Development and Validation of a Risk Score predicting substantial Weight Gain over 5 Years in middle-aged European Men and Women

vorgelegt von

Diplom-Oecotrophologin (FH),

Master of Science in Epidemiology

Annika Steffen

Von der Fakultät VII – Wirtschaft und Management der Technischen Universität Berlin

zur Erlangung des akademischen Grades

Doktorin der Gesundheitswissenschaften/ Public Health

– Dr. P.H. –

genehmigte Dissertation

Promotionsausschuss:

Vorsitzende: Prof. Dr. Jacqueline Müller-Nordhorn

Gutachter: Prof. Dr. Reinhard Busse

Gutachter: Prof. Dr. Heiner Boeing

Tag der wissenschaftlichen Aussprache: 26.01.2012

Berlin 2012

D83

“Prediction is very difficult, especially about the future.”

Niels Bohr

Table of Content

Index of Tables

Index of Figures

Glossary of Abbreviations

ACSM	American College of Sports Medicine
aROC	Area under a receiver operating characteristic curve
BMI	Body Mass Index
BMR	Basal Metabolic Rate
CDC	Centers for Disease Control and Prevention
CHD	Coronary heart disease
CHO	Carbohydrate
CI	Confidence Interval
Cox PH regression	Cox Proportional Hazards regression
c-statistic	Concordance statistic
CVD	Cardiovascular disease
DiOGenes	Diet Obesity and Genes project
DK-CopAa	Copenhagen/Aarhus, Denmark
EPIC	European Prospective Investigation into Cancer and Nutrition
FFQ	Food frequency questionnaire
FTO	Fat mass and obesity-associated gene
GER-Pot	Cohort of Potsdam, Germany
HLT	Hosmer-Lemeshow Test
HR	Hazard ratio
IT-Flo	Cohort of Florence, Italy
NL-AmMa	Cohort of Amsterdam/Maastricht, The Netherlands
NL-Doe	Cohort of Doetinchem, The Netherlands
PY	Person-years
ROC	Receiver operating characteristic curve
RR	Relative Risk
RS	Risk Score
SAS	Statistical Analysis System
SWG	Substantial weight gain
TEF	Thermic effect of food
UK-Nor	Cohort of Norfolk, United Kingdom
WCRF	World Cancer Research Fund
WHO	World Health Organization

1 Introduction

1.1 Background and aim of the thesis

The prevalence of overweight and obesity has increased dramatically during the last decades and excess body weight has become a major public health problem worldwide [1]. According to the World Health Organization (WHO), over 1.5 billion adults were overweight in 2008 globally, including more than 500 million obese [2]. In Europe, the prevalence of overweight among adults ranges between 30-80% and up to 36% of all adults are classified as obese [3, 4]. A substantial body of evidence has emerged to show that excess body fat is associated with higher risks for chronic diseases, including type 2 diabetes, cardiovascular diseases, and certain types of cancer [5]. According to the World Health Report 2002, approximately 58% of diabetes, 21% of ischemic heart disease and 8-42% of certain cancers are attributable to a Body-Mass-Index (BMI) above 21 kg/m² [6]. Consequently, obesity places a substantial economic burden on societies. In Europe, for example, the relative economic burden of obesity-related healthcare ranges from 0.09% to 0.61% of the national gross domestic product [7].

Given the increasing prevalence of overweight and obesity in most Western populations and its significant cost to society, strategies to combat the rapid rise in obesity prevalence are urgently needed. Although intervention studies report effective strategies for weight loss in the short-term, the long-term effect of most obesity treatments is limited [8]. Therefore, in the first place, it is preferable to avoid (substantial) weight gain that may lead to overweight and obesity. Even though excess weight is in principle a matter of energy balance, susceptibility to weight gain appears to be determined by a complex interaction between genetic, environmental, socio-economic, cultural and behavioural factors [1]. Up to now, much effort has been devoted to the identification of single risk factors for weight gain or risk of obesity; however, understanding the magnitude of effects of these risk factors, particularly in combination, is fundamental in order to identify priorities for public health efforts. Also, in view of limited resources, obesity prevention efforts may be targeted specifically to those individuals who are at high risk for gaining substantial amounts of weight and thus might benefit most from prevention programmes. One way to summarise the impact of multiple risk factors and to identify high-risk groups is the development of risk prediction models. Such statistical models are being increasingly used in the medical field to estimate an individual's risk of disease on the basis of established biological or behavioural risk factors and to stratify apparently healthy individuals into risk categories [9]. During the last decades, numerous risk assessment tools have been developed in the context of cardiovascular diseases (CVD) [10], type 2 diabetes [11] and cancer [12-15]. These prediction models have been mainly developed in rather homogenous study populations, in which they generally performed well,

and a number of them have been validated in external populations, mainly with less success [10, 11]. Hence, these findings challenge the idea of the existence of one universal, disease-specific prediction model applicable to all populations.

Apart from a recently published study among newborns presenting a risk score to predict overweight at the age of 8 years [16], there is currently no risk score available to predict absolute risk of weight gain or overweight/obesity among adults. Therefore, the aim of the present thesis is twofold. First, it is attempted to develop a simple, practical and informative risk score model predicting risk of substantial weight gain (SWG) within the following 5 years based on the current risk factor profile using data of the large-scale, multi-centre European Prospective Investigation into Cancer and Nutrition (EPIC) study. Second, it will be investigated whether this overall risk score is equally well applicable to different European populations in order to evaluate the idea of one universal, transnational risk prediction model for weight gain.

1.2 Overweight and obesity

1.2.1 Definition

Obesity and to a lesser extent overweight is defined as a condition of excessive fat accumulation in adipose tissue to the extent that health may be adversely affected [1]. The most commonly used marker of body fatness is the BMI. It is a simple index of weight relative to height, calculated as body weight in kilograms divided by height in metres squared (kg/m²). It is used to classify underweight (BMI<18.5 kg/m²), overweight (25≤BMI<30 kg/m²) and obesity (BMI≥30 kg/m²) in adults [1] and is considered to be the most useful population-level measure of obesity. It should be noted, though, that it does not distinguish between weight associated with muscle and weight associated with fat and hence does not account for the variation in the nature of obesity between different individuals and populations [1].

1.2.2 Human evolution

Only fairly recently in human history, overweight and obesity became a public health issue [17]. Before industrialisation, underweight and weight loss due to periodic episodes in deficiency of energy from food, e.g. resulting from seasonal variation, constituted the main nutritional problems and overweight has been considered a sign of wealth, wellbeing and social advantage. In order to cope with circular energy deficiencies, the human body has developed excellent physiological strategies in that it favoured the deposition of body fat when (good-quality) food was readily available [18]. Nowadays, seasonal variation in food supply resulting in cycles of positive and negative energy balance, are still a feature of life for the remaining gatherer-hunter and many agricultural communities in adverse social or

environmental circumstances [17]. In Europe and the United States, however, food insecurity disappeared with industrialisation and urbanisation from the 19th century onwards and food became abundant, inexpensive and in general more energy-dense (increasingly processed and rich in fat and sugar, low in complex carbohydrates and fibre). These shifts in the structure of diet have been widely referred to as the "nutrition transition" [19]. At the same time, the overall lifestyle became more sedentary due to mechanisation and computerisation of workplaces, increase in motor vehicle dependence, widely spread lifts and escalators in public places, and increase in sedentary pursuits such as watching television, surfing the internet or playing video games. Consequently, overweight and obesity became fairly common in industrialised countries, particularly from the middle of the 20th century. Within the last two decades, however, there has been a dramatic increase in the prevalence of overweight and obesity in developed and also developing countries [1] and excess body fat is increasingly recognised as a serious, global public health problem.

1.2.3 Public health relevance

According to the WHO, more than 10% of the world's adult population was obese in 2008 [2]. Based on latest data from the US National Health and Nutrition Examination Survey in 2003-2004, a relative increase of 18% and 41% in prevalence of overweight and obesity, respectively, was observed in comparison to the preceding survey conducted between 1988 and 1994 [20]. Likewise, the prevalence of obesity has risen threefold or more since the 1980s in European countries [4].

In recent years, a large body of evidence has accumulated to show that excess body fat is associated with higher risks for diseases, disorders or disabilities. **Table 1** summarises the approximate relative risk (RR) of health problems associated with obesity. A recent meta-analysis of 89 studies reported on the association of overweight and obesity with the incidence of various co-morbidities and found obesity most strongly associated with risk of type 2 diabetes compared to other co-morbidities [5]. In that study, the pooled relative risks (95% Confidence Interval, CI) for obese subjects compared to normal-weight subjects were 6.74 (5.55-8.19) among men and 12.41 (9.03-17.06) among women. Excess body fat also predisposes to dyslipidaemia and hypertension, which are established risk factors for cardiovascular diseases [21]. Further, a number of epidemiological studies have shown that obesity is associated with higher risks for certain types of cancer. The expert panel of the World Cancer Research Fund (WCRF) has recently judged that there was convincing evidence for overweight and obesity to increase risk of cancers of the colorectum, breast (postmenopausal), oesophagus, pancreas, kidney, and the endometrium [17]. Finally, there is an exhaustive body of literature suggesting excess body weight to be related to the

development of other health problems, including chronic kidney diseases [22], musculoskletal disorders [23], respiratory diseases [24], and psychological problems [1].

Table 1. Approximate relative risk of physical health problems associated with obesity

Relative Risk > 3	Relative Risk 2 – 3	Relative Risk 1 – 2
Type 2 diabetes	Coronary heart disease	Cancer
Gallbladder disease	Hypertension	Reproductive hormone abnormalities
Dyslipidaemia	Osteoarthritis	Polycystic ovary syndrome
Insulin resistance	Hyperuricaemia and gout	Impaired fertility
Breathlessness		Low back pain
Sleep apnoea		Increased risk of anaesthesia complications
		Foetal defects (associated with maternal obesity)

Adapted from WHO [1].

Beyond its impact on health and well-being of individuals, obesity imposes a substantial economic burden on society. This burden is not limited to the direct costs associated with medical treatment of obesity-related co-morbidities, but also includes indirect costs measured as loss of production due to absenteeism, early retirement or premature death [1]. In addition, obesity is accompanied by intangible costs arising for the individual with respect to quality of life. In the United States, the obesity-related costs were estimated to account for 1.2% of the gross domestic product [25]. For Europe, a recent systematic review reported the estimated relative economic burden of obesity to range between 0.09% and 0.61% of the national gross domestic product, reflecting a maximum in absolute costs of 10.4 billion Euros in Germany [7].

1.2.4 Energy balance and the regulation of body weight

A simplistic overview of major influences on energy balance and weight gain is illustrated in **Figure 1**. In simple terms, energy balance, thereby stability of body weight, is achieved when energy intake matches energy expenditure over longer periods of time. Hence, weight gain results from a prolonged positive energy balance, while weight loss occurs when energy intake is lower than energy expenditure [1]. Energy intake refers to all energy consumed as food and drink that can be metabolised inside the body, with the constituent macronutrients fat, carbohydrate and protein providing approximately 9 kcal/g, 4 kcal/g and 4 kcal/g, respectively [1]. Further, alcohol contributes about 7 kcal/g. On the other side of the energy balance equation, total energy expenditure is accounted for by basal metabolic rate (BMR), (diet-induced) thermogenesis and the non-resting energy expenditure (physical activity) [1].

The BMR reflects the energy needed to sustain the metabolic activities of cells and tissues, plus the energy to maintain blood circulation and respiration in the awake state and is measured while the subject is supine and motionless in a thermoneutral environment [26]. It accounts for approximately 60% of total energy expenditure in sedentary individuals and is affected by age, gender, body composition, health and nutrition status [1]. Thermic energy comprises the increase in energy expenditure related to the digestion and absorption of food and processing or storing of nutrients, or related to cold and heat exposure [26]. It accounts for about 10% of total energy expenditure. Finally, the remaining 30% are accounted for by physical activity. Physical activity is the most variable component of total energy expenditure and offers the largest potential to increase energy output.

Under normal circumstances, energy balance oscillates from day to day and week to week without any lasting changes in body fat stores to occur, a regulatory process referred to as energy homeostasis [1]. Therefore, body fat mass reflects the long-term balance between energy expenditure and energy intake, whereby the latter appears to have the predominant role in maintaining this balance [27]. Humans have many physiological control mechanisms that are primarily responsible for body weight regulation. Although this complex regulatory process is still incompletely understood, it is currently well-accepted that food intake, thereby energy balance, is determined by a complex interplay of hormonal and neural mechanisms coordinated by the central nervous system [27, 28]. Briefly, in the arcuate nucleus of the hypothalamus two opposing sets of neuronal circuitry, an appetite-stimulating and an appetite-inhibitory, are housed [28]. Those two circuits receive signals from peripheral hormones expressed in the stomach, intestine, pancreas and adipose tissue. These peripheral hormones are thought to monitor the inflow of nutrients, their distribution and metabolism and/or storage. For example, several short-term acting hormones are expressed in the intestine immediately in response to the presence of food to determine meal size and terminate meals. Leptin and insulin, in contrast, are released in proportion to body fat stores, thereby monitoring the long-term nutritional status of the body and regulating individual meal size in the service of overall energy homeostasis.

In view of this complex and exquisite physiological regulation to maintain energy homeostasis, long-lasting changes in body weight appear to be difficult to achieve. Nevertheless, the increasing prevalence of overweight and obesity suggests that powerful environmental and psycho-social forces may influence energy intake and energy expenditure and may overwhelm the tightly regulated physiological control systems [1].

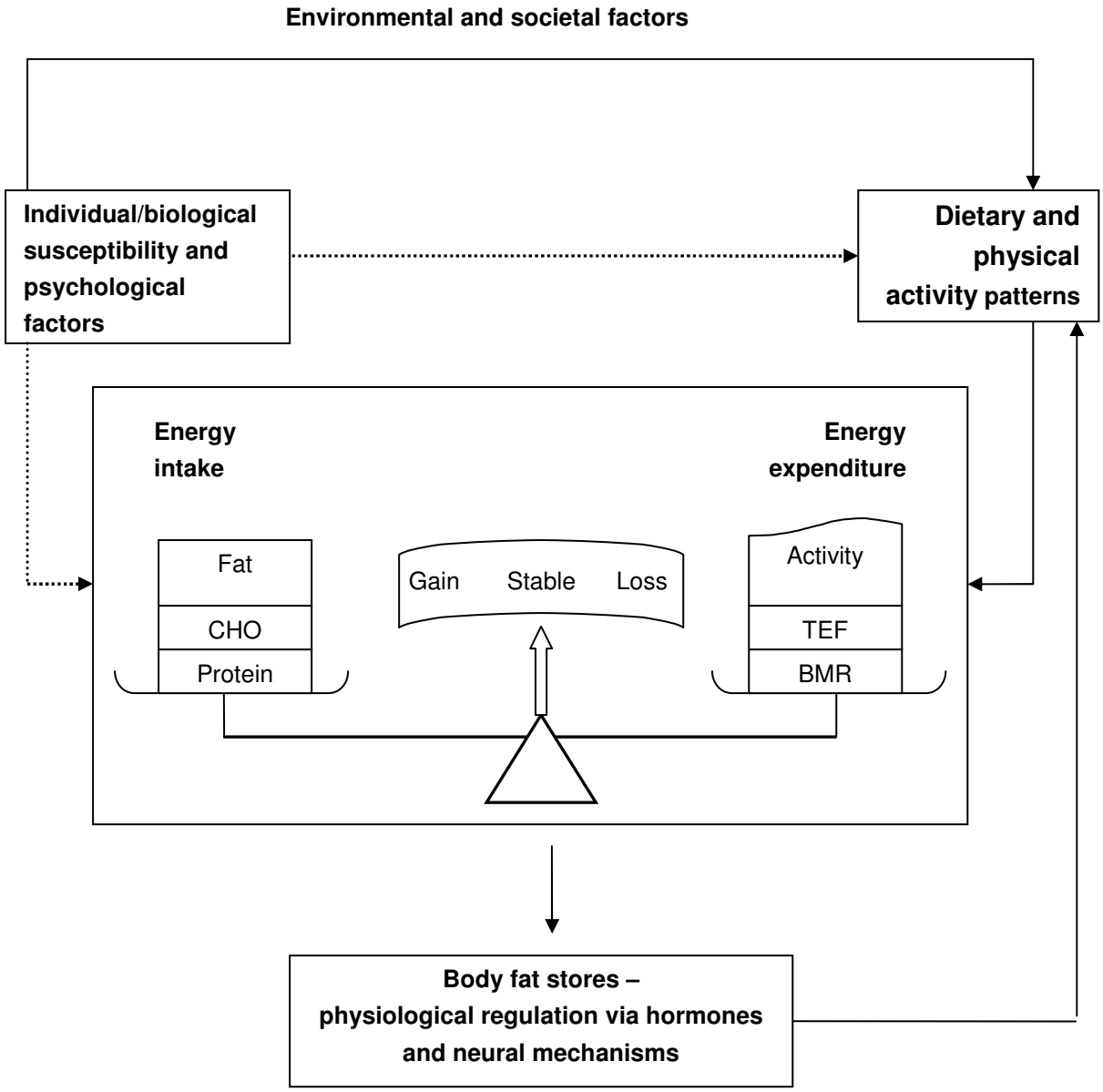


Figure 1. Schematic overview of major influences on energy balance and body weight regulation

BMR = Basal metabolic rate, CHO = carbohydrate, TEF = thermic effect of food. Modified from WHO 2000 [1].

1.2.5 Major modifiable risk factors for weight gain, overweight and obesity

Although genetics clearly contribute to individual differences in anthropometric measures, the genetic milieu is unlikely to have changed during recent decades and the inexorable increase in prevalence of overweight and obesity has been mainly attributed to changes in diet and physical activity [29]. It is currently well-accepted that body weight is ultimately determined by a complex interaction of genetic, environmental and psycho-social factors acting through

several physiological mediators of energy intake and energy expenditure. The environment, for instance, includes several extremely potent factors to overcome the physiological mechanisms operating to maintain body weight, e.g. readily available and cheap high energy-dense food, mechanised transport, non-manual labour ("obesogenic environment"). Further, humans do not eat only to satisfy their appetite but also for many other reasons such as sensory hedonics, sensory stimulation, tension reduction, social pressure, emotional problems, and boredom [30]. Hence, psychological, social, environmental and cultural factors are important underlying influences on dietary patterns and physical activity, finally determining weight gain. **Table 2** shows the most important factors related to diet and physical activity that promote or protect against weight gain and obesity as suggested by the WCRF in 2007 [17]. In the following section, the main modifiable risk factors will be discussed briefly.

