might explain the quadratic trends observed for cancer and all-cause mortality.

Second, the EWS+ aggregate score, by including the positive components only, aimed at removing the confounding effect of differential under-reporting of unhealthy foods on the association between negative components and prospective health status. The estimates from table 9.2 confirmed that the combination of the WXYfm positive components alone could predict reduced incidence of all outcomes. Model 1 results followed the original hypothesis, with significant risk reduction for all outcomes. Despite the attenuation in the fully-adjusted model, these results showed that the NP approach could be predictive of lower chronic disease risk, and that NP could therefore represent a relevant public health tool. The results further confirmed chapter 8 observations that the weighting of negative components in

WXYfm and the derived EWS aggregate score emphasised the information bias within the Whitehall II data, entailing the U-shaped associations of chapter 6. The EW(SAIN) aggregate score displayed weaker associations and no significant linear trends. This suggested that the positive components and associated target values included in the WXYfm model might be more appropriate for the British Whitehall II population. In particular, the fruit, vegetable, and nuts component which predicted independently lower risk of adverse health outcomes (appendix 7).

However, the analyses conducted in this chapter were subject to some limitations. The WXYfm NP model had to be altered for the implementation of the EWS+ aggregate score. This new algorithm yielded a new classification of foods quite different from the original one (appendix 6.3), and therefore did not reflect the WXYfm model. The same limitation applied to the EW(SAIN) score which did not use the full SAIN,LIM algorithm. The results for the WXYfm positive components did indicate that a NP model could predict prospective health outcomes as hypothesised without including negative nutrients.

The RFS aggregate score was essentially a diet variety score, and it was not possible to determine the extent to which the Cox regressions results depended either on the selection of healthier foods or on the increased variety. The results obtained for the Food Variety Score in chapter 8 suggest that most of the protective effect of RFS was linked to diet variety. Also, The RFS algorithm relied on the arbitrary healthiness thresholds of the WXYfm and SAIN,LIM models (chapter 2). For WXYfm, the threshold used by Ofcom and the Food Standards Agency was implemented, but a more or less restrictive value could have been chosen. Such an analysis was carried out using the more restrictive "healthier" threshold (i.e. foods scoring below 0 on the overall score scale—chapter 2) and results are presented in appendix 6.4. Such RFS(healthier) aggregate score was very highly correlated to RFS(WXYfm) (r=0.99), which was translated into similar survival analysis estimates. Stronger associations were obtained for cancer and all-cause mortality with the RFS (healthier), suggesting that the reduced number of healthy foods allowed identifying better healthier dietary patterns.

9.4 Conclusion

Results from this chapter brought more evidence that the factors identified in chapters 8 were all explaining part of the U-shapes. The influence of the aggregating method was well illustrated in tables 9.1 and 9.2 which displayed different trends compared to the original EWS aggregate score. The consequences of energy misreporting on the negative components particularly affected the WXYfm model which put more weight on these negative components. The estimates obtained with the positive components of the WXYfm model confirmed that the NP approach could predict adequately future health status and therefore represent a relevant public health tool.

The information gathered in all the results chapters gave considerable insight into the mechanisms linking the two NP models to prospective health outcomes. All the elements of conclusions presented above and in the preceding chapters were used to discuss the potential predictive validity of the WXYfm and SAIN,LIM NP models.

Chapter 10: Discussion

Food nutrient profiling (NP) aims at "categorising foods according to their nutritional content" (Rayner *et al.*, 2004a), using selected 'positive' nutrients e.g. fibre, protein, and 'negative' nutrients e.g. sodium, saturated fat. Applications of NP are numerous and aim at helping consumers make healthier choices. This food-based concept could represent the key missing link between general dietary guidelines and nutrient recommendations (Darmon, 2009). Yet, only one patented and not-publicly available model, ONQI, was related to prospective chronic disease risk (Chiuve *et al.*, 2011). The results obtained with the ONQI model need confirmation using publicly available models if NP is to become a recognised public health tool. The WXYfm and SAIN,LIM NP models were developed for the British and French food safety agencies, respectively (chapter 2). Their development was an open peer review process and their algorithm is publicly available. British civil servants of the Whitehall II cohort study have completed detailed dietary assessment questionnaires at baseline (1991-93), and have been followed-up until December 2009 for incident CHD and diabetes, and until January 2010 for mortality events (chapter 4).

The main aim of the project was therefore to assess the predictive validity, i.e. associations with prospective health outcomes, of the WXYfm and SAIN,LIM nutrient profiling (NP) models within the Whitehall II study dataset. The hypothesis was that diets containing higher proportion of healthy foods as defined by the NP models would be associated with lower incidence of chronic disease.

Five research objectives were defined to assess the predictive validity of the two NP models (chapter 3). Results and conclusions for each research objective are summarised in this final chapter. Through the analysis of the project's strengths and limitations and comparisons with existing evidence, it is possible to identify the implications of the project's results, as well as the steps needed to further develop and validate NP as a public health tool.

10.1 Summary of results

The first step of the project consisted in the application of the two NP models to the Whitehall II data: WXYfm and SAIN,LIM scores were calculated for all non-alcohol items in the food frequency questionnaire (FFQ). In order to test for predictive validity of the two models, a summary measure (or index) had to be created for each participant since foods are not directly related to health outcomes. This aggregated NP score, referred to in this project as "aggregate score", was meant to reflect the food-based NP concept at the whole diet level. In particular, the aggregate score should discriminate individuals according to the variations in FFQ-items intake. Using this rationale, two energy-weighted aggregate scores were designed based on the NP score and the reported intake of each FFQ-item: EWS for the WXYfm model, and PES(Q1) for the SAIN,LIM (chapter 4).

Prior to testing for predictive validity of the two NP models, via their respective aggregate scores, it was necessary to verify that the EWS and PES(Q1) translated adequately the NP concept at diet level. The results discussed in chapters 5 and 7 showed that the two aggregate scores were positively and significantly associated with improved dietary patterns, with particularly strong relationships for the fruit and vegetables. A strong positive association was also observed for dairy products, while an inverse one was obtained for the snacks and sweets food group. The associations between aggregate scores and reported intakes were shown to be at the FFQ-item level, thus confirming that EWS and PES(Q1) were adequate to assess the predictive validity of their respective NP model.

The improved dietary patterns observed in participants classified as healthiest by the aggregate scores resulted in positive associations with existing dietary quality measures (dietary clusters and Alternative Healthy Eating Index). These associations were relatively weak indicating that the aggregate scores were not simple copies of the existing measures. Participants classified in the 'healthier' quartiles of EWS and PES(Q1) tended to have higher BMI, blood pressure, and blood levels of inflammatory biomarkers. These surprising results called attention to the need of

carefully considering potential confounding factors that may affect the predictive validity testing process.

Both aggregate scores were included in multiply-adjusted Cox regression models to assess the predictive validity of the WXYfm and SAIN,LIM NP models—the main aim of the project (chapters 6 and 7). The hypothesised risk reduction was observed in participants classified in the middle quartiles of the aggregate scores. The significant PES(Q1) estimates suggested an almost 30% reduced risk of incident CHD and of all-cause mortality. Unexpected findings were obtained for participants classified as healthiest by the two aggregate scores whose risk estimates were not different from the least healthy individuals. This led to U-shaped associations well illustrated by the quadratic trend tests that were significant for all outcomes except diabetes. Therefore, predictive validity of the WXYfm and SAIN,LIM models was only partly established using the energy-weighted aggregate scores. Similar results were obtained with aggregate scores weighted by portion size or weight of intake. These unexpected results guided the last steps of the project. Three potential explanatory factors for the observed quadratic trends were assessed thoroughly (chapter 8).

First, energy misreporting was detected using the Goldberg cut-off method. Almost 30% of participants were found to be energy misreporters, most of them being energy under-reporters. Because low energy reporting was associated with healthier reported intakes, low energy reporters were more likely to be *mis*classified as healthiest by the EWS and PES(Q1) aggregate scores. Low energy reporting was also associated with less favourable risk factor profiles, mainly high BMI and hypertension. A strong association between BMI and energy reporting had been previously observed in the Whitehall II cohort (Stallone *et al.*, 1997; Brunner *et al.*, 2001) and in other studies (Macdiarmid & Blundell, 1998; Livingstone & Black, 2003). Individuals with higher BMI necessitate higher energy intake to maintain their body weight, at constant physical activity level. Yet, social acceptance norms and self-perception of what constitutes a healthy diet usually leads to under-reporting of food intake by over-weight or obese participants (Livingstone & Black, 2003). These individuals, with higher vascular risk, are therefore more likely to be low-energy reporters. Further, Whitehall II low-energy reporters were more likely to be

of low employment grade (Brunner *et al.*, 1997; Brunner *et al.*, 2001) which had not been consistently observed before (Macdiarmid & Blundell, 1998). Within the Whitehall II cohort, low employment grades were associated with higher prevalence of vascular risk (Marmot & Brunner, 2005). This would further explain the association between energy misreporting and vascular risk which would have confounded the prospective associations between the aggregate scores and health outcomes. This assumption was confirmed by sensitivity analyses in which energy misreporters were excluded. Compared to the original Cox regressions, the sensitivity analyses yielded attenuated U-shapes with lower hazard ratio estimates for the fourth quartile of both aggregate scores.

The misclassification of low energy reporters by the aggregate scores was mainly due to differential under-reporting of energy-dense and less healthy foods, as shown by the diet diary data. Being energy-weighted, the EWS and PES(Q1) were particularly sensitive to the exact reported amounts of these energy-dense foods. It was not possible to fully correct for this systematic differential misreporting of foods since some assumptions necessary for the regression calibration model were likely to be flawed (e.g. normal distribution of intakes for foods rarely consumed).

Second, the EWS and PES(Q1) aggregate scores were not designed to take into account diet variety because this is a characteristic of the diet, independent from the food-based NP concept. Yet, participants with healthier aggregate scores had slightly lower diet variety, which could have entailed the U-shapes. Diet variety did not confound the predictive validity results. Instead, it acted as an effect modifier since a relationship between the aggregate scores and chronic disease appeared only among participants with low diet variety. For a separate analysis of the prospective effects of the aggregate scores and diet variety it would be necessary to work with controlled environments, i.e. in which either the aggregate score or diet variety is fixed. The Whitehall II is a free-living cohort, and controlling for such factors was only possible in a statistical way, further showing the difficulty of translating the NP concept at the diet level.

Third, the nutrients or food characteristics included in the WXYfm and SAIN,LIM models, referred to as components, were analysed separately. For most components,

crude intake was associated with prospective chronic disease risk as hypothesised (i.e. protective effect of the positive components and increased risk with the negative components). The exceptions were the protein and calcium positive components moderately associated with increased risk of CHD and diabetes, and the negative sugar component associated with a non-significant reduced risk of cancer and all-cause mortality. However, when applying the aggregate scores algorithms to individual components, i.e. calculating energy-weighted "component scores", protective associations were suggested between the negative "component scores" and prospective chronic disease. In particular, the sodium and saturated fat component scores were protective against incident CHD. These unexpected results further showed the influence of the aggregating algorithm on prospective associations between dietary intake and health outcomes. Analyses in chapter 8 also showed that the negative component scores were strongly associated with energy under-reporting, i.e. participants with lower negative component scores were more likely to be energy under-reporters. Therefore, the unexpected results for the negative component scores appeared to be artefacts due to their high sensitivity to energy under-reporting—itself associated with increased vascular risk—which is consistent with the results obtained with the CHD outcome.

In addition, the aggregate scores were not correlated evenly with all their components. Similarly to the WXYfm algorithm that puts more emphasis on the negative components, the EWS relied mainly on the sodium, saturated fat, and sugar components. It was not related to the fibre and protein components, and moderately to the fruit, vegetable, and nuts component. Hence, the unexpected results observed above for the negative components were emphasised by the EWS aggregate score, which would have entailed the U-shaped associations. Contrary to EWS, the PES(Q1) relied more evenly on all its components. This would have explained the stronger risk reductions observed in chapter 7 for participants in the middle quartiles of PES(Q1). The quadratic trends for PES(Q1) would have been linked to the unexpected results for both the negative components and the protein and calcium components.

In chapter 9, the original NP models and respective aggregate scores were modified to verify the hypothesised effect of dietary misreporting, diet variety, and the

unexpected effect of the negative components. The first alternative aggregate score, the Recommended Food Score (RFS), was very similar to a variety score but it only included healthy foods as defined by the NP models. It was not weighted by energy intake and did not rely on the exact reported amounts for each FFQ-item. It was therefore less sensitive to dietary misreporting. When applied to WXYfm and SAIN,LIM, the RFS suggested a linear risk reduction of all-cause mortality, but null-associations could not be rejected. Quadratic trends were not significant for CHD. The RFS results confirmed that (i) diet variety played a crucial role in the prediction of future health status, and (ii) the original aggregate scores were over-sensitive to dietary misreporting. The second alternative, EWS+ for WXYfm and EW(SAIN) for SAIN,LIM, was similar to the EWS except that it included only the positive components of the NP models. For EWS+, clear protective trends were observed for all outcomes in the least adjusted Cox model. The association was robust to adjustment for the CHD outcome. These results confirmed that when the effect of differential under-reporting of unhealthy foods—which mainly affected the negative components—was removed, the NP approach could predict adequately future health status. Given such information bias within the Whitehall II data, the EWS relied too much on its negative components, which led to the unexpected U-shaped associations. The results were not so conclusive for EW(SAIN): no significant linear trend was observed. This coincides with the results obtained for individual components and suggested that the WXYfm positive components were more adequate to predict health outcomes in the Whitehall II data.

All the results presented above brought considerable insight into the mechanisms linking NP to prospective health outcomes within the Whitehall II data. The findings were consistent with a possible protective effect of NP-derived aggregate scores on disease risk, but the ultimate hypothesis of the project (that consumption of healthy foods as defined by NP models WXYfm and SAIN,LIM is protective against adverse health outcomes) was only partly established. Information bias within the Whitehall II data, in particular the association between energy under-reporting and vascular risk, appeared to be the main reason for this. The NP models which include energy density in their calculations and the aggregation techniques weighted by energy intake emphasised differential misreporting of energy dense unhealthy foods,

resulting in misclassification of some participants and quadratic trends due to the higher vascular risk profile of energy under-reporters.

10.2 Strengths and limitations

This section presents the strengths and limitations of four key aspects of the project: data, NP models, aggregate scores, and analysis strategy.

10.2.1 Whitehall II data

(i) Cohort study and predictive validity

The Whitehall II data used for this project were ideally suited to answer the initial research question regarding the predictive validity of NP. The study's longitudinal design, with dietary assessment at baseline and almost complete follow-up of incident events, was its main strength. Moreover, regular contact with participants and their perceived benefit of regular health check-ups kept attrition relatively low (Marmot & Brunner, 2005). Follow-up for incident cases followed rigorous protocols using self-reported doctors' diagnoses and oral glucose tolerance tests for diabetes; self-report, doctors' diagnoses, and electro-cardiograms for CHD; and the National Health Services death and electronic patient records for fatal events. The study was updated recently: data for diabetes and CHD were censored in December 2009 and for mortality outcomes in January 2010. Numerous covariates were measured using standard procedures, and the use of well-specified civil service grades made the measure of socio-economic position reliable.

However, the target population of the Whitehall II study—middle-aged white-collar civil servants—is not fully representative of the British population (Marmot & Brunner, 2005). The Whitehall II is an occupational cohort in which fitter individuals are more likely to be over-represented, as suggested by the healthy worker effect (Li & Sung, 1999) and as confirmed by the selection bias observed in chapter 4. Therefore, the sample used in our analysis was likely to be more homogeneous

compared to the general population and the generalisation of the present results would need to be confirmed, in particular using data less prone to dietary misreporting. Yet, previous investigations of the diet-disease relationship in the Whitehall II study were in line with the vast body of evidence linking dietary patterns to health outcomes (Brunner *et al.*, 2008; Akbaraly *et al.*, 2011), which suggested that the unexpected associations of the NP models with prospective disease risk may be replicated in alternative datasets.

(ii) Post hoc power and sample size calculations

The Cox proportional hazards regressions implemented in chapters 6 to 9 yielded wide confidence intervals which suggested a low statistical power, i.e. the ability to reject the null hypothesis when the null hypothesis is false. To assess the power of the present analyses, the SAS power procedure was used by including the hazard ratio estimates obtained in tables 6.1 and 7.11, the survival rates for the four outcomes, and the total number included in each model. Considering the quartile analysis, statistical power ranged from 0.06, for the EWS aggregate score and cancer mortality, to 0.76 for PES(Q1) and all-cause mortality (table 10.1). Such figures were relatively low compared to the 0.9 targeted by most study designs and indicated that the Whitehall II sample used for the present analysis was too small to detect the hypothesized effect with sufficient power. The SAS power procedure was further used to estimate the total sample size needed to achieve a power of 0.9 with the present Cox regression results. The sample size estimates presented in table 10.1 followed the power calculations, with the lowest requested sample size being for all-cause mortality and the PES(Q1) aggregate score (n=10,824). Such sample size was just above the total number of participants recruited in the Whitehall II study (n=10,308). Dietary assessments were only included at phase 3 of the study (chapter 4) which limited the potential number of participants to those still followed-up at phase 3. Power to detect weaker relationships within the Whitehall II data would therefore be higher with variables measured from phase 1.

Table 10.1: Post hoc power and sample size calculations

Outcome	CHD		Diabetes		Cancer mortality		All-cause mortality	
(events / n)	(318 / 7,174)		(754 / 6,868)		(251 / 7,235)		(524 / 7,242)	
	EWS	PES(Q1)	EWS	PES(Q1)	EWS	PES(Q1)	EWS	PES(Q1)
Power*	0.34	0.54	0.21	0.24	0.06	0.39	0.26	0.76
Sample[$]	32,108	17,732	54,684	45,692	610,000	26,800	43,912	10,824

* Based on the strongest hazard ratio estimate (model 1) of table 6.1 for EWS and table 7.11 for PES(Q1).
[$] Estimated sample size to achieve a power of 0.9.

(iii) Food frequency questionnaire

The most important limitation of the Whitehall II data was the dietary assessment tool. Food frequency questionnaires (FFQ) have been the most commonly used method in large scale nutritional studies for efficiency and practical reasons. They have been validated to assess usual dietary patterns and have been shown to relate well to alternative dietary assessment methods (chapter 2). Yet, for the specific analysis of NP, the FFQ may not be the best tool.

Participants reported intake of items rather than individual foods. Items regrouped several foods of similar characteristics but their nutrient composition could vary greatly, for instance, the saturated fat content of the "Yoghurt" and "Chicken and other poultry" items. It is therefore possible that the calculated NP scores of the FFQ-items did not reflect the actual foods consumed by participants. In which case, aggregate scores were likely to have high random error. Such random error could differ between NP components since the FFQ put more emphasis on fruit and vegetables (34 items) compared to other groups, for example, meat and fish (16 items) or snacks and sweets (12 items). This would have affected particularly the energy, saturated fat, sodium, sugar, and protein NP components whose content is highly variable in the meat and snacks group. In contrast, the error for the fibre component may have been smaller since it was mainly present in the fruit and vegetables or in wholemeal products, which were distinguished more precisely in the FFQ. In addition, there was no specific item for "mixed dishes" and their reported intake depended on the participants' interpretation. Some individuals might have

reported all the ingredients of a beef stew, while others would only report the beef, or just the vegetables. As a result, aggregate scores were likely to have a relatively high random error, which could have led to underestimate some associations.

Furthermore, the uneven amount of FFQ-items in the different food groups could be an explanatory factor for the widespread misreporting of intakes. The emphasis on fruit and vegetables might have encouraged participants to report a higher intake for this food group because many individuals were likely to consume more than once a month all 34 fruit and vegetables items. In comparison, if a participant reported having consumed all twelve items under snacks and sweets, it would have resulted in a smaller total intake for this food group. If assumed to be constant among participants, the aggregate scores rankings may not have been modified by this systematic bias.

(iv) 7-day diary data

Misreporting of dietary intakes in the FFQ was confirmed by the use of the 7-day diary (7DD) data (chapter 8). Results were in line with the validation study of Brunner et al. (2001) which indicated that the FFQ over-estimated plant-based micro-nutrients, i.e. fruit and vegetable intake.

Diet diaries have been shown to relate better to true intake (Kipnis *et al.*, 2003; Prentice *et al.*, 2011) and to be more adequate to identify existing or non-existing diet-disease relationships (Day *et al.*, 2001; Bingham *et al.*, 2003; Freedman *et al.*, 2006; Spencer *et al.*, 2010; Hutchinson *et al.*, 2011; Key *et al.*, 2011). In addition, dietary intakes are reported at the food level, rather than for pre-specified items, which makes diet diaries an ideal tool for NP validation. Yet, only 1,350 diet diaries were coded by the Whitehall II study team (appendix 7) and survival analysis models could not be conducted on such a limited sample.

Instead, the 7-day diary (7DD) data were used in regression calibration models to evaluate and correct differential misreporting for each of the FFQ-items. Most results were coherent and in line with previous observations (chapter 8), but the 7DD and the FFQ are measuring different aspects of dietary intake, which led to some

inconsistencies. In particular, many items included in the FFQ were not necessarily consumed by participants within the weekly period of the 7DD. For such items, the 7DD data included a majority of non-consumers, resulting in non-normal distributions. The assumptions beneath the regression calibration model were flawed and the method could not be retained.

10.2.2 Nutrient profiling models WXYfm and SAIN,LIM

The two NP models used in the project were mainstream models designed by national food safety agencies for regulatory purposes. They had been previously included in many validation studies (chapter 2). Testing their predictive validity appeared most relevant both in terms of scientific and public health interest.

The WXYfm and SAIN,LIM models are publicly available and do not require extensive nutrient composition tables to apply. It should be possible to apply these models on most existing datasets, including alternatives of the Whitehall II study. As shown in chapter 9, the two models were easily modified. The WXYfm model has already been adapted for use in Australia and New Zealand for the regulation of health claims, an application it was not originally intended for (chapter 2).

WXYfm and SAIN,LIM are both across-the-board NP models. While this may reflect general dietary guidelines, it leads to a systematic association between the aggregate scores and diet variety (chapter 8). As highlighted in chapter 2, healthy dietary patterns always include foods considered as less healthy. This has been underlined by several validation studies in which unhealthy foods were consumed in association with healthier ones (e.g. butter and jam with wholemeal bread). Therefore, category-specific NP models, which select healthier options within food categories, might be a more favourable approach to identify realistic healthy dietary patterns. Yet, the number of food categories should be carefully determined. A model with too many food groups may not be able to correctly identify healthy foods, whereas a model with a limited number of categories may be a realistic and practical approach to promote healthy and varied diets (Scarborough *et al.*, 2010).

The developers of the SAIN,LIM model recently designed an alternative category-specific SAIN sub-score (Lesturgeon *et al.*, 2011). Testing for predictive validity of such a category-specific model could have confirmed the above assumption.

WXYfm and SAIN,LIM used thresholds determined arbitrarily to define the foods which could be advertised or could carry a claim. The WXYfm model developers did not explain how the healthiness categories thresholds were determined. For such reason, the thresholds were not used in the EWS aggregate score that relied on the "overall score" scale of WXYfm. In chapter 9, the RFS aggregate score which used the thresholds was applied to the WXYfm model. A more systematic approach, for instance using sensitivity analysis, could have tested the validity of these thresholds. Yet, this was not prioritised considering the unexpected findings obtained with EWS. The bi-dimensional aspect of the SAIN,LIM model made it necessary to use the "quadrants" thresholds to combine the SAIN and LIM sub-scores. The developers of the SAIN,LIM model did justify the thresholds (chapter 2): they were based on the assumption that if a food represented the whole dietary intake, it needed to reach the average of 100% of the recommended intake for positive components (SAIN threshold), and be under the maximum limit for the negative components (LIM threshold). Such a rationale could be argued since no single food could realistically represent 100% of one's intake. Aggregate scores similar to the WXYfm-derived EWS were applied separately to the SAIN and LIM sub-scores. Their combination would have required alteration of the original NP algorithm.

Further, the WXYfm model puts more weight on its negative components and was therefore more affected than SAIN,LIM by the unexpected associations obtained with the negative components (chapter 8). While differential misreporting of unhealthy foods explained these artefacts, they raised the issue of the balance between positive and negative nutrients in NP algorithms. Negative nutrients have been the focus of many NP models and associated public health policies, but the present results suggest that the focus cannot be put on negative nutrients alone.

10.2.3 The EWS and PES(Q1) aggregate scores

The choice of an aggregating algorithm revealed to be a crucial step, necessary to test the predictive validity of NP models. Therefore, the validation of the WXYfm and SAIN,LIM models was an indirect process, and the first research objective of the project consisted in designing aggregate scores for both NP models. The EWS and PES(Q1) aggregate scores used straightforward algorithms which followed approaches used in previous studies (Arambepola *et al.*, 2008; Fulgoni *et al.*, 2009; Chiuve *et al.*, 2011). They were shown to discriminate participants with respect to their reported intake at the FFQ-item level, in line with the NP concept. The EWS and PES(Q1) algorithms were adaptable and could have been used with many different NP models.

However, the choice of the aggregate scores algorithms was arbitrary, and these were shown to influence the predictive validity results. In particular, the implementation of the variety-oriented RFS aggregate score did lead to very different prospective associations compared to the original EWS and PES(Q1). The RFS(WXYfm) and the RFS(SAIN,LIM) were highly correlated, as were EWS and PES(Q1); whereas the correlations between the EWS and the RFS(WXYfm), and between the PES(Q1) and the RFS(SAIN,LIM), were weaker. This indicated that the nature of the aggregating algorithm had more influence on the rankings of participants than the differences between the two NP models.

In addition, and as mentioned above (section 10.2.2), the EWS algorithm was not similarly applicable to the two NP models, and it may not be possible to define a universal aggregating method.

10.2.4 Analysis strategy and design

The analysis framework was a clear strength of this project, and it could serve as basis for the subsequent investigation of NP validity. The design of the necessary aggregate score was presented transparently. Several aspects of the validity of

WXYfm and SAIN,LIM were assessed. Identification of potential confounders was done through a systematic approach, and information bias was explored carefully. The alternative aggregate scores of chapter 9 confirmed the results presented in chapter 8, and showed that longitudinal data could be very useful in designing new models.

However, the statistical methods used throughout this study may have affected the results. First, predictive validity was tested on a complete-case sample, which led to selection bias since participants with missing covariate information were more likely to be in poor health (chapter 4). Working with a reduced size sample may have also resulted in limited variations in the exposure variables. While the use of imputation models could have increased the number of observations included in the analyses (Sterne *et al.*, 2009), this was not a priority and time constraints did not allow for a full investigation of this method.

Second, FFQs were completed at phases 3, 5, and 7 but only phase 3 (baseline for the project) data were used. The implementation of Cox regressions with time-varying aggregate scores reflecting the most recent contemporary diet during follow-up may have resulted in stronger associations (Hu *et al.*, 1999). However, the nutrient composition of foods was likely to change in the 10-year period between phase 3 and phase 7, and no updated nutrient composition table was available.

Last, all the statistical methods used in the project were subject to a range of assumptions. Effort was made to make sure the assumptions were met (e.g. proportional hazards for the Cox regressions), but some were likely to be flawed. In particular, the diabetes outcome variable was interval censored, i.e. event information was only available by time intervals. The SAS statistical software used the method developed by Breslow (1974) to approximate the likelihood function in the presence of tied events (i.e. events occurring at the same time).

10.3 Comparison with existing literature

To date, only one other published study has analysed the prospective relationship between NP and health outcomes (Chiuve *et al.*, 2011). The Chiuve study used US-based cohort data. Its study design was essentially the same, therefore this section will focus on the comparison of the NP models, the datasets, and the results of the two studies.

10.3.1 Nutrient profiling models: WXYfm, SAIN,LIM, and ONQI

The use of the government-endorsed WXYfm and SAIN,LIM NP models is one of the strengths of our project. Their validation with respect to prospective health outcomes is an essential step to ensure the public health relevance of these two regulation-oriented models.

The Overall Nutritional Quality Index (ONQI) is the underlying NP model for the commercial NuVal food logo (www.nuval.com). It is aimed at food manufacturers who buy the right to display the NuVal label and score (ranging from 1 to 100) on their food packaging. The ONQI algorithm is patent protected and not publicly available. It has been included in several validation studies which partly described the components of the ONQI model (chapter 2), but no details were given on several key aspects such as the reference amount. It was impossible to assess the model using the Whitehall II data without the prior consent of the ONQI developers. In the same issue in which the Chiuve *et al.* study was published, two commentaries focused on this major limitation, which shows that journals are cautious when publishing results obtained with patented methods (Jakicic, 2011; Reedy & Kirkpatrick, 2011). Reedy and Fitzpatrick indicated that the proprietary nature of ONQI meant that "the tool [could not] be considered as a potential option for public policy intervention" and that such an approach did not promote further research to improve the model. In addition, the ONQI model includes 30 nutrients and adjusting factors (Katz *et al.*, 2010). Extensive composition tables would be required to apply

the ONQI algorithm to the Whitehall II FFQ data. Such a limitation did not apply to WXYfm and SAIN,LIM.

10.3.2 Longitudinal data: Whitehall II, Nurses' Health Study, and Health Professionals Follow-up Study

The Whitehall II study, the Nurses' Health Study, and the Health Professionals Follow-up Study are all occupational cohorts with dietary assessment at baseline. Therefore, these datasets shared the same strengths and limitations associated with the study design. They were particularly suited to test for predictive validity of NP.

The Nurses' Health Study was set up in the United States in 1976: 121,700 registered female nurses aged 30 to 55 years were enrolled (Willett *et al.*, 1987). As for Chiuve's study, the baseline was the FFQ assessment made in 1986. 62,284 women were included in the study, with a total of 20,004 chronic disease events. 1986 was also the baseline for the Health Professionals Follow-up Study. A mailed FFQ was returned by 51,529 men aged 40 to 75 years (Colditz *et al.*, 1991). From this sample, Chiuve included in her analysis 42,382 participants with 13,520 chronic disease events in total. The population size of the two US studies therefore largely exceeds the 7,251 Whitehall II civil servants included in the analyses of our project. Likewise, the statistical power of Chiuve's study was without comparison. The range of confidence intervals between the two projects clearly illustrated these differences.

Additionally, the dietary assessment tools used in the two US studies have been thoroughly validated (appendix 1) and were used as basis for the Whitehall II FFQ (Willett, 1998; Brunner *et al.*, 2001). The nutrient content data was more detailed, which allowed Chiuve and colleagues to apply the refined ONQI algorithm.

Both US-based longitudinal datasets were at the heart of the development of chronic disease based nutritional epidemiology. They were used in many occasions including the link between dietary fat and CHD (Hu *et al.*, 1997), the validation of the Healthy Eating Index (HEI) and the development of the AHEI (McCullough *et al.*, 2000a; McCullough *et al.*, 2000b; McCullough *et al.*, 2002), the development of energy

adjustment methods to deal with global energy misreporting (Willett & Stampfer, 1998), and the use of repeated dietary assessments (Hu *et al.*, 1999). The use of the Nurses' Health Study and the Health Professionals Follow-Up study therefore appeared as a specific strength of Chiuve's investigation.

10.3.3 Predictive validity results

The age of Nurses and Health professionals at the baseline of both US cohorts (between 40 and 75 years) was reasonably similar to the age range in the Whitehall II sample (39 to 63 years). Despite very different environments, the results from both studies could therefore be compared to some extent.

With the ONQI-f (the ONQI aggregate score weighted by portions of intake), protective and linear trends were obtained in both men and women for cardiovascular disease, diabetes, and total mortality, but not for cancer. The size of the protective effect was modest but significant, with the hazard ratios point estimates for the fifth quintile of the ONQI-f ranging from 0.77 to 0.91 for all outcomes except cancer. Such point estimates were in the range of those observed for the middle quartiles of EWS and PES(Q1), though few estimates were significant. When using an aggregate score weighted by energy intake, and therefore more similar to EWS and PES(Q1), Chiuve and colleagues did not obtain significant results (results were not shown by the authors). Implementing an aggregate score weighted by portion size did not change our results (not shown).

Significant or borderline significant risk reduction of total mortality was observed for all the middle quintiles (i.e. second to fourth quintiles) of ONQI-f, replicating the second and third quartiles results of the EWS and PES(Q1) aggregate scores for all-cause mortality. Point estimates obtained in the present analyses indicated greater risk reduction, but the smaller number of events led to wider confidence intervals and non-significant results.

Results for diabetes and cancer did not converge between the two studies. This could be due to the definition of cases which were slightly different. For cancer, our project

included fatal events only whereas Chiuve and colleagues included all diagnosed cancers; both excluded non-melanoma skin cancer. Concerning diabetes, the present project used the 1999 WHO classification while the National Diabetes Data Group and the American Diabetes Association definitions were used in the US study.

Overall the power of the data used by Chiuve and colleagues was a strength of their project which enabled obtaining robust estimates. But the tested ONQI model did not allow for full comparison and remained a strong limitation of Chiuve's study.

10.4 Implications and meanings of project's results

The unexpected results and their explanation highlighted that testing for predictive validity of NP was not a straightforward process. This section presents the possible implications of the project's results for all the actors linked to the development, validation, and application of NP.

10.4.1 Implications for scientists and model developers

The main task for scientists and NP model developers is to verify the present results (see "Areas for further research" section), ideally on alternative data less prone to dietary misreporting. For this purpose, the framework used in the project signposted the aspects which need to be given particular attention.

First, an aggregation method must be determined prior to testing for predictive validity. As shown in chapters 8 and 9, the aggregate score algorithm can influence the final results, and scientists need to be transparent on the chosen algorithms to allow comparisons with other studies. Section 10.2 further suggested that there might not be a universal aggregation method, and comparing the results obtained with different aggregation algorithms could enable understanding better the impact of the aggregate scores in the relationship between NP and prospective health status.

Second, reporting the associations between aggregate scores and food intake (i.e. construct validity) could also be necessary, to confirm that the aggregation methods do reflect the food-based NP concept.

Third, scientists interested in the predictive validity of NP must identify the potential factors that could confound and/or bias the prospective associations. In particular, the aggregation methods should not be systematically associated with energy misreporting which is quite common in large scale nutritional studies.

Fourth, analysing the effects of each one of the NP components may be very helpful to design more efficient NP models. Results from chapter 9 confirmed the observations made in chapter 8 and indicated that the choice of components was an essential aspect of NP models.

In summary, testing for predictive validity entails aggregating the food-based NP concept at diet level which is a challenging task. The aggregation method further needs to circumvent the potential information bias within the dietary assessment data. The use of "results-driven" models (chapter 9) was shown to be an effective way of deriving alternative aggregate scores and/or NP models. Scientists and developers should therefore be encouraged to use longitudinal data when possible.

10.4.2 Implications for regulators and public health policies

NP is currently used as a regulatory tool in several countries including the UK. In the project, the construct and convergent validity results did show that both WXYfm and SAIN,LIM were associated with healthier dietary choices, confirming that NP could contribute towards better intake for key public health nutrients. The predictive validity results previously obtained with the ONQI model further confirmed that NP could be a relevant public health lever to help lower chronic disease health outcomes.

NP is a food-based concept whose rationale lies in the relationship between intake of individual foods and healthiness of the diet as a whole. Global dietary intake has been consistently shown to be related to health status, and the definition of healthy dietary patterns were derived from a considerable body of evidence (chapter 1), including the results of the Whitehall II study (chapter 2). Therefore, the findings of this project do not undermine the validity of the relationship between NP and health outcomes as confirmed by the results obtained in chapter 9 and for the ONQI NP model (Chiuve *et al.*, 2011). They do reveal inherent methodological difficulties in translating the food-based concept at a diet level. Aggregate scores algorithms were shown to influence the predictive validity results, notably via their association with energy misreporting. The NP concept, and the WXYfm and SAIN,LIM NP models which were shown to reduce risk of prospective chronic disease for participants in middle quartiles, could therefore not be invalidated based on the present results. Rather, testing for their predictive validity should be done in alternative datasets less prone to misreporting of dietary intakes.

Regulators also need to take into consideration the population's perception of public health messages. Enforcing NP models which were not proved to predict chronic disease as expected may raise scepticism towards public health policies. The implementation of a regulatory NP model would require that its predictive validity were confirmed to ensure coherent guidelines and acceptance by the general public.

10.4.3 Implications for the food industry

The food industry is likely to be interested in the applications of NP, in particular those concerning consumers' behaviours with respect to food labelling using NP-based criteria. This project has shown that large scale epidemiological studies with semi-quantitative dietary assessment may not be the best tool to validate NP models. As a result, smaller scale studies with more precise dietary assessment methods (e.g. diet diaries or 24h recalls) may be more effective—and less costly than large cohorts—to validate NP-based applications. The food industry should therefore be encouraged to promote the application of NP-based policies via such small scale and

more affordable studies, directly linked to the application of a NP model. Since small-scale studies would not have sufficient power in relation to disease events, case-control designs or alternative markers of health (e.g. contemporary risk factor status) could be used to further investigate the association between NP and health outcomes.

Similarities in the results obtained for the two NP models suggest that other across-the-board models could produce similar results. This would need to be confirmed, but could indicate that category-specific models may be more effective than across-the-board models to promote healthier dietary intakes (see section 10.2). Food manufacturers often produce a range of products belonging to the same food category, and category-specific NP models may be better suited to their portfolios. Therefore, these results should be used by food companies to promote predictive validity research focusing on such category-specific NP models.

10.4.4 Implications for the general public

The general population may be confused by the predictive validity results which did not concur completely with a vast body of evidence and associated messages concerning diet and chronic disease. NP is aimed at classifying foods, not diets, and the communication of the project's results needs to focus on this key difference. Existing public health advice on healthy eating is not invalidated, and consumers must not take these new results as totally conclusive. Also, the two NP models under investigation were not designed to be consumer-facing, and the implications with respect to consumers would be more linked to the applications of the models.

10.5 Areas for further research

Testing for predictive validity is a new step in the assessment of NP models. Hence, further investigations are needed to fully establish predictive validity of the WXYfm and SAIN,LIM models. Specific limitations identified throughout the project need to be addressed by future research.

10.5.1 Data

First, reproducibility of the present results needs to be assessed in alternative longitudinal datasets. An ideal option would be to run similar analyses in the Nurses' Health Study and the Health Professionals Follow-up Study, for cross-comparison with ONQI and for increased statistical power. Other datasets could include the European Prospective Investigation into Cancer and Nutrition (EPIC) that gathered more than half a million participants across ten European countries (International Agency for Research on Cancer), or the UK Women's Cohort Study which includes more than 30,000 middle-aged women (Cade *et al.*, 2004). The assessment of predictive validity in particular regions or countries should ideally be done using local longitudinal data.

Second, the FFQ used in the present project and by Chiuve and colleagues was prone to dietary misreporting and was shown to be a limited tool in the context of NP validation. Contrary to FFQs, diet diaries or 24h recalls provide detailed information on specific foods. If NP models were applied to such type of data, the random error linked to the scoring of FFQ-items by NP models could be limited. Diet diaries are further thought to be less prone to misreporting of intakes than FFQs. However, due to the coding burden, only limited longitudinal data with diet diary or 24h recall assessments are available. Some possible datasets were created by the UK dietary cohort consortium which regrouped data from seven UK cohorts with diet diary assessment (Dahm *et al.*, 2010). Nested case-control datasets were produced to assess relationship between diet and breast, colorectal, and prostate cancers, and could be used to assess predictive validity of NP models with respect to these events.

Third, in the absence of biomarker data, more refined techniques such as structural equation modelling could not be used to correct for dietary misreporting (appendix 7). The availability of a wider range of nutrient biomarkers, and more specifically of recovery biomarkers such as urinary nitrogen or doubly labelled water, could allow recalibrating the reported dietary intakes to obtain prospective associations less confounded by misreporting of intakes (Kaaks, 1997; Subar *et al.*, 2003; Rosner *et al.*, 2008). Yet, suitable biomarkers do not exist for all nutrients or food groups

(Prentice, 2003), and cannot be considered as markers of the intake of individual foods. Testing for predictive validity of NP using existing biomarker data could therefore reduce the random and systematic error associated with self-reported intakes, but it would not replace the use of food-based diet diary or 24h recall data.

In addition, the concurrent validity results of chapters 5 and 7 did indicate that the WXYfm and SAIN,LIM models were not associated with improved risk factors profiles, in line with the predictive validity conclusions. If testing for predictive validity is not feasible due to lack of data or resources, the assessment of NP against biological risk factors should therefore be investigated, though this would not replace the "ideal" predictive validity (Drewnowski & Fulgoni, 2008). Notably, many national dietary surveys now include some cross-sectional measurements, e.g. the "Etude nationale nutrition santé" in France, the National Dietary and Nutrition Survey in the UK, and the NHANES studies in the US (Unité de surveillance et d'épidémiologie nutritionnelle, 2007; Bates *et al.*, 2010; Centers for Disease Control and Prevention & National Center for Health Statistics, 2012). If such cross-sectional associations were to be investigated, potential information bias and confounding factors would need to be given particular attention.

10.5.2 Nutrient profiling models

Both WXYfm and SAIN,LIM models were clearly relevant in terms of public health implications and potential applications, but more research is needed to assess other algorithms. In particular, the assessment of category-specific models would allow testing the hypothesis that such models would promote more effectively varied diets. The Whitehall II data could be used for such analysis, for cross-comparison with the present project. Implementation of category-specific models must be cautious given that the definition of food categories may not always be easily translated at dataset level. Also, the use of a FFQ may not be appropriate if some food categories contain too few items (or no items since FFQs do not necessarily include the full range of available foods).

The construct validity results for the EWS and PES(Q1) aggregate scores confirmed that NP models did not need including many nutrients to be associated with healthier nutrient intakes overall. The ONQI contains 30 components and such increased complexity may have explained the diverging predictive validity results. A detailed investigation into the impacts of different aspects of NP algorithms (e.g. components, reference amount, thresholds for healthiness categories) would allow understanding better which type of model may be more effective to predict health outcomes. A model containing more nutrients would be more costly to implement, but it may be preferable for regulators to have a costly model proven to be linked to prospective health status rather than a non-validated simple model. Optimising the efficiency of a model (i.e. simplicity of the algorithm in conjunction with prediction of prospective health status) would be an interesting step in the development of NP.

10.5.3 Statistical analysis and study design

Chapter 9 showed that results from prospective associations could serve as a basis for model improvement. The development of such results-driven models must be conducted in several cohort studies for generalisation purposes.

Also, the use of repeated dietary assessments should allow obtaining stronger prospective associations (Hu *et al.*, 1999). Yet, the food market being in constant evolution—especially for the manufactured goods, updated nutrient content tables would be necessary to calculate updated NP scores and associated aggregate scores.

Chapters 8 and 9 showed that diet variety acted as an effect modifier in the associations between aggregate scores and health outcomes, playing a crucial role in the prediction of future health status. As a result, the relationship between diet variety and NP needs further investigation. In particular, the impact of across-the-board vs. category-specific NP models on diet variety should indicate which type of algorithm would be more likely to promote healthier dietary patterns, without the use of longitudinal data.

10.5.4 Validity of nutrient profiling models

In the project, the definition of "predictive validity" was limited to prospective associations between a NP model and health outcomes, which may have appeared as the best possible validation (chapters 1 and 2). Yet, some scientists did not necessarily consider such step as essential because it is not related to the application of NP models (Reedy & Kirkpatrick, 2011).

Potential applications of NP are numerous and aim to help consumers adopt healthier diets. For predictive validity to be complete, further research would need to assess the impact of the use of a NP model on the intended application. For example, Ofcom did assess the impact of introducing WXYfm to regulate TV advertising though WXYfm itself was not validated (Office of communications, 2007a); and the analysis of the effect on consumers' behaviours of a NP model appearing on food labels should be considered (Epstein *et al.*, 2010; Katz *et al.*, 2010; Vyth *et al.*, 2010; Muller & Ruffieux, 2011; Temple *et al.*, 2011; Vyth *et al.*, 2011b). Such an investigation would require behavioural data not available in the Whitehall II study.

The present project considered that the first step in predictive validity was to verify whether diets with a higher content of healthy foods (as defined by the NP model) were beneficial. The following step in predictive validity research should be to analyse how NP could facilitate such shifts in consumption in the general population.

10.6 Concluding remarks

This project aimed at assessing the predictive validity of WXYfm and SAIN,LIM nutrient profiling models. A protective effect of diets containing higher amounts of healthy foods was hypothesised. Aggregate scores were defined to classify participants according to their relative consumption of healthy and unhealthy foods. The aggregate scores were associated with better food intake profiles, associations were weaker with diet quality indices. Survival analyses yielded U-shaped

associations; risk reduction was observed for participants with middle aggregate scores rankings, significant for the SAIN,LIM model.

The unexpected quadratic trends were better explained by misreporting of dietary intakes. Low-energy reporters differentially under-reported the energy-dense unhealthy foods, and were therefore more likely to be classified as healthy by the energy-weighted aggregate scores. Low-energy reporting was further associated with increased vascular risk, and this association confounded the predictive validity results. The role of misreporting was confirmed by the implementation of alternative aggregating method and the modification of the original NP models.

Aggregation of NP scores for individual foods to produce an aggregate score that indexes the nutritional quality of the diet is problematic, particularly as a consequence of information bias. The analysis presented in this thesis indicates that methods of dietary assessment that more accurately reflect true intake would further confirm the predictive validity of NP.

References

Agence française de sécurité sanitaire des aliments (2008) Définition de profils nutritionnels pour l'accès aux allégations nutritionnelles et de santé : propositions et arguments. Paris: Afssa.

Akbaraly TN, Brunner EJ, Ferrie JE, Marmot MG, Kivimaki M & Singh-Manoux A (2009a) Dietary pattern and depressive symptoms in middle age. *Br J Psychiatry* **195**, 408-413.