Table 2. Summary of strength of evidence on factors that might promote or protect against weight gain and obesity

Evidence	Decreased risk	Increased risk
Convincing	Physical activity	Sedentary lifestyle
Probable	Low energy-dense foods	High intake of energy-dense foods
	Breastfeeding	High intake of sugar-sweetened soft dinks
		Fast foods (energy-dense, large portion sizes)
		Television viewing
Possible	-	-
Insufficient	Refined cereals; starchy roots, tubers, and plantains; fruits; meat; fish; milk and dairy products; fruit juices; alcohol drinks	

Adapted from WCRF [17].

1.2.5.1 Dietary factors

Fruit and vegetables

Due to their high content of water and their low content of energy, fruit and vegetables have a low energy density which has been shown to increase satiety and decrease energy intake [31]. Fruit and vegetables are also rich in dietary fibre which may slow gastric emptying and result in elevated stomach distention, thereby increasing satiety and reducing hunger. Although not completely consistent, results from intervention and observational studies generally support a beneficial, albeit small, effect of a diet high in fruit and vegetables in the prevention of weight gain [32-38].

Meat intake

Particularly because of its high energy density and fat content, meat has been suggested to be a risk factor for weight gain. On the other hand, meat is a relevant source of protein and diets high in protein intake have been suggested to be beneficial in the prevention of weight gain because of increased thermogenesis and satiety [39]. Unlike expected, the majority of the rather few previous cohort studies did not observe an association between protein intake and subsequent weight gain [40]; two studies, however, found positive associations, one of them particularly for animal protein from red and processed meat as well as poultry [41]. Although the evidence is not abundant in respect of meat intake, most prospective observational studies reported a positive association between intake of meat and subsequent weight gain [34, 35, 42-46]. Short-term weight loss intervention studies specifically on meat intake, however, yielded mix results [47-52], with most studies showing no difference in weight loss between meat-based or plant-based diets [49-52].

Whole-grain cereals and cereal products

Whole-grains have been suggested to facilitate weight management due to different physiological mechanisms [53]. Among other valuable constituents such as vitamins, minerals, phytoestrogens and antioxidants, whole-grain foods provide an abundant source of dietary fibre which has been shown to promote satiation and satiety by increasing chewing, slowing gastric emptying, elevating stomach distention, and stimulating gut hormones such as CCK [53, 54]. Also, the increased viscosity of soluble fibre may reduce the overall rate and extent of digestion which may reduce enzymatic digestion of other macronutrients such as fat and protein. Finally, the slower digestion and absorption rate of carbohydrates in high fibre foods would lead to a reduced postprandial blood glucose response, which increases satiety and could improve insulin sensitivity over the long-term and increases fat oxidation. Although the epidemiological data that directly investigates whole-grain intake on weight gain is sparse, results of the few studies generally show a decreased risk of weight gain and/or obesity for higher intake of wholegrain products [17, 55]. With respect to dietary fibre, findings from most well-controlled intervention trials and large cohort studies indicate a beneficial role in body weight control [17, 55, 56].

Milk and dairy products

The consumption of dairy foods has been hypothesised to play a beneficial role in the regulation of body weight, mainly due to the anti-obesity effects suggested for dietary calcium [57]. Despite intensive research efforts and a convincing biological mechanisms, the possible link between dairy products/calcium and weight change remains controversial though. While no relation was found in the Health Professionals Follow-up Study [58], results from other prospective studies indicate different associations across subgroups of the population and

according to type of dairy product [59, 60]. Intervention studies also yielded mixed results. Of nine randomised dairy product supplementation trials reviewed by Barr [61], seven did not find a significant difference between treatment and control groups.

Sweets (cake, cookies, chocolate)

Because sweets tend to be energy dense (rich in sugar and fat), higher intake of these foods may cause many people to exceed daily energy requirements. Evidence supporting this hypothesis, however, is sparse and the few available studies yielded contradictory findings. While Parker et al. did not observe an association between intake of sweets (not defined) and changes in BMI over twelve years [33], Sammel reported an inverse association between intake of desserts and candy with body weight gain over four years among American women [62]. Finally, a cohort of German adults found higher intake of sweets (chocolate, pralines, candy bars, ice-cream, and sugar) to be related to higher risks of large weight gain among men and to lower risks of large weight loss among women [35].

Sugar-sweetened beverages ("soft drinks")

Sugar-sweetened beverages, including sodas such as coke, fruit drinks, lemonade, and iced tea have been suggested to promote weight gain mainly due to increasing overall energy intake [63]. In line with this, short-term feeding studies in humans have shown that energy provided by sugar-sweetened beverages is not adequately compensated for in subsequent meals in contrast to energy consumed from solid foods [63-65]. In recent years, a huge amount of evidence has accumulated to report on the role of sugar-sweetened beverages in weight gain or development of overweight and obesity [66, 67]. Although there is a paucity of high-quality intervention studies, the evidence in general support that consumption of soft drinks promotes weight gain and the WCRF judged the epidemiological and mechanistic evidence for a positive association to be "probable" [17].

Alcohol use

Similar to sugar-sweetened beverages, alcohol is a form of liquid energy which might lead to excess energy intake and thus predispose to (larger) weight gain in the long-term [63]. The drinking pattern of the moderate alcohol consumer is characterised by the addition of alcoholic beverages to usual food intake leading to a positive energy balance [68]. In addition, alcohol cannot be stored in the human body and its oxidation takes precedence over other substrates which may result in suppressed fat oxidation, thereby promoting fat storage. Finally, alcohol has been shown to stimulate appetite, though the underlying mechanisms are unclear. So far, epidemiological studies have not provided consistent evidence for alcohol use as a risk factor for weight gain and obesity. Findings from prospective studies are conflicting [17, 68], showing, for example, positive [69, 70] or null [35, 71] associations in men and positive [69], null [35, 72] or inverse [71, 73] associations in

women. In 2007, the WCRF judged the evidence for an association of alcohol use with subsequent weight change or the development of overweight/obesity to be “insufficient” [17]. A recently published review including 13 prospective studies also concluded that results were contradictory, but also suggested that particularly heavy drinking might be associated with larger weight gain and that relations may differ according to type of alcoholic beverage [74].

Other factors related to diet

Many other factors related to diet have been suggested to play a role in the regulation of energy intake and body weight. For example, increasing portion size, e.g. served at restaurants and produced by manufacturers, could promote weight gain by increasing total energy intake [75, 76]. Likewise, eating outside home has been proposed to be a risk factor for weight gain due to the generally higher fat and total energy content of foods eaten outside home, however, evidence from longitudinal studies is scarce and inconsistent [77]. Further, a higher eating frequency might prevent weight gain as it appears to be associated with metabolic advantages, including reduced serum levels of total cholesterol, LDL cholesterol, and insulin [78]. Eating frequently may further increase food-induced thermogenesis, decrease efficiency of energy utilisation and suppress hunger; however, evidence is sparse [79, 80]. Finally, being breastfed has been consistently shown to protect against childhood obesity [17].

1.2.5.2 Physical activity

As physical activity is the most variable component of total daily energy expenditure, it qualifies as key lifestyle intervention strategy to prevent weight gain and thus to combat the increasing prevalence of overweight and obesity and associated co-morbidities. Apart from solely increasing energy expenditure, physical activity has been shown to involve several metabolic and physiological benefits. For example, regular (moderate-intensity) exercise increases fat oxidation presumed to result from improved insulin sensitivity and has been shown to affect appetite control by increasing sensitivity of satiety signals, altering food choices and macronutrient preference, and modifying the pleasure response to food [17, 81].

A number of studies have reported on the role of physical activity in successful weight loss, in the prevention of weight regain after weight loss, and in the primary prevention of weight gain up to now [82]. In 2007, the WCRF systematically summarised the evidence on physical activity and subsequent weight gain from 16 cohort studies and from the few available randomized controlled trials on weight loss maintenance [17]. The expert panel concluded that there was substantial and convincing evidence that all types of physical activity protect against weight gain. Only recently, results from the Nurses’ Health Study and from the CARDIA study underlined this conclusion by reporting that greater duration of physical activity was associated with less weight gain over 8 years of follow-up [83] and that subjects

who consistently maintained high levels of physical activity over 20 years gained significantly lower amounts of weight than subjects who maintained lower physical activity [84], respectively.

However, despite the large body of evidence indicating an inverse association between physical activity and weight gain, the amount of physical activity that is necessary to prevent weight gain has been widely debated [85, 86]. In 1995, the US Centers for Disease Control and Prevention (CDC) and the American College of Sports Medicine (ACSM) jointly recommended 30 minutes or more of moderate-intensity activity (such as brisk walking) on most, preferably on all days of the week to improve health (in particular cardiovascular health) [87]. This recommendation was confirmed in the US Surgeon General's report in 1996 [88, 89] and has been generally accepted as the physical activity guideline for the public worldwide. In 2007, the ACSM and the American Heart Association updated the recommendation from 1995 and specified these recommendations with regard to type and amounts of physical activity by recommending a minimum of 30 minutes moderate-intensity aerobic activity on five days each week or vigorous-intensity aerobic activity for a minimum of 20 minutes at three days of the week to promote and maintain overall health [90]. With respect to the primary prevention of weight gain, however, the adequacy of these recommendations has been debated [85, 86, 91]. Although acknowledging that "definitive data are lacking", the authors of a consensus statement published in 2003 concluded that physical activity of moderate intensity for 45 to 60 minutes per day may be required to avoid the transition to overweight or obesity [85]. Erlichman and colleagues even suggested 60-90 min of moderate-intensity activity to maintain a stable weight [86]. These higher recommendations are underlined by recent findings from the Women's Health Study showing that 60 minutes of moderate-intensity activity each day were needed to successfully maintain or to gain only few amounts of weight over a period of 13 years [92]. Nevertheless, irrespective of the exact duration of physical activity necessary for weight gain prevention, vigorous activity appears to be more clearly linked to weight stability, it allows a higher intensity of exercise for general activities and reduces the time needed for achieving the recommended energy expenditure through physical activity.

Apart from physical activity, associations of sedentary living or single sedentary behaviours like television watching with weight gain are increasingly being investigated. With respect to television watching, for instance, the evidence is mostly consistent showing a positive association with weight gain and overweight [17]. Mechanism to explain this relationship refer to reduced time available for physical activity, reduced resting metabolic rate and increased energy intake as television watching appears to be associated with the consumption of energy-dense foods and drinks [93].

1.2.5.3 Other modifiable factors

Duration of sleep

Parallel to the dramatic increase in prevalence of overweight and obesity in Western societies during the last decades, a steady and rapid decline in time spent sleeping has been documented [94, 95]. One of the strongest hypotheses for a link between sleep deprivation and higher risk for weight gain is that sleep restriction leads to increased dietary intake. In animal studies, sleep deprivation has been consistently found to produce hyperphagia. In line with this finding, sleep restriction was associated with decreased leptin levels and increased appetite-stimulating ghrelin levels also in human cohort studies, suggesting an effect on peripheral regulators of dietary intake. Further hypotheses refer to reduced impulse control and difficulties with delaying gratification following chronic sleep deprivation which may result in increased hedonistic eating and the suggestion that calorie intake may be directly proportional to time spent awake. A recent systematic review concluded that short sleep duration appears to be associated with weight gain [94]; however, it should be noted that evidence from prospective studies is sparse.

Status and change in smoking habits

Current smoking has generally been associated with lower BMI, but increased abdominal fat distribution [96, 97]. Another widely observed phenomenon is weight gain after cessation of smoking [96, 98]. About 80% of all smokers who quit smoking are estimated to gain weight, particularly in the first year after cessation [96]. Data from the NHANES study have shown that weight gain attributable to smoking cessation was 4.4 kg among men and 5.0 kg among women over a 10-year period [99]. Although the evidence for a strong association between smoking cessation and subsequent weight gain is convincing, the underlying biological mechanisms are not yet clear. Hypothesised mechanisms include increased energy intake, metabolic changes, i.e. decreased resting metabolic rate, and increased lipoprotein lipase activity [96, 98]. Nicotine has been suggested to influence levels and expression of peptide hormones and neurotransmitters, such as leptin and neuropeptide Y, which are involved in the regulation of food intake. With respect to adipose tissue metabolism, it has been observed that fat oxidation increases with increasing nicotine uptake and that smoking cessation is related to increases in lipoprotein lipase activity in adipose tissue, which may contribute to post-cessation weight gain.

1.3 The risk score approach: application and methodological aspects

Prediction models are important statistical tools in various fields, including meteorology, physics and finance [100]. In the medical field, much emphasis has traditionally been given to the identification of single risk factors that aetiologically relate to the development of diseases or other adverse events and measures of associations such as regression coefficients or relative risks were commonly reported [100]. However, in recent years, prediction models are increasingly being developed to summarise the effect of single risk factors (predictors) and to estimate an individual's absolute risk for the outcome of interest given a specific risk factor profile. Importantly, in contrast to models aiming to explain aetiological associations between risk factors and the outcome, prediction models aim to develop a good predictor (e.g. risk score) and do only marginally consider the model structure [101]. In chronic disease epidemiology, logistic or Cox Proportional Hazards (PH) regression models are most frequently used to identify relevant predictors of the outcome. Predictors are usually retained in the model if they are statistically significantly associated with the outcome and a risk score is commonly calculated for each individual by assigning weights to the predictors based on their effect size, with larger risk scores relating to higher risks for the outcome. Different approaches exist for selecting candidate predictors to be included in the final model, including causal vs. non-causal, modifiable vs. non-modifiable and easily assessable vs. more sophisticated measurements of risk factors.

Prediction models are considered valuable tools for public health, clinical practice, and medical research. In public health, the key purpose of prediction models is to estimate an individual's risk of developing a disease within a specified time period and to stratify apparently healthy individuals into clinically relevant or other meaningful risk categories [9, 100]. This information may then be used to target preventive interventions particularly to those subjects who are at high risk for developing the disease of interest. For example, statin therapy is only considered for those subjects at relatively high risk for cardiovascular disease [100] and subjects with high risk for certain cancers may consider chemoprevention intervention [102] or may undergo a program of screening surveillance that might not be appropriate for individuals with lower disease risk.

In clinical practice, prediction models may inform patients and physicians on the probability of an existing, underlying disease (diagnosis) or a prognostic outcome (e.g. mortality risk after diagnosis, weighing of harms vs. individual benefit) and thus may help to decide on further testing and may guide therapeutic decision-making [100].

In medical research, prediction models may for instance be used for designing, planning and establishing eligibility criteria for intervention studies or stratification of individuals to obtain

balanced treatment groups with respect to the main prognostic factors in a randomised trial [100]. In observational studies, prediction models may be used to control for confounding, for example using propensity scores.

Importantly, a clear distinction has to be made between diagnostic and prognostic models. While in the diagnostic setting, models are concerned with accurately determining the current, but unknown, disease state of a patient, in prognostic modelling, the disease outcome has not yet developed at the time when predictors are assessed and future disease status is determined by stochastic processes and estimated as a probability [103]. In the following, the term “prediction model” will refer to prognostic models unless otherwise stated.

Risk assessment tools may help to correctly and easily identify subjects at high risk for a specific disease in order to deliver targeted intervention advice and treatment to them. Accurate risk assessment will avoid over-treatment of those individuals with lower risk of the disease and allows an efficient use of medical resources [10]. Therefore, appropriate and careful evaluation of the accuracy of a prediction model is essential before it can be recommended for practical use. The two main aspects of the performance (validity) of a statistical prediction model are discrimination and calibration. Within the following, these two performance measures will be described. Further, the importance of evaluating the model’s accuracy in other populations (external validity) will be highlighted. Finally, a brief overview about some important and most well-known risk prediction models in the field of major chronic diseases, CVD, cancer and type 2 diabetes, is given.

1.3.1 Measures of model performance

Discrimination

Discrimination refers to the ability of a diagnostic test or a risk prediction model to distinguish between those individuals with and without the outcome or between those at high and low risk of the disease, respectively. In the case of a simple binary (diagnostic) test, subjects are classified into two groups, those with the outcome and those without the outcome. Subsequently, a 2-by-2 table (**Table 3**) can be used to evaluate how well the test assesses the outcome in comparison to the truth which is usually determined using a “gold standard”, a more definitive and often more invasive test [104]. Ideally, all subjects would fall into the two light blue-shaded cells in the upper left and lower right on the table, implying that people would be correctly classified as diseased or healthy. In reality, this is very rarely if ever the case and some people are classified erroneously as diseased although they are healthy and vice versa (referred to as “false-positives” and “false-negatives”, respectively).

The two basic measures of quantifying the discriminatory accuracy of a test are the sensitivity and specificity [104]. Sensitivity of a test, or true positive rate, is defined as the

probability of a positive test result among those subjects with the outcome (TP/P). Specificity, or true negative rate, refers to the probability of a negative test result among those subjects without the outcome (TN/N). In comparing tests, those that are higher in both sensitivity and specificity are preferred; however, there is a trade-off between both measures, as sensitivity increases, specificity decreases and vice versa.

Table 3. Comparison of a dichotomous test result with true disease status

Disease / Test	Positive	Negative	
Positive	True positive (TP)	False Positive (FP)	Test positive (P')
Negative	False Negative (FN)	True Negative (TN)	Test negative (N')
	Disease (P)	Healthy (N)	Total

In most circumstances, however, the result of a test is not a simple binary one, but may be rather a continuous measure, such as blood pressure, plasma glucose or a risk score derived from a multivariate prediction model. In this case, designation of a cut-off point for distinguishing test positive (high risk) versus test negative (low risk) in order to compute sensitivity and specificity is arbitrary [105]. Alternatively, the whole range of a continuous predictor and all combinations of sensitivity and specificity may be considered. Such a summary of sensitivity and specificity across the whole range of a continuous predictor can be illustrated using a receiver operating characteristic (ROC) curve (**Figure 2**).