Akbaraly TN, Ferrie JE, Berr C, Brunner EJ, Head J, Marmot MG, Singh-Manoux A, Ritchie K, Shipley MJ & Kivimaki M (2011) Alternative Healthy Eating Index and mortality over 18 y of follow-up: results from the Whitehall II cohort. *Am J Clin Nutr* **94**, 247-253.

Akbaraly TN, Singh-Manoux A, Marmot MG & Brunner EJ (2009b) Education attenuates the association between dietary patterns and cognition. *Dement Geriatr Cogn Disord* **27**, 147-154.

Akbaraly TN, Singh-Manoux A, Tabak AG, Jokela M, Virtanen M, Ferrie JE, Marmot MG, Shipley MJ & Kivimaki M (2010) Overall diet history and reversibility of the metabolic syndrome over 5 years: the Whitehall II prospective cohort study. *Diabetes Care* **33**, 2339-2341.

American Heart association (AHA) (2009) Nutritional Criteria for Certified Foods. http://www.americanheart.org/presenter.jhtml?identifier=4973 Accessed: 06/06/2009.

Arambepola C, Scarborough P & Rayner M (2008) Validating a nutrient profile model. *Public Health Nutr* **11**, 371-378.

Azais-Braesco V, Goffi C & Labouze E (2006) Nutrient profiling: comparison and critical analysis of existing systems. *Public Health Nutr* **9**, 613-622.

Bates B, Lennox A & Swan G (2010) National Diet and Nutrition Survey - Headline results from Year 1 of the Rolling Programme (2008/2009). London: Food Standards Agency and Department of Health.

Bingham SA & Day NE (1997) Using biochemical markers to assess the validity of prospective dietary assessment methods and the effect of energy adjustment. *Am J Clin Nutr* **65**, 1130S-1137S.

Bingham SA, Gill C, Welch A, Day K, Cassidy A, Khaw KT, Sneyd MJ, Key TJ, Roe L & Day NE (1994) Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records. *Br J Nutr* **72**, 619-643.

Bingham SA, Luben R, Welch A, Wareham N, Khaw KT & Day N (2003) Are imprecise methods obscuring a relation between fat and breast cancer? *Lancet* **362**, 212-214.

Black AE (2000a) 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. *Int J Obes Relat Metab Disord* **24**, 1119-1130.

Black AE (2000b) The sensitivity and specificity of the Goldberg cut-off for EI:BMR for identifying diet reports of poor validity. *Eur J Clin Nutr* **54**, 395-404.

Blaikie N (2003) *Analyzing Quantitative Data*. London: SAGE.

Block G (1982) A review of validations of dietary assessment methods. *Am J Epidemiol* **115**, 492-505.

Bouillon K, Singh-Manoux A, Jokela M, Shipley MJ, Batty GD, Brunner EJ, Sabia S, Tabak AG, Akbaraly T, Ferrie JE & Kivimaki M (2011) Decline in low-density lipoprotein cholesterol concentration: lipid-lowering drugs, diet, or physical activity? Evidence from the Whitehall II study. *Heart* **97**, 923-930.

Bourdillon F, Cano N, Delarue J & Turck D (2010) Propositions pour le PNNS 2011-2012 des Sociétés savantes et d'experts en nutrition: Société française de santé publique.

Breslow N (1974) Covariance analysis of censored survival data. *Biometrics* **30**, 89-99.

Brownell KD, Farley T, Willett WC, Popkin BM, Chaloupka FJ, Thompson JW & Ludwig DS (2009) The public health and economic benefits of taxing sugar-sweetened beverages. *N Engl J Med* **361**, 1599-1605.

Brunner E (1997) Inequalities in diet and health. In *Diet, nutrition & chronic disease : lessons from contrasting worlds*, pp. 77-94 [SS Prakash and K McPherson, editors]. Chichester: John Wiley & Sons.

Brunner E, Stallone D, Juneja M, Bingham S & Marmot M (2001) Dietary assessment in Whitehall II: comparison of 7 d diet diary and food-frequency questionnaire and validity against biomarkers. *Br J Nutr* **86**, 405-414.

Brunner EJ, Marmot MG, Nanchahal K, Shipley MJ, Stansfeld SA, Juneja M & Alberti KG (1997) Social inequality in coronary risk: central obesity and the metabolic syndrome. Evidence from the Whitehall II study. *Diabetologia* **40**, 1341-1349.

Brunner EJ, Marmot MG, White IR, O'Brien JR, Etherington MD, Slavin BM, Kearney EM & Smith GD (1993) Gender and employment grade differences in blood cholesterol, apolipoproteins and haemostatic factors in the Whitehall II study. *Atherosclerosis* **102**, 195-207.

Brunner EJ, Mosdol A, Witte DR, Martikainen P, Stafford M, Shipley MJ & Marmot MG (2008) Dietary patterns and 15-y risks of major coronary events, diabetes, and mortality. *Am J Clin Nutr* **87**, 1414-1421.

Burk MC & Pao EM (1976) Methodology for Large-Scale Surveys of Households and Individual Diets. In *Home Economics Research Report No. 40*. Washington D.C.: Agricultural Research Service, USDA.

Cade JE, Burley VJ & Greenwood DC (2004) The UK Women's Cohort Study: comparison of vegetarians, fish-eaters and meat-eaters. *Public Health Nutr* **7**, 871-878.

Centers for Disease Control and Prevention & National Center for Health Statistics (2012) National Health and Nutrition Examination Survey. Accessed: 24/11/2011. Hyattsville, MD: US Department of Health and Human Services.

Chan W, Brown J & Buss DH (1994) *Miscellaneous Foods. Fourth supplement to the 5th edition of McCance and Widdowson's The Composition of Foods.* Cambridge: Royal Society of Chemistry.

Chan W, Brown J & Lee SM (1995) *Meat, Poultry and Game. Supplement to the 5th edition of McCance and Widdowson's The Composition of Foods.* Cambridge: Royal Society of Chemistry.

Chiuve SE, Sampson L & Willett WC (2011) The association between a nutritional quality index and risk of chronic disease. *Am J Prev Med* **40**, 505-513.

Cho S, DeVries JW & Prosky L (2007) *Dietary Fiber Analysis and Applications*. Gaithersburg: AOAC international.

Choices International Foundation (2009) PRODUCT CRITERIA - 20 October 2009 - Version 2.1. http://www.choicesinternational.org/images/choices%20product%20criteria%20v2_1.pdf Accessed: 15/12/2009 [Choices International Foundation, editor]. Brussels.

Chriqui JF, Eidson SS, Bates H, Kowalczyk S & Chaloupka FJ (2008) State sales tax rates for soft drinks and snacks sold through grocery stores and vending machines, 2007. *J Public Health Policy* **29**, 226-249.

Clarke R, Shipley M, Armitage J, Collins R & Harris W (2009) Plasma phospholipid fatty acids and CHD in older men: Whitehall study of London civil servants. *Br J Nutr* **102**, 279-284.

Cohen J (1968) Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* **70**, 213-220.

Colditz GA, Rimm EB, Giovannucci E, Stampfer MJ, Rosner B & Willett WC (1991) A prospective study of parental history of myocardial infarction and coronary artery disease in men. *Am J Cardiol* **67**, 933-938.

Cox DR (1972) Regression Models and Life-Tables. *Journal of the Royal Statistical Company, Series B* **34**, 187-220.

Cox DR & Oakes D (1984) *Analysis of Survival Data*. London: Chapman & Hall.

Cronbach LJ & Meehl PE (1955) Construct validity in psychological tests. *Psychol Bull* **52**, 281-302.

Dahm CC, Keogh RH, Lentjes MA, Spencer EA, Key TJ, Greenwood DC, Cade JE, Burley VJ, Shipley MJ, Brunner EJ, Stephen AM, Mishra G, Kuh D, Fentiman IS, White IR, Luben R, Khaw KT & Rodwell Bingham SA (2010) Intake of dietary fats and colorectal cancer risk: prospective findings from the UK Dietary Cohort Consortium. *Cancer Epidemiol* **34**, 562-567.

Darmon N (2009) The good, the bad, and the ultra-processed. *Public Health Nutr* **12**, 1967-1968; author reply 1968-1969.

Darmon N, Darmon M, Maillot M & Drewnowski A (2005) A nutrient density standard for vegetables and fruits: nutrients per calorie and nutrients per unit cost. *J Am Diet Assoc* **105**, 1881-1887.

Darmon N, Vieux F, Maillot M, Volatier JL & Martin A (2009) Nutrient profiles discriminate between foods according to their contribution to nutritionally adequate diets: a validation study using linear programming and the SAIN,LIM system. *Am J Clin Nutr* **89**, 1227-1236.

Day N, McKeown N, Wong M, Welch A & Bingham S (2001) Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *Int J Epidemiol* **30**, 309-317.

Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom*. London: HMSO.

Department of Health (1994) *Nutritional Aspects of Cardiovascular Disease*. London: HMSO.

Department of Health (1998) *Nutritional Aspects of the Development of Cancer*. London: The Stationery Office.

Department of Health & Food Standards Agency (2011) National Diet and Nutrition Survey. Headline results from Years 1 and 2 (combined) of the Rolling Programme (2008/2009 - 2009/2010) [B Bates, A Lennox, C Bates and G Swan, editors]. London.

Diabetes UK (2010) Diabetes in the UK 2010: Key statistics on diabetes.

Drewnowski A (2005) Concept of a nutritious food: toward a nutrient density score. *Am J Clin Nutr* **82**, 721-732.

Drewnowski A & Fulgoni V, 3rd (2008) Nutrient profiling of foods: creating a nutrient-rich food index. *Nutr Rev* **66**, 23-39.

Drewnowski A & Fulgoni V, 3rd (2011) Comparing the nutrient rich foods index with "Go," "Slow," and "Whoa," foods. *J Am Diet Assoc* **111**, 280-284.

Drewnowski A, Fulgoni VL, Young MK & Pitman S (2008) Nutrient-rich foods: applying nutrient navigation systems to improve public health. *J Food Sci* **73**, H222-228.

Drewnowski A, Henderson SA, Driscoll A & Rolls BJ (1997) The Dietary Variety Score: assessing diet quality in healthy young and older adults. *J Am Diet Assoc* **97**, 266-271.

Drewnowski A, Henderson SA, Shore AB, Fischler C, Preziosi P & Hercberg S (1996) Diet quality and dietary diversity in France: implications for the French paradox. *J Am Diet Assoc* **96**, 663-669.

Drewnowski A, Maillot M & Darmon N (2009) Should nutrient profiles be based on 100 g, 100 kcal or serving size? *Eur J Clin Nutr* **63**, 898-904.

Drummond SE, Crombie NE, Cursiter MC & Kirk TR (1998) Evidence that eating frequency is inversely related to body weight status in male, but not female, non-obese adults reporting valid dietary intakes. *Int J Obes Relat Metab Disord* **22**, 105-112.

Dubois L, Girard M & Bergeron N (2000) The choice of a diet quality indicator to evaluate the nutritional health of populations. *Public Health Nutr* **3**, 357-365.

Elovainio M, Ferrie JE, Singh-Manoux A, Gimeno D, De Vogli R, Shipley M, Vahtera J, Brunner E, Marmot MG & Kivimaki M (2010) Organisational justice and markers of inflammation: the Whitehall II study. *Occup Environ Med* **67**, 78-83.

Epstein LH, Dearing KK, Roba LG & Finkelstein E (2010) The influence of taxes and subsidies on energy purchased in an experimental purchasing study. *Psychol Sci* **21**, 406-414.

Eurodiet Core Report (2000) Nutrition & Diet for Healthy Lifestyles in Europe. Science & Policy Implications: Health & Consumer Protection DG.

FAO/WHO/UNU (1985) Energy and protein requirements. In *Technical Report Series n. 724*. Geneva: World Health Organization.

Food and Drug Administration (FDA) (2008) Code of Federal Regulations Title 21, Volume 2, Chapter I, Part 101, Sec. 101.14 - 21CFR101.14, revised as of April 1, 2008 [U.S. Department of Health and Human Services, editor.

Food Standards Australia New Zealand CALCULATION METHOD FOR DETERMINING FOODS ELIGIBLE TO MAKE HEALTH CLAIMS - Nutrient Profiling Calculator. http://www.foodstandards.gov.au/_srcfiles/method%20to%20determine%20eligibility%20_final_.pdf Accessed: 15/12/2009.

Freedman LS, Potischman N, Kipnis V, Midthune D, Schatzkin A, Thompson FE, Troiano RP, Prentice R, Patterson R, Carroll R & Subar AF (2006) A comparison of two dietary instruments for evaluating the fat-breast cancer relationship. *Int J Epidemiol* **35**, 1011-1021.

Fulgoni VL, 3rd, Keast DR & Drewnowski A (2009) Development and validation of the nutrient-rich foods index: a tool to measure nutritional quality of foods. *J Nutr* **139**, 1549-1554.

Garsetti M, de Vries J, Smith M, Amosse A & Rolf-Pedersen N (2007) Nutrient profiling schemes: overview and comparative analysis. *Eur J Nutr* **46 Suppl 2**, 15-28.

Gazibarich B & Ricci PF (1998) Towards better food choices: the nutritious food index. *Aust J Nutr Diet* **55**, 10-20.

Goldberg GR, Black AE, Jebb SA, Cole TJ, Murgatroyd PR, Coward WA & Prentice AM (1991) Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. *Eur J Clin Nutr* **45**, 569-581.

Gonzalez EL, Johansson S, Wallander MA & Rodriguez LA (2009) Trends in the prevalence and incidence of diabetes in the UK: 1996-2005. *J Epidemiol Community Health* **63**, 332-336.

Grambsch PM & Therneau TM (1994) Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* **81**, 515-526.

Greenland S (1995) Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology* **6**, 356-365.

Guthrie HA (1977) Concept of a nutritious food. *J Am Diet Assoc* **71**, 14-18.

Hamer M, Witte DR, Mosdol A, Marmot MG & Brunner EJ (2008) Prospective study of coffee and tea consumption in relation to risk of type 2 diabetes mellitus among men and women: the Whitehall II study. *Br J Nutr* **100**, 1046-1053.

Hansen RG (1973) An index of food quality. *Nutr Rev* **31**, 1-7.

Hatloy A, Torheim LE & Oshaug A (1998) Food variety--a good indicator of nutritional adequacy of the diet? A case study from an urban area in Mali, West Africa. *Eur J Clin Nutr* **52**, 891-898.

Hoffmann K, Schulze MB, Schienkiewitz A, Nothlings U & Boeing H (2004) Application of a new statistical method to derive dietary patterns in nutritional epidemiology. *Am J Epidemiol* **159**, 935-944.

Holland B, Brown J & Buss DH (1993) *Fish and fish products. Third supplement to the 5th edition of McCance and Widdowson's The Composition of Foods.* Cambridge: Royal Society of Chemistry.

Holland B, Welch AA & Buss DH (1988) *Cereals and cereal products. Third supplement to the 4th edition of McCance and Widdowson's The Composition of Foods.* Cambridge: Royal Society of Chemistry.

Holland B, Welch AA & Buss DH (1989) *Milk products and eggs. Fourth supplement to the 4th edition of McCance and Widdowson's The Composition of Foods.* Cambridge: Royal Society of Chemistry.

Holland B, Welch AA & Buss DH (1991a) *Vegetables, herbs and spices. Fifth supplement to the 4th edition of McCance and Widdowson's The Composition of Foods.* Cambridge: Royal Society of Chemistry.

Holland B, Welch AA & Buss DH (1992a) *Fruit and nuts. First supplement to the 5th edition of McCance and Widdowson's The Composition of Foods.* Cambridge: Royal Society of Chemistry.

Holland B, Welch AA & Buss DH (1992b) *Vegetable dishes. Second supplement to the 5th edition of McCance and Widdowson's The Composition of Foods.* Cambridge: Royal Society of Chemistry.

Holland B, Welch AA, Unwin ID, Buss DH, Paul AA & Southgate DAT (1991b) *McCance and Widdowson's The Composition of Foods, 5th ed.* Cambridge: Royal Society of Chemistry.

Hu FB, Stampfer MJ, Manson JE, Rimm E, Colditz GA, Rosner BA, Hennekens CH & Willett WC (1997) Dietary fat intake and the risk of coronary heart disease in women. *N Engl J Med* **337**, 1491-1499.

Hu FB, Stampfer MJ, Rimm E, Ascherio A, Rosner BA, Spiegelman D & Willett WC (1999) Dietary fat and coronary heart disease: a comparison of approaches for adjusting for total energy intake and modeling repeated dietary measurements. *Am J Epidemiol* **149**, 531-540.

Hutchinson J, Lentjes MA, Greenwood DC, Burley VJ, Cade JE, Cleghorn CL, Threapleton DE, Key TJ, Cairns BJ, Keogh RH, Dahm CC, Brunner EJ, Shipley MJ, Kuh D, Mishra G, Stephen AM, Bhaniani A, Borgulya G & Khaw KT (2011) Vitamin C intake from diary recordings and risk of breast cancer in the UK Dietary Cohort Consortium. *Eur J Clin Nutr*.

International Agency for Research on Cancer EPIC Project. http://epic.iarc.fr/ Accessed: 22/11/2011.

Jacobson MF & Brownell KD (2000) Small taxes on soft drinks and snack foods to promote health. *Am J Public Health* **90**, 854-857.

Jakicic JM (2011) Proprietary information considerations in health, activity, and dietary research: implications of knowing what is in the "black box". *Am J Prev Med* **40**, 583-584.

Kaaks R, Riboli E, Esteve J, van Kappel AL & van Staveren WA (1994) Estimating the accuracy of dietary questionnaire assessments: validation in terms of structural equation models. *Stat Med* **13**, 127-142.

Kaaks RJ (1997) Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. *Am J Clin Nutr* **65**, 1232S-1239S.

Kaluza J, Hakansson N, Brzozowska A & Wolk A (2009) Diet quality and mortality: a population-based prospective study of men. *Eur J Clin Nutr* **63**, 451-457.

Kant AK (2004) Dietary patterns and health outcomes. *J Am Diet Assoc* **104**, 615-635.

Kant AK, Graubard BI & Schatzkin A (2004) Dietary patterns predict mortality in a national cohort: the National Health Interview Surveys, 1987 and 1992. *J Nutr* **134**, 1793-1799.

Kant AK, Schatzkin A, Graubard BI & Schairer C (2000) A prospective study of diet quality and mortality in women. *Jama* **283**, 2109-2115.

Katz DL, Njike VY, Rhee LQ, Reingold A & Ayoob KT (2010) Performance characteristics of NuVal and the Overall Nutritional Quality Index (ONQI). *Am J Clin Nutr* **91**, 1102S-1108S.

Kennedy E, Racsa P, Dallal G, Lichtenstein AH, Goldberg J, Jacques P & Hyatt R (2008) Alternative approaches to the calculation of nutrient density. *Nutr Rev* **66**, 703-709.

Kennedy ET, Ohls J, Carlson S & Fleming K (1995) The Healthy Eating Index: design and applications. *J Am Diet Assoc* **95**, 1103-1108.

Keogh RH, White IR & Rodwell SA (2010) Measurement error models for FFQ and food records using a surrogate biomarker: Fruit and vegetable intake and plasma vitamin C in EPIC-Norfolk.

Key TJ, Appleby PN, Masset G, Brunner EJ, Cade JE, Greenwood DC, Stephen AM, Kuh D, Bhaniani A, Powell N & Khaw KT (2011) Vitamins, minerals, essential fatty acids and colorectal cancer risk in the United Kingdom Dietary Cohort Consortium. *Int J Cancer*.

Kipnis V, Midthune D, Buckman DW, Dodd KW, Guenther PM, Krebs-Smith SM, Subar AF, Tooze JA, Carroll RJ & Freedman LS (2009) Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics* **65**, 1003-1010.

Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano RP, Bingham S, Schoeller DA, Schatzkin A & Carroll RJ (2003) Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol* **158**, 14-21; discussion 22-16.

Kirkwood BR & Sterne JA (2003) *Essential medical statistics - 2nd edition*. Oxford: Blackwell Science.

Kumari M, Marmot M & Brunner E (2000) Social determinants of von willebrand factor: the Whitehall II study. *Arterioscler Thromb Vasc Biol* **20**, 1842-1847.

Labouze E, Goffi C, Moulay L & Azais-Braesco V (2007) A multipurpose tool to evaluate the nutritional quality of individual foods: Nutrimap. *Public Health Nutr* **10**, 690-700.

Lesturgeon A, Vieux F, Georgé S, Rouveyrol C, Amiot M & Darmon N (2011) 27/370 - Taking into account the nutritional specificities of food categories in an across-the-board nutrient profile system. *Ann Nutr Metab* **58**, 107.

Li CY & Sung FC (1999) A review of the healthy worker effect in occupational epidemiology. *Occup Med (Lond)* **49**, 225-229.

Livingstone MB & Black AE (2003) Markers of the validity of reported energy intake. *J Nutr* **133 Suppl 3**, 895S-920S.

Ludbrook J (2002) Statistical techniques for comparing measurers and methods of measurement: a critical review. *Clin Exp Pharmacol Physiol* **29**, 527-536.

Macdiarmid J & Blundell J (1998) Assessing dietary intake: Who, what and why of under-reporting. *Nutr Res Rev* **11**, 231-253.

Macfarlane PW, Devine B, Latif S, McLaughlin S, Shoat DB & Watts MP (1990) Methodology of ECG interpretation in the Glasgow program. *Methods Inf Med* **29**, 354-361.

Maillot M, Darmon N, Darmon M, Lafay L & Drewnowski A (2007) Nutrient-dense food groups have high energy costs: an econometric approach to nutrient profiling. *J Nutr* **137**, 1815-1820.

Maillot M, Drewnowski A, Vieux F & Darmon N (2011) Quantifying the contribution of foods with unfavourable nutrient profiles to nutritionally adequate diets. *Br J Nutr* **105**, 1133-1137.

Mantel N (1966) Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* **50**, 163-170.

Marmot M & Brunner E (2005) Cohort Profile: the Whitehall II study. *Int J Epidemiol* **34**, 251-256.

Marmot MG, Rose G, Shipley M & Hamilton PJ (1978) Employment grade and coronary heart disease in British civil servants. *J Epidemiol Community Health* **32**, 244-249.

Marmot MG, Shipley MJ, Hemingway H, Head J & Brunner EJ (2008) Biological and behavioural explanations of social inequalities in coronary heart disease: the Whitehall II study. *Diabetologia* **51**, 1980-1988.

Marr JW (1971) Individual dietary surveys: purposes and methods. *World Rev Nutr Diet* **13**, 105-164.

Martikainen P, Brunner E & Marmot M (2003) Socioeconomic differences in dietary patterns among middle-aged men and women. *Soc Sci Med* **56**, 1397-1410.

Martin A (2001) *Apports nutritionnels conseillés pour la population française. 3ème édition*. Paris: Lavoisier.

McClellan MB (2003) Speech before Harvard School of Public Health, "Changing the American Diet", July 1, 2003. http://www.fda.gov/NewsEvents/Speeches/ucm053648.htm Accessed: 15/12/2009.

McCullough ML, Feskanich D, Rimm EB, Giovannucci EL, Ascherio A, Variyam JN, Spiegelman D, Stampfer MJ & Willett WC (2000a) Adherence to the Dietary Guidelines for Americans and risk of major chronic disease in men. *Am.J.Clin.Nutr.* **72**, 1223-1231.

McCullough ML, Feskanich D, Stampfer MJ, Giovannucci EL, Rimm EB, Hu FB, Spiegelman D, Hunter DJ, Colditz GA & Willett WC (2002) Diet quality and major chronic disease risk in men and women: moving toward improved dietary guidance. *Am J Clin Nutr* **76**, 1261-1271.

McCullough ML, Feskanich D, Stampfer MJ, Rosner BA, Hu FB, Hunter DJ, Variyam JN, Colditz GA & Willett WC (2000b) Adherence to the Dietary Guidelines for Americans and risk of major chronic disease in women. *Am.J.Clin.Nutr.* **72**, 1214-1222.

McNaughton SA, Mishra GD & Brunner EJ (2008) Dietary patterns, insulin resistance, and incidence of type 2 diabetes in the Whitehall II Study. *Diabetes Care* **31**, 1343-1348.

McNaughton SA, Mishra GD & Brunner EJ (2009) Food patterns associated with blood lipids are predictive of coronary heart disease: the Whitehall II study. *Br J Nutr* **102**, 619-624.

McNulty H, Eaton-Evans J, Cran G, Woulahan G, Boreham C, Savage JM, Fletcher R & Strain JJ (1996) Nutrient intakes and impact of fortified breakfast cereals in schoolchildren. *Arch Dis Child* **75**, 474-481.

Michels KB, Bingham SA, Luben R, Welch AA & Day NE (2004) The effect of correlated measurement error in multivariate models of diet. *Am J Epidemiol* **160**, 59-67.

Michels KB & Wolk A (2002) A prospective study of variety of healthy foods and mortality in women. *Int J Epidemiol* **31**, 847-854.

Mosdol A, Witte DR, Frost G, Marmot MG & Brunner EJ (2007) Dietary glycemic index and glycemic load are associated with high-density-lipoprotein cholesterol at baseline but not with increased risk of diabetes in the Whitehall II study. *Am J Clin Nutr* **86**, 988-994.

Muller L & Ruffieux B (2011) Modification des achats en réponse à l'apposition de différents logos d'évaluation nutritionnelle de face avant des emballages. http://www.grenoble.inra.fr/Docs/pub/A2011/gael2011-07.pdf Accessed: 12/12/2011.

Nabi H, Singh-Manoux A, Shipley M, Gimeno D, Marmot MG & Kivimaki M (2008) Do psychological factors affect inflammation and incident coronary heart disease: the Whitehall II Study. *Arterioscler Thromb Vasc Biol* **28**, 1398-1406.

National Health Service (2011) The eatwell plate. http://www.nhs.uk/Livewell/Goodfood/Pages/eatwell-plate.aspx Accessed: 30/01/2012.

National Heart Foundation of Australia (AHF) (2009) Heart Foundation Tick. http://www.heartfoundation.org.au/sites/tick/Health_Professionals/Pages/TickCriteria.aspx Accessed: 16/12/2009.

National Heart Lung and Blood Institute (2006) Your guide to lowering your blood pressure with DASH: DASH Eating Plan. Bethesda, MD: National Institutes of Health.

National Obesity Observatory (2011) Trends in obesity prevalence. Accessed: 28/11/2011: NHS.

Netherlands Nutrition Centre (NNC) (2005) Criteria for the nutritional evaluation of foods. The Netherlands tripartite classification model for foods. Den Haag: NNC.

Ocke MC & Kaaks RJ (1997) Biochemical markers as additional measurements in dietary validity studies: application of the method of triads with examples from the European Prospective Investigation into Cancer and Nutrition. *Am J Clin Nutr* **65**, 1240S-1245S.

Office for National Statistics (2011) Births and Deaths in England and Wales, 2010. Newport.

Office of communications (2007a) Impact Assessment - Food Advertising to Children. London: Ofcom.

Office of communications (2007b) Television Advertising of Food and Drink Products to Children - Final statement. London: Ofcom.

Office of communications (2011) The Ofcom broadcasting code (Incorporating the Cross-promotion Code) - 28 February 2011. http://stakeholders.ofcom.org.uk/binaries/broadcast/831190/broadcastingcode2011.pdf Accessed: 28/11/2011. London: Ofcom.

Parkin DM, Boyd L & Walker LC (2011) The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010 - Summary and conclusions. *Brit J Cancer* **105**, S77-S81.

Patterson RE, Haines PS & Popkin BM (1994) Diet quality index: capturing a multidimensional behavior. *J Am Diet Assoc* **94**, 57-64.

Paul AA & Southgate DAT (1978) *McCance and Widdowson's The Composition of Foods, 4th ed.* London: HMSO.

Pekkarinen M (1970) Methodology in the collection of food consumption data. *World Rev Nutr Diet* **12**, 145-171.

Pomerleau J, Ostbye T & Bright-See E (1999) Potential underreporting of energy intake in the Ontario Health Survey and its relationship with nutrient and food intakes. *Eur J Epidemiol* **15**, 553-557.

Poppitt SD, Swann D, Black AE & Prentice AM (1998) Assessment of selective under-reporting of food intake by both obese and non-obese women in a metabolic facility. *Int J Obes Relat Metab Disord* **22**, 303-311.

Prentice AM, Black AE, Coward WA, Davies HL, Goldberg GR, Murgatroyd PR, Ashford J, Sawyer M & Whitehead RG (1986) High levels of energy expenditure in obese women. *Br Med J (Clin Res Ed)* **292**, 983-987.

Prentice RL (2003) Dietary assessment and the reliability of nutritional epidemiology reports. *Lancet* **362**, 182-183.

Prentice RL, Mossavar-Rahmani Y, Huang Y, Van Horn L, Beresford SA, Caan B, Tinker L, Schoeller D, Bingham S, Eaton CB, Thomson C, Johnson KC, Ockene J, Sarto G, Heiss G & Neuhouser ML (2011) Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. *Am J Epidemiol* **174**, 591-603.

Pryer J, Brunner E, Elliott P, Nichols R, Dimond H & Marmot M (1995) Who complied with COMA 1984 dietary fat recommendations among a nationally representative sample of British adults in 1986-7 and what did they eat? *Eur J Clin Nutr* **49**, 718-728.

Pryer JA, Vrijheid M, Nichols R, Kiggins M & Elliott P (1997) Who are the 'low energy reporters' in the dietary and nutritional survey of British adults? *Int J Epidemiol* **26**, 146-154.

Quinio C, Biltoft-Jensen A, De Henauw S, Gibney MJ, Huybrechts I, McCarthy SN, O'Neill JL, Tetens I, Turrini A & Volatier JL (2007) Comparison of different nutrient profiling schemes to a new reference method using dietary surveys. *Eur J Nutr* **46 Suppl 2**, 37-46.

Rayner M & Scarborough P (2005) The burden of food related ill health in the UK. *J Epidemiol Community Health* **59**, 1054-1057.

Rayner M, Scarborough P, Boxer A & Stockley L (2005a) Nutrient profiles: Development of Final Model. Final Report. London: Food Standards Agency.

Rayner M, Scarborough P & Stockley L (2004a) Nutrient profiles: Options for definitions for use in relation to food promotion and children's diets. London: Food Standards Agency.

Rayner M, Scarborough P & Stockley L (2005b) Nutrient profiles: Applicability of currently proposed model for uses in relation to promotion of food to children aged 5-10 and adults. London: Food Standards Agency.

Rayner M, Scarborough P, Stockley L & Boxer A (2005c) Nutrient profiles: Further refinement and testing of Model SSCg3d - Final report. London: Food Standards Agency.

Rayner M, Scarborough P & Williams C (2004b) The origin of Guideline Daily Amounts and the Food Standards Agency's guidance on what counts as 'a lot' and 'a little'. *Public Health Nutr* **7**, 549-556.

Reedy J & Kirkpatrick SI (2011) The use of proprietary nutrient profiling tools in nutrition science and policy: a commentary. *Am J Prev Med* **40**, 581-582.

Roodenburg AJ, Schlatmann A, Dotsch-Klerk M, Daamen R, Dong J, Guarro M, Stergiou M, Sayed N, Ronoh E, Jansen L & Seidell JC (2011) Potential effects of nutrient profiles on nutrient intakes in the Netherlands, Greece, Spain, USA, Israel, China and South-Africa. *PLoS One* **6**, e14721.

Roodenburg AJC, Temme EHM, Howell-Davies O & Seidell JC (2009) Potential impact of the Choices Programme on nutrient intakes in the Dutch population. *Nutr Bull* **34**, 318-323.

Rose G (1982) Incubation period of coronary heart disease. *Br Med J (Clin Res Ed)* **284**, 1600-1601.

Rose G & Marmot MG (1981) Social class and coronary heart disease. *Br Heart J* **45**, 13-19.

Rosner B & Gore R (2001) Measurement error correction in nutritional epidemiology based on individual foods, with application to the relation of diet to breast cancer. *Am J Epidemiol* **154**, 827-835.

Rosner B, Michels KB, Chen YH & Day NE (2008) Measurement error correction for nutritional exposures with correlated measurement error: use of the method of triads in a longitudinal setting. *Stat Med* **27**, 3466-3489.

Rosner B, Willett WC & Spiegelman D (1989) Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Stat Med* **8**, 1051-1069; discussion 1071-1053.

SAS Institute Inc. SAS OnlineDoc 9.1.3. http://support.sas.com/onlinedoc/913/docMainpage.jsp Accessed: 2011. Cary, NC.

Savy M, Martin-Prevel Y, Sawadogo P, Kameli Y & Delpeuch F (2005) Use of variety/diversity scores for diet quality measurement: relation with nutritional status of women in a rural area in Burkina Faso. *Eur J Clin Nutr* **59**, 703-716.

Scarborough P, Arambepola C, Kaur A, Bhatnagar P & Rayner M (2010) Should nutrient profile models be 'category specific' or 'across-the-board'? A comparison of the two systems using diets of British adults. *Eur J Clin Nutr* **64**, 553-560.

Scarborough P, Boxer A, Rayner M & Stockley L (2007a) Testing nutrient profile models using data from a survey of nutrition professionals. *Public Health Nutr* **10**, 337-345.

Scarborough P, Rayner M, Boxer A & Stockley L (2005) 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. London: Food Standards Agency.

Scarborough P, Rayner M & Stockley L (2007b) Developing nutrient profile models: a systematic approach. *Public Health Nutr* **10**, 330-336.

Scarborough P, Rayner M, Stockley L & Black A (2007c) Nutrition professionals' perception of the 'healthiness' of individual foods. *Public Health Nutr* **10**, 346-353.

Schectman G, Byrd JC & Gruchow HW (1989) The influence of smoking on vitamin C status in adults. *Am J Public Health* **79**, 158-162.

Scheidt DM & Daniel E (2004) Composite index for aggregating nutrient density using food labels: ratio of recommended to restricted food components. *J Nutr Educ Behav* **36**, 35-39.

Schoenfeld D (1982) Partial Residuals for the Proportional Hazards Regression-Model. *Biometrika* **69**, 239-241.

Schofield WN, Schofield C & James WPT (1985) Basal metabolic rate - review and prediction, together with an annotated bibliography of source material. *Hum Nutr Clin Nutr* **39C Suppl 1**, 1-96.

Shortt CT, Duthie GG, Robertson JD, Morrice PC, Nicol F & Arthur JR (1997) Selenium status of a group of Scottish adults. *Eur J Clin Nutr* **51**, 400-404.

Sorenson AW, Wyse BW, Wittwer AJ & Hansen RG (1976) An Index of Nutritional Quality for a balanced diet. New help for an old problem. *J Am Diet Assoc* **68**, 236-242.

Spencer EA, Key TJ, Appleby PN, Dahm CC, Keogh RH, Fentiman IS, Akbaraly T, Brunner EJ, Burley V, Cade JE, Greenwood DC, Stephen AM, Mishra G, Kuh D, Luben R, Mulligan AA, Khaw KT & Rodwell SA (2010) Meat, poultry and fish and risk of colorectal cancer: pooled analysis of data from the UK dietary cohort consortium. *Cancer Causes Control* **21**, 1417-1425.

Spiegelman D, Schneeweiss S & McDermott A (1997) Measurement error correction for logistic regression models with an "alloyed gold standard". *Am J Epidemiol* **145**, 184-196.

Stallone DD, Brunner EJ, Bingham SA & Marmot MG (1997) Dietary assessment in Whitehall II: the influence of reporting bias on apparent socioeconomic variation in nutrient intakes. *Eur J Clin Nutr* **51**, 815-825.

Stefanik PA & Trulson MF (1962) Determining the frequency intakes of foods in large group studies. *Am J Clin Nutr* **11**, 335-343.

Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM & Carpenter JR (2009) Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj* **338**, b2393.

Stockley L, Rayner M & Kaur A (2007) Nutrient profiles for use in relation to food promotion and children's diet: Update of 2004 literature review. London: Food Standards Agency.

Sturm R, Powell LM, Chriqui JF & Chaloupka FJ (2010) Soda taxes, soft drink consumption, and children's body mass index. *Health Aff (Millwood)* **29**, 1052-1058.

Subar AF, Kipnis V, Troiano RP, Midthune D, Schoeller DA, Bingham S, Sharbaugh CO, Trabulsi J, Runswick S, Ballard-Barbash R, Sunshine J & Schatzkin A (2003) Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. *Am J Epidemiol* **158**, 1-13.

Swedish National Food Administration (2009) National Food Administration's Regulations on the Use of a Particular Symbol (LIVSFS 2005:9). http://www.slv.se/upload/nfa/documents/food_regulations/keyhole_2009.pdf Accessed: 15/12/2009.

Temple JL, Johnson KM, Archer K, Lacarte A, Yi C & Epstein LH (2011) Influence of simplified nutrition labeling and taxation on laboratory energy intake in adults. *Appetite* **57**, 184-192.

The European Parliament and the Council of the European Union (2006) Regulation (EC) n. 1924/2006 of the European Parliament and the Council of the European Union of 20 December 2006 on nutrition and health claims made on food. *Off J eur Union* **L404**, 9-25.

Tooze JA, Midthune D, Dodd KW, Freedman LS, Krebs-Smith SM, Subar AF, Guenther PM, Carroll RJ & Kipnis V (2006) A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *J Am Diet Assoc* **106**, 1575-1587.

Trichterborn J, Harzer G & Kunz C (2011) Fine bakery wares with label claims in Europe and their categorisation by nutrient profiling models. *Eur J Clin Nutr* **65**, 307-312.

Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, Arveiler D, Rajakangas AM & Pajak A (1994) Myocardial infarction and coronary deaths in the World Health Organization MONICA Project. Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents. *Circulation* **90**, 583-612.

Unité de surveillance et d'épidémiologie nutritionnelle (2007) Étude nationale nutrition santé - ENNS, 2006. Saint-Maurice, FR: Institut de veille sanitaire.

US Department of Agriculture (2012) ChooseMyPlate.gov. http://www.choosemyplate.gov/ Accessed: 1/2/2012.

US Department of Health and Human Services - National Heart Lung and Blood Institute We can! Go, Slow and Whoa foods. http://www.nhlbi.nih.gov/health/public/heart/obesity/wecan/downloads/gswtips.pdf Accessed: 11/11/11.

US Department of Health and Human Services and US Department of Agriculture (2005) Dietary guidelines for Americans, 2005. 6th ed. [US Government Printing Office, editor. Washington, DC.

Volatier JL (2000) *Enquête INCA (individuelle et nationale sur les consommations alimentaires).* Paris: Lavoisier.

Volatier JL, Biltoft-Jensen A, De Henauw S, Gibney MJ, Huybrechts I, McCarthy SN, O'Neill JL, Quinio C, Turrini A & Tetens I (2007) A new reference method for the validation of the nutrient profiling schemes using dietary surveys. *Eur J Nutr* **46 Suppl 2**, 29-36.

Vyth EL, Hendriksen MA, Roodenburg AJ, Steenhuis IH, van Raaij JM, Verhagen H, Brug J & Seidell JC (2011a) Consuming a diet complying with front-of-pack label criteria may reduce cholesterol levels: a modeling study. *Eur J Clin Nutr*.

Vyth EL, Steenhuis IH, Heymans MW, Roodenburg AJ, Brug J & Seidell JC (2011b) Influence of placement of a nutrition logo on cafeteria menu items on lunchtime food Choices at Dutch work sites. *J Am Diet Assoc* **111**, 131-136.

Vyth EL, Steenhuis IH, Vlot JA, Wulp A, Hogenes MG, Looije DH, Brug J & Seidell JC (2010) Actual use of a front-of-pack nutrition logo in the supermarket: consumers' motives in food choice. *Public Health Nutr* **13**, 1882-1889.

Waijers PM, Feskens EJ & Ocke MC (2007) A critical review of predefined diet quality scores. *Br J Nutr* **97**, 219-231.

Wiehl DG & Reed R (1960) Development of new or improved dietary methods for epidemiological investigations. *Am J Public Health Nations Health* **50**, 824-828.

Willett WC (1998) *Nutritional epidemiology, 2nd edition*. New York: Oxford University Press.

Willett WC, Green A, Stampfer MJ, Speizer FE, Colditz GA, Rosner B, Monson RR, Stason W & Hennekens CH (1987) Relative and absolute excess risks of coronary heart disease among women who smoke cigarettes. *N Engl J Med* **317**, 1303-1309.

Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH & Speizer FE (1985) Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* **122**, 51-65.

Willett WC & Stampfer MJ (1998) Implications of Total Energy Intake for Epidemiologic Analyses. In *Nutritional Epidemiology, 2nd edition* [WC Willett, editor]. New York: Oxford University Press.

World Cancer Research Fund and American Institute for Cancer Research (2007) Food, Nutrition, Physical Activity and the Prevention of Cancer: a Global Perspective. Washington D.C.

World Health Organization (1999) *Definition, diagnosis and classification of diabetes mellitus and its complications. Report of a WHO consultation, Part 1: Diagnosis and classification of diabetes mellitus* Geneva: World Health Organization.

World Health Organization (2003) Diet, nutrition and the prevention of chronic diseases. Report of a joint WHO/FAO Expert Consultation. In *WHO technical report series 916* [World Health Organization, editor. Geneva.

World Health Organization (2004) Global strategy on diet and physical activity [World Health Organization, editor. Geneva.

World Health Organization (2006) Marketing of food and non-alcoholic beverages to children : report of a WHO forum and technical meeting, Oslo, Norway, 2-5 May 2006.

World Health Organization (2011) Nutrient profiling - Report of a WHO/IASO technical meeting - London, United Kingdom, 4-6 October 2010. Geneva: WHO.

Zelman K & Kennedy A (2005) Naturally nutrient rich ... putting more power on American's plates. *Nutr Today* **40**, 60-68.

Zhang S, Midthune D, Guenther PM, Krebs-Smith SM, Kipnis V, Dodd KW, Buckman DW, Tooze JA, Freedman L & Carroll RJ (2011) A New Multivariate Measurement Error Model with Zero-Inflated Dietary Data, and Its Application to Dietary Assessment. *Ann Appl Stat* **5**, 1456-1487.

Appendices

Appendix 1: Validation of food-frequency questionnaires

(i) Dietary assessment in nutritional epidemiology: 24-hour recalls, diet diaries, and FFQ

Two common methods were developed based on report of recent intake: 24-hour recalls and diet diaries. The first method, 24-hour recalls, is based on an interview conducted by a trained dietary interviewer. The participant is asked to recall all foods and drinks consumed in the last 24 hours, or in the day preceding the interview (Pekkarinen, 1970; Burk & Pao, 1976). The second method, diet diaries, relies on the listing of all foods and drinks consumed on a defined period of time (usually less than a week) (Pekkarinen, 1970; Block, 1982). The participant is asked to report consumption at the time the foods are eaten, portion sizes being directly weighted or estimated. Both methods are open-ended and participants can therefore describe precisely the composition of their diets. However, because of the burden on participants and scientists and the difficulty to estimate actual intake, 24h recalls and diet diaries are not the most convenient methods for large scale epidemiological studies.

Food-frequency methods, which measure the usual intake over a longer past period (e.g. last month/6 months/year), were mainly developed during the 1950s and 60s (Wiehl & Reed, 1960; Stefanik & Trulson, 1962; Marr, 1971). The basic questionnaire consists of two parts: a food list and a frequency response section where participants indicate the frequency of consumption of each food. Some questionnaires may also include a section on the usual portion size consumed of the relative foods. The underlying principle of the food-frequency approach is that long-term average diet is more important than intakes on a few specific days in relation to health outcomes. Since these questionnaires are developed in order to be machine-readable, the burden on the scientists is considerably reduced, and raw data can be available very quickly. Food-frequency questionnaires (FFQs) require less

investment from the participants and the list of foods can be adapted to the research question and/or the population of interest.

FFQs have therefore become the primary method for measuring dietary intakes in epidemiological studies, and the Whitehall II study has been using an anglicised version of the Willett food-frequency questionnaire (Willett *et al.*, 1985). More detail on this questionnaire is given in chapter 4. Assessing the criterion validity of FFQs would require an independent measure of true intake. Observing or measuring true dietary intake in free living individuals is practically impossible and another gold-standard was necessary.

(ii) Validity of FFQs against alternative dietary assessment methods

The comparison of intakes derived from a FFQ with intakes derived from independent dietary assessment methods has been the main approach used to validate FFQ data. Reporting errors associated with FFQs are due to a variety of factors including memory, perception of portion sizes, balance of foods in the FFQ, calculation of mean intake of seasonally variable foods, interpretation of questions, conscious and unconscious bias related to perceived health desirability, BMI, and socio-economic position (Willett, 1998). Diet diaries are the dietary assessment method most different to FFQs: they are open-ended; foods are recorded when consumed—hence no memory reliance and direct assessment of portion size; and interpretation errors are usually made by coders and not participants. Reporting errors are therefore more likely to be independent from the FFQ ones. An alternative to diet diaries may be 24h recalls, but the reliance on memory may cause errors to be more correlated with the FFQ. In both cases, a sufficient number of days is needed to represent average intake. A detailed investigation of the 1980 Nurses' Health Study dietary assessment tools combined the use of two similar FFQ and four 1-week diet diaries. It showed that the relatively cheap questionnaire could capture almost as much information as the diet diary (Willett *et al.*, 1985).

Diet diaries and 24h recalls have further been used to correct the risk estimates between FFQ derived intakes and outcomes (Rosner *et al.*, 1989). However, it has been observed that errors in FFQs and other dietary assessment methods are often correlated (Michels *et al.*, 2004), and more independent measures are needed.

(iii) Validity of FFQs against biomarkers

Biomarkers are another alternative for an independent measure of dietary intake. They have been considered as the gold-standard as all measurement errors associated with the estimation of their concentrations are supposed to be uncorrelated with errors of any dietary questionnaire (Bingham & Day, 1997; Livingstone & Black, 2003).

Plasma β-carotene which was shown to be well related to dietary intakes (Willett, 1998) was measured in almost all Whitehall II participants. It was used in an attempt to validate fruit and vegetable reported intake (appendix 7).