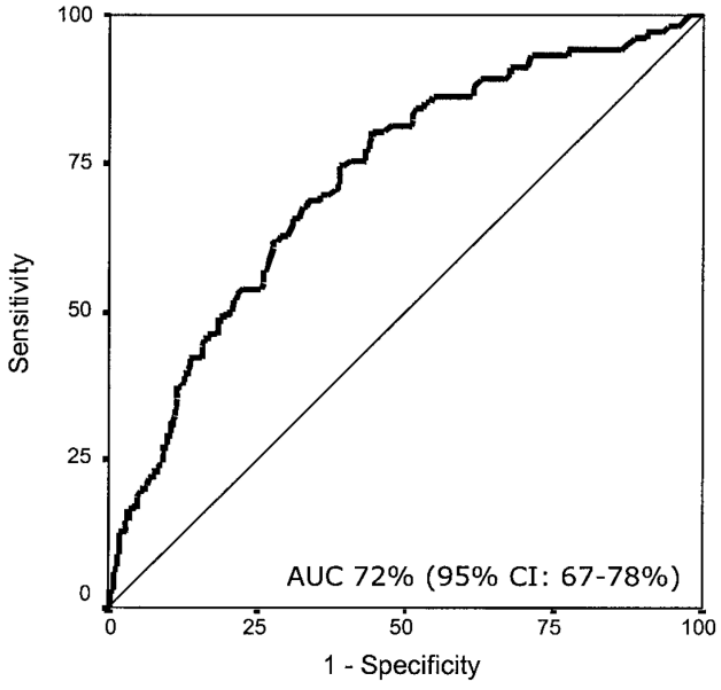


Figure 2. Example of a receiver operating characteristic curve.
This figure is taken from [106]. AUC = area under the receiver operating characteristic curve.

The ROC curve is a plot of the sensitivity versus 1 – specificity (false-positive rate) where the different points on the curve correspond to different cut-off points of the continuous measure used to designate test positives/high-risk individuals [105]. The discriminatory ability of a continuous test or risk score is most commonly quantified by the concordance statistic (c-statistic) which is identical to the area under the ROC curve (aROC) [100]. In the case of risk prediction, it is equivalent to the probability that the predicted risk is higher for a randomly selected individual who will develop the disease within a specific time period than for a randomly drawn non-case who will not develop the disease within the same period of time [100]. The aROC ranges from a minimum of 0.5 when the ROC curve lies on the 45° reference line to a theoretical maximum of 1.0 when the curve reaches the upper left corner. A value of 0.5 indicates that the model is uninformative, it has no discriminatory ability, and is not superior to a random guess; a value of 1.0 represents perfect discrimination. Perfect discrimination is achieved if the predicted risk for all cases is higher than for non-cases, with no overlap.

Because sensitivity and specificity reflect the ability of a test to distinguish between individuals with and without the outcome, they are specifically important for public health, i.e. for screening of free-living populations [104]. Basically, the question that can be answered using sensitivity and specificity is: “If we screen a population, what proportion of people will be correctly identified as diseased or non-diseased, respectively?”. In clinical practice, however, it is also important to know, how good the test is at giving the correct outcome. Specifically, a physician may be particularly interested in the question “If the test result is positive, what is the probability that the person has or will develop the disease of interest?”. Likewise, a parallel question about negative test results may be asked: “If the test result is negative, what is the probability that the person does not have or will not develop the disease?”. These aspects of a test are captured by the concept of predictive values. The positive predictive value (PPV) is the probability that an individual has (or will develop) the disease or condition of interest given that the test result is positive (TP/P’). Hence, the negative predictive value (NPV) is the probability that an individual does not (will not) have the disease/condition given that the test result is indeed negative (TN/N’). Importantly, the predictive value is affected by the prevalence of the disease in the population tested [104]. The higher the prevalence, the higher the PPV, with the gain in PPV being particularly high in the lower ranges of disease prevalences (<20%). Therefore, the result of any test result must be interpreted in the context of the disease prevalence in the population from which the person originates (high-risk vs. low-risk population) [104]. Additionally, the PPV may be improved by increasing the specificity of a test [104].

With respect to the application of a risk score in practice, an appropriate cut-off value for the continuous risk score needs to be chosen in order to discriminate between test positive/high-

risk and test negative/low-risk individuals. Mathematically, the optimal threshold in ROC curves is defined as the 45 tangent in the upper left corner of the plot, i.e., the point maximising sensitivity and specificity. The Youden's index is a simple measure finding the "optimal" threshold value for which sensitivity and specificity are maximised across a range of possible cut-off values [107, 108]. It is defined as $J = sensitivity + specificity - 1$ and ranges from 0 to 1, with 1 implying perfect separation of diseased and non-diseased by the continuous marker [108]. One drawback of the Youden's index is that sensitivity and specificity are considered equally important in the calculation which might not hold true in practice. In some circumstances, for example, a false-negative finding may be more critical than a false-positive one or vice versa. False positives may be an important issue because individuals with a positive test may be advised to undergo more sophisticated and more expensive tests placing a considerable burden on the health care system [104]. Further, anxiety and worry may be induced by a positive test result and create a stigma that is difficult to erase. In contrast, false negative test results might be of particular concern when the disease is serious and effective intervention is available, but may only be effective in early stages of the disease. Therefore, in clinical practice and public health, designation of a cut-off value depends in the end on the importance attached to false-positives and false-negatives and misclassification-costs need to be taken into account.

Calibration

Calibration is a measure of how reliable the predictions are, that is how well predicted probabilities agree with actual observed risks [100]. Thus, a model is well calibrated when the average predicted risk matches the proportion that actually develops the disease under study. Usually, subjects are placed within categories of predicted risk and the category values are compared with the observed incidence in each category. More formally, the Hosmer-Lemeshow goodness-of-fit test (HLT) compares observed with predicted risk, typically across deciles of the distribution of estimated risk [109]. The null hypothesis is "the model is well calibrated" and thus a low p-value indicates lack of calibration. A graphical illustration of the HLT is often presented in terms of calibration plots, which plot the observed risk by deciles of the predicted risk (**Figure 3**). The closeness of the points to a 45° line is then evaluated. If the points are lying above the straight line, the model underestimates the true risk, while it overestimates the true risk when the points are lying below the line. However, it should be noted that the p-value of the HLT is highly influenced by sample size and sensitive to the way groups are formed [110]. In this respect, it might be reasonable to form categories on the basis of predicted probabilities that are of clinical relevance or have a more intuitive meaning than deciles (such as >0–<5%, 5–<10%, etc.) [103].

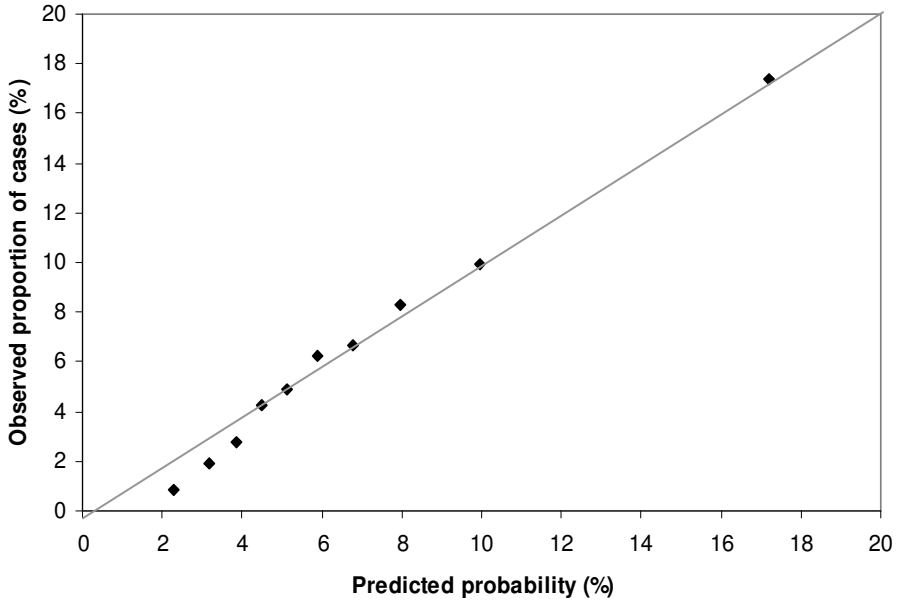


Figure 3. Hypothetical example of a calibration plot

The calibration plot indicates that the risk score overestimates the true risk in the lower range of the score points while it accurately quantifies the true risk in the upper range of the score points.

It is important to note that a risk prediction model cannot be both perfectly discriminatory and perfectly calibrated. It has been shown that a model which maximises discrimination does so at the expense of calibration and vice versa [111]. Diamond et al. demonstrated that a perfectly calibrated model, in which the observed risk matches the predicted risk for all subgroups, cannot achieve a c-statistic of 1.00 in usual settings [111]. In this regard, it was also shown that the maximum attainable c-statistic varies with the distribution of risk in the population [112]. Further, the c-statistic is based on the ranks of the predicted probabilities of cases and non-cases, but it is not a function of the actual predicted probabilities, which means that it is insensitive to errors in calibration [9]. For instance, a model that assigns all cases a value of 0.56 and all non-cases a value of 0.54 would have perfect discrimination, although the probabilities it assigns may not be meaningful. Whether discrimination or calibration may be of prime importance depends on the purpose of the risk score. In the diagnostic setting where classification into groups of disease state is the main purpose, discrimination is of higher interest than calibration. In prognostic modelling where risk of future diseases is estimated and with respect to public health decision making, the actual risk itself may be of major concern and calibration accuracy may play a pivotal role [103]. Also, precise computation of the absolute risk is particularly essential when the likely absolute benefit from an intervention for cost-benefit analyses needs to be estimated [11]. Further, the decision of an individual to participate in an intervention program might be influenced by providing information on the expected benefit. In this case, accurate estimation of absolute

risk is also highly important. On the contrary, if a prognostic risk score solely serves the purpose to identify high-risk individuals to undergo prevention interventions, the risk score would need to accurately rank individuals according to their absolute risk, but would not necessarily need to provide accurate estimates of the risk itself.

Validation of a risk score

Because the purpose of a prediction model is to provide valid outcome predictions in practical settings outside the sample it was developed in, validation is a crucial step in predictive modelling [100]. An important distinction has to be made between internal and external validation. Internal validation refers to the assessment of the validity (discrimination, calibration) of a predictive model for the population it was developed in (training sample). In contrast, external validation determines whether the model is generalisable to other populations that are fully independent from the development data and originate from different but related settings (same underlying source population). A key threat to validity of a prediction model is overfitting/optimism, which means that the data under study are well described but that predictions do not generalise to new subjects outside the sample [100]. Overfitting is likely to occur in data-dependent analyses and may result from fitting specifics and idiosyncrasies of the sample rather than generalisable patterns [100]. It leads to a too optimistic impression of model performance that can be achieved in new subjects from the underlying population. Although several techniques do exist to correct for optimism, including cross-validation and bootstrapping methods, validation in external, independent study populations is the most stringent type of validation and of crucial importance if a prediction model should be used in the clinical or public health setting [100]. The more often a model is externally validated and the more diverse these validation settings, the higher the confidence with regard to general applicability of the model.

1.3.2 Overview about existing risk scores

During the past three decades, numerous risk prediction models have been developed in the context of cardiovascular diseases, type 2 diabetes and cancer. A number of these prediction algorithms have been transferred to simplified score sheets that allow physicians to easily identify subjects at high risk of the disease and to deliver targeted individualised intervention advice [113-116]. In the field of weight management, there is currently no risk score available to predict weight gain or overweight/obesity among adults; nevertheless, a risk score to predict overweight among children has been published recently [16] and is presented below. Also, **Table 1 in the appendix** gives a selective overview about some important, widely-known risk prediction models in the field of major chronic diseases, including information on the predictors included in the respective score, the corresponding performance measures and on external validation.

The first risk prediction model for a chronic disease was published in 1976 in the field of **cardiovascular diseases** [117]. This risk prediction model used data from the Framingham Heart Study to assess an individual's risk for developing CVD during the next 8 years using information on a number of clinical and biological factors [117, 118]. Recently, a modified version of this risk assessment tool to predict 10-year risk of CVD was published [119]. Further, data from the Framingham study were used to develop risk prediction models to predict specific components of CVD, including coronary heart disease (CHD) [113], heart failure [120] and stroke [121]. For instance, the Framingham risk score to predict risk of developing CHD within the next 10 years is a simple algorithm using information on age, prevalence of diabetes, smoking, blood pressure and cholesterol and was incorporated into the Third Report of the Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults [122]. This CHD risk assessment tool showed a discriminatory ability ranging between 0.73 and 0.77, depending on sex and whether categorical or continuous variables for cholesterol were used, and has been validated in whites and blacks in the United States [123]. Because the Framingham risk equations were only transportable to other populations after recalibration, several CVD prediction models have been developed in Europe, such as the SCORE equation [115] recommended by the Third Joint European Task Force on cardiovascular prevention [124], the PROCAM score developed in Germany [125], the QRISK algorithm developed in the UK [126], and the ASSIGN score developed in Scotland [127]. Similarly, risk prediction models for CVD have been constructed in other parts of the world, including China, Japan, New Zealand, and Australia [10].

In the late 1980s and early 1990s, the first risk prediction models in the field of **cancer** were published. The best known and most widely applied risk prediction model is that developed by Gail et al. in 1989 to predict breast cancer risk within the following 5 years [12]. The Gail model is based on risk factor information collected in a case-control study nested in the Breast Cancer Detection Demonstration Project (BCDDP). Estimates of relative risk from the case-control data were combined with age-specific breast cancer rates in the entire BCDDP cohort to estimate absolute risk developing breast cancer within the next 5 years. The model includes standard risk factors known at that time, including current age, age at menarche, age at birth of first child, number of first-degree relatives with breast cancer, and number of previous breast biopsy examinations. This original model was recalibrated to national cancer incidence rates for estimating 5-year risks of invasive breast cancer and this modified version is implemented in the National Cancer Institute's Breast Cancer Risk Assessment Tool (BCRAT) [128]. The model has been used to design prevention trials such as the Breast Cancer Prevention Trial to evaluate the benefits of tamoxifen in a population of women with an elevated risk of breast cancer and to assist in clinical decisions. Further, women with a high risk score are encouraged to undergo mammographic screening or genetic evaluation.

The Gail model has been shown to reliably predict risk at the population level (well calibrated), but its discriminatory accuracy at the individual level is only fair [129-133]. For instance, in a prospective cohort study of the San Francisco Mammography Registry [131] and in the Nurses' Health Study [130], the aROCs were 0.67 and 0.58, respectively. In recent years, the Gail model was modified by adding newly identified risk factors such as breast density [131, 134] or by including genetic information [135, 136]. However, improvements in discriminatory ability were generally modest, e.g. the aROC increased from 0.607 to 0.632 when risk factor information of 7 genetic variants was added [135]. Next to breast cancer, an increasing number of risk prediction tools have also been developed for other types of cancer, including cancers of the colorectum [13, 14, 137], ovary [138], prostate [139], lung [102, 140-144], and skin [145, 146]. The discriminatory ability was mainly fair, with aROCs ranging between 0.60 and 0.75, with the exception of recently published results on lung cancer coming from a large randomized screening trial showing aROCs of 0.86 and 0.81 among the whole control arm and among smokers only, respectively [144]. In external validation studies of cancer risk prediction models, calibration is mostly good and discriminatory accuracy is generally modest. For instance, the aROC was 0.61 for colorectal cancer [147] and 0.69 for lung cancer [148]. Although these risk prediction models were developed on the basis of well-established risk factors, the comparatively poor discriminatory ability suggests the need to find additional strong risk factors.

As recently summarised in a systematic review, a variety of risk prediction models has also been developed in the field of **type 2 diabetes** [11]. Overall, 46 studies presenting risk prediction models for diabetes were identified, of which 10 were validated in external populations, including the Finnish diabetes risk score [149], the Atherosclerotic Risk in Communities study (ARIC) diabetes risk score from the US [150], the Cambridge diabetes risk score [151], and the German diabetes risk score developed in EPIC-Potsdam [114]. Risk prediction models for diabetes can be broadly divided into three categories: models solely based on non-invasive measurements (e.g. age, measures of anthropometry, diet and lifestyle factors), models including classical biochemical measures such as fasting plasma glucose and lipids along with non-invasively assessed variables, and finally those models additionally containing novel biomarkers (e.g. C-reactive protein or adiponectin) and/or genetic information. The discriminatory accuracy of risk scores solely based on non-invasive measurements generally ranged from 0.70 to 0.80, the Finnish Diabetes Risk Score and the German Diabetes Risk Score even reported aROCs of 0.84 and 0.82, respectively. In general, the addition of biochemical measures, in particular fasting plasma glucose, could strongly improve the performance of models only containing non-invasive measures. For example, the German Diabetes Risk Score based on age, waist circumference, height, hypertension, physical activity, smoking, and consumption of whole-grain bread, red meat,

coffee and moderate alcohol improved after inclusion of measures of fasting plasma glucose, glycated haemoglobin, lipids, and liver enzymes to an aROC of 0.90 [114]. In contrast, the addition of multiple genetic markers only marginally improved discrimination beyond non-invasive characteristics in EPIC-Potsdam and in other studies [11, 152]. Similar as with prediction models in the context of CVD and cancer, the discriminatory ability was usually reduced in external populations. For instance, the Finnish diabetes risk score, which is the most frequently validated risk score, was tested in eight independent cohorts and yielded good discrimination among a similar Finnish population (aROC=0.87), though somewhat lower discrimination among other cohorts (aROC ranged from 0.65 to 0.81) [11]. The majority of studies did not report measures of calibration.

With regard to **weight gain or overweight**, only fairly recently a risk score predicting the risk of newborn children to be overweight at 8 years of age was developed [16]. In this birth cohort study from the Netherlands, a total of 1,687 newborn children were followed until the age of 8 years to undergo a medical examination. Overweight was defined according to age- and sex-specific cut-off points for BMI. A risk score comprising information on paternal BMI, maternal BMI, gender, smoking in the parental house, birth weight of the child and hospital delivery was able to adequately predict risk of overweight at the age of 8 years. The aROC was 0.75, and the Hosmer-Lemeshow Test implied adequate calibration after adjustment for optimism ($p=0.30$). A validation of the risk score in an external study population has not yet been performed.

1.4 Challenges and research questions of the thesis

Given the rapid rise in obesity prevalence around the world in recent years and the considerable strains involved with the treatment of overweight and obesity, primary prevention of weight gain that may lead to overweight and obesity is of predominant public health importance. Because resources for prevention are limited, a simple and pragmatic tool that allows for identifying individuals who are at high risk of experiencing substantial weight gain and who may benefit most from timely intervention seems appealing. As no such tool has been developed among adults so far, the aim of the present thesis is to investigate whether it is possible to develop a risk score predicting 5-year risk of SWG in the context of the multi-centre EPIC study.