Nonetheless, not all nutrients have a marker in blood or other tissue, and the measured concentrations are not necessarily meaningful in regards to long-term intakes. Also, levels of biomarkers can be linked to other non-dietary covariates (Schectman *et al.*, 1989), and their collection can be prohibitively expensive.

Appendix 2: Classification of all FFQ-items according to WXYfm and SAIN,LIM nutrient profiling models

Table A0.1: WXYfm and SAIN,LIM scores for phase 3 FFQ-items

	WXYfm		SAIN,LIM		
FFQ-item	**Score**	**Category**	**SAIN**	**LIM**	**Q**
Beer, lager or cider	N/A		N/A		
Liqueurs	N/A		N/A		
Port, cherry, vermouth	N/A		N/A		
Spirits	N/A		N/A		
Wine	N/A		N/A		
Peas	-13	Healthy	14.62	0.56	1
Green beans, broad beans, runner beans	-9	Healthy	32.86	0.20	1
Spring greens, kale	-9	Healthy	106.06	0.29	1
Dried lentils, beans, peas	-9	Healthy	8.38	0.24	1
Garlic	-8	Healthy	9.17	0.27	1
Mushrooms	-8	Healthy	28.22	0.25	1
Parsnips, turnips, swedes	-8	Healthy	11.97	0.45	1
Spinach	-8	Healthy	56.46	1.44	1
Brussels sprouts	-8	Healthy	43.22	0.55	1
Cabbage	-7	Healthy	43.73	0.24	1
Cauliflower	-7	Healthy	28.93	0.39	1
Leeks	-7	Healthy	24.41	0.28	1
Tofu or soya bean curd	-7	Healthy	17.79	0.13	1
Baked beans	-6	Healthy	12.54	9.69	3
Broccoli	-6	Healthy	47.68	0.49	1
Carrots	-6	Healthy	16.51	0.73	1
Marrow, courgettes	-6	Healthy	21.55	0.01	1
Peaches, plums, apricots	-6	Healthy	26.00	0.01	1
Sweet peppers	-6	Healthy	126.49	0.38	1
Green salad	-6	Healthy	28.12	0.23	1
Strawberries, raspberries	-6	Healthy	62.38	0.06	1
Tomatoes	-6	Healthy	31.19	0.20	1
Apples	-5	Healthy	6.52	0.03	1
Coffee, regular	-5	Healthy	N/A		
Coffee, decaffeinated	-5	Healthy	N/A		
Grapefruit	-5	Healthy	29.14	0.03	1
Melon	-5	Healthy	34.51	0.08	1
Onions	-5	Healthy	9.76	0.03	1
Oranges, satsuma, mandarins	-5	Healthy	34.60	0.05	1
Wholemeal pasta	-5	Healthy	6.31	1.57	1
Crisp bread	-4	Healthy	5.85	5.10	1
Other white fish, fresh or frozen	-4	Healthy	7.36	0.99	1
Bananas	-3	Healthy	5.75	0.18	1
Brown bread/rolls	-3	Healthy	6.00	8.65	3
Chips or French fries	-3	Healthy	3.58	1.50	2
Real fruit juice	-3	Healthy	7.61	6.62	1
Liver, liver pate, sausage	-3	Healthy	13.64	0.87	1
Pears	-3	Healthy	4.37	0.03	2

(Continued)

	WXYfm		SAIN,LIM		
FFQ-item	**Score**	**Category**	**SAIN**	**LIM**	**Q**
Tinned fruit	-3	Healthy	3.93	9.90	4
Grapes	-2	Healthy	3.46	0.02	2
Oily fish, fresh or canned	-2	Healthy	42.01	1.69	1
White or green pasta	-2	Healthy	3.37	0.55	2
Roast potatoes	-2	Healthy	3.73	1.15	2
White bread/rolls	-2	Healthy	4.52	7.84	4
Wholemeal bread/rolls	-2	Healthy	6.64	7.83	3
Brown rice	-1	Healthy	2.05	0.84	2
Chicken or other poultry	-1	Healthy	6.09	4.72	1
Pork: roast, chops or stew[#]	-1	Healthy	7.80	4.03	1
Skimmed milk	-1	Healthy	12.00	0.72	1
Soya milk	-1	Healthy	5.69	1.24	1
White rice	-1	Healthy	1.68	0.54	2
Yoghurt	-1	Healthy	8.48	3.74	1
Fried fish in batter	0	Healthy	3.09	4.12	2
Beef: roast, steak etc[#]	0	Healthy	8.44	5.87	1
Boiled, mashed, instant or jacket potatoes	0	Healthy	3.87	2.46	2
Cocoa, hot chocolate	0	Healthy	7.73	7.09	1
Coleslaw	0	Healthy	2.58	10.20	4
Lamb: roast, chops or stew[#]	0	Healthy	8.40	6.42	1
Low calorie or diet fizzy drinks	0	Healthy	N/A		
Pork: roast, chops or stew[#]	0	Healthy	6.60	6.57	1
Semi-skimmed milk	0	Healthy	8.66	2.20	1
Sterilized milk	0	Healthy	6.29	4.21	1
Tea	0	Healthy	N/A		
Vegetable soup	0	Healthy	2.79	7.95	4
Beef: roast, steak etc[#]	1	Inter.	7.61	6.85	1
Fizzy soft drinks	1	Less heal.	0.51	9.85	4
Lamb: roast, chops or stew[#]	1	Inter.	9.43	7.84	3
Lamb: roast, chops or stew[#]	1	Inter.	10.85	9.25	3
Fruit squash or cordial	1	Less heal.	4.90	3.35	2
Beef: roast, steak etc[#]	2	Inter.	6.67	8.61	3
Channel Islands milk	2	Less heal.	5.61	6.00	1
Cottage cheese, low fat soft cheese	2	Inter.	6.27	7.77	3
Dried fruit, e.g. raisins, prunes	2	Inter.	4.71	0.63	2
Full cream milk	2	Less heal.	5.97	4.52	1
Horlicks, Ovaltine	2	Less heal.	7.50	7.84	3
Lasagne	3	Inter.	3.95	10.24	4
Meat soup	3	Inter.	7.33	6.16	1
Milk puddings	3	Inter.	3.56	12.04	4
Pork: roast, chops or stew[#]	3	Inter.	5.74	9.12	3
Soya meat, TVP, vegeburger	3	Inter.	6.09	9.88	3
Fish fingers, fish cakes	4	Less heal.	3.24	10.21	4
Potato salad	4	Less heal.	1.47	6.65	2
Fruit pies, tarts, crumbles	5	Less heal.	3.12	17.40	4
Peanuts and other nuts	6	Less heal.	2.65	19.43	4
Peanut butter	6	Less heal.	2.66	27.53	4

(Continued)

	WXYfm		SAIN,LIM		
FFQ-item	**Score**	**Category**	**SAIN**	**LIM**	**Q**
Sauces, e.g. white/cheese sauce, gravy	6	Less heal.	4.35	11.30	4
Porridge, Readybrek	7	Less heal.	4.78	10.06	4
Shredded wheat, Weetabix etc	7	Less heal.	5.19	13.02	3
Muesli, Fruit'n' Fibre, etc	9	Less heal.	5.85	21.95	3
All-Bran, Bran Flakes etc	11	Less heal.	21.24	23.05	3
Cream crackers, cheese biscuits	11	Less heal.	2.95	14.66	4
Ice cream, choc ices	11	Less heal.	2.48	23.14	4
Jam, marmalade, honey	11	Less heal.	1.30	46.31	4
Shellfish	11	Less heal.	14.80	41.20	3
Corn Flakes, Rice Krispies, Special K	12	Less heal.	6.83	17.57	3
Ham	12	Less heal.	6.48	14.99	3
Tomato ketchup	12	Less heal.	1.41	35.57	4
Marmite, Bovril	12	Less heal.	10.40	46.53	3
Pizza	12	Less heal.	4.70	15.63	4
Single cream	12	Less heal.	1.88	18.93	4
Buns and pastries	13	Less heal.	2.57	24.41	4
Eggs	14	Less heal.	7.90	16.03	3
Savoury pies	14	Less heal.	2.88	15.97	4
Sweets, toffees, mints	14	Less heal.	0.23	58.06	4
Double or clotted cream	15	Less heal.	0.72	50.98	4
Sponge puddings	15	Less heal.	3.76	23.95	4
Sugar added to tea, coffee, cereal	15	Less heal.	0.14	70.05	4
Corned beef, spam, luncheon meats	16	Less heal.	8.59	18.02	3
Quiche	16	Less heal.	4.94	21.56	4
Pickles, chutney	17	Less heal.	1.90	39.62	4
Beef burgers	19	Less heal.	6.42	20.60	3
Bacon[#]	20	Less heal.	4.19	35.51	4
Coffee whitener	20	Less heal.	0.17	59.77	4
Dried milk	20	Less heal.	14.58	6.42	1
Frosties, Ricicles, Sugar Puffs, Coco Pops	20	Less heal.	5.82	36.59	3
Salad cream	20	Less heal.	0.71	27.11	4
Sausages	20	Less heal.	3.17	24.35	4
Bacon[#]	21	Less heal.	3.61	35.03	4
Biscuits	21	Less heal.	2.19	36.82	4
Crisps or other packet snacks	21	Less heal.	3.46	28.63	4
Low fat spread	21	Less heal.	8.88	24.96	3
Cheese, e.g. Cheddar, Brie, Edam	22	Less heal.	6.16	36.25	3
Bacon[#]	23	Less heal.	3.13	34.54	4
Cakes	23	Less heal.	1.37	41.12	4
Chocolates, chocolate bars	26	Less heal.	1.27	69.79	4
Butter	28	Less heal.	1.93	86.85	4
Hard margarine	28	Less heal.	8.68	61.65	3
Polyunsaturated margarine	28	Less heal.	8.68	34.84	3
Other soft margarine	28	Less heal.	8.68	50.34	3
French dressing, vinaigrette	28	Less heal.	4.83	25.68	4

Q, quadrant; Inter., intermediate; Less heal., less healthy. [#] For meats, participants were asked if they consumed all, some, or none of the visible fat.

Appendix 3: Chapter 8 results for the PES(Q1) aggregate score

Appendix 3.1: Goldberg cut-off, energy misreporting and PES(Q1)

Table A0.2: Mean PES(Q1) by reporting level

	Men				Women			
	Under	**Acceptable**	**Over**	**p***	**Under**	**Acceptable**	**Over**	**p***
PES(Q1)	32.9	28.5	28.9	<.001	38.9	35.9	39.0	<.001

*Heterogeneity ANOVA across reporting levels

Table A0.3: Distribution of reporting levels across PES(Q1) quartiles (4: healthier)

	Men				Women			
Column %	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
% Under	14.1	15.4	23.4	33.2	11.0	12.1	16.8	17.8
% Acceptable	79.4	79.7	72.8	62.4	73.5	78.3	74.8	62.9
% Over	6.50	4.89	3.74	4.49	15.5	9.66	8.39	19.4

χ^2 p<0.001 for both sexes

Table A0.4: Hazard ratio estimates for sensitivity analyses excluding energy misreporters, PES(Q1) quartiles (4: healthier)

Outcome, cases/total (numbers for acceptable reporters only).		Model 3			Model 3, acceptable reporters only		
		HR	**95**	**% CI**	**HR**	**95**	**% CI**
CHD, 318 / 7,174 (220 / 5,263)	1	Ref			Ref[#]		
	2	0.79	0.58	1.09	0.87	0.60	1.27
	3	**0.71**	**0.51**	**0.99**	0.79	0.53	1.17
	4	1.21	0.89	1.64	1.41	0.96	2.06
Linear		1.05	0.95	1.17	1.09	0.96	1.24
p quadratic trend		**0.010**			**0.042**		
Diabetes, 754 / 6,868 (511 / 5,060)	1	Ref*			Ref*		
	2	0.90	0.72	1.11	0.85	0.66	1.10
	3	1.02	0.82	1.25	1.00	0.78	1.28
	4	1.06	0.85	1.31	1.10	0.84	1.43
Linear		1.03	0.96	1.10	1.04	0.96	1.14
p quadratic trend		0.803			0.278		
Cancer mortality, 251 / 7,235 (185 / 5,309)	1	Ref			Ref		
	2	0.79	0.55	1.11	0.81	0.54	1.21
	3	0.73	0.51	1.05	0.73	0.48	1.11
	4	0.69	0.48	1.01	0.69	0.44	1.09
Linear		0.89	0.79	1.00	0.88	0.77	1.02
p quadratic trend		**0.027**			0.288		
All-cause mortality, 524 / 7,242 (372 / 5,312)	1	Ref			Ref[§]		
	2	**0.71**	**0.56**	**0.92**	**0.72**	**0.54**	**0.97**
	3	0.87	0.68	1.10	0.94	0.70	1.25
	4	0.79	0.61	1.02	0.82	0.60	1.12
Linear		0.95	0.87	1.03	0.97	0.87	1.07
p quadratic trend		**0.035**			0.559		

Model 3 adjusted for age, sex, ethnicity, marital status, employment grade, smoking status, physical activity level, energy and alcohol intake, BMI, hypertension and dyslipidaemia status, and prevalence of longstanding illness * Stratified for sex. [#] Stratified for BMI categories. [§] Stratified for longstanding illness and dyslipidaemia. HR, hazard ratio; CI, confidence interval

Appendix 3.2: Differential misreporting of foods and PES(Q1)

Table A0.5: Summary statistics for original PES(Q1) and regression calibration derived PES_C

Variable	N	Mean	Std Dev	Minimum	Median	Maximum
PES(Q1)	7,463	31.6	11.4	2.48	30.1	92.9
PES_C	7,463	16.1*	12.4	0.04	13.0	90.2

* Significantly different from original score (paired t-test p<0.001)

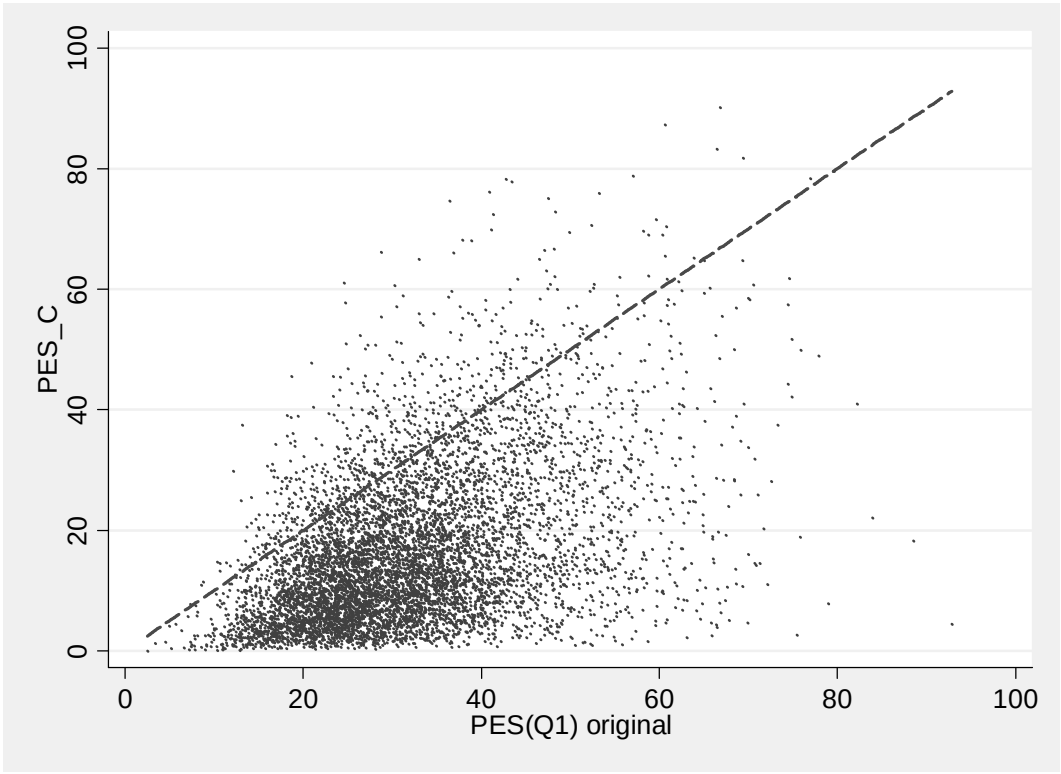


Figure A1: Regression calibration derived PES_C vs. original PES(Q1)
The dashed line represents the y=x function.

Table A0.6: Cox regression estimates across PES(Q1) and PES_C quartiles (4: healthier)

Outcome (cases / n)	Quartile/ trend	Original PES(Q1) HR	95	% CI	Corrected PES_C HR	95	% CI
CHD (318 / 7,174)	1	Ref			Ref[#]		
	2	0.79	0.58	1.09	1.30	0.95	1.76
	3	**0.71**	**0.51**	**0.99**	1.01	0.72	1.43
	4	1.21	0.89	1.64	1.24	0.87	1.75
	Linear	1.05	0.95	1.17	1.04	0.93	1.16
	p quadratic	**0.010**			0.746		
Diabetes (754 / 6,868)	1	Ref*			Ref*		
	2	0.90	0.72	1.11	1.18	0.96	1.44
	3	1.02	0.82	1.25	0.93	0.74	1.16
	4	1.06	0.85	1.31	0.99	0.79	1.23
	Linear	1.03	0.96	1.10	0.97	0.91	1.04
	p quadratic	0.803			0.096		
Cancer mortality (251 / 7,235)	1	Ref			Ref		
	2	0.79	0.55	1.11	0.98	0.68	1.43
	3	0.73	0.51	1.05	1.19	0.82	1.72
	4	0.69	0.48	1.01	1.19	0.81	1.74
	Linear	0.89	0.79	1.00	1.07	0.95	1.21
	p quadratic	**0.027**			0.618		
All-cause mortality (524 / 7,242)	1	Ref			Ref[§]		
	2	**0.71**	**0.56**	**0.92**	1.03	0.81	1.31
	3	0.87	0.68	1.10	0.87	0.67	1.13
	4	0.79	0.61	1.02	0.96	0.74	1.25
	Linear	0.95	0.87	1.03	0.97	0.89	1.06
	p quadratic	**0.035**			0.641		

Models adjusted for age, sex, ethnicity, marital status, employment grade, smoking status, physical activity level, energy and alcohol intake, BMI, hypertension and dyslipidaemia status, and prevalence of longstanding illness * Stratified for sex. [#] Stratified for BMI categories. [§] Stratified for longstanding illness and dyslipidaemia. HR, hazard ratio; CI, confidence interval

Appendix 3.3: Food variety score and PES(Q1) aggregate score

Table A0.7: Mean food variety score (FVS) across PES(Q1) quartiles (4=healthier)

	Men				Women			
	1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
FVS	43.6	44.9	43.9	41.1	39.0	43.8	43.4	41.3

p<.001 in both sexes for heterogeneity ANOVA across quartiles

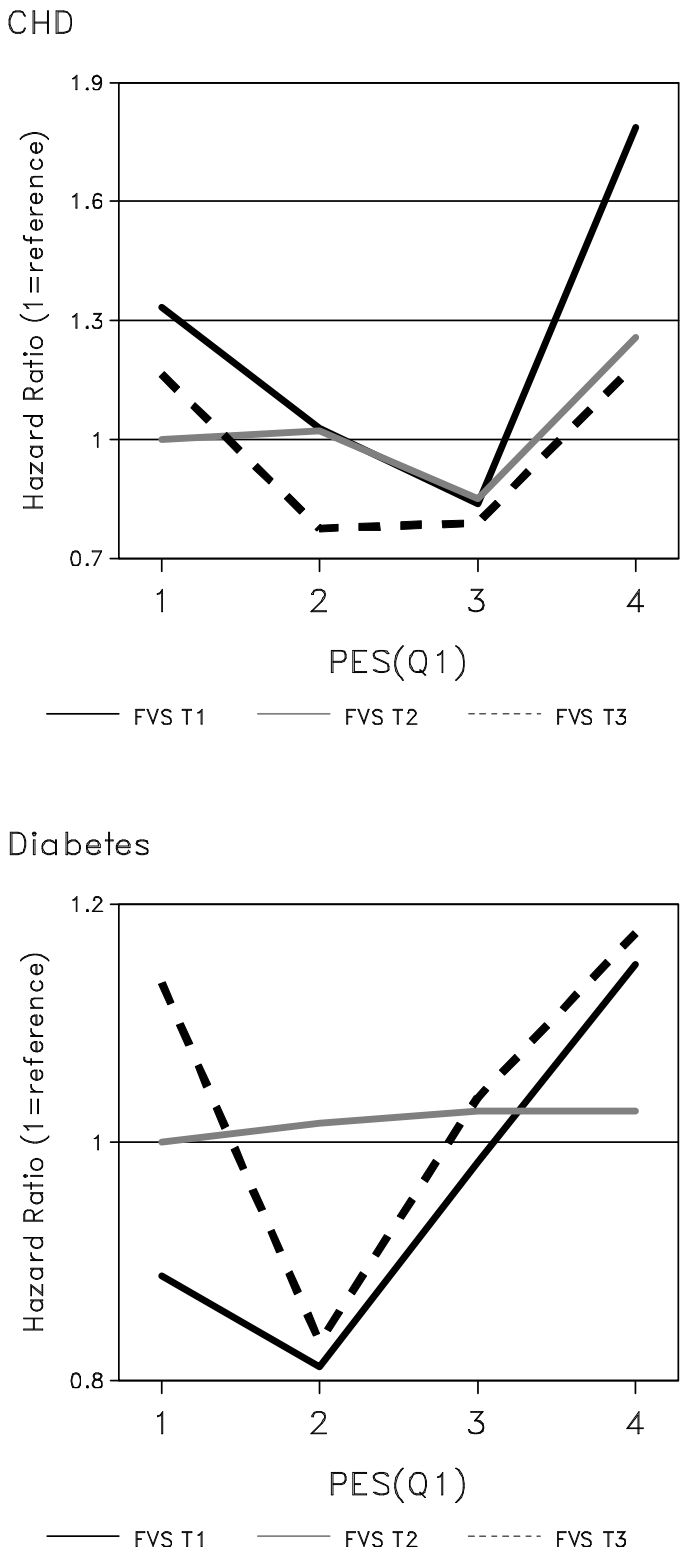


Figure A2: Hazard ratio estimates across PES(Q1) quartiles (4: healthier), stratified by FVS tertiles

The reference group was the 1st quartile of PES(Q1) with the 2nd tertile of FVS. Models were adjusted for age, sex, and ethnicity. FVS, Food variety score