Given the continuous nature of the outcome variable weight gain and the multi-centric design of the EPIC study, the current thesis is faced with two major methodological challenges that need to be accomplished. First, the question arises of how to model the continuous measure weight gain in order to estimate absolute risks analogously to risk prediction model building for hard clinical endpoints, while accounting for varying follow-up times between study participants. Second, care needs to be taken of heterogeneity, originating from differences in questionnaire design, anthropometric measurements and follow-up times, across single centres in the statistical modelling process in a way that the final risk score is independent of any effects of the study centres themselves.

In particular, the following research questions will be addressed:

Is it possible to develop a risk score based on readily available factors to predict substantial weight gain over 5 years using data of the multi-centre EPIC study?

- How well does the risk score discriminate between cases and non-cases?
- How well does the score quantify absolute risk?
- How well does the score perform in an independent population?
- How well does the score perform across diverse European populations?

2 Material and Methods

2.1 The EPIC study

The EPIC study is an ongoing large multi-centre prospective cohort study designed primarily to investigate the relationship between diet, lifestyle and genetic factors and the incidence of cancer [153, 154]. It was initiated in 1992 within the framework of the "Europe against Cancer" programme of the European Union and is coordinated by the International Agency for Research on Cancer (IARC) of the World Health Organization (WHO) in Lyon, France. Between 1992 and 2000, a total of 521,448 participants (~70% women and 30% men) has been recruited in 23 administrative centres located in 10 European countries: Denmark, Sweden, Norway, the United Kingdom, France, Germany, The Netherlands, Spain, Italy and Greece (**Figure 4**).

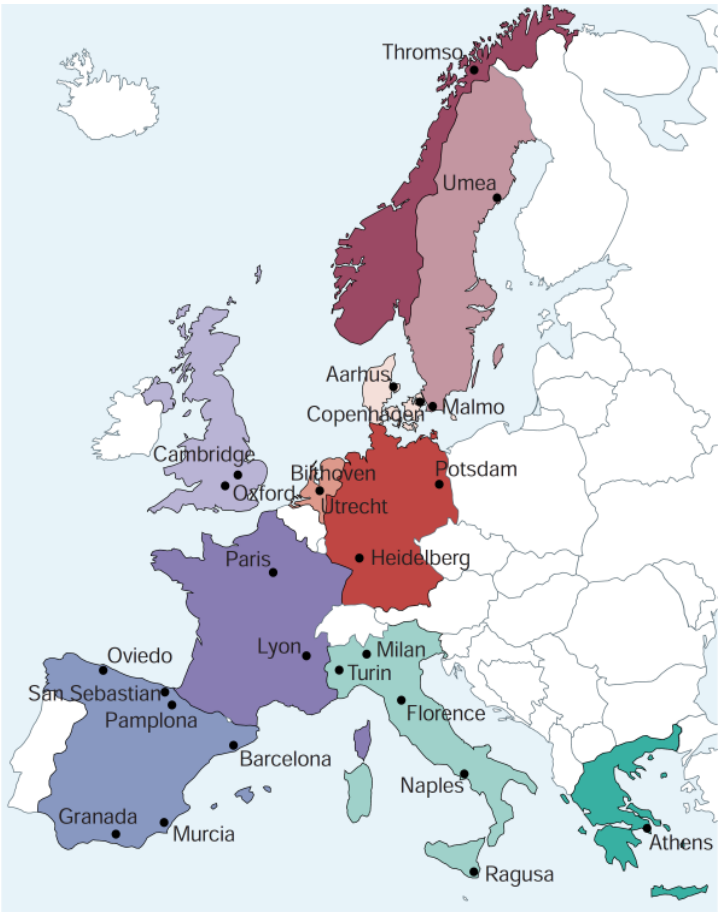


Figure 4. Map of cohorts participating in the EPIC study
Source: Bingham and Riboli, 2004 [155]

Eligible male and female study participants were generally aged between 35 and 70 years at enrolment and, in the majority of study centres, they were invited from the general population residing in a given geographical area, a town, or a province. Exceptions were the French cohort (based on members of the health insurance for teachers), the cohorts of Utrecht (The

Netherlands) and Florence (Italy), which are based on women attending local population-based breast cancer screening programmes, components of the Italian and Spanish cohorts (including members of local blood donor associations), and most of the Oxford (UK) cohort (comprising health-conscious subjects, mainly vegetarians). In France, Norway, Utrecht, and Naples (Italy) only women were recruited. All participants gave written informed consent and approval for this study was obtained by the local ethics committee in the participating countries and the International Review Board of IARC.

Baseline examination

In the majority of centres, individuals who agreed to participate were mailed a questionnaire on diet and a questionnaire on lifestyle, which they completed at home. Subjects were then invited to a study centre to turn in the questionnaires, to provide a blood sample and to have anthropometric measurements taken.

Follow-up data collection

EPIC participants are passively followed for vital status, cause of death and disease occurrence. In most of the centres, these data are obtained by regular record linkage, with the exception of Greece and Germany where an active follow-up is used. For updating lifestyle and anthropometric data, most participants were contacted for a second time several years after recruitment. Assessment was conducted through mailed questionnaires, with exception of Spain and Greece who contacted all participants by phone, Varese (Italy) who used a combination of questionnaires and telephone, and Cambridge (UK) and Doetinchem (The Netherlands) where individuals were invited to a study centre for a second measurement of anthropometric parameters.

2.1.1 Assessment of relevant variables

Diet

Usual dietary intake over the past 12 months was assessed at baseline by means of country-specific dietary questionnaires that were designed to capture local dietary habits and to provide high compliance [154, 156]. These dietary questionnaires had been developed and validated in a series of studies within the various source populations participating in EPIC and they had been shown to be able to rank subjects according to dietary intake within centres [156, 157]. Participants were asked to report their average consumption of each food item over the previous 12 months, according to pre-coded categories ranging from never or less than once per month to five or more times per day. In the majority of countries, extensive quantitative food frequency questionnaires (FFQ) containing up to 260 food items and estimating individual portion sizes were used [154]. In Denmark, Norway, Naples, and Umea (Sweden), semi-quantitative FFQ assigning the same standard portion(s) to all

participants were administered. In most countries, dietary questionnaires were self-administered, with the exception of Greece, Spain, and southern Italy (Naples and Ragusa) where questionnaires were administered at a personal interview in order to increase compliance of study participants. In Malmö (Sweden) and the UK, a combination of dietary methods (semi-quantitative FFQ and diet record) was used. Food intake was calculated by multiplying food frequency and portion size and expressed as grams per day.

Anthropometry

In all EPIC centres, except France, part of the Oxford cohort, and Norway, height and weight were measured at baseline on all subjects by trained personnel according to standardised procedures with subjects wearing no shoes [158]. Body weight was corrected to reduce heterogeneity due to protocol differences in clothing worn during measurement by subtracting 1.5kg in those individuals who were normally dressed and 1kg in those participants who wore light clothing (**Table 4**). For part of the Oxford cohort where only self-reported data were available, linear regression models were used to predict sex- and age-specific values from subjects with both measured and self-reported body measures (referred to as Oxford prediction equations in the remainder of the text, equations see below) [158]:

$$\text{Men}:$$
$$\text{Weight} = 0.561 + (1.012 \cdot \text{sr\_weight}) + (0.006 \cdot \text{age})$$
$$\text{Height} = 15.032 + (0.923 \cdot \text{sr\_height}) + (0.052 \cdot \text{age})$$

$$\text{Women}:$$
$$\text{Weight} = 0.444 + (1.010 \cdot \text{sr\_weight}) + (0.006 \cdot \text{age})$$
$$\text{Height} = 27.096 + (0.853 \cdot \text{sr\_height}) + (0.069 \cdot \text{age})$$

sr = self-reported

At follow-up, body weight was self-reported by the participants in all centres, except for participants in the cohorts of Cambridge (UK-Nor) and Doetinchem (NL-Doe), in whom measurements were performed by trained staff following the same protocol as during baseline measurements. Self-reported body weight at follow-up was corrected for potential underreporting by using the Oxford prediction equations. BMI was calculated by dividing weight in kilograms by height in metres squared (kg/m^2).

Socio-demographic and lifestyle factors

At baseline, information on a large number of lifestyle and health factors was collected by means of questionnaires and/or face-to-face interviews including questions on education and socio-economic status, occupation, physical activity (occupational and recreational activity), history of previous illnesses, menstrual and reproductive history, consumption of alcoholic beverages, and tobacco smoking [154]. Updated information on smoking status was also collected at follow-up at the same time as the second anthropometric assessment.

Table 4. Body weight assessment at baseline in the centres participating in EPIC

	Method	In light underwear	In light clothing *	Normally dressed	Correction for clothing
Greece	M		+		- 1.0 kg
Spain (all centres)	M	+			none
Italy					
Ragusa	M	+			none
Naples	M	+			none
Florence	M	+			none
Turin	M			+	- 1.5 kg
Varese	M	+			none
France (all centres)	SR			+	-
Germany					
Heidelberg	M	+			none
Potsdam	M	+			none
The Netherlands					
Doetinchem	M		+		- 1.0 kg
Amsterdam/Maastricht	M		+		- 1.0 kg
Utrecht	M			+	- 1.5 kg
United Kingdom					
Cambridge (Norfolk)	M		+		- 1.0 kg
Oxford (GP)	M		+		- 1.0 kg
Oxford (HC)	SR/M		+		- 1.0 kg
Denmark (both centres)	M	+			none
Sweden					
Malmö	M		+		- 1.0 kg
Umea	M			+	- 1.5 kg
Norway	SR				-

M = measured, SR = self-report, GP = General population, HC = health conscious.
* After removal of shoes, heavier sweaters, indoor jackets and heavier objects from pockets.
Modified from Haftenberger et al. (2002).

2.1.2 Analytical study populations

Given the importance of validating a risk prediction model in an external study population, the total study population of EPIC was non-randomly divided into two samples: a training sample to guide model development and an independent validation sample to test the model. As illustrated in **Figure 5**, the model was developed using data of those EPIC cohorts participating in the Diet, Obesity and Genes (**DiOGenes**) project which is a multi-disciplinary, integrated European research programme focussing on the identification of key dietary, psychological, genetic and behavioural factors in the prevention of weight gain [159, 160]. The development of the risk score model was therefore based on data from five EPIC countries which are the United Kingdom (Norfolk, UK-Nor), the Netherlands (Doetinchem, Amsterdam and Maastricht), Denmark (Copenhagen and Aarhus, DK-CopAa), Germany (Potsdam, GER-Pot), and Italy (Florence, IT-Flo). Because of differences in the assessment of follow-up anthropometric measurements in the Netherlands (see 2.1.1), the cohort of Doetinchem (NL-Doe) and the cohort of Amsterdam/Maastricht (NL-AmMa) were treated as

separate cohorts; thus, a total of six EPIC cohorts constituted the training sample. Because the definition of the training sample was determined by participation in the DiOGenes project and not by making a random split of the total EPIC cohort, the remaining, but independent and geographically separate, EPIC cohorts could be used to assess the external validity of the prediction model. The remaining EPIC cohorts are hereafter referred to as the (external) validation sample.

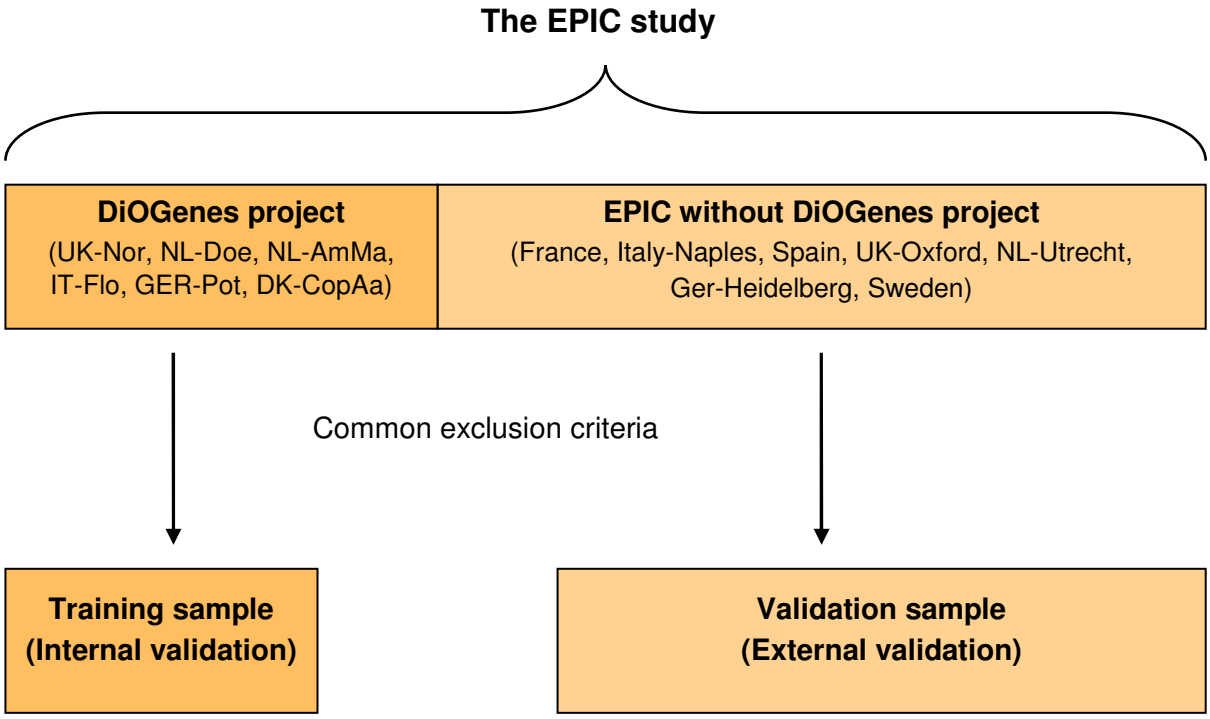


Figure 5. Definition of the training and validation sample

The exclusion criteria applied to both the training and validation sample and corresponding numbers of included participants are displayed in a flow diagram below (**Figure 6**). A number of exclusion criteria were agreed upon among EPIC centres beforehand and were applied in the present analysis. For instance, to reduce the impact of implausible extreme dietary values, subjects who were in the sex- and centre-specific top 1% and bottom 1% of the distribution of the ratio of reported energy intake to estimated energy requirement (energy requirement = basal metabolic rate calculated from age, body weight and height [161] x 1.55) were excluded. Further exclusions refer to participants who provided no or unrealistic information on anthropometrics at either baseline or follow-up or who reported prevalent CVD, diabetes or cancer to avoid influences of these conditions on body weight changes. Additionally, to maintain the same age range in all cohorts and to minimise confounding from changes in body composition and shape occurring in older age [162] or from undiagnosed chronic disease, the present study was restricted to participants aged ≥35 years at baseline and <65 years at follow-up. Since prevention efforts should particularly focus on individuals

at high-risk of (substantial) weight gain to avoid obesity, the present study was restricted to non-obese (BMI<30) individuals. Finally, individuals who had missing values in any of the candidate predictors were excluded. From the 146,543 men and women who took part in the baseline examination across the 6 training cohorts during 1992-1998, data of 47,203 male and female participants were finally used to guide model development. For the validation sample, all individuals participating in any of the six DiOGenes cohorts were additionally excluded in order to obtain two completely independent samples. Also, it may be noteworthy that the complete cohorts of Norway, Greece and Varese (Italy) had to be excluded from the validation sample due to missing information on physical activity or smoking at follow-up. In Ragusa (Italy) and Turin (Italy) the follow-up assessment of body weight was ongoing when the dataset was compiled and therefore data from Ragusa and Turin were not included in the present dataset. The final validation sample thus consisted of 115,099 men and women.

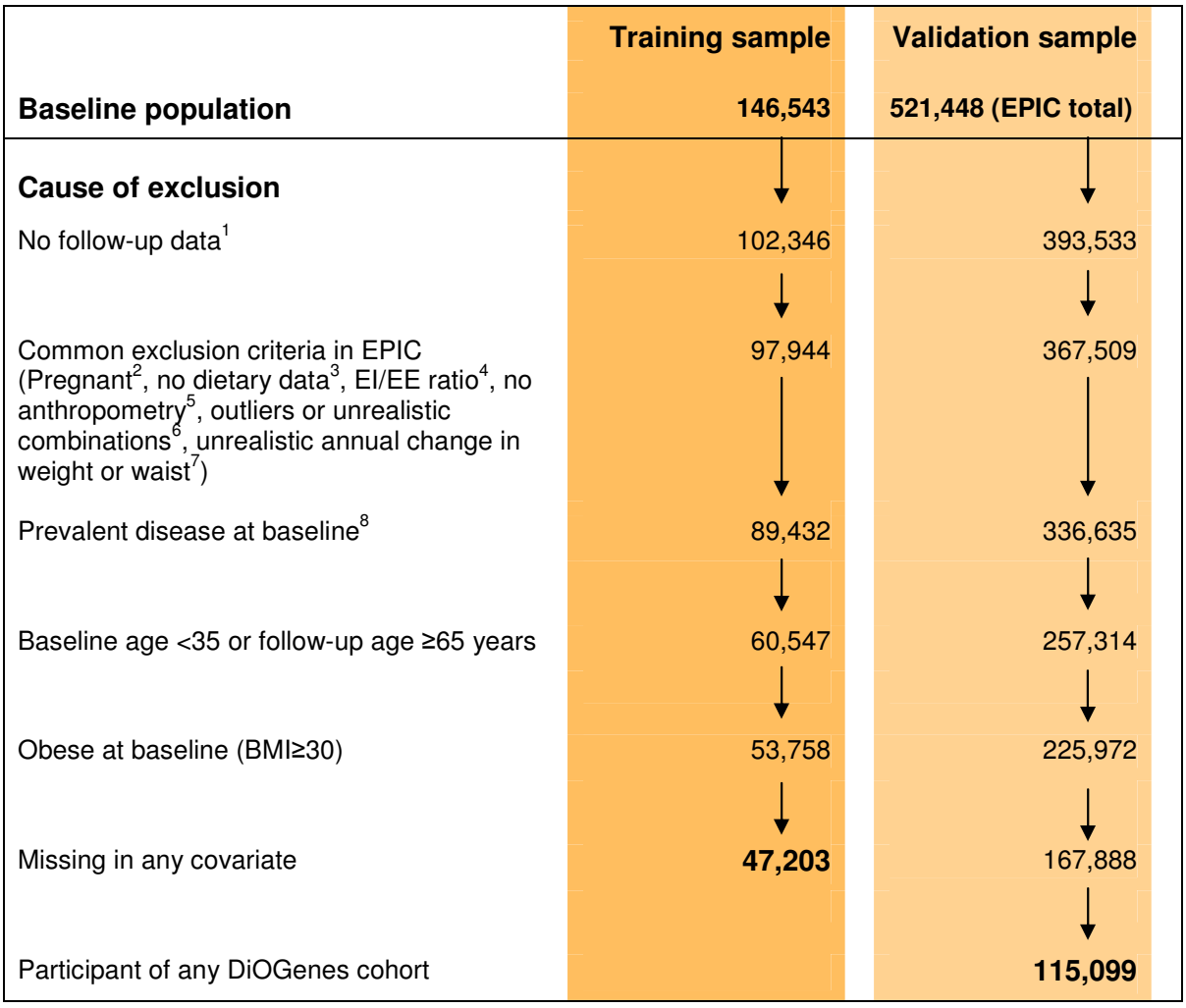


Figure 6. Flow diagram of participants excluded from the present study

[1] No follow-up questionnaire. [2] Pregnant at baseline or follow-up. [3] 10% missing items on FFQ. [4] Ratio of energy intake (EI) to energy expenditure (EE) estimated from predicted resting energy expenditure. [5] Missing data on baseline or follow-up weight, waist or height, missing follow-up time. [6] Baseline height<130 cm, BMI<16 kg/m^2, 0<waist <40 cm, waist>160 cm, follow-up weight>700 kg. Combination of waist<60 cm and BMI>25 kg/m^2. [7] Annual weight change>5 kg (either direction) or annual waist change>7 cm (either direction). [8] Baseline cancer, diabetes or cardiovascular disease.