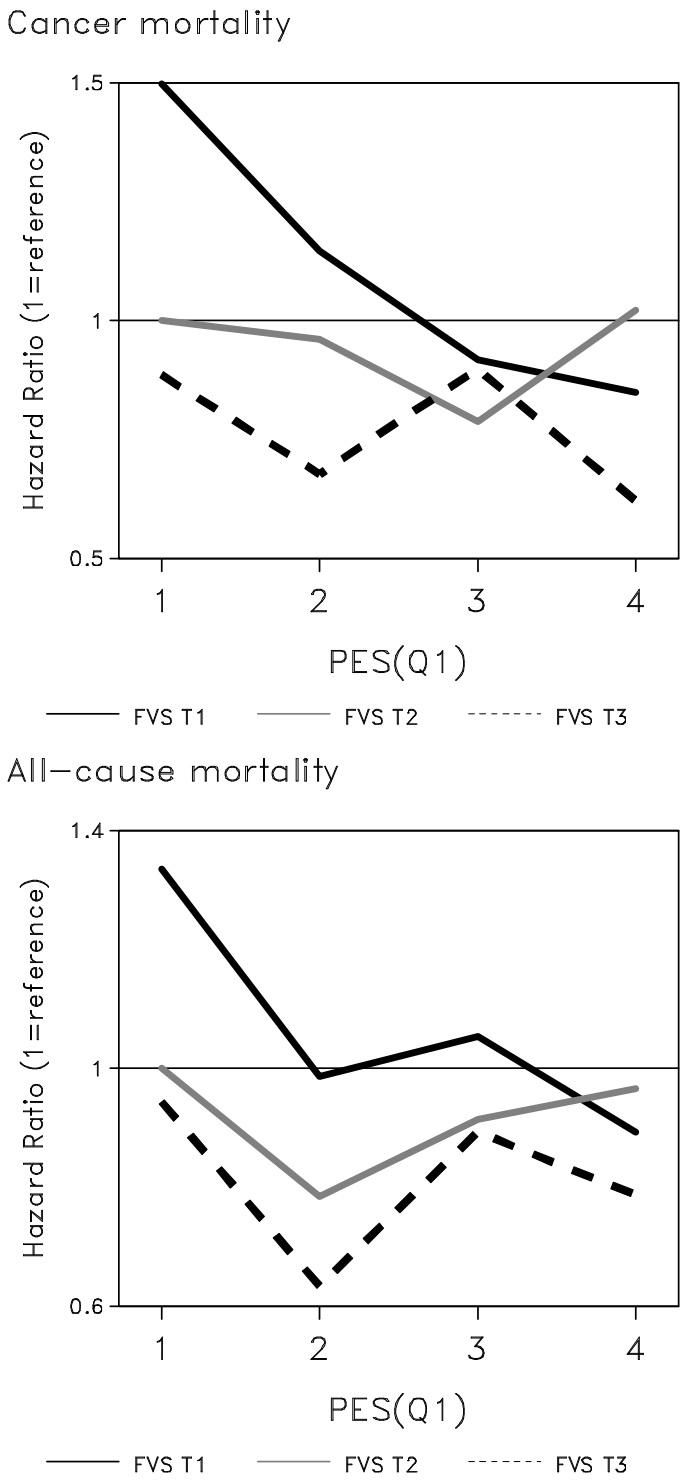


Figure A2(continued): Hazard ratio estimates across PES(Q1) quartiles (4: healthier), stratified FVS tertiles

The reference group was the 1st quartile of PES(Q1) with the 2nd tertile of FVS. Models were adjusted fir age, sex, and ethnicity. FVS, Food variety score

Appendix 3.4: SAIN,LIM component analysis

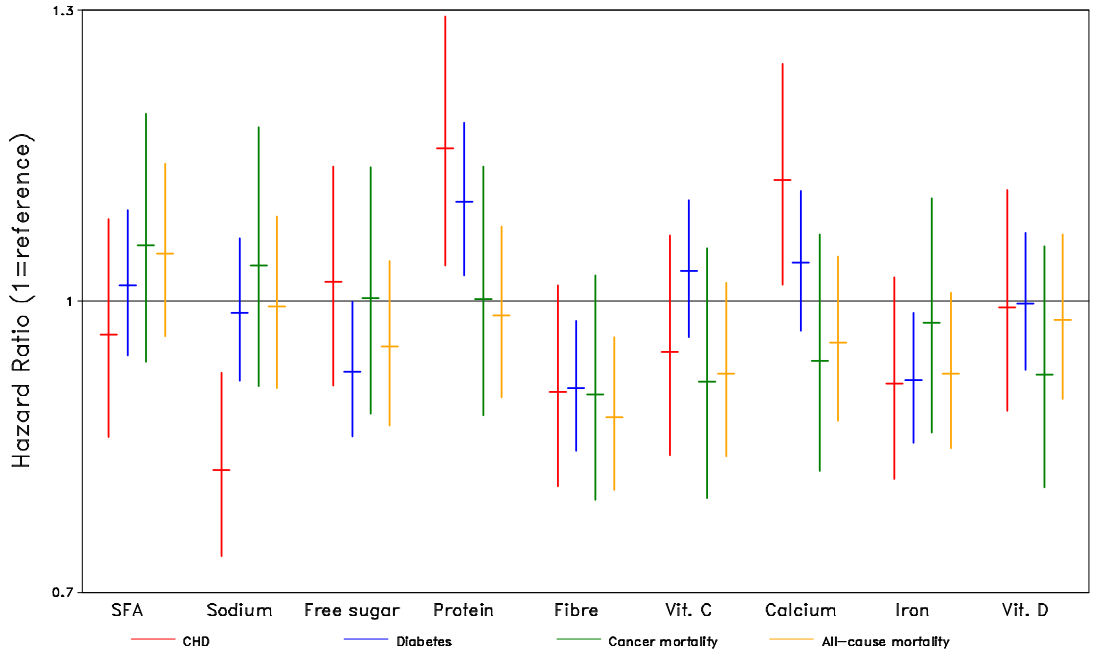


Figure A3: Hazard ratio estimates and 95% CI for SAIN,LIM "component score" Z-scores (n=7251)

Models adjusted for age, sex, and ethnicity. Diabetes models were stratified for sex.
SFA, Saturated fatty acid.

The results for the SAIN,LIM components presented in figure A3 were similar to the WXYfm ones (figure 8.4) for sodium, saturated fats, sugars, and fibre. Indeed, the negative components were not associated with risk increase, and fibre was protective for all outcomes. In addition, the protein and calcium positive components were associated with increased incidence of CHD and diabetes. The unexpected results for the negative components appeared to be explained by the association between lox energy reporting and vascular risk (table A8).

Table A9 showed that the PES(Q1) aggregate score was mainly correlated with the negative components and with the protein and calcium positive components. Likewise EWS, the U-shaped associations obtained with PES(Q1) could have been explained by the unexpected results obtained for the negative components. The associations observed for protein and calcium would have strengthened the quadratic trend, particularly for CHD and diabetes, resulting in the J-shapes of table 7.11.

Table A9 also showed that the other positive components (vitamins C and D, iron, and fibre) were more associated with PES(Q1) than the protein and fibre components were with EWS (table 8.14). The SAIN,LIM model appeared to be more associated with the positive components than WXYfm. These stronger correlations could have explained the stronger risk reduction observed for some individual quartiles of PES(Q1) in table 7.11.

Table A0.8: SAIN,LIM component scores and energy reporting

Component	Rank correlation with reported energy intake	Mean component score by reporting level			
		Low	Acceptable	High	p*
Saturated fat	0.20	19.4	22.5	23.4	<.001
Sodium	0.18	8.66	9.44	9.07	<.001
Sugar	0.20	13.9	16.1	16.5	<.001
Protein	-0.21	7.09	6.69	6.70	<.001
Fibre	-0.16	4.79	4.65	4.21	<.001
Vitamin C	-0.22	7.05	6.14	5.99	<.001
Calcium	-0.03	6.00	6.09	7.62	<.001
Iron	-0.22	4.80	4.64	4.19	<.001

*Heterogeneity ANOVA across reporting level

Table A0.9: Rank correlations between SAIN,LIM component scores, nutrient intake and PES(Q1)

	SFA	Na	Free sugars	Protein	Fibre	Vit. C	Ca	Fe	Vit. D
Nutrient intake	0.59	0.44	0.71	0.37	0.64	0.82	0.67	0.40	0.86
PES(Q1)	-0.52	-0.64	-0.46	0.65	0.11	0.45	0.58	0.14	0.08*

SFA, saturated fatty acids; Vit. Vitamin. Na, sodium; Ca, calcium; Fe, iron. *Vitamin D was an optional nutrient, hence the low correlation.

Appendix 4: T-tests and regression calibration estimates between FFQ and diet diaries food items

Table A0.10: T-tests and regression calibration estimates for all food items, by food group

	T-test		Regression calibration estimates			
FFQ/7DD item	**Mean difference[#]**	**p**	**Intercept (λ_0)**	**λ_{REG}**	**R^2**	**n. param**
Apples	25.0	<.001	-0.18	0.710	0.343	19
Bananas	8.26	<.001	-2.61	0.628	0.324	19
Grapefruit	1.86	<.001	-5.17	0.392	0.236	16
Oranges	33.4	<.001	-3.61	0.544	0.227	17
Pears	11.7	<.001	-6.06	0.325	0.189	18
Dried fruit	-2.06	<.001	-3.36	0.366	0.158	16
Grapes	9.39	<.001	-5.55	0.309	0.162	17
Nuts	-0.62	0.020	-2.64	0.360	0.157	17
Tinned fruit	3.77	<.001	-7.98	0.198	0.095	16
Peaches	10.6	<.001	-5.30	0.120	0.084	19
Melon	5.29	<.001	-5.09	0.116	0.072	16
Strawberries	23.6	<.001	-4.50	0	0.045	15
Tomatoes	12.0	<.001	0.65	0.569	0.188	17
Soya meat	0.56	<.001	-3.05	0.213	0.201	18
Green salad	-1.72	<.001	0.03	0.455	0.186	18
Baked beans	5.76	<.001	-5.17	0.383	0.169	19
Peppers	3.41	<.001	-2.27	0.255	0.188	20
Parsnips	1.65	<.001	-3.71	0.290	0.139	17
Mushrooms	-2.78	<.001	-3.07	0.463	0.128	16
Coleslaw	1.15	<.001	-4.39	0.252	0.114	17
Vegetable soup	-0.24	0.864	-3.00	0.324	0.157	21
Broccoli	2.79	<.001	-1.43	0.330	0.113	17
Peas	5.98	<.001	-1.98	0.423	0.138	19
Carrots	5.80	<.001	-0.68	0.490	0.111	18
Marrow	2.97	<.001	-3.44	0.192	0.106	16
Onions	12.3	<.001	-2.15	0.333	0.100	17
Spinach	2.78	<.001	-3.39	0.156	0.086	16
Leeks	2.67	<.001	-4.36	0.141	0.092	17
Sprouts	8.17	<.001	-7.65	0.232	0.090	17
Cauliflower	3.38	<.001	-3.35	0.336	0.087	17
Lentils	-1.38	<.001	-2.48	0.154	0.105	16
Garlic	0.50	<.001	-3.87	0.154	0.098	16
Green beans	11.0	<.001	-3.36	0.358	0.088	16
Cabbage	6.87	<.001	-4.20	0.264	0.072	17
Spring greens	6.98	<.001	-6.33	0.080	0.057	17
Tofu	0.18	0.019	-5.99	0.056	0.041	17

n. param: Number of independent variables included in the model after stepwise selection (p=0.01).
[#]Mean (FFQ – 7DD) difference in g/d.

Table (continued)

FFQ/7DD item	T-test		Regression calibration estimates			
	Mean difference[#]	p	Intercept (λ_0)	λ_{REG}	R^2	n. param
Eggs	-2.66	<.001	-3.72	0.603	0.220	16
Poultry	6.28	<.001	0.78	0.654	0.194	17
Luncheon meats	-0.34	0.166	-4.52	0.339	0.180	17
Ham	-1.34	<.001	-1.09	0.435	0.169	18
Pork	1.43	0.002	-3.08	0.387	0.148	16
Bacon	-4.25	<.001	-5.58	0.354	0.197	20
Sausages	-3.36	<.001	-3.85	0.321	0.170	19
Savoury pies	-11.2	<.001	-2.72	0.281	0.166	20
Beef	2.00	0.021	-4.01	0.466	0.136	19
Liver	-1.72	<.001	-5.71	0.291	0.117	17
Lamb	1.31	0.009	-2.29	0.344	0.110	17
Beef burgers	-0.27	0.095	-4.35	0.187	0.089	16
Meat soup	6.98	<.001	-6.00	0.060	0.030	16
Oily fish	-4.85	<.001	-4.21	0.439	0.190	17
Shellfish	-2.42	<.001	-0.90	0.400	0.203	17
Fish fingers	0.23	0.134	-6.36	0.190	0.089	16
White fish	4.15	<.001	-4.99	0.284	0.100	18
Fried fish	1.34	0.002	-3.83	0.186	0.093	19
Muesli	-1.05	0.024	-1.16	0.704	0.532	17
Shredded cereals	2.68	<.001	-1.42	0.608	0.472	17
Brans	2.94	<.001	-1.27	0.550	0.417	18
Corn flakes	0.10	0.655	-3.21	0.564	0.370	17
Porridge	3.85	<.001	-3.16	0.455	0.325	15
Wholemeal bread	8.37	<.001	-0.08	0.472	0.277	18
Frosties	0.14	0.167	-3.79	0.333	0.199	15
Crispbread	0.22	0.201	-3.93	0.403	0.208	17
White bread	-17.8	<.001	3.07	0.365	0.223	18
Chips	0.96	0.234	-3.24	0.354	0.244	22
Roast potatoes	5.03	<.001	-2.67	0.384	0.179	18
Brown rice	6.09	<.001	-2.87	0.215	0.189	21
Boiled potatoes	14.4	<.001	2.11	0.458	0.175	18
Crackers	-0.75	<.001	-6.81	0.330	0.147	18
Pasta	9.39	<.001	-0.66	0.294	0.126	18
White rice	-1.60	0.045	-0.56	0.298	0.169	20
Brown bread	1.37	0.267	-1.80	0.269	0.120	20
Wholemeal pasta	4.53	<.001	-5.69	0.099	0.074	17
Potato salad	1.96	<.001	-5.95	0.088	0.039	16

n. param: Number of independent variables included in the model after stepwise selection (p=0.01).
[#]Mean (FFQ – 7DD) difference in g/d.

Table (continued)

FFQ/7DD item	T-test		Regression calibration estimates			
	Mean difference[#]	p	Intercept (λ_0)	λ_{REG}	R^2	n. param
Channel Island milk	1.25	0.008	-3.19	0.527	0.513	16
Skimmed milk	63.1	<.001	-1.47	0.629	0.494	20
Semi-skimmed milk	137	<.001	-1.33	0.559	0.490	22
Soya milk	1.74	0.115	-3.67	0.433	0.396	15
Coffee whitener	0.63	<.001	-3.17	0.478	0.366	16
Yoghurt	16.4	<.001	-2.04	0.630	0.379	18
Dried milk	0.12	0.188	-3.48	0.552	0.367	18
Whole milk	57.6	<.001	0.24	0.441	0.389	18
Cheese	-5.69	<.001	1.56	0.515	0.247	20
Cottage cheese	-0.78	0.033	-4.66	0.220	0.169	19
Single cream	-1.70	<.001	-2.86	0.253	0.137	18
Sterilised milk	7.23	0.018	-6.23	0.112	0.078	15
Double cream	0.00	0.994	-5.37	0.152	0.074	16
Jam	1.98	<.001	1.52	0.652	0.377	17
Polyunsaturated margarine	3.17	<.001	-0.65	0.544	0.386	21
Low-fat spread	-0.40	0.041	-3.71	0.547	0.345	17
Marmite	0.01	0.822	-4.68	0.492	0.334	17
Butter	-1.65	<.001	-1.06	0.499	0.363	20
Peanut butter	0.28	0.002	-2.58	0.403	0.273	15
Vinaigrette	0.80	<.001	-3.30	0.260	0.159	18
Pizza	1.68	<.001	-3.57	0.272	0.146	16
Pickles	3.15	<.001	-4.63	0.254	0.145	19
Mayonnaise	1.20	<.001	-0.63	0.269	0.127	16
Sauces	3.20	<.001	0.93	0.194	0.105	19
Ketchup	1.90	<.001	-6.34	0.107	0.069	17
Quiche	1.35	<.001	-4.92	0.156	0.062	15
Soft margarine	-0.42	0.001	-5.91	0.215	0.062	16
Lasagne	4.05	<.001	-5.06	0.132	0.060	17
Hard margarine	-0.20	0.086	-6.31	0.111	0.019	16
Sugar (in drinks, cereals)	-2.30	<.001	-0.04	0.656	0.484	19
Crisps	2.14	<.001	-1.53	0.408	0.202	15
Biscuits	-4.32	<.001	-0.18	0.506	0.211	17
Chocolate	2.62	<.001	0.50	0.382	0.216	19
Ice cream	-2.57	<.001	-5.05	0.410	0.149	15
Sweets	-0.33	0.120	-6.20	0.366	0.156	17
Cakes	-1.08	0.105	-3.08	0.422	0.144	17
Tarts	1.90	0.002	-3.09	0.311	0.165	18
Milk pudding	-8.39	<.001	-2.44	0.222	0.117	19
Buns & pastries	-4.62	<.001	0.44	0.139	0.127	20
Sponge puddings	-0.28	0.489	-4.54	0.138	0.060	16

n. param: Number of independent variables included in the model after stepwise selection (p=0.01).
[#]Mean (FFQ – 7DD) difference in g/d.

Table (continued)

FFQ/7DD item	T-test		Regression calibration estimates			
	Mean difference[#]	p	Intercept (λ_0)	λ_{REG}	R^2	n. param
Tea	-12.0	0.097	1.62	0.839	0.660	18
Beer	-44.9	0.000	3.24	0.603	0.436	17
Horlicks	-4.05	0.002	-5.29	0.619	0.429	18
Wine	-13.7	0.000	3.04	0.594	0.445	20
Spirits	-0.81	0.005	-5.24	0.559	0.351	18
Cocoa	2.02	0.192	-4.71	0.444	0.308	18
Coffee	-96.9	0.000	4.13	0.471	0.418	21
Fruit juice	25.2	0.000	-1.75	0.576	0.308	16
Port	0.11	0.692	-5.00	0.347	0.210	18
Squash	11.9	0.000	-2.00	0.321	0.182	18
Fizzy drinks	-6.66	0.000	-5.82	0.301	0.132	18
Liqueurs	0.07	0.139	-5.96	0.118	0.055	15

n. param: Number of independent variables included in the model after stepwise selection (p=0.01).
[#]Mean (FFQ – 7DD) difference in g/d

Appendix 5: Food variety score as a confounding variable

Table A11 estimates revealed that additional adjustment for diet variety changed very slightly the original estimates. No new trends appeared and conclusions from chapters 6 and 7 were not altered. Results for diabetes were equally similar to the original models (not shown).

Table A0.11: Cox regressions estimates across EWS and PES(Q1) quartiles (4: healthier)

Outcome, quartile/trend		Model 1 HR	95	% CI	Model 1 + FVS HR	95	% CI
EWS							
CHD	1	Ref			Ref		
	2	0.78	0.56	1.10	0.80	0.57	1.12
	3	1.06	0.77	1.45	1.08	0.79	1.48
	4	1.31	0.96	1.79	1.29	0.95	1.76
Linear		**1.12**	**1.01**	**1.25**	**1.12**	**1.01**	**1.24**
Quadratic		**<.001**			**0.001**		
Cancer mortality	1	Ref			Ref		
	2	0.94	0.66	1.35	0.97	0.68	1.39
	3	1.01	0.71	1.43	1.03	0.72	1.46
	4	0.95	0.66	1.36	0.94	0.65	1.36
Linear		0.99	0.88	1.11	0.99	0.88	1.11
Quadratic		**0.032**			0.075		
All-cause mortality	1	Ref			Ref		
	2	0.85	0.66	1.09	0.86	0.67	1.11
	3	0.89	0.69	1.14	0.91	0.71	1.16
	4	1.04	0.81	1.33	1.03	0.81	1.32
Linear		1.02	0.94	1.10	1.02	0.94	1.10
Quadratic		**<.001**			**0.002**		
PES(Q1)							
CHD	1	Ref			Ref		
	2	0.80	0.58	1.09	0.81	0.59	1.12
	3	**0.71**	**0.51**	**0.98**	0.72	0.52	1.00
	4	1.23	0.91	1.67	1.22	0.90	1.65
Linear		1.06	0.95	1.17	1.05	0.95	1.17
Quadratic		**0.002**			**0.008**		
Cancer mortality	1	Ref			Ref		
	2	0.80	0.56	1.13	0.82	0.58	1.17
	3	0.76	0.53	1.08	0.78	0.55	1.11
	4	0.73	0.51	1.06	0.74	0.51	1.06
Linear		0.90	0.80	1.02	0.91	0.81	1.02
Quadratic		**0.029**			0.081		
All-cause mortality	1	Ref			Ref		
	2	**0.71**	**0.56**	**0.92**	**0.73**	**0.57**	**0.94**
	3	0.86	0.68	1.09	0.88	0.69	1.12
	4	0.80	0.62	1.03	0.80	0.63	1.03
Linear		0.95	0.88	1.03	0.95	0.88	1.03
Quadratic		**0.011**			**0.042**		

Model 1 adjusted for age, sex, and ethnicity. HR, hazard ratio; CI, confidence interval; FVS, food variety score. The diabetes estimates not shown in this table were also very slightly influence by FVS.

Appendix 6: Chapter 9 results for SAIN,LIM nutrient profiling model, and alternative WXYfm algorithms

Appendix 6.1 SAIN,LIM alternative aggregating algorithms

As for WXYfm, two new aggregate score were implemented:

- EW(SAIN) which was the energy weighted score for the SAIN sub-score. It therefore counted positive nutrients only and was comparable to the EWS+ algorithm presented in chapter 9. It was quite highly related to the PES(Q1) aggregate score used previously (rank correlation 0.63).
- RFS(Q1) which counted the number of foods and drinks from quadrant 1 reported to be consumed once a week or more. This diversity and semi-quantitative score was moderately linked to PES(Q1), with a rank correlation of 0.29.