2.2 Statistical approaches

Statistical analyses were performed using SAS (Statistical Analysis System, version 9.2; SAS Institute Inc, Cary, NC). Meta-analyses were conducted using the package “meta” developed by Schwarzer [163] for R software (version 2.12.1, http://www.r-project.org).

2.2.1 Definition of study variables

2.2.1.1 Case status

In order to develop a risk score predicting absolute risks of SWG within the next 5 years, a meaningful threshold for weight gain representing substantial and, most likely, unhealthy weight gain over 5 years needed to be defined. Most previous studies on a dichotomous endpoint related to weight gain either used established BMI categories and estimated risk of obesity or defined major weight gain on the basis of absolute changes in weight, e.g. 25 kg over 12 years [32] or 15 kg over 16 years of follow-up [164]. However, absolute weight gains are difficult to compare across persons with different initial body weights. Thus, in order to facilitate more fair comparisons between persons, weight change was used on a relative scale and the outcome was defined as weight change in percent of baseline weight (p_i) (Equation 1).

$$p_i = \frac{W_{T_i} - W_{0_i}}{W_{0_i}} \qquad \begin{matrix} , W_{Ti} = \text{weight at follow-up} \\ W_{0i} = \text{weight at baseline} \end{matrix} \qquad \text{(Equation 1)}$$

Substantial weight gain was defined as gaining ≥10% of baseline weight during follow-up ($p \geq 0.1$) which was considered major weight gain with respect to the time period of 5 years the risk score will be tailored to. The threshold was mainly chosen because it is high enough to exclude random variation in body weight and, additionally, allows for some weight gain as natural part of the aging process while identifying those subjects who gain major and possibly unhealthy amounts of weight over a relatively short time period. Given the large variation in follow-up times between individuals, methods of survival analysis appeared most appropriate for statistical analysis. Thus, each participant was followed for incidence of SWG from study entry to the second assessment of body weight (end of follow-up) and all participants gaining ≥10% of their baseline weight during follow-up constituted the set of cases. Those subjects not experiencing SWG within the period of observation were censored at time of their second weight assessment. Because it was only possible to determine case status at the end of follow-up, the exact time point when the threshold of $p \geq 0.1$ was crossed and thus the exact survival time was unknown for the cases. Therefore, the theoretical time point when the threshold of ≥10% baseline-based weight gain had been crossed was

estimated. For this, weight gain was assumed to be linear and the mean annual proportion of baseline-based weight gain was determined (Equation 2).

$$p_{annual\,i} = \frac{W_{T_i} - W_{0_i}}{T_i * W_{0i}} \qquad \text{(Equation 2)}$$

The time period theoretically needed to gain ≥10% of baseline weight ("survival time") among cases was subsequently calculated by modifying the observed follow-up time (Equation 3).

$$\text{If } p_i \geq 0.1 \text{ then} \qquad T_{ci} = \frac{0.1}{p_{annual\,i}} \qquad \text{(Equation 3)}$$

Exemplarily, **Figure 7** illustrates the estimation of the survival time for a case with baseline weight W_0=64kg and follow-up weight after 7 years W_{7years}=73kg. Assuming linear weight gain, this particular case theoretically crossed the threshold of ≥10% baseline-based weight gain after approximately 5 years.

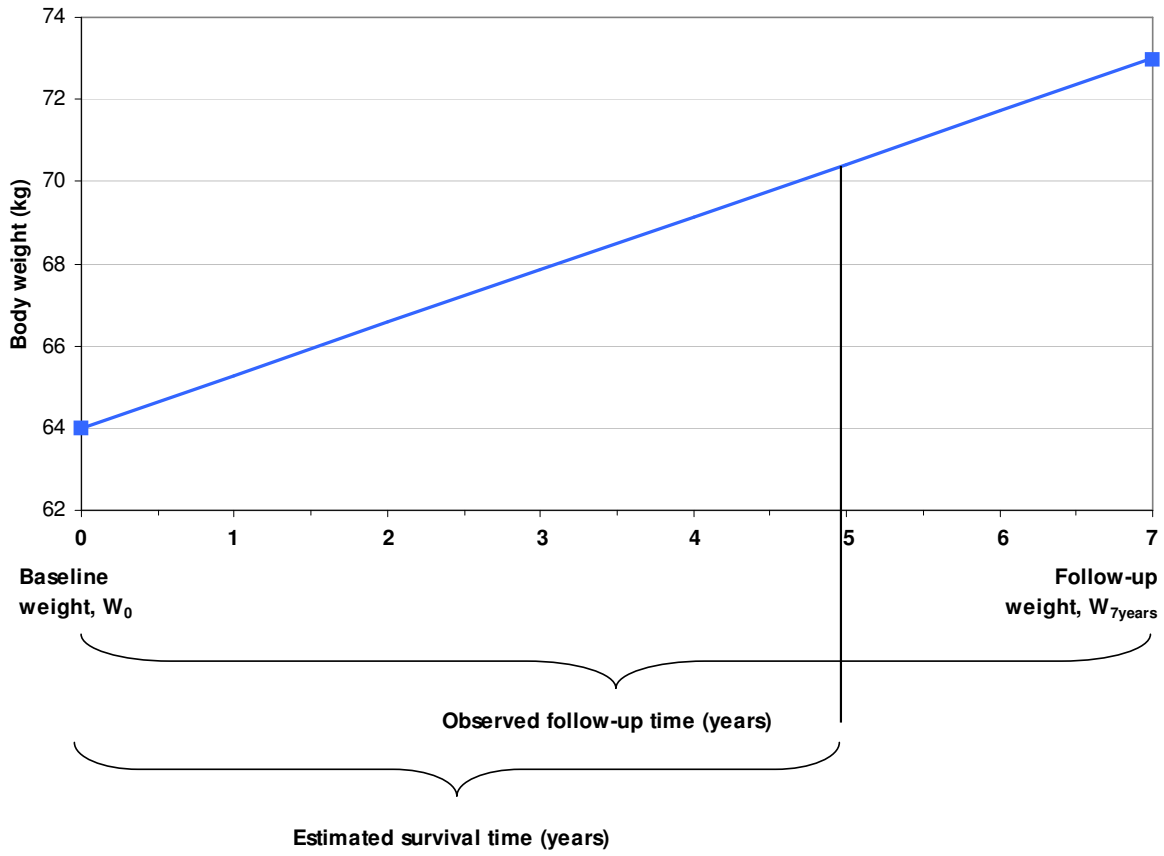


Figure 7. Example for the estimation of survival time among cases

For a hypothetical case with W_0=64kg and W_{7years}=73kg (weight gain in percent of baseline weight = 14%).

2.2.1.2 Potential predictor variables

For this study, a total of 21 variables were selected to be included in the risk prediction model as candidate predictors of SWG (**Table 5**). Selection of candidate predictors was primarily based on reported or hypothesised associations with weight change and risk of overweight or obesity in the literature (see chapter 1.2.5). In view of a possible application of the risk score in practice, candidate predictors should also be easy to assess; therefore, the selection of dietary factors was restricted to main food groups and nutrients were not considered. As available in the data set, the following candidate predictors were finally considered: age, sex, baseline weight and height, sports, occupational physical activity, smoking habits, alcohol consumption, education, and several dietary factors, such as intake of fruits and vegetables, red and processed meat, bread, and soft drinks. All dietary factors were treated continuously per increase of a portion size mainly following the specifications in the FFQ and the actually observed intake values in the study population. Updated information on smoking status was used to classify participants into one of the following four categories: stable non-smokers (never smokers and those who stopped smoking before baseline), stables smokers (those who reported smoking at baseline and follow-up), quitters (those who stopped smoking during follow-up), and started smokers (those who started smoking during follow-up). Smoking was the only variable for which updated information was used in the present study. However, change in smoking status, particularly smoking cessation, is one of the strongest determinants of subsequent weight change that should be accounted for if information is available. The decision to deviate from the otherwise systematic line of action of solely including baseline data was additionally underlined by the finding that analyses exclusively using baseline smoking status showed a positive association between current smoking at baseline and future risk of SWG, which was found to be due to the strong positive effect among those baseline smokers who stopped smoking during follow-up.

Table 5. Candidate predictors to be included in the prediction model building and their corresponding scale

Candidate predictors	Scale
Socio-demographic and anthropometric factors	
Age	continuous (per year)
Sex	Dichotomous
Baseline weight	continuous (per kg)
Baseline height	continuous (per cm)
Education	4 categories: none or primary school, technical school, secondary school, university
Occupational physical activity	4 categories: sedentary, standing, manual, and non-work
Lifestyle factors	
Sports	continuous (per hour/week)
Smoking habits	4 categories: stable non-smoking during follow-up, stable smoking, start smoking, smoking cessation
Alcohol consumption	6 categories: non-consumer, >0-<6g/d, 6-<18 g/d, 18-<30 g/d, 30-<60 g/d, and ≥60 g/d
Dietary factors	
Fruits and vegetables	continuous (per 125g/d)
Red and processed meat	continuous (per 100g/d)
Poultry	continuous (per 50g/d)
Fish	continuous (per 100g/d)
Milk and yogurt	continuous (per 150g/d)
Pasta and rice	continuous (per 50g/d)
Bread	continuous (per 50g/d)
Vegetable oil	continuous (per 20g/d)
Butter and margarine	continuous (per 20g/d)
Chocolate	continuous (per 25g/d)
Cake and cookies	continuous (per 50g/d)
Soft drinks	continuous (per 250g/d)

2.2.2 Descriptive statistics

For descriptive purposes, general characteristics were computed across the training and validation sample. For continuous variables, arithmetic means and standard deviations were presented; proportions were computed for categorical variables. Further, absolute weight gain was calculated across subgroups of the training and validation sample and the incidence of SWG in the two study populations was displayed graphically.

2.2.3 Risk prediction model building

Cox Proportional Hazards regression was used to identify significant predictors for substantial weight gain by estimating relative risks as incidence rate ratios (PROC PHREG in SAS) [165]. In contrast to the logistic regression model which simply considers whether an event occurs, in Cox PH regression time to event is taken into account. In this way, varying follow-up times between individuals could be accounted for, while extrapolation of actually observed weight changes to a fixed non-observed follow-up time was avoided.

Figure 8 illustrates the statistical analysis in a flow-diagram. In terms of practical usefulness, the prediction model may only include relevant predictor variables. Therefore, the first step of the model building encompassed the selection of the most significant predictors to be retained in the final model. For this, the set of candidate predictor variables was entered into a Cox PH regression model using “stepwise selection”, an automatic variable selection procedure implemented in the PHREG procedure in SAS [166]. In this stepwise selection process, variables are entered into and removed from the model by repeated application of a selection-deletion cycle based on a pre-specified level of significance. In the present study, the p-values for entering and staying in the model were both set to 0.1. The stepwise selection process terminates if no further variable can be added to the model or if the variable just entered into the model is the only variable removed in the subsequent backward elimination. Interaction terms were not included in order to keep the model parsimonious and easy to use. In view of the aim to develop a universal risk score predicting SWG among European middle-aged adults based on data from six EPIC cohorts, the prediction model should be independent of any effects of the study centres themselves. In order to account for heterogeneity between centres due to differences in questionnaire design, follow-up procedures, and other non-measured centre effects, stepwise Cox regression was conducted separately by centre. Variables that appeared to be statistically significantly associated with SWG in the same direction in at least two cohorts were retained as predictors for construction of the final prediction model. Ultimately, centre-specific regression coefficients were obtained for all retained predictor variables and random-effects meta-analysis was used to calculate combined estimates by weighting the centre-specific predictor-SWG effects by the inverse of their variances [167]. In contrast to the fixed-effect model, the random-effects

model has the advantage that it allows for variation of effect size between studies [167, 168]. I^2, describing the percentage of variation across centres that is due to heterogeneity rather than chance, with 95% confidence intervals (CI), as well as Cochran's Q-statistic and corresponding p-value were reported as measures to quantify the amount of heterogeneity [169]. The results of these meta-analyses were graphically presented on forest plots using the package "meta" for R software [163].

To estimate absolute risks of experiencing SWG within the following 5 years (**Step II**), the survival function from the Cox PH model was used (Equation 4) [170].

$$S(t, X) = S_0(t)^{\exp(\sum_{i=1}^{m} \beta_i * x_i)} \quad \text{(Equation 4)}$$

, m = number of predictors
X = x_i, …, x_m, set of predictors

This survival function formula says that the estimated survival probability at time t for a subject with a given specification of a set of predictors, denoted by the bold **X**, is given by a baseline survival function $S_0(t)$ raised to a power equal to the exponential of the sum of β_i times x_i [170]. Score points for each predictor were assigned based on the value of the corresponding pooled β-coefficients obtained from separate meta-analyses multiplied by the constant factor of 100 and rounded to two decimal places. For each individual, a risk score (RS_i) was computed as a linear function of those score points. To avoid negative score values, the score was rescaled by adding a constant of 500.

$$RS_i = 500 + \sum_{i=1}^{m} (100 * \beta_i) * x_i \quad \text{(Equation 5)}$$

As can be seen from **Equation 4**, for the calculation of individual survival probabilities S(t,X), and thus the calculation of absolute risks of SWG for the next 5 years, it was necessary to estimate the baseline survival probability of SWG at 5 years $S_0(5y)$. By definition, baseline survival probability refers to the probability of survival (probability of not having gained ≥10% of baseline weight) at a specific time point when all covariates are equal to zero. However, when $S_0(5y)$ was estimated with all covariates equal to zero separately by centre, there was large heterogeneity in the 5-year baseline survival probability across the six training cohorts (**App. Figure 1**) and pooling was impossible. Therefore, the mean value of each predictor over all subjects in the total training sample ($\bar{x}_i$) was chosen to compute the "baseline" survival function at 5 years ($S_M(5y)$) (Equation 6). As for the predictors, centre-specific values of $S_M(5y)$ were computed first and then pooled using random-effects meta-analysis.

$$S_M(5y) = S(5y, \overline{X}) = S_0(t)^{\exp(\sum_{i=1}^{m} \beta_i * \bar{x}_i)} \quad \text{(Equation 6)}$$

Next, the individual risk score points (Equation 5) were inserted into the survival function and the probability of experiencing SWG within the following 5 years was finally calculated by subtracting the survival probability $S(5, RS_i)$ from 1 (Equation 7). Importantly, in order to account for the modification in estimating $S_0(5y)$, the linear predictor in the survival function formula was corrected for the averages of the participants' predictors (RS_M). It can be shown mathematically that this approach and the approach of setting all covariates to zero yield equivalent results in terms of absolute risks (**App. Figure 2**).

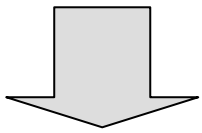
$$P(SWG,5y)_i = 1 - S_M(5y)^{\exp((RS_i - RS_M)/100)} \quad \text{(Equation 7)}$$

$$\text{with} \quad RS_M = 500 + \sum_{i=1}^{m} (100 * \beta_i) * \bar{x}_i$$

Subsequent to the development of the risk score, its predictive performance was evaluated by means of discrimination and calibration in the training sample (internal validation) and in the external validation sample (**Step III**). The discriminatory accuracy of the risk score was assessed by plotting the ROC curve and computing the corresponding c-statistic. 95% confidence intervals for the c-statistic were estimated according to the algorithm developed by DeLong and colleagues [171] using the macro provided by the SAS support page on the internet (http://support.sas.com/kb/25/017.html, last access: May 23rd, 2011). For determining an appropriate cut-off value of the score, the Youden's index was used. Calibration of the model was evaluated by comparing the observed incidence of SWG within the first 5 years of follow-up with the corresponding predicted probability across meaningful categories of predicted risk. Additionally, a graphical presentation was chosen to best capture the intuitive meaning of calibration by plotting observed versus predicted risk across deciles of predicted risk. As a formal test, the Hosmer-Lemeshow Test (HLT) was performed.

Step I: Selection of predictor variables

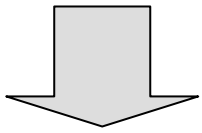
a) Identification of predictors for substantial weight gain (SWG) separately by centre in the six cohorts of the training sample using proportional hazards regression (stepwise, slentry=0.1, slstay=0.1)
b) Selection of common predictor variables across the six centres (to be retained variables should be predictors in at least two centres; association in the same direction)
c) Estimation of centre-specific regression coefficients for each retained predictor and pooling of these estimates using a random-effects meta-analytical approach



Step II: Construction of the risk score and estimation of absolute risks

d) Estimation of the background survival probability for the following 5 years given all covariates take on average values of risk factors (S_M (5y))
e) Calculation of the score points as a linear function of the predictors weighted by the regression coefficients (multiplied by 100)
f) Estimation of the absolute risk of SWG during the following 5 years after insertion the individual score (RS_i) into a survival function corrected for the averages of participants' values (RS_M):

$$P(SWG,5y)_i = 1 - S_M(5y)^{\exp((RS_i - RS_M)/100)}$$



Step III: Evaluation of the predictive performance of the risk score

g) Internal validation (Discrimination, Calibration)
h) External validation (Discrimination, Calibration)

Figure 8: Overview of the statistical analysis to develop a risk score predicting substantial weight gain in the multi-centre EPIC study

Substantial weight gain was defined as gaining ≥10% of baseline weight during follow-up.

3 Results

3.1 Description and comparison of the study populations

3.1.1 General characteristics

General characteristics for each cohort of the training sample and the total validation population are presented in **Table 6**. Cohort-specific results for the validation set can be found in the appendix (**App. Table 2**).

A total of 47,203 subjects were included in the training sample, 41% of whom were men. The largest contribution in terms of participants came from Denmark already constituting half (51%) of the total training sample, followed by GER-Pot (19%) and IT-Flo (12%). Mean age at baseline was 50.4 years, with participants from the Netherlands being the youngest on average and the Danish people being the oldest. Due to logistical reasons, follow-up time differed considerably between centres with an average (±SD) follow-up of 6.2±2 years, ranging from an average of 3.6 years in UK-Nor to an average of 8.7 years in IT-Flo (total variation: 1.2–12.2 years). Mean annual weight gain was 387g, with the lowest annual weight gain observed in the Danish cohort and the largest annual weight gain found in UK-Nor. On average, subjects gained 3.7% of their baseline weight during follow-up, representing a mean annual proportion of baseline-based weight gain of 0.6%. The overall percentage of obese individuals at follow-up mounts up to 4.9%. Almost one third of all individuals had a university degree, varying from 18% to 44% across cohorts. About 7% of all subjects stopped smoking during follow-up, whereas the proportion varied according to cohort from 2% in UK-Nor to 13% in NL-AmMa. Among baseline smokers only, 24% quit smoking on average, ranging between 19% in NL-Doe and 37% in NL-AmMa (data not shown). With regard to dietary habits, intake of fruits and vegetables, pasta and rice and vegetable oil was substantially higher in the Italian cohort than in all other cohorts, reflecting a typical Mediterranean dietary pattern. The highest consumption of meat was found in the Danish, Dutch and German cohorts, consumption of soft drinks was highest in Denmark, the Netherlands and the UK-Nor.

In comparison to the training sample, the proportion of men was substantially lower in the validation sample (19.5 vs. 41.1%) which was due to the cohorts of France, IT-Naples and NL-Utrecht that only consisted of women. The cohorts of France, Spain and GER-Heidelberg contributed the largest proportion of subjects with 40%, 21% and 13%, respectively. The average (±SD) duration of follow-up was 3.6±1.2 years and thus considerable lower than in the training population. Conversely, mean annual weight gain was almost 150g higher in the validation sample. With regard to lifestyle and dietary factors, some differences between the two study populations became apparent. First, subjects in the validation set were more likely

to have a higher education and more likely to be a constant non-smoker during follow-up than individuals in the training set. Second, the proportion of non- and low-consumers of alcohol was substantially higher in the validation population than in the training set. Finally, while mean intake of fruits and vegetables was considerably higher in the validation than in the training sample, the intake of red and processed meat was lower. The average consumption of soft drinks in the validation population was less than half of the average consumption in the training set, which was attributable to the cohorts of France, IT-Naples and Spain showing a very low consumption of sugar-sweetened beverages (9–26g/d) while the consumption in the other cohorts varied between 60 and 113g/d.

Table 6. General characteristics of the training and validation population

	Training population							Validation population
	UK-Nor	NL-Doe	NL-AmMa	IT-Flo	GER-Pot	DK-CopAa	All	
N	3,930	2,088	2,327	5,512	9,104	24,242	47,203	115,099
Men (%)	42.3	49.0	44.4	22.5	35.5	46.2	41.1	19.5
Age at baseline (y)	51.5 (5.2)	45.9 (6.7)	45.3 (5.8)	47.5 (5.7)	45.2 (6.3)	53.7 (2.6)	50.4 (5.9)	49.5 (6.3)
Duration of follow-up (y)	3.61 (0.81)	4.89 (0.46)	8.58 (1.8)	8.74 (1.8)	7.97 (1.5)	5.22 (0.53)	6.18 (1.99)	3.56 (1.15)
Anthropometry								
Weight								
Baseline (kg)	69.5 (11.1)	73.8 (11.5)	71.1 (11.4)	64.5 (10.3)	69.3 (11.4)	72.4 (11.4)	70.6 (11.5)	65.0 (10.4)
Follow-up (kg)	71.2 (11.6)	76.0 (12.1)	75.2 (12.4)	68.2 (11.4)	73.2 (12.2)	74.2 (12.0)	73.2 (12.1)	66.9 (11.0)
Absolute change (kg)	1.67 (3.60)	2.23 (4.1)	4.10 (5.5)	3.63 (4.9)	3.84 (5.0)	1.82 (4.1)	2.54 (4.5)	1.94 (3.7)
Annual change (g/y)	469.4 (1019)	444.7 (809)	440.0 (598)	385.0 (518)	447.7 (583)	340.3 (771)	386.5 (734)	522.6 (1054)
% Change	2.48 (5.2)	3.10 (5.6)	5.97 (8.0)	5.79 (7.7)	5.74 (7.4)	2.61 (5.7)	3.76 (6.6)	3.05 (5.6)
BMI								
Baseline (kg/m²)	24.5 (2.6)	24.6 (2.6)	24.1 (2.8)	24.0 (2.7)	24.4 (2.8)	24.7 (2.7)	24.5 (2.7)	24.0 (2.8)
Follow-up (kg/m²)	25.1 (2.9)	25.4 (2.9)	25.5 (3.2)	25.4 (3.2)	25.8 (3.2)	25.3 (2.9)	25.4 (3.0)	24.7 (3.1)
Obese at follow-up (%)	5.8	6.9	6.6	6.2	6.7	3.4	4.9	4.8
Physical activity								
At Work (%)								
Sedentary	30.6	32.5	42.5	46.1	57.9	43.3	44.9	32.0
Standing	22.3	20.2	16.9	18.5	28.3	19.5	21.1	38.5
Manual	22.4	18.3	15.6	9.2	4.8	26.6	19.1	6.6
Non-workers	24.7	29.0	25.0	26.2	9.1	10.6	14.8	22.9
Sports (hours/week)	1.48 (2.6)	1.79 (2.3)	1.95 (2.7)	1.30 (2.1)	1.08 (1.8)	1.45 (2.1)	1.40 (2.1)	1.39 (2.2)
Education (%)								
No school / primary school	24.2	6.1	6.5	35.3	9.2	25.6	21.7	24.1
Techn. / profess. school	48.5	44.9	31.1	12.9	40.1	41.2	37.9	14.4
Secondary school	9.2	25.0	24.7	29.2	7.2	11.5	13.8	27.9
University degree	18.1	24.0	37.7	22.6	43.5	21.7	26.6	33.6

Table 6 continued.

	Training population							Validation population
	UK-Nor	NL-Doe	NL-AmMa	IT-Flo	GER-Pot	DK-CopAa	All	
Smoking habits (%)								
Non-smokers	89.4	67.4	62.7	69.8	78.4	64.2	69.7	79.4
Stable smokers	7.7	24.4	22.4	19.2	14.1	26.5	21.4	14.2
Started smokers	0.8	2.5	1.9	1.8	2.0	2.1	2.0	3.1
Quitters	2.0	5.7	13.1	9.3	5.6	7.2	6.9	3.3
Alcohol use (%)								
No alcohol	14.7	9.3	6.6	10.1	2.1	1.5	4.3	14.6
> 0 - ≤ 6g/d	38.3	35.9	33.2	38.6	37.2	18.5	27.6	33.4
> 6 - ≤ 18g/d	33.8	30.1	28.5	23.2	33.2	38.0	34.2	28.3
> 18 - ≤ 30g/d	5.9	12.5	15.7	14.3	14.8	14.2	13.7	11.6
> 30 - ≤ 60g/d	6.6	10.6	12.7	11.3	10.1	20.8	15.6	9.4
> 60g/d	0.6	1.6	3.4	2.5	2.6	6.8	4.6	2.6
Dietary factors (g/d)								
Fruits and vegetable	476.0 (218)	318.4 (138)	312.6 (149)	509.3 (229)	289.5 (134)	364.4 (199)	371.6 (202)	504.8 (268)
Red and processed meat	64.9 (41)	103.5 (49)	94.4 (55)	79.2 (42)	98.2 (57)	104.1 (51)	96.3 (52)	74.2 (50)
Poultry	27.0 (20)	10.5 (9)	12.7 (12)	26.9 (20)	11.8 (11)	22.0 (18)	20.0 (18)	20.7 (21)
Fish	36.5 (25)	10.1 (9)	11.7 (12)	29.7 (21)	22.3 (23)	41.9 (24)	33.3 (25)	37.5 (33)
Milk and yogurt	387.7 (177)	349 (263)	296.0 (285)	188.8 (173)	180.9 (208)	325.6 (290)	286.4 (264)	247.8 (207)
Pasta and rice	44.3 (41)	54.5 (45)	63.9 (57)	167.5 (112)	17.7 (15)	49.6 (38)	57.7 (67)	62.7 (47)
Bread	88.0 (58)	153.9 (65)	153.1 (73)	160.2 (90)	179.3 (79)	147.2 (66)	150.5 (75)	125.9 (78)
Vegetable oil	4.3 (3)	2.9 (3)	4.6 (4)	31.2 (14)	4.2 (3)	2.7 (4)	6.6 (11)	9.7 (11)
Butter and margarine	20.9 (16)	25.4 (16)	22.2 (16)	2.3 (4)	26.8 (17)	20.8 (15)	20.1 (16)	12.9 (15)
Chocolate	13.3 (17)	9.8 (12)	10.2 (12)	4.3 (8)	12.4 (15)	8.0 (12)	9.0 (13)	8.5 (16)
Cake and cookies	68.0 (62)	30.3 (23)	27.0 (22)	51.7 (49)	61.0 (56)	20.1 (21)	36.5 (43)	43.3 (43)
Soft drinks	121.7 (195)	110 (129)	124 (147)	23.0 (70)	46.4 (135)	131.0 (227)	100.0 (194)	41.0 (119)

DK-CopAa = Denmark (Copenhagen and Aarhus) cohort; GER-Pot = Germany (Potsdam) cohort; IT-Flo = Italy (Florence) cohort; NL-AmMa = the Netherlands (Amsterdam and Maastricht) cohort; NL-Doe = Doetinchem cohort; UK-Nor = United Kingdom (Norfolk) cohort. Data are means (SD) or percentages.

3.1.2 Weight gain across subgroups of the study populations

Mean annual weight gain according to subgroups of the training and validation population is presented in **Table 7**. In both training and validation set, mean absolute weight gain per year was higher among men and decreased with increasing duration of follow-up. Further, weight gain constantly decreased across groups of age and baseline BMI. Large differences were found across categories of smoking habits. Weight gain was twice as large among individuals who stopped smoking than among stable non-smokers.

Table 7. Annual weight gain across subgroups of the training and validation set

	Training population		Validation population	
		Weight gain (g/y) *		Weight gain (g/y) *
All, N	47,203	386.5 (734)	115,099	522.6 (1054)
	%		%	
Sex				
Men	41.1	414.5 (738)	19.5	611.8 (1145)
Women	58.9	367.0 (730)	80.5	500.9 (1029)
Duration of follow-up (y)				
<3	3.3	1383.3 (1472)	29.0	788.1 (1489)
3 – 5	14.5	792.5 (943)	59.0	421.0 (820)
5 – 8	59.1	289.4 (635)	11.9	382.1 (588)
≥8	23.0	237.5 (386)	0.2	285.1 (303)
Baseline age group (y)				
35 – < 45	18.3	501.3 (655)	21.7	515.1 (1128)
45 – < 55	56.2	377.3 (741)	53.6	543.4 (1036)
55 – < 65	25.5	324.5 (760)	24.7	483.9 (1022)
Baseline BMI category				
<23 kg/m²	30.7	442.9 (613)	52.8	513.3 (954)
23 – <27 kg/m²	49.1	390.1 (728)	31.6	576.1 (1085)
27 – <30 kg/m²	20.2	292.3 (889)	15.7	446.4 (1280)
Physical activity at work				
Sedentary	44.9	381.7 (704)	32.0	543.9 (1078)
Standing	21.1	397.1 (713)	38.5	507.7 (1018)
Manual	19.1	389.6 (771)	6.6	680.5 (1070)
Non-workers	14.8	381.9 (797)	22.9	471.8 (1069)
Education				
No school/primary school	21.7	399.3 (792)	24.1	507.1 (1117)
Techn./profess. school	37.9	389.1 (760)	14.4	610.7 (1115)
Secondary school	13.8	369.6 (687)	27.9	513.6 (1006)
University degree	26.6	381.2 (666)	33.6	503.3 (1016)
Smoking habits				
Non-smokers	69.7	368.9 (697)	79.4	514.9 (1021)
Stable smokers	21.4	333.1 (759)	14.2	497.7 (1133)
Started smokers	2.0	173.7 (820)	3.1	361.9 (1097)
Quitters	6.9	789.3 (850)	3.3	965.3 (1292)

* All values are arithmetic means (SD).

3.1.3 Incidence of substantial weight gain

During an average follow-up of 6.2 years (291,748 person-years, PY), a total of 6,471 men and women gained ≥10% of baseline weight in the training population. In the validation sample, a total of 10,785 participants experienced substantial weight gain during a mean follow-up of 3.6 years (409,909 PY). Incidence rates (per 10,000 PY) for the training and validation population are shown in **Figure 9**. The highest incident rates were observed in the cohorts of NL-AmMa, IT-Flo and GER-Pot, while the lowest was found in the Danish cohort. However, it has to be noted that despite rigorous exclusion criteria in respect of age, the age distribution differs in part between study centres, e.g. the Danish cohort only included subjects aged 50 to 65 years at recruitment. Because weight gain is related to age (see **Table 6**), this may explain the comparably low incidence of substantial weight gain observed in the Danish cohort. Also, none of the cohorts is a representative sample of the underlying source population and direct comparison of country-specific crude incidence rates may warrant caution. Incidence of substantial weight gain was consistently higher in women than in men, for instance, 262 vs. 162 per 10,000 PY in the total training sample. Overall, the largest contribution of cases was from GER-Pot (32%), DK-CopAa (31%) and IT-Flo (20%).

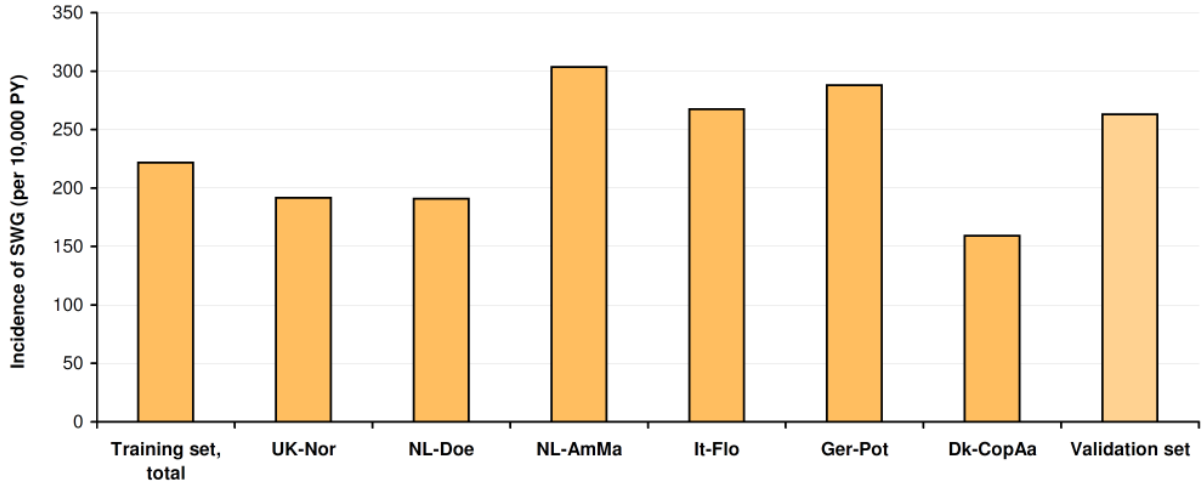


Figure 9. Incidence rates of substantial weight gain (per 10,000 person-years) across cohorts of the training and in the validation sample

Substantial weight gain was defined as gaining ≥10% of baseline weight during follow-up.

3.2 Risk score for substantial weight gain

3.2.1 Selection of predictor variables and computation of the risk score

To account for heterogeneity between centres with respect to the effect size of some predictors, stratified Cox regression models were conducted (**App. Table 3**) and variables that were predictive of SWG in at least two cohorts in the same direction were retained for the final model. The final prediction model was then defined on the basis of the following factors: age, sex, baseline body weight and height, technical school, secondary school, university, cessation of smoking, sports, non-consumption of alcohol, moderate consumption of alcohol, intake of fruits and vegetables, red and processed meat, poultry, bread, butter and margarine, cake and cookies, and soft drinks. This final regression model was run in each of the six cohorts and the combined effect across cohorts was estimated using random-effects meta-analysis. Exemplarily, the forest plots for the relation of sports and intake of red and processed meat with risk of SWG are displayed in **Figure 10**. The remaining graphs can be found in the appendix (**App. Figure 3**). Heterogeneity was found between predictors and study centre, as observed by the range in hazard ratios (HR) among centres. For instance, the HRs ranged from 1.07 in UK-Nor to 1.46 in NL-Doe for each daily increase in red and processed meat intake of 100g. However, confidence intervals did overlap for most, though not all, predictors.