Appendix 6.2 Survival analyses results for EW(SAIN) and RFS(Q1)

Table A0.12: Cox regression estimates across quartiles of the EW(SAIN) aggregate score (4: healthier)

Outcome (cases / n)	Quartile/ trend	Model 1 HR	95	% CI	Model 3 HR	95	% CI
CHD (318 / 7,174)	1	Ref			Ref[#]		
	2	**0.65**	**0.47**	**0.89**	**0.68**	**0.49**	**0.94**
	3	0.84	0.62	1.13	0.89	0.66	1.21
	4	0.90	0.67	1.23	0.94	0.69	1.30
	Linear	0.99	0.89	1.09	1.00	0.90	1.11
	p quadratic	0.056			0.242		
Diabetes (754 / 6,868)	1	Ref*			Ref*		
	2	0.92	0.76	1.13	0.98	0.80	1.20
	3	**0.80**	**0.65**	**0.99**	0.82	0.66	1.01
	4	0.95	0.78	1.17	0.98	0.79	1.21
	Linear	0.97	0.91	1.04	0.98	0.91	1.05
	p quadratic	0.098			0.247		
Cancer mortality (251 / 7,235)	1	Ref			Ref		
	2	0.94	0.66	1.34	1.00	0.70	1.42
	3	0.83	0.58	1.18	0.84	0.58	1.21
	4	0.85	0.59	1.22	0.84	0.58	1.22
	Linear	0.94	0.84	1.05	0.93	0.83	1.05
	p quadratic	**<.001**			**0.003**		
All-cause mortality (524 / 7,242)	1	Ref			Ref[§]		
	2	0.80	0.63	1.02	0.84	0.66	1.08
	3	**0.77**	**0.60**	**0.98**	0.81	0.63	1.03
	4	0.82	0.64	1.04	0.85	0.66	1.09
	Linear	0.94	0.86	1.01	0.95	0.87	1.03
	p quadratic	**0.008**			**0.025**		

Model 1 adjusted for age, sex, and ethnicity. Model 3 further adjusted for marital status, employment grade, smoking status, physical activity level, and energy and alcohol intake, BMI, hypertension and dyslipidaemia status, and prevalence of longstanding illness.
[#] Stratified for BMI categories * Stratified for sex. [§] Stratified for longstanding illness and dyslipidaemia. HR, hazard ratio; CI, confidence interval

Table A0.13: Cox regression estimates across quartiles of the RFS(Q1) aggregate score (4: healthier)

Outcome	Quartile/ trend	Model 1 HR	95	% CI	Model 3 HR	95	% CI
CHD (318 / 7,174)	1	Ref			Ref[#]		
	2	0.91	0.67	1.24	0.96	0.70	1.32
	3	0.90	0.65	1.24	1.00	0.72	1.40
	4	0.77	0.56	1.07	0.89	0.62	1.27
	Linear	0.92	0.84	1.02	0.97	0.87	1.08
	p quadratic	0.703			0.715		
Diabetes (754 / 6,868)	1	Ref*			Ref*		
	2	1.03	0.84	1.27	1.07	0.87	1.33
	3	0.92	0.74	1.15	1.00	0.80	1.26
	4	1.11	0.90	1.37	1.18	0.94	1.49
	Linear	1.02	0.96	1.09	1.05	0.97	1.12
	p quadratic	0.611			0.953		
Cancer mortality (251 / 7,235)	1	Ref			Ref		
	2	0.71	0.50	1.00	0.75	0.53	1.07
	3	0.73	0.51	1.04	0.78	0.54	1.14
	4	**0.69**	**0.48**	**0.98**	0.74	0.50	1.09
	Linear	0.90	0.80	1.01	0.92	0.81	1.04
	p quadratic	**0.020**			**0.048**		
All-cause mortality (524 / 7,242)	1	Ref			Ref[§]		
	2	**0.76**	**0.60**	**0.97**	0.81	0.64	1.04
	3	**0.74**	**0.57**	**0.95**	0.82	0.63	1.07
	4	**0.75**	**0.59**	**0.96**	0.84	0.65	1.10
	Linear	**0.92**	**0.85**	**0.99**	0.96	0.88	1.04
	p quadratic	0.055			0.079		

Model 1 adjusted for age, sex, and ethnicity. Model 3 further adjusted for marital status, employment grade, smoking status, physical activity level, and energy and alcohol intake, BMI, hypertension and dyslipidaemia status, and prevalence of longstanding illness.
[#] Stratified for BMI categories * Stratified for sex. [§] Stratified for longstanding illness and dyslipidaemia. HR, hazard ratio; CI, confidence interval

Appendix 6.3: FFQ-items classification using WXYfm positive components only

Item	Score*
Salad cream	0
French dressing, vinaigrette	0
Butter	0
Hard margarine	0
Polyunsaturated margarine	0
Other soft margarine	0
Sweets, toffees, mints	0
Sugar added to tea, coffee, cereal	0
Tea	0
Fizzy soft drinks	0
Low calorie or diet fizzy drinks	0
Fruit squash or cordial	0
Porridge, Readybrek	1
Potato salad	1
Full cream milk	1
Soya milk	1
Coffee whitener	1
Single cream	1
Double or clotted cream	1
Meat soup	1
Pickles, chutney	1
Boiled, mashed, instant or jacket potatoes	2
White rice	2
Brown rice	2
Semi-skimmed milk	2
Skimmed milk	2
Channel Islands milk	2
Sterilized milk	2
Milk puddings	2
Ice cream, choc ices	2
Chocolates, chocolate bars	2
Sauces, e.g. white/cheese sauce, gravy	2
Jam, marmalade, honey	2
Cocoa, hot chocolate	2
Horlicks, Ovaltine	2
Roast potatoes	3
White or green pasta	3
Yoghurt	3
Low fat spread	3
Frosties, Ricicles, Sugar Puffs, Coco Pops	4
Lasagne	4
Cakes	4
Fruit pies, tarts, crumbles	4
Sponge puddings	4
Beef: roast, steak etc	5
Beefburgers	5
Pork: roast, chops or stew	5
Lamb: roast, chops or stew	5
Chicken or other poultry	5
Bacon	5
Ham	5
Corned beef, spam, luncheon meats	5
Sausages	5
Savoury pies	5
Liver, liver pate, sausage	5
Fried fish in batter	5
Fish fingers, fish cakes	5
Other white fish, fresh or frozen	5
Oily fish, fresh or canned	5
Shellfish	5
Chips or French fries	5
Dried milk	5
Cheese, e.g. Cheddar, Brie, Edam	5
Cottage cheese, low fat soft cheese	5
Eggs	5
Quiche	5
Vegetable soup	5
Tomato ketchup	5
Marmite, Bovril	5
Coffee, regular	5
Coffee, decaffeinated	5
Real fruit juice	5
Pears	5
Grapes	5
Melon	5
Soya meat, TVP, vegeburger	5
Wholemeal pasta	6
Pizza	6
Biscuits	6
Buns and pastries	6
Oranges, satsumas, mandarins	6
Grapefruit	6
Tinned fruit	6
Broccoli	6
Marrow, courgettes	6
Onions	6
Sweet peppers	6
Green salad	6
Tomatoes	6
Coleslaw	6
Corn Flakes, Rice Krispies, Special K	7
Apples	7
Peaches, plums, apricots	7

Item	Score*
Strawberries, raspberries	7
Carrots	7
Cabbage	7
Cauliflower	7
Leeks	7
Crisps or other packet snacks	8
Bananas	8
Brussels sprouts	8
Mushrooms	8
White bread/rolls	9
Spinach	9
Spring greens, kale	9
Green beans, broad beans, runner beans	9
Parsnips, turnips, swedes	9
Garlic	9
Brown bread/rolls	10
Wholemeal bread/rolls	10
Cream crackers, cheese biscuits	10
Crispbread	10
Shredded wheat, Weetabix etc	10
Muesli, Fruit'n' Fibre, etc	10
All-Bran, Bran Flakes etc	10
Dried lentils, beans, peas	10
Tofu or soya bean curd	10
Dried fruit, e.g. raisins, prunes	11
Peas	13
Baked beans	13
Peanuts and other nuts	15
Peanut butter	15

*score = protein points + fibre points + fruit, vegetable and nuts points

Appendix 6.4: RFS(*healthier*) results

In order to assess the influence of the healthiness threshold value chosen for the WXYfm model, an alternative RFS selecting "healthier" foods only (i.e. threshold of 0 on the overall score scale) was implemented. This new RFS(*healthier*) was highly related to the RFS(WXYfm), with a rank correlation of 0.99. This was reflected in the Cox regression estimates which displayed very similar trends (table A14). Yet, the associations observed for cancer and all-cause mortality were stronger.

Table A0.14: Cox regression estimates across quartiles of the RFS(*healthier*) aggregate score (4: healthier)

Outcome	Quartile/ trend		Model 1 HR	95	% CI	Model 3 HR	95	% CI
CHD (318 / 7,174)		1	Ref			Ref[#]		
		2	0.76	0.55	1.05	0.79	0.57	1.09
		3	0.76	0.56	1.03	0.85	0.61	1.17
		4	0.86	0.63	1.17	0.98	0.69	1.39
	Linear		0.95	0.86	1.06	1.00	0.90	1.12
Diabetes (754 / 6,868)		1	Ref*			Ref*		
		2	0.90	0.73	1.12	0.92	0.74	1.15
		3	0.83	0.67	1.02	0.88	0.71	1.09
		4	1.01	0.83	1.24	1.02	0.82	1.28
	Linear		1.00	0.93	1.07	1.01	0.93	1.08
Cancer mortality (251 / 7,235)		1	Ref			Ref		
		2	**0.62**	**0.43**	**0.89**	**0.65**	**0.45**	**0.94**
		3	**0.60**	**0.43**	**0.85**	**0.64**	**0.45**	**0.92**
		4	0.72	0.51	1.01	0.78	0.53	1.14
	Linear		0.90	0.80	1.01	0.92	0.81	1.05
All-cause mortality (524 / 7,242)		1	Ref			Ref[§]		
		2	**0.66**	**0.51**	**0.84**	**0.70**	**0.54**	**0.91**
		3	**0.65**	**0.51**	**0.82**	**0.72**	**0.56**	**0.92**
		4	**0.76**	**0.60**	**0.97**	0.86	0.66	1.12
	Linear		**0.92**	**0.85**	**0.99**	0.96	0.87	1.04

Model 1 adjusted for age, sex, and ethnicity. Model 3 further adjusted for marital status, employment grade, smoking status, physical activity level, and energy and alcohol intake, BMI, hypertension and dyslipidaemia status, and prevalence of longstanding illness.
[#] Stratified for BMI categories * Stratified for sex. [§] Stratified for longstanding illness and dyslipidaemia. HR, hazard ratio; CI, confidence interval

Appendix 7: Fruit and vegetable case study for implementation of measurement error models

This appendix presents the case study conducted for the selection of the regression calibration model of chapter 8. Several measurement error models were tested on fruit and vegetable intake derived from both food frequency questionnaire (FFQ) and diet diary data. Total intake of fruit and vegetable was considered so that plasma β-carotene could be used as biomarker of dietary intake. In this case study, diet diary data coded in Cambridge using the DINER programme for the UK Dietary Cohort Consortium could also be used. It was further referred to as CNC-coded data. Diet diary data coded by the Whitehall II study team was referred to as UCL-coded data.

Appendix 7.1: Material and Methods

(i) Measures of fruit and vegetable intake

Table A15 presents the summary statistics of all the main variables: food frequency questionnaire (FFQ), 7-day diary (7DD, UCL and CNC coded), and β-carotene measurements (main sample and repeats). All these variables were positively skewed and log-transformations were applied to obtain distribution closer to normality. Distributions of log-transformed variables were close to normal (figure A4).

Table A0.15: Summary statistics for fruit and vegetable reported intake and beta-carotene measurements

		N	Mean	SD	Min	Med	Max
Fruit & Vegetable (g/d)	**FFQ**	8225	513	291	0	458	5503
	7DD – UCL*	1350	284	146	0	257	1071
	7DD – CNC*	570	340	159	46.5	309	1181
β -carotene (mmol/L)	**Main sample**	6418	0.84	0.58	0.02	0.75	7.34
	Repeats	397	0.85	0.58	0.02	0.74	6.57

* Centre of coding using different programs: WFOOD (UCL) and DINER (CNC). SD, standard deviation; Med, median value.

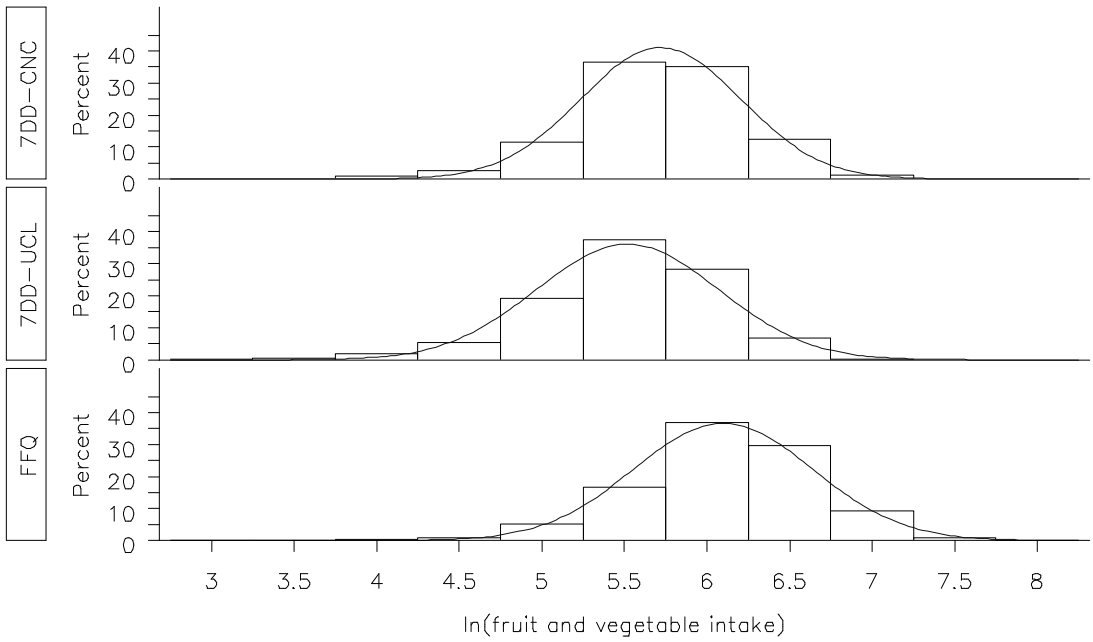


Figure A4: Log-transformed distribution of fruit and vegetable intake

Table A15 and figure A4 illustrated well the difference in reported intake between the two dietary assessment methods, with higher reported values for the FFQ. This followed the expectations that participants tended to over-report their fruit and vegetable consumption with the FFQ. Repeated β-carotene measurements yielded similar results to the main sample ones (not shown).

(ii) Measurement error models

Two sets of correcting techniques were implemented. The first one was based on regression calibration and its derivative and used only one alternative for dietary intake, the 7DD data. The second method was derived from the Triads model (structural equation modelling) and included biomarkers as another external measure of food intake. Both models are presented in the general case, for implementation of a full set of foods. There were applied on total fruit and vegetable intake only.

Terminology and notations

Let consider T_{ij} as the *true* intake of food i in participant j, which cannot be observed. The food record (7-day diary), R_{ij} , and FFQ, Q_{ij} , are two surrogate measures of T_{ij} and are measured with some error.

The 7DD data was considered as the closest measure to true food intake and followed a simple random error model:

$$R_{ij} = T_{ij} + \varepsilon_{Rij} \quad [1]$$

Where the errors are independent of T_{ij} and of each other and $\varepsilon_{Rij} \sim N(0, \sigma^2)$.

FFQ data were likely to be biased measures of true intake. Therefore, FFQ measures were assumed to follow the linear model defined below:

$$Q_{ij} = \alpha_{Qi} + \beta_{Qi}T_{ij} + \varepsilon_{Qij} \quad [2]$$

Where ε_{Qij} has the same properties as above; α_{Qi} is the systematic bias and β_{Qi} is the scaling bias of the FFQ, for food i.

In order to estimate the systematic and scaling bias parameters, the FFQ variable of interest must be regressed on another "reference" measure following model [1]. The first method presented here, regression calibration, used 7DD as this reference measure; and relied on the strong assumption that their respective errors were independent. The inclusion of a third variable, following measurement error model [2] and having errors independent of the two other types of measurement (FFQ and 7DD), allowed considering true intake as a latent and unobserved variable. The second method, the method of triads, used biomarkers (M_{ij}) as the third measure:

$$M_{ij} = \alpha_{Mi} + \beta_{Mi}T_{ij} + \varepsilon_{Mij} \quad [3]$$

All measurement error models implemented were linear models assuming normality of all variables. As a result, all intake and biomarkers variables were log-transformed.

Both methods allowed for the introduction of non-dietary covariates (Z_{ij}) associated with true intake. Categorical variables were recoded as dummy variables for the linear models. The inclusion of the covariates made the assumptions related to the different models more plausible as they were conditional on the covariates (e.g. reporting error was related to BMI when BMI is included in the model).

Regression calibration and "Rosner & Gore" method

a. General model

With this method, true intake was represented by the diet diary reported value, and equation [2] became:

$$Q_{ij} = \alpha_{Qi} + \beta_{Qi}R_{ij} + \varepsilon_{Qij} \quad [4]$$

Under the assumption that random errors of both methods are not correlated:

$$\begin{cases} \text{cov}(\varepsilon_{Rij}, T_{ij}) = 0 \\ \text{cov}(\varepsilon_{Rij}, \varepsilon_{Qij}) = 0 \end{cases} \quad [5]$$

The goal was to predict the diet diary value for all participants, including those not in the validation sub-sample. This was achieved by implementing a linear model between FFQ and 7DD reported values in the validation sub-sample:

$$R_{ij} = \lambda_{0i} + \lambda_{REGi} Q_{ij} + \varepsilon_{ij} \quad [6]$$

Where λ_{REGi} is often referred to as the regression dilution ratio (RDR).

The regression parameters would then be used in the main study population to calculate predicted diet diary intake—assumed to represent the true intake. The inclusion of different covariates in the model allowed for three models to be designed, resulting in three sets of predicted intakes.

b. Model 1

The base model included only the FFQ and 7DD measures, and was similar to the single imputation method introduced by Rosner et al. (1989). Assuming a linear relationship, the following model was fitted within the validation sub-sample:

$$R_{ij} = \lambda_{01i} + \lambda_{REG1i} Q_{ij} + \varepsilon_{ij} \quad [7]$$

Where Q_{ij} and R_{ij} represent intake of item i in participant j from the FFQ and the 7DD, respectively. Once parameter estimates were obtained using general least square method, the true (diet diary) intakes ($\hat{R}_{ij}$) could be calculated in the whole population:

$$\hat{R}_{ij} = \hat{\lambda}_{01i} + \hat{\lambda}_{REG1i} Q_{ij} \quad [8]$$

c. Model 2

The second model further included non-dietary covariates (Z_{ij} and the associated γ_{2i} regression parameters) associated with both true and reported intake: sex, age, BMI, physical activity, employment grade, ethnicity, and smoking status. The model became:

$$R_{ij} = \lambda_{02i} + \lambda_{REG2i}Q_{ij} + \Sigma\, \gamma_{2i}Z_{ij} + \varepsilon_{ij} \quad [9]$$

The following steps were similar to model 1.

d. Model 3

The third and final model used Rosner and Gore's approach by further including all FFQ-items as potential predictors of the diet diary value (Rosner & Gore, 2001). This was done since it was observed that some FFQ-items were more associated to a diet diary food then the respective FFQ-item (e.g. FFQ hamburger was a better predictor of diet diary chips than FFQ chips). The model was:

$$R_{ij} = \lambda_{03i} + \Sigma\, \lambda_{REG3i}Q_{ij} + \Sigma\, \gamma_{3i}Z_{ij} + \varepsilon_{ij} \quad [10]$$

The large number of predictors due to the inclusion of all non-dietary covariates and all FFQ-items may have led to poor predicting power. As a result, a stepwise selection of FFQ-items variables was done to retain only variables which contributed significantly to the model ($p<0.01$), with all non-dietary covariates forced in the model.

The regression calibration method was quite straightforward and suited well the goal of correcting individual intakes. However, it relied on the fundamental assumptions that errors in the FFQ and the 7DD were independent and that 7DD represented true intake. Such assumptions were likely to be flawed, and a method that could relax such assumptions was investigated.

Structural equation modelling: method of triads

Structural equation modelling treats true intake as an unobserved latent variable. This is assumed to be more accurate because diet diaries can be subject to systematic and scaling bias. Yet, with FFQ and diet records data only, models are unidentifiable and parameters of interest cannot be estimated, making a third variable necessary (Kaaks *et al.*, 1994). The additional variable can be a repeat of the reference

measure, i.e. the diet diary, or any other measure associated with true intake. "Recovery" biomarkers such as urinary nitrogen or doubly labelled water have usually been used in the "method of triads" since they are good markers of usual intake. None of these being available in the Whitehall II data, β-carotene was used as the third variable, as a biomarker of fruit and vegetable intake.

The original method of triads which used three variables (FFQ, 7DD and biomarker) is first presented. Similarly to the regression calibration method, a model including non-dietary covariates was also implemented.

a. Model 1

Kaaks and colleagues introduced a structural equation modelling approach to investigate the error structure of the FFQ (Kaaks *et al.*, 1994). For participant j and considering a single nutrient or food the structural equation model is as follows:

$$\begin{cases} Q_j = \alpha_Q + \beta_Q T_j + \varepsilon_{Qj} \\ R_j = T_j + \varepsilon_{Rj} \\ M_j = \alpha_M + \beta_M T_j + \varepsilon_{Mj} \end{cases} \quad [11]$$

Where $\varepsilon_{xi} \sim N(0; \sigma_{\varepsilon x})$ are assumed to be independent of the true intake (linear model assumption) and of each other; all variables are assumed to be normally distributed ($X_i \sim N(\mu_X, \sigma_{eX})$); diet records are assumed to follow measurement error model [1].

The model parameters μ_T, α_Q, α_M, $\sigma_{\varepsilon Q}$, $\sigma_{\varepsilon M}$, $\sigma_{\varepsilon R}$, β_Q, and β_M could be estimated with the moments approach using the following covariance matrix:

Covariance matrix Means

$$\begin{array}{c} R \\ Q \\ M \end{array} \begin{bmatrix} \sigma_T^2 + \sigma_{eR}^2 & & \\ \beta_Q \sigma_T^2 & \beta_Q \sigma_T^2 + \sigma_{eQ}^2 & \\ \beta_M \sigma_T^2 & \beta_Q \beta_M \sigma_T^2 & \beta_M \sigma_T^2 + \sigma_{eM}^2 \end{bmatrix} \quad \begin{matrix} \mu_T \\ \alpha_Q + \beta_Q \mu_T \\ \alpha_M + \beta_M \mu_T \end{matrix}$$

The estimation of all parameters allowed obtaining an estimate for true intake:

$$T_i = \lambda_0 + \lambda_{QT} Q_i + \varepsilon_i \quad \text{where } \varepsilon_i \sim N(0, \sigma^2) \quad [12]$$

Where λ_{QT}, the regression dilution ratio (RDR), is the equivalent of λ_{REG1}, λ_{REG2}, and λ_{REG3} parameters of the regression calibration models.

The RDR being the slope parameter of model [12], its estimate was:

$$\hat{\lambda}_{QT} = \hat{\text{cov}}(T,Q) \Big/ \hat{\text{var}}(Q) \quad [13]$$

Which is:

$$\hat{\lambda}_{QjT} = \frac{\hat{\beta}_Q \hat{\sigma}_T^2}{\hat{\beta}_Q \hat{\sigma}_T^2 + \hat{\sigma}_{Qj}^2} \quad [14]$$

Computation of this model was done using the CALIS procedure of the SAS software.

In Kaaks' model, the additional measurement—the biomarker—allows for all parameters to be identified. However, as stated above, it is still assumed that the random errors of each measurement method are independent. In order to relax this independence assumption, an alternative constraint must be set (e.g. fixing some other parameters) or another measure must be added. This measure can be a repeated biomarker measurement. At phase 3, repeated measurement of β-carotene was available on 406 participants. However, complete case analysis including FFQ and 7DD data yielded only 86 participants, which was not enough to obtain robust estimates for these extended models, and results were not presented.

b. Model 2

Model [11] was extended to include non-dietary covariates.

Let Z_i denote the covariates associated with true dietary intake. The structural equation model is given by:

$$\begin{cases} T_i = \mu_{T|Z_T} + \Sigma\gamma_T Z_i + \varepsilon_{Ti} \\ Q_i = \alpha_Q + \beta_Q T_i + \Sigma\gamma_Q Z_i + \varepsilon_{Qi} \\ R_i = T_i + \Sigma\gamma_R Z_i + \varepsilon_{Ri} \\ M_i = \alpha_M + \beta_M T_i + \Sigma\gamma_M Z_i + \varepsilon_{Mi} \end{cases} \quad [15]$$

Model [15] is bound to the same assumptions than model [11]. It was fitted in the same way as model [11], but including the residuals from linear regressions of dietary measurements on the covariates in place of the dietary measurements.

Survival analysis

Once parameter estimates were obtained for all models, corrected fruit and vegetable intakes were derived for the whole study sample. The corrected intakes were then included in Cox proportional hazards regressions as main exposures and compared to non-corrected intakes and energy adjusted intakes (including either energy in the model as a covariate, or using energy residuals). Cox models were adjusted for age, sex, and ethnicity. Participants were classified into quartiles of corrected or non-corrected fruit and vegetable intake, the first quartile with lowest intake was the reference group. Outcomes included CHD, diabetes, cancer mortality, and all-cause mortality. For the regression calibration and structural equation models, results using both the UCL and the CNC coded 7DD data were presented.

Appendix 7.2 Results

a. Regression calibration methods

Table A16 summarises parameter estimates from all three models [7], [8], and [10]. Results from the first model indicated that FFQ-reported fruit and vegetables intake explained more variance of CNC-coded diaries than of UCL-coded ones, though the scaling effect was stronger as the slope parameter (λ_{REG}) was closer to 0. This was balanced by a higher intercept with the CNC data. The inclusion of non-dietary covariates in the 2nd model resulted in higher R^2 for both coding centres. All non-dietary covariates with significant parameter estimates predicted lower 7DD reported intake; these were current smoking, low employment grade, and Asian ethnicity for UCL-coded data, and BMI and current smoking for the CNC fitted model (not shown). Model 3 confirmed that fruit and vegetable consumption as reported in the FFQ contributed more towards the model R^2 than all other dietary and non-dietary covariates (results not shown). The inclusion of new dietary predictors increased the R^2 with battered fish, white rice, margarine, ketchup, carrots, beans, and tofu retained by the stepwise selection for the UCL data; and roast potatoes, apples, soft drinks, and tomatoes retained with CNC data. Overall, models fitted with both the UCL and the CNC data delivered the same conclusions: there was a scaling bias

between the two dietary assessment methods, with FFQ over-estimating the reference 7DD intake data.

Table A0.16: Parameters estimates from the regression calibration models

Data*	Model [equation]	n	R^2	n pred.	Intercept (λ_0)			Slope (λ_{REG})		
					Est	95% CI		Est	95% CI	
UCL	1 [7]	1328	0.188	1	1.83	1.42	2.25	0.600	0.533	0.668
	2 [8]	1328	0.233	15	2.32	1.77	2.87	0.577	0.507	0.647
	3 [10]	1328	0.286	23	1.95	1.37	2.53	0.568	0.496	0.641
CNC	1 [7]	560	0.298	1	2.53	2.12	2.94	0.513	0.447	0.578
	2 [8]	560	0.365	15	2.90	2.34	3.47	0.512	0.448	0.577
	3 [10]	560	0.425	19	3.32	2.76	3.88	0.410	0.338	0.481

* Centre of coding using different programs: WFOOD (UCL) and DINER (CNC). n pred. number of predictors included in the model, more FFQ foods were used with UCL data than CNC for model 3.

b. Structural equation modelling: Method of triads

Table A17 summarises the parameter estimates for both models and both 7DD data coding centres. Compared to the previous approach, estimates for the RDRs and the intercepts were in the same ranges. The inclusion of a biomarker did not dramatically change the results. The further inclusion of a repeated beta-carotene measure would have allowed relaxing more assumption and obtain a greater insight into the error structure of the two dietary assessment techniques, but the lack of observations prevented from doing so.

c. Application to survival analysis

Tables A18-A to A18-D present all hazard ratios estimates together with their confidence intervals for the four outcomes. The model using non-corrected intakes (model 1) is displayed first, followed by energy adjusting methods and the measurement error models.

The non-corrected intakes were significantly associated with linear risk reduction of cancer and all-cause mortality, confirming the general public health recommendations made on fruit and vegetable intake. Including total energy intake in the model or using energy residuals slightly attenuated these trends which became borderline significant. The trends were not modified for CHD and diabetes, but

individual quartiles estimates were affected. Overall, the use of energy residuals had more effect on the estimates than the inclusion of energy intake in the model.

The use of corrected intakes using either regression calibration or structural equation modelling (SEM) did not always modify the estimates. Instead, there were exactly similar to the non-corrected ones for all regression calibration models 1 and SEM models. This was explained by the fact that the corrected intakes for these models were proportional, except for the intercept, to the original FFQ reported intakes (Corrected = λ_0 + $\lambda_{REG/QT}$ FFQ). As a result, the ranking of participants was almost not modified and the resulting Cox models gave exactly the same estimates.

Regression models 2 and 3 did include further covariates, and the resulting corrected intakes did not follow proportionally the original ones. Hazard ratio estimates for such models did differ and some new trends appeared: linear risk reduction was observed for all outcomes, including CHD and diabetes (significant only with CNC data). These trends were mainly due to changes in the individual quartiles estimates for CHD and diabetes, suggesting that the reporting error was strongly linked to the non-dietary covariates included in the regression calibration models. The effect was smaller on the cancer and all-cause mortality outcomes and the linear risk reduction was confirmed.

Table A0.17: Parameters estimates for all structural equations models

Model	UCL				CNC			
	1		2		1		2	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
n	1033		1033		462		462	
$\overline{Q}$	6.13	0.016	6.13	0.016	6.23	0.024	6.23	0.024
β_Q	0.841	0.210	0.754	0.258	0.878	0.243	1.10	0.332
$\mu_T = \overline{R}$	5.54	0.024	5.54	0.023	5.74	0.023	5.74	0.021
β_M	0.282	0.058	0.206	0.060	0.299	0.096	0.300	0.099
σ_T^2	0.193	0.052	0.189	0.067	0.162	0.046	0.120	0.038
σ_{eQ}^2	0.139	0.034	0.144	0.037	0.137	0.035	0.108	0.044
σ_{eR}^2	0.421	0.051	0.379	0.067	0.069	0.044	0.085	0.037
σ_{eM}^2	0.295	0.014	0.283	0.013	0.313	0.021	0.300	0.020
α_Q	1.47	1.16	-0.004	0.019	1.19	1.40	0.028	0.022
α_M	-1.77	0.320	0.012	0.017	-1.84	0.550	0.037	0.026
Slope (λ_{QT})	**0.590**		**0.568**		**0.542**		**0.523**	
Intercept (λ_0)	**1.92**		**2.06**		**2.37**		**2.49**	

See covariance matrix in the methods section for significance of parameters symbols.
Est. Parameter estimate; SE. Standard error of the mean.

Table A0.18: Cox regression estimates for fruit and vegetables measurement error models, by quartile of intake

Table A: CHD

	Quartile 1	Quartile 2			Quartile 3			Quartile 4			Trend		
Model (7DD data)		HR	95	% CI	HR	95	% CI	HR	95	% CI	HR	95	% CI
Model 1	Ref	0.85	0.59	1.22	0.75	0.51	1.09	0.83	0.57	1.20	0.93	0.83	1.05
Model 1 + energy	Ref	0.84	0.58	1.21	0.74	0.51	1.09	0.82	0.56	1.20	0.93	0.82	1.05
Energy residuals	Ref	**0.61**	**0.42**	**0.89**	**0.66**	**0.46**	**0.96**	0.74	0.51	1.06	0.91	0.81	1.03
RegCal 1 (UCL)	Ref	0.85	0.59	1.22	0.75	0.51	1.09	0.83	0.57	1.20	0.93	0.83	1.05
RegCal 1 (CNC)	Ref	0.85	0.59	1.22	0.75	0.51	1.09	0.83	0.57	1.20	0.93	0.83	1.05
RegCal 2 (UCL)	Ref	**0.53**	**0.36**	**0.77**	**0.68**	**0.48**	**0.97**	**0.60**	**0.41**	**0.86**	**0.86**	**0.76**	**0.98**
RegCal 2 (CNC)	Ref	**0.53**	**0.36**	**0.77**	**0.68**	**0.48**	**0.97**	**0.60**	**0.41**	**0.86**	**0.86**	**0.76**	**0.98**
RegCal 3 (UCL)	Ref	**0.60**	**0.42**	**0.88**	0.79	0.56	1.12	**0.57**	**0.39**	**0.84**	**0.86**	**0.77**	**0.97**
RegCal 3 (CNC)	Ref	**0.60**	**0.42**	**0.88**	**0.54**	**0.38**	**0.77**	**0.57**	**0.39**	**0.84**	**0.86**	**0.77**	**0.97**
SEM 1 (UCL)	Ref	0.85	0.59	1.22	0.75	0.51	1.09	0.83	0.57	1.20	0.93	0.83	1.05
SEM 1 (CNC)	Ref	0.85	0.59	1.22	0.75	0.51	1.09	0.83	0.57	1.20	0.93	0.83	1.05
SEM 2 (UCL)	Ref	0.85	0.59	1.22	0.75	0.51	1.09	0.83	0.57	1.20	0.93	0.83	1.05
SEM 2 (CNC)	Ref	0.85	0.59	1.22	0.75	0.51	1.09	0.83	0.57	1.20	0.93	0.83	1.05

Model 1 included ln(fruit and vegetables) adjusted for age, sex and ethnicity. The energy residuals model included fruit and vegetables energy residuals as exposure. RegCal 1/2/3 Exposures derived from regression calibration models 1, 2, and 3. SEM 1/2 Exposures derived from structural equation models 1 and 2 (see methods section for more details). HR, hazard ratio; CI, confidence interval.

Table B: Diabetes

	Quartile 1	Quartile 2			Quartile 3			Quartile 4			Trend		
Model (7DD data)		HR	95	% CI	HR	95	% CI	HR	95	% CI	HR	95	% CI
Model 1	Ref	1.00	0.81	1.23	0.92	0.75	1.14	1.10	0.90	1.35	1.03	0.96	1.09
Model 1 + energy	Ref	0.99	0.80	1.22	0.91	0.73	1.13	1.07	0.87	1.33	1.02	0.95	1.09
Energy residuals	Ref	0.94	0.76	1.16	0.93	0.76	1.15	1.02	0.83	1.26	1.01	0.94	1.08
RegCal 1 (UCL)	Ref	1.00	0.81	1.23	0.92	0.75	1.14	1.10	0.90	1.35	1.03	0.96	1.09
RegCal 1 (CNC)	Ref	1.00	0.81	1.23	0.92	0.75	1.14	1.10	0.90	1.35	1.03	0.96	1.09
RegCal 2 (UCL)	Ref	0.98	0.81	1.19	**0.75**	**0.61**	**0.93**	0.89	0.73	1.09	0.94	0.88	1.01
RegCal 2 (CNC)	Ref	0.93	0.77	1.13	**0.75**	**0.61**	**0.93**	**0.76**	**0.62**	**0.93**	**0.90**	**0.84**	**0.96**
RegCal 3 (UCL)	Ref	0.86	0.70	1.05	0.86	0.71	1.05	0.82	0.67	1.00	0.94	0.88	1.00
RegCal 3 (CNC)	Ref	0.87	0.72	1.06	**0.76**	**0.62**	**0.92**	**0.73**	**0.60**	**0.90**	**0.90**	**0.84**	**0.96**
SEM 1 (UCL)	Ref	1.00	0.81	1.23	0.92	0.75	1.14	1.10	0.90	1.35	1.03	0.96	1.09
SEM 1 (CNC)	Ref	1.00	0.81	1.23	0.92	0.75	1.14	1.10	0.90	1.35	1.03	0.96	1.09
SEM 2 (UCL)	Ref	1.00	0.81	1.23	0.92	0.75	1.14	1.10	0.90	1.35	1.03	0.96	1.09
SEM 2 (CNC)	Ref	1.00	0.81	1.23	0.92	0.75	1.14	1.10	0.90	1.35	1.03	0.96	1.09

Model 1 included ln(fruit and vegetables) adjusted for age, sex and ethnicity. The energy residuals model included fruit and vegetables energy residuals as exposure. RegCal 1/2/3 Exposures derived from regression calibration models 1, 2, and 3. SEM 1/2 Exposures derived from structural equation models 1 and 2 (see methods section for more details). HR, hazard ratio; CI, confidence interval.

Table C: Cancer mortality

	Quartile 1	Quartile 2			Quartile 3			Quartile 4			Trend		
Model (7DD data)		HR	95	% CI	HR	95	% CI	HR	95	% CI	HR	95	% CI
Model 1	Ref	**0.65**	**0.46**	**0.93**	0.76	0.55	1.07	**0.62**	**0.43**	**0.88**	**0.87**	**0.78**	**0.98**
Model 1 + energy	Ref	**0.67**	**0.47**	**0.96**	0.80	0.57	1.12	**0.66**	**0.46**	**0.96**	0.90	0.79	1.01
Energy residuals	Ref	0.83	0.59	1.18	0.84	0.60	1.19	0.69	0.48	1.00	0.90	0.80	1.01
RegCal 1 (UCL)	Ref	**0.65**	**0.46**	**0.93**	0.76	0.55	1.07	**0.62**	**0.43**	**0.88**	**0.87**	**0.78**	**0.98**
RegCal 1 (CNC)	Ref	**0.65**	**0.46**	**0.93**	0.76	0.55	1.07	**0.62**	**0.43**	**0.88**	**0.87**	**0.78**	**0.98**
RegCal 2 (UCL)	Ref	**0.66**	**0.48**	**0.92**	**0.46**	**0.32**	**0.67**	**0.64**	**0.46**	**0.89**	**0.84**	**0.75**	**0.94**
RegCal 2 (CNC)	Ref	**0.66**	**0.48**	**0.92**	**0.46**	**0.32**	**0.67**	**0.64**	**0.46**	**0.89**	**0.84**	**0.75**	**0.94**
RegCal 3 (UCL)	Ref	0.88	0.63	1.22	0.71	0.50	1.01	0.71	0.50	1.00	**0.88**	**0.79**	**0.99**
RegCal 3 (CNC)	Ref	0.85	0.61	1.17	**0.61**	**0.43**	**0.87**	**0.65**	**0.46**	**0.92**	**0.88**	**0.79**	**0.99**
SEM 1 (UCL)	Ref	**0.65**	**0.46**	**0.93**	0.76	0.55	1.07	**0.62**	**0.43**	**0.88**	**0.87**	**0.78**	**0.98**
SEM 1 (CNC)	Ref	**0.65**	**0.46**	**0.93**	0.76	0.55	1.07	**0.62**	**0.43**	**0.88**	**0.87**	**0.78**	**0.98**
SEM 2 (UCL)	Ref	**0.65**	**0.46**	**0.93**	0.76	0.55	1.07	**0.62**	**0.43**	**0.88**	**0.87**	**0.78**	**0.98**
SEM 2 (CNC)	Ref	**0.65**	**0.46**	**0.93**	0.76	0.55	1.07	**0.62**	**0.43**	**0.88**	**0.87**	**0.78**	**0.98**

Model 1 included ln(fruit and vegetables) adjusted for age, sex and ethnicity. The energy residuals model included fruit and vegetables energy residuals as exposure. RegCal 1/2/3 Exposures derived from regression calibration models 1, 2, and 3. SEM 1/2 Exposures derived from structural equation models 1 and 2 (see methods section for more details). HR, hazard ratio; CI, confidence interval.

Table D: All-cause mortality

	Quartile 1	Quartile 2			Quartile 3			Quartile 4			Trend		
Model (7DD data)		HR	95	% CI	HR	95	% CI	HR	95	% CI	HR	95	% CI
Model 1	Ref	**0.67**	**0.53**	**0.86**	0.80	0.63	1.01	**0.70**	**0.55**	**0.89**	**0.91**	**0.84**	**0.99**
Model 1 + energy	Ref	**0.68**	**0.53**	**0.87**	0.81	0.64	1.03	**0.72**	**0.56**	**0.92**	0.92	0.85	1.00
Energy residuals	Ref	**0.69**	**0.54**	**0.88**	0.79	0.63	1.00	**0.76**	**0.59**	**0.96**	0.93	0.86	1.01
RegCal 1 (UCL)	Ref	**0.67**	**0.53**	**0.86**	0.80	0.63	1.01	**0.70**	**0.55**	**0.89**	**0.91**	**0.84**	**0.99**
RegCal 1 (CNC)	Ref	**0.67**	**0.53**	**0.86**	0.80	0.63	1.01	**0.70**	**0.55**	**0.89**	**0.91**	**0.84**	**0.99**
RegCal 2 (UCL)	Ref	**0.67**	**0.53**	**0.84**	**0.54**	**0.42**	**0.69**	**0.66**	**0.52**	**0.83**	**0.85**	**0.79**	**0.92**
RegCal 2 (CNC)	Ref	**0.67**	**0.53**	**0.84**	**0.54**	**0.42**	**0.69**	**0.66**	**0.52**	**0.83**	**0.85**	**0.79**	**0.92**
RegCal 3 (UCL)	Ref	0.92	0.73	1.15	**0.73**	**0.57**	**0.93**	**0.68**	**0.53**	**0.87**	**0.87**	**0.80**	**0.94**
RegCal 3 (CNC)	Ref	**0.77**	**0.61**	**0.96**	**0.73**	**0.57**	**0.93**	**0.68**	**0.53**	**0.87**	**0.87**	**0.80**	**0.94**
SEM 1 (UCL)	Ref	**0.67**	**0.53**	**0.86**	0.80	0.63	1.01	**0.70**	**0.55**	**0.89**	**0.91**	**0.84**	**0.99**
SEM 1 (CNC)	Ref	**0.67**	**0.53**	**0.86**	0.80	0.63	1.01	**0.70**	**0.55**	**0.89**	**0.91**	**0.84**	**0.99**
SEM 2 (UCL)	Ref	**0.67**	**0.53**	**0.86**	0.80	0.63	1.01	**0.70**	**0.55**	**0.89**	**0.91**	**0.84**	**0.99**
SEM 2 (CNC)	Ref	**0.67**	**0.53**	**0.86**	0.80	0.63	1.01	**0.70**	**0.55**	**0.89**	**0.91**	**0.84**	**0.99**

Model 1 included ln(fruit and vegetables) adjusted for age, sex and ethnicity. The energy residuals model included fruit and vegetables energy residuals as exposure. RegCal 1/2/3 Exposures derived from regression calibration models 1, 2, and 3. SEM 1/2 Exposures derived from structural equation models 1 and 2 (see methods section for more details). HR, hazard ratio; CI, confidence interval.

Appendix 7.3: Discussion

This fruit and vegetable case study was implemented to assess the feasibility and relevance of two measurement error methods: regression calibration and structural equation modelling (SEM). The implementation of both methods was done using published models and adapting them to the Whitehall II data. Due to the number of participants who attended a repeated clinic visit, the SEM models could only be run with a single β-carotene measurement. This did not allow relaxing the assumption of independence of FFQ and 7DD errors, but it did allow gaining more insight on the relationship between the two measures. The two sets of coded 7DD were used separately to verify whether a centre effect could be detected.

Overall, measurement error models results were consistent with both approaches and both centres of 7DD coding, suggesting that fruit and vegetable intake, as reported in the FFQ, was overestimating true intake assumed to be closer to the 7DD estimates. It was not possible to say which model performed best as no true reference was available. Slope estimates (or "regression dilution ratio") were in line with previously published results and suggested that the scaling bias was lower than in Keogh's study (Kaaks *et al.*, 1994; Ocke & Kaaks, 1997; Rosner *et al.*, 2008; Keogh *et al.,* 2010). For the regression calibration models, observed R^2 between FFQ and 7DD were in the range of R^2 obtained for individual fruit and vegetables (Rosner & Gore, 2001).

Corrected (i.e. predicted true) intakes were then included in Cox regression models and results were compared with non-corrected intakes and two usual ways of adjusting for total energy intake. The first regression calibration model and the SEM method yielded exactly the same estimates as the non-corrected intakes, highlighting that the proportional correction of fruit and vegetable intake did not modify the rankings of participants. This illustrated well the need to estimate correctly the standard errors of the new parameters when implementing such regression calibration or SEM methods to one variable only. It also showed that these methods would be more efficient if a dose-effect relationship was investigated.

When the corrected intakes depended on other covariates (non-dietary and dietary), as in regression calibration models 2 and 3, hazard ratio estimates were modified for all outcomes. Linear risk reduction of CHD and diabetes was suggested, and similar trends for cancer and all-cause mortality were confirmed.

The survival analysis results therefore indicated that fruit and vegetables may be protective against all the outcomes considered in this analysis, and mainly cancer and all-cause mortality. For CHD and diabetes, there was an indication that the true and protective association may be biased by reporting error on the FFQ items.

The implementation of the two approaches for correction of individual FFQ items, to derive corrected aggregate scores, must take into account the limitations that appeared in this case study. Of the two methods, the structural equation modelling was initially thought to be the better option as it included an external measure of food intake and considered true intake as a latent and unobserved variable. With repeated biomarker measurement, it would allow more reliable estimates by relaxing the assumption that errors in the FFQ and 7DD are independent. However, few participants went twice to the clinic at phase 3, and only few repeated measurements of β-carotene were obtained. As a result, the SEM model could only be applied with a single β-carotene measure for each participant, and the error independence assumption could not be relaxed. The use of a single biomarker measure led to parameter estimates close to the regression calibration models. Further, the use of the β-carotene biomarker was feasible for fruit and vegetables as a food group. Conducting SEM models individually for each fruit or vegetable may not be possible as the relationship with β-carotene may not be shown. Also, no other biomarker was available for the whole Whitehall II population. As a result, the implementation of SEM models on other food groups, let alone individual items, was not feasible.

The regression calibration approach, despite its limitations was therefore the sole technique which could be applied uniformly across all FFQ items. The algorithms implemented in this case study are easily transferable to any FFQ item, at the condition that an equivalent "7DD item" exists. The ranking issue observed with the first model and the SEM approach was not a major concern as corrected aggregate scores would rely on the corrected intakes of all FFQ items. The precise estimation

of the regression calibration parameters variance was therefore not investigated in further details.

Regression calibration relied on two fundamental assumptions: (i) that random errors in the 7DD and FFQ tools were independent; and (ii) that 7DD reported intakes represented true intakes. As a result, the implementation of such method using 7DD as reference measure did not allow concluding that the corrected estimates reflected better the true relationship. They have to be used as an indication that misreporting may bias the observed results, but must be taken with extreme caution.