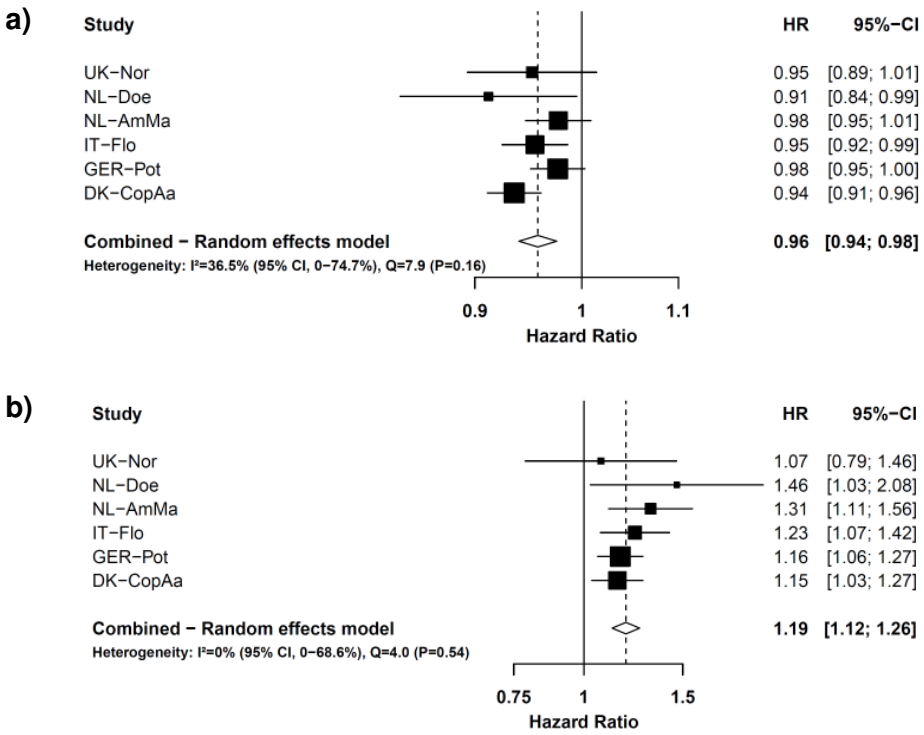


Figure 10. Association of a) sports (per hour/week) and b) intake of red and processed meat (per 100g/d) with risk of substantial weight gain across cohorts of the training sample and combined

The pooled estimates of relative risk for the association of all retained predictors with risk of SWG are presented in **Table 8**. Baseline age and weight as well as education, sports and alcohol consumption were inversely related to risk of SWG. Women had higher risks of gaining ≥10% of baseline weight than men and individuals who stopped smoking were also at higher risk of SWG in comparison to all other categories of smoking habits. With respect to diet, intake of fruits and vegetables, bread, butter and margarine as well as cake and cookies showed an inverse association with risk of SWG, while higher intake of meat and soft drinks was related to higher risks.

Table 8. Pooled estimates of relative risk from random-effects meta-analyses for the association of retained predictors with substantial weight gain in the training sample*

Predictor	β	Hazard Ratio (95% CI)	Points allocated
Age (years)	-0.035731	0.96 (0.96 – 0.97)	-3.57
Sex (Female vs. Male)	0.208152	1.23 (1.07 – 1.41)	20.82
Baseline weight (kg)	-0.022348	0.98 (0.97 – 0.99)	-2.23
Baseline height (cm)	0.003793	1.004 (0.99 – 1.02)	0.38
Technical school	-0.174234	0.84 (0.78 – 0.90)	-17.42
Secondary school	-0.196867	0.82 (0.68 – 0.99)	-19.69
University	-0.319379	0.73 (0.63 – 0.84)	-31.94
Smoking cessation	1.090748	2.98 (2.21 – 4.01)	109.07
Sports (h/week)	-0.043325	0.96 (0.94 – 0.98)	-4.33
No alcohol	0.159991	1.17 (1.03 – 1.34)	16.00
Alcohol >6 to ≤ 18g/d	-0.210474	0.81 (0.74 – 0.89)	-21.05
Alcohol >18 to ≤ 30g/d	-0.242326	0.78 (0.66 – 0.94)	-24.23
Alcohol >30 to ≤ 60g/d	-0.251157	0.78 (0.64 – 0.95)	-25.12
Fruits and vegetables (per 125g/d)	-0.030047	0.97 (0.95 – 0.99)	-3.00
Red and processed meat (per 100g/d)	0.172187	1.19 (1.12 – 1.26)	17.22
Poultry (per 50g/d)	0.182238	1.20 (1.11 – 1.29)	18.22
Bread (per 50g/d)	-0.036353	0.96 (0.95 – 0.98)	-3.64
Butter and margarine (per 20g/d)	-0.034902	0.97 (0.90 – 1.04)	-3.49
Cake and cookies (per 50g/d)	-0.065713	0.94 (0.88 – 0.995)	-6.57
Soft drinks (per 250g/d)	0.093126	1.10 (1.06 – 1.14)	9.31

*Substantial weight gain was defined as ≥10% of baseline weight during follow-up.

Next, β-coefficients were used to assign score points for each variable and, for each individual, a risk score (RS_i) was calculated as a linear combination of the single predictors according to the formula below.

$$
\begin{aligned}
RS_i = \; & 500 \text{ (constant)} \\
& - 3.57 * \text{ age (y)} \\
& + 20.82 * \text{ sex (f vs. m)} \\
& - 2.23 * \text{ weight (kg)} \\
& + 0.38 * \text{ height (cm)} \\
& - 17.42 * \text{ technical school} \\
& - 19.69 * \text{ secondary school} \\
& - 31.94 * \text{ university} \\
& + 109.07 * \text{ smoking cessation} \\
& - 4.33 * \text{ sports (hours/week)} \\
& + 16.00 * \text{ no alcohol consumption} \\
& - 21.05 * \text{ alcohol consumption (>6 to ≤18g/d)} \\
& - 24.23 * \text{ alcohol consumption (>18 to ≤30g/d)} \\
& - 25.12 * \text{ alcohol consumption (>30 to ≤60g/d)} \\
& - 3.00 * \text{ fruits and vegetables (per 125 g/d)} \\
& + 17.22 * \text{ red and processed meat intake (per 100g/d)} \\
& + 18.22 * \text{ poultry (per 50g/d)} \\
& - 3.64 * \text{ bread (per 50g/d)} \\
& - 3.49 * \text{ butter/margarine (per 20g/d)} \\
& - 6.57 * \text{ cake and cookies (per 50g/d)} \\
& + 9.31 * \text{ soft drinks (per 250g/d)}
\end{aligned}
$$

The pooled estimate of the background survival probability at 5 years estimated at average values of the predictors was 0.9413. This means that the “average” individual had a probability of 94% to survive the following 5 years without gaining ≥10% of his/her baseline body weight. Conversely, the 5-year risk of experiencing SWG was 6% for this hypothetical subject. The corresponding risk score points were 208.65 points. For each individual, the probability of ≥10% baseline-based weight gain within the next 5 years [P(SWG,5y)] was then calculated by inserting the individual’s risk score into the following survival function while correcting for the averages of the participants’ risk factors:

$$P\,(SWG,5y)_i = 1 - 0.9413^{\exp((RS_i - 208.65)/100)}$$

3.2.2 Description of the risk score

Figure 11 shows the absolute risk of substantial weight gain over 5 years across the whole range of observed risk score points in the training sample. A risk score of 100 and 400 points corresponded to a 5-year absolute risk of 2.0% and 33.6%, respectively. The lowest score observed in the training population was 9 (0.1% risk), while the highest was 450 (49% risk). The relative risks (95% CIs) of SWG according to quintiles of the risk score with the first quintile as the reference category were 1.67 (1.49–1.87), 2.07 (1.85–2.30), 2.71 (2.44–3.01), and 4.23 (3.84–4.67), respectively (data not shown). The corresponding cut-off points for the risk score across quintiles were <165, ≥165–<193, ≥193–<218, ≥218–<249, and ≥249, respectively.

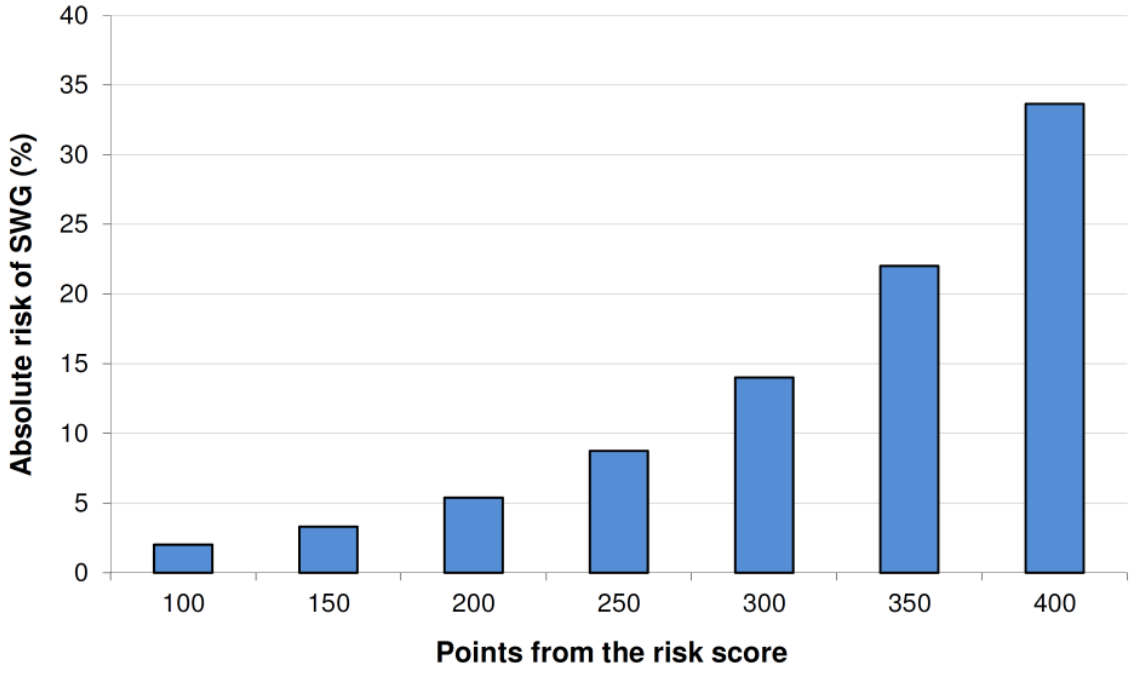


Figure 11. Absolute risk of substantial weight gain according to points from the score
Substantial weight gain was defined as gaining ≥10% of baseline weight during follow-up.

An illustrative example for the calculation of the risk score and the corresponding 5-year risk of SWG for two hypothetical individuals is given in **Table 9**. Individual 1 is a slightly overweight man, 39 years old, who obtained a technical school degree, gives up smoking, does no sports and follows a diet high in meat (225 g/day) and low in vegetables (125 g/d) and bread (50 g/d). For this hypothetical person, the probability of experiencing substantial weight gain within the following 5 years was 31.9%. In contrast, the second individual, same age and sex, is a stable non-smoker, has a university degree, exercises 2.5 hours per week and follows a diet high in fruits and vegetables (500 g/d) and bread (150 g/d), but low in meat

(50 g/d). The person has a risk of 5.5% to gain ≥10% of his current body weight during the next 5 years.

Table 9. Example for the calculation of the individual risk of experiencing substantial weight gain within the following 5 years for two hypothetical individuals*

Predictor	Points from the risk score	Individual 1		Individual 2	
		Value †	Individual points ‡	Value †	Individual points ‡
Constant			500		500
Age (years)	-3.57	39	-139.2	39	-139.2
Sex (Female vs. Male)	20.82	0	0	0	0
Baseline weight (kg)	-2.23	72	-160.6	64	-142.7
Baseline height (cm)	0.38	170	64.6	169	64.2
Technical school	-17.42	1	-17.4	0	0
Secondary school	-19.69	0	0	0	0
University	-31.94	0	0	1	-31.9
Smoking cessation	109.07	1	109.1	0	0
Sports (h/week)	-4.33	0	0	2.5	-10.8
No alcohol	16.00	0	0	0	0
Alcohol >6 to ≤ 18g/d	-21.05	0	0	1	-21.1
Alcohol >18 to ≤ 30g/d	-24.23	0	0	0	0
Alcohol >30 to ≤ 60g/d	-25.12	0	0	0	0
Fruits and vegetables (per 125g/d)	-3.00	1	-3.0	4.0	-12.0
Red and processed meat (per 100g/d)	17.22	2	34.4	1	8.6
Poultry (per 50g/d)	18.22	0.5	9.1	0	0
Bread (per 50g/d)	-3.64	1.5	-3.6	3.0	-10.9
Butter/margarine (per 20g/d)	-3.49	0	0	0.1	-0.3
Cake and cookies (per 50g/d)	-6.57	0	0	0.3	-2.0
Soft drinks (per 250g/d)	9.31	0	0	0	0
Sum of risk score points (RS)			393.4		201.9
5-year absolute risk of SWG (%)¶			**31.9**		**5.5**

* Substantial weight gain (SWG) was defined as gaining ≥10% of baseline weight during follow-up.
† For dietary factors, the value equals the number of portions of the defined size.
‡ Individual points were calculated by multiplying the points from the risk score with the value of the respective variable.

¶ 5-year risk of SWG calculated using the following formula: $P(SWG,5y) = 1 - 0.9413^{\exp((RS - 208.65)/100)}$

Table 10 depicts absolute risks of SWG across various categories of predictors using cross-classification of sex, age group and categories of sports and intake of red and processed meat. The remaining predictors in the score were held constant at mean values (continuous variables) or reference category (categorical variables). Apart from absolute risks of SWG according to different exposure strata, changes in risk with changing exposure can be read from this table. For instance, a man, aged 40 years, doing no sports, and consuming 150g of red and processed meat per day had a 10.8% risk of experiencing SWG within the next 5 years. If he reduced meat intake by 50g/d and started doing sports for 2 hours/week, he would lower his risk to 9.2%, indicating a difference in risk of 1.6 percentage point, with all other predictors held constant. Overall, the impact of changes in these two modifiable risk factors was rather low compared to the decreasing risk accompanied by the natural process of aging. Women had generally higher risks than men in all combinations of risk factors.

Table 10. Predicted 5-year risk of substantial weight gain for men and women across categories of age, sport and red and processed meat intake*

		Men						Women					
		Red and processed meat intake (g/d)					Age (y)	Red and processed meat intake (g/d)					
		25	50	100	150	200		25	50	100	150	200	
Sports (h/wk)	0	8.8	9.2	10.0	10.8	11.8		10.8	11.2	12.2	13.2	14.3	
	1	8.5	8.8	9.6	10.4	11.3		10.3	10.8	11.7	12.7	13.7	
	2	8.1	8.5	9.2	10.0	10.8	40	9.9	10.3	11.2	12.2	13.2	
	3	7.8	8.1	8.8	9.6	10.4		9.5	9.9	10.8	11.7	12.7	
	4	7.5	7.8	8.5	9.2	10.0		9.1	9.5	10.3	11.2	12.2	
	0	6.3	6.5	7.1	7.7	8.4		7.7	8.0	8.7	9.4	10.2	
	1	6.0	6.3	6.8	7.4	8.0		7.4	7.7	8.3	9.0	9.8	
	2	5.8	6.0	6.5	7.1	7.7	50	7.1	7.4	8.0	8.7	9.4	
	3	5.5	5.8	6.3	6.8	7.4		6.8	7.1	7.7	8.3	9.0	
	4	5.3	5.5	6.0	6.5	7.1		6.5	6.8	7.4	8.0	8.7	
	0	4.4	4.6	5.0	5.5	5.9		5.4	5.7	6.2	6.7	7.3	
	1	4.2	4.4	4.8	5.2	5.7		5.2	5.4	5.9	6.4	7.0	
	2	4.1	4.2	4.6	5.0	5.5	60	5.0	5.2	5.7	6.2	6.7	
	3	3.9	4.1	4.4	4.8	5.2		4.8	5.0	5.4	5.9	6.4	
	4	3.7	3.9	4.2	4.6	5.0		4.6	4.8	5.2	5.7	6.1	

* Substantial weight gain (SWG) was defined as gaining ≥10% of baseline weight during follow-up. All other predictors were held constant at mean values (continuous variables) or reference category (categorical variables).

Risk of SWG
< 4
4.1 - < 6
6.1 - < 8
8.1 - < 10
≥ 10

3.2.3 Predictive performance of the risk score

3.2.3.1 Internal validation

The discriminatory ability of the score assessed by the aROC (95% CI) was 0.67 (0.66–0.68) in the whole training sample (**Figure 12**). This implies that individuals who experienced SWG within 5 years of follow-up had higher predicted risks than persons not experiencing SWG approximately 67% of the time. The discriminatory accuracy showed considerable variation across single study centres of the training sample. In UK-Nor, NL-Doe, NL-AmMa, IT-Flo, GER-Pot, and DK-CopAa, the area under the ROC curve (95% CI) was 0.65 (0.62–0.68), 0.70 (0.66–0.74), 0.76 (0.72–0.80), 0.66 (62–0.69), 0.69 (0.66–0.71), and 0.69 (0.67–0.70), respectively. In addition to between-cohort differences, the score generally performed better among men than among women; specifically, in the total training sample, the aROCs (95% CI) were 0.71 (0.69–0.73) and 0.63 (0.62–0.64) for men and women, respectively. In UK-Nor, NL-Doe, NL-AmMa, IT-Flo, GER-Pot, and DK-CopAa, the aROCs (95% CI) for men vs. women were 0.65 (0.58–0.73) vs. 0.58 (0.54–0.62), 0.71 (0.64–0.79) vs. 0.64 (0.58–0.69), 0.76 (0.67–0.85) vs. 0.72 (0.67–0.77), 0.62 (0.50–0.74) vs. 0.61 (0.59–0.66), 0.73 (0.69–0.78) vs. 0.62 (0.60–0.65), and 0.72 (0.70–0.74) vs. 0.65 (0.63–0.67), respectively.

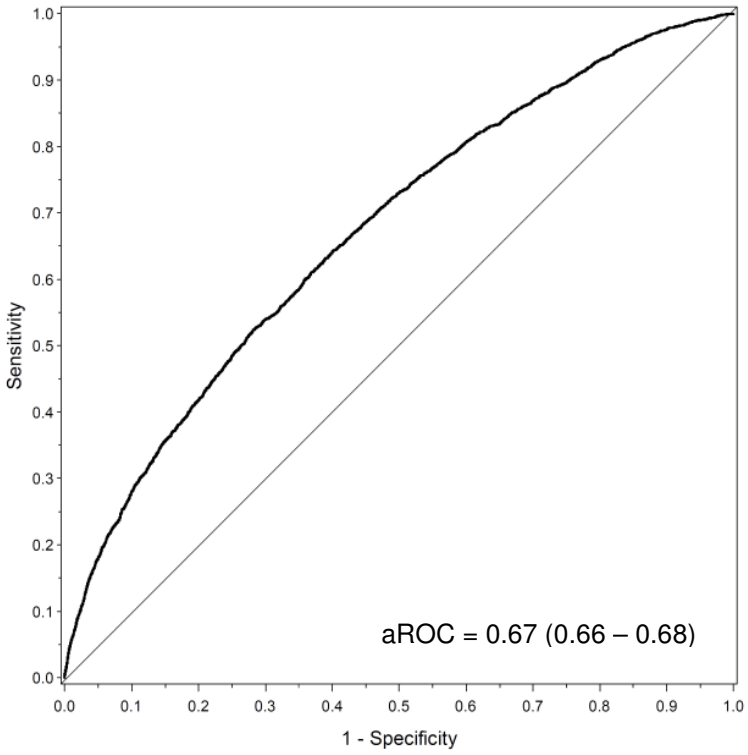


Figure 12. Receiver operating characteristic curve for the prediction of substantial weight gain (gaining ≥10% of baseline weight) over 5 years in the training sample

The proportion of incident cases of SWG and non-cases across categories of the risk score is illustrated in **Figure 13**. In case of perfect discrimination, all cases would be assigned a

higher score value than non-cases and two distinct distributions would be obtained. In line with the ROC curve, the figure shows that the score's ability to distinguish between cases and non-cases was moderate. Specifically, the more one gets to the extreme values of the score, the better the discrimination was while in the middle range of the score, the score was not able to adequately discriminate cases and non-cases.

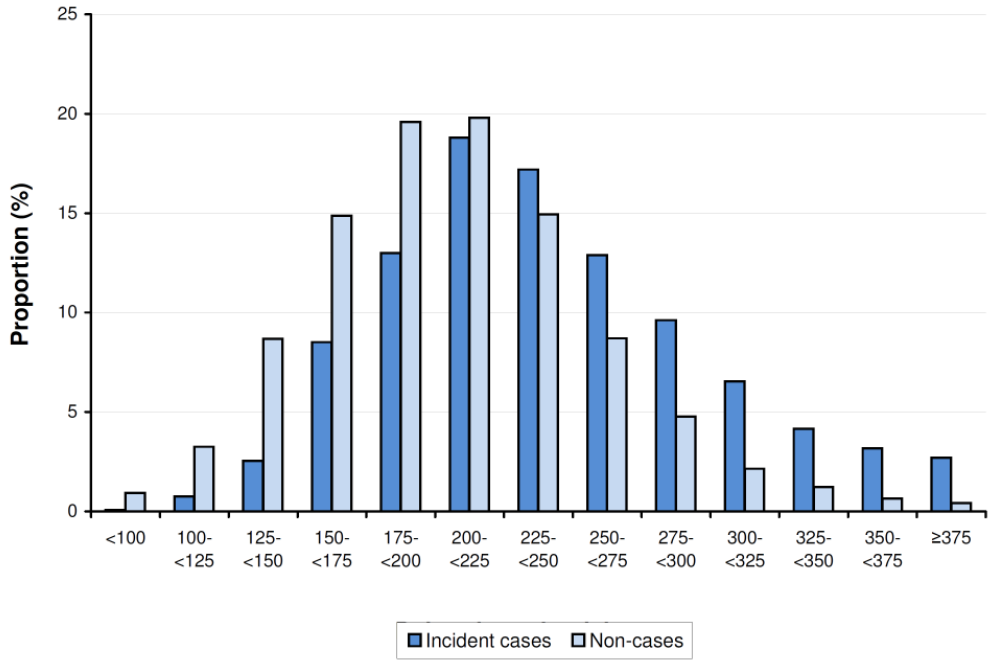


Figure 13. Proportion of cases and non-cases across score points in the training set

A plot of sensitivity and specificity against the points from the risk score illustrating the trade-off between both measures is shown in **Figure 14**. A cut-off value maximising sensitivity did so at the expense of specificity and vice versa. At a score value of 218, sensitivity and specificity were both 62%.

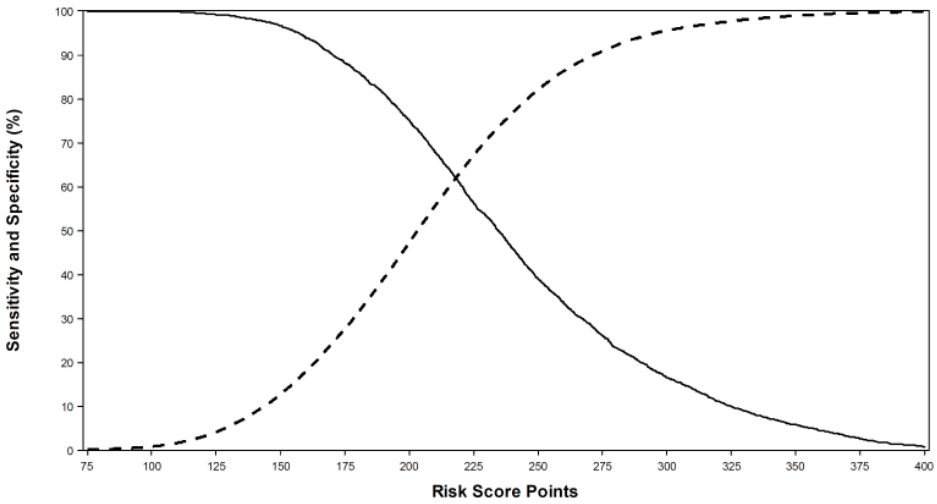


Figure 14. Plot of sensitivity and specificity across points from the risk score

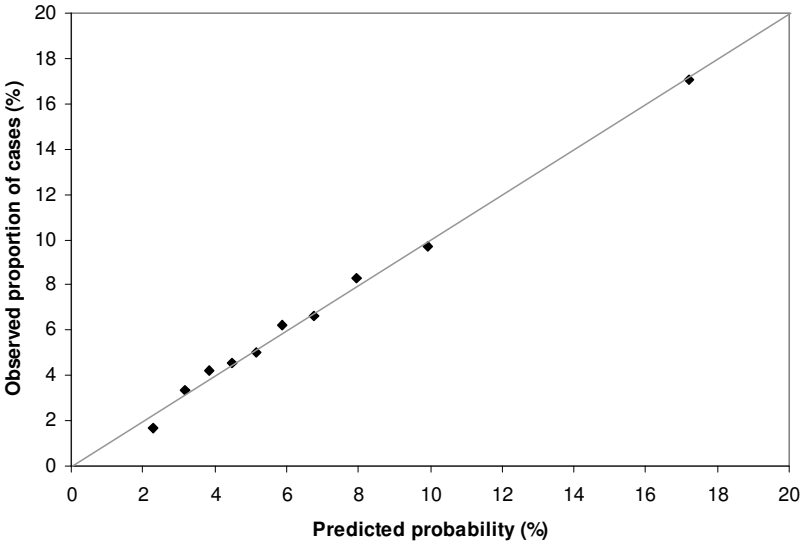
In order to identify high-risk individuals to undergo intervention programmes or to be included in clinical trials, designation of a cut-off point to define high-risk individuals is necessary. The decision on such an appropriate cut-off value may be taken by evaluating measures of the model's discriminatory accuracy, including sensitivity, specificity, positive and negative predictive values across a wide range of possible cut-offs. These test characteristics according to various cut-off points of the risk score are presented in **Table 11**. For example, a cut-off point of ≥175 points corresponded to a sensitivity and specificity of 88.1% and 27.7%, respectively, implying that the probability of having a score value of 175 or above among those individuals who actually experience SWG is 88.1%. Likewise, the probability of having a risk score below this threshold among those subjects who do not experience SWG is 27.7%. The corresponding Youden's index was J=0.158. The probability of experiencing SWG given that the risk score is above the threshold of ≥175 (PPV) and the probability of not experiencing SWG given that the test is below this threshold (NPV) were 8.0% and 95%, respectively. All NPV's lay above 90% implying a high ability of the score to rule out substantial weight gain within the following 5 years. From the table, the best cut-off value was ≥225 score points, with J=0.234. Of the total population, 34% had a score of 225 or higher. A score of ≥225 captured 56% of the cases who will experience SWG. Furthermore, 67% of the persons who do not experience SWG had a score <225. The corresponding PPV and NPV were 10.9% and 95.6%, respectively.

Table 11. Sensitivity, specificity, positive and negative predictive value for various cut-off points of the risk score in the training sample

Score points	Percentage of the population	Sensitivity (%)	Specificity (%)	Youden's index (J)	PPV (%)	NPV (%)
≥100	99.1	99.9	0.9	0.008	6.7	99.5
≥125	96.0	99.2	4.2	0.034	6.9	98.7
≥150	87.8	96.6	12.9	0.095	7.3	98.2
≥175	73.3	88.1	27.7	0.158	8.0	97.0
≥200	45.2	75.1	47.3	0.224	9.2	96.4
≥225	34.4	56.3	67.1	0.234	10.9	95.6
≥250	19.4	39.1	82.1	0.212	13.5	95.0
≥275	10.4	26.2	90.8	0.170	16.9	94.5
≥300	5.3	16.6	95.5	0.121	21.0	94.1
≥325	2.8	10.0	97.7	0.077	23.7	93.8
≥350	1.4	5.9	98.9	0.048	28.2	93.6
≥375	0.6	2.7	99.6	0.023	29.2	93.5

PPV = positive predictive value; NPV = negative predictive value. Youden's index was calculated according to the following formula: J = (sensitivity (%) + specificity (%) – 100)/100.

Apart from the score's ability to discriminate between cases and non-cases, its ability to quantify absolute risk was evaluated. Calibration of the score was evaluated by comparing observed and predicted probabilities across deciles of predicted risk. In general, the estimated probability of experiencing SWG agreed very well with the observed proportion of incident cases across deciles (**Figure 15**), indicating good calibration of the score in the total training population (HLT: p=0.15). Although slight differences in the ability of quantifying risk were observed across cohorts, the HLT indicated no evidence of miscalibration in any of them (**App. Figure 4**).



p-value for miscalibration: p=0.15

Figure 15. Calibration plot showing observed proportion of cases across deciles of predicted risk in the training sample

Hosmer-Lemeshow Test: p=0.15.

In addition to the HLT, **Table 12** compares the observed proportion of cases with the predicted risk across more intuitive categories of predicted risk. The table shows that the observed incidence increased with increasing score points and lay within the range predicted by the score, also indicating that the score quantifies risk very well.

Table 12. Incidence of substantial weight gain within the first 5 years of follow-up across categories of predicted risk in the training sample

Estimated probability (%)	Score points	Total N	Cases (n)	Observed incidence (%)
<5	<192	18,588	633	3.4
5 – <10	192 – <264	22,235	1,533	6.9
10 – <15	264 – <307	4,292	517	12.1
15 – <20	307 – <339	1,174	233	19.9
20 – <25	339 – <364	515	107	20.8
≥25	≥364	399	126	31.6

3.2.3.2 External validation

In the independent study sample, the area under the ROC curve (95% CI) implied a poor discriminatory ability of the risk score, with a c-statistic of 0.57 (0.566–0.578) (**Figure 16**). This means that the risk score performed only slightly better than a random guess in distinguishing cases from non-cases in this external population. To rule out influences of differences in the structure of the population with respect to sex between the training and validation sample, the performance of the risk score was assessed after exclusion of those cohorts consisting only of women (France, Utrecht and Naples); however, the discriminatory ability only slightly improved to 0.59 (0.58 to 0.59). Additional exclusions refer to all subjects for whom only self-reported body weight at baseline was available, but results remained virtually unchanged. Importantly, as in the training sample, there were considerable differences in discriminatory accuracy of the score between the separate cohorts of the validation sample. Specifically, the aROCs (95% CIs) for France, IT-Naples, Spain, UK-General Population, UK-Health Conscious, NL-Utrecht, GER-Heidelberg, and Sweden were 0.56 (0.55–0.57), 0.60 (0.50–0.70), 0.62 (0.60–0.63), 0.65 (0.61–0.68), 0.58 (0.57–0.59), 0.62 (0.61–0.64), 0.66 (0.63–0.68), and 0.66 (0.64–0.67), respectively.

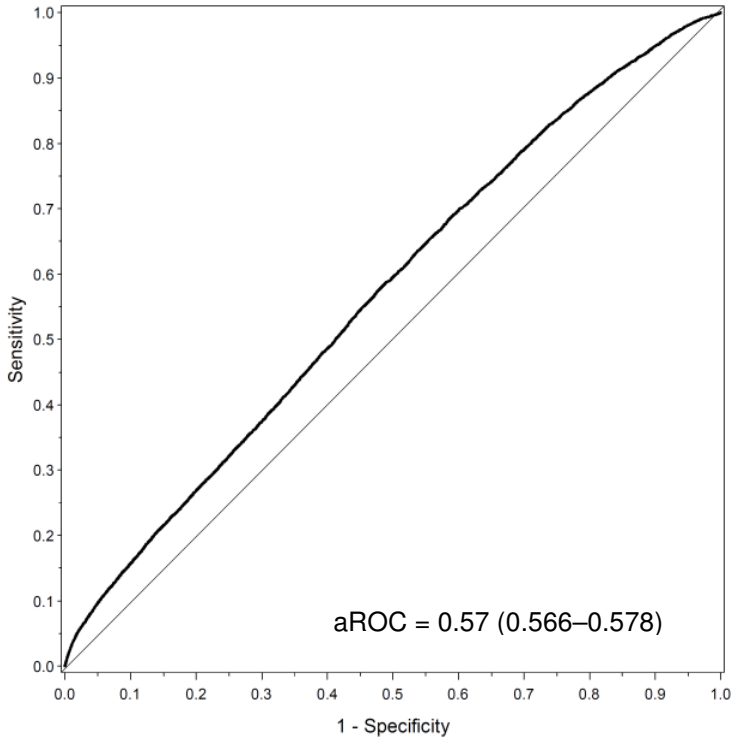


Figure 16. Receiver operating characteristic curve for the prediction of substantial weight gain (gaining ≥10% of baseline weight) over 5 years in the external validation sample

The calibration plot (**Figure 17**, HLT: p<0.0001) as well as the comparison of observed incidence with predicted risk across the chosen categories of predicted risk (**Table 13**)

showed that the score was able to adequately quantify absolute risk in the external validation sample, though not as well as in the training sample. The calibration plot indicated that the score somewhat overestimated risk in the lower and upper range of the score while it slightly underestimated risk in the middle range of the score. Cohort-stratified analyses, however, did not reveal a particular centre to be responsible for the slight evidence of miscalibration (all p-values >0.05).

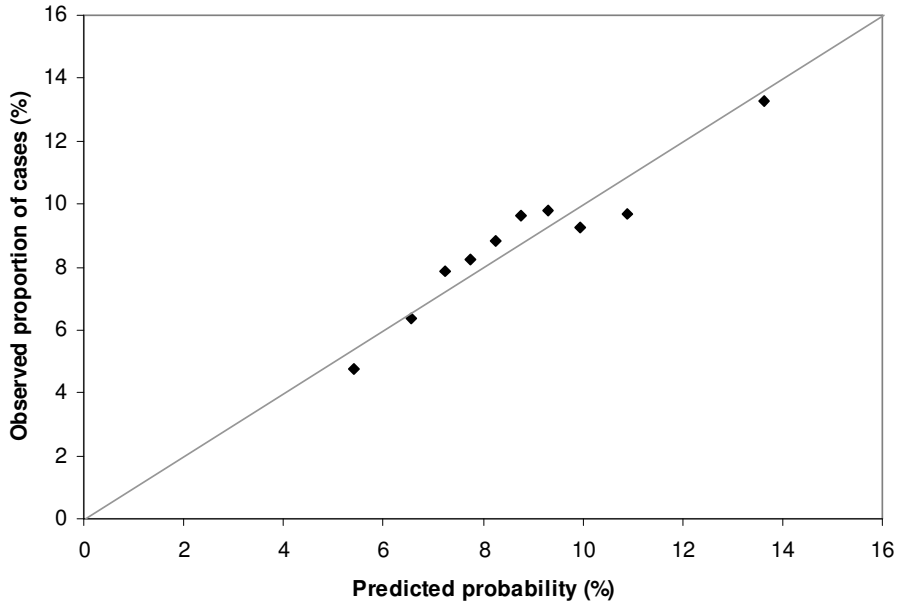


Figure 17. Calibration plot showing observed proportion of cases across deciles of predicted risk in the external validation sample

Table 13. Incidence of gaining ≥10% of baseline weight within the first 5 years of follow-up across categories of predicted risk in the validation sample

Estimated probability (%)	Score points	Total N	Cases (n)	Observed incidence (%)
< 5	< 192	31,111	1,899	6.1
5 - < 10	192 to < 264	65,227	5,950	9.1
10 - < 15	264 to < 307	14,231	1,481	10.4
15 - < 20	307 to < 339	2,889	398	13.8
20 - < 25	339 to < 364	882	180	20.4
≥ 25	≥ 364	759	190	25.0

3.2.4 Robustness of the risk score

Because there were considerable differences in the predictive accuracy of the risk score between centres, both within the training sample and within the validation sample, the question arose whether centre-specific models might yield better predictive performances than the overall, universal model. Hence, to further investigate this issue, two approaches were followed for all cohorts of the training and validation sample. First, the overall risk score was re-estimated in each cohort by using all predictors included in the overall risk score but assigning weights to the predictors based on centre-specific regression coefficients. Second, centre-specific predictors were identified by stepwise Cox PH regression and used to construct population-specific risk scores (for the training sample, centre-specific predictors can be found in **App. Table 3**). Results in terms of discriminatory ability and calibration for these two approaches in comparison to the overall model among centres of the training and validation sample are presented in **Table 14**. For the majority of centres in the training set, model performance changed only marginally when the overall risk score was re-estimated using centre-specific regression coefficients of all predictors or when centre-specific risk scores were derived. An exception was the cohort of UK-Norfolk for which the re-estimated overall model and the centre-specific prediction model obtained moderately higher aROCs than the overall model (0.69 and 0.68 vs. 0.65). Further, it seems noteworthy that in both Dutch cohorts almost the same discriminatory ability was achieved in the centre-specific prediction models in comparison to the overall model although a remarkably reduced number of variables were included (9 vs. 20 predictors for NL-Doe and 11 vs. 20 predictors for NL-AmMa). Also among most centres of the validation population, only slight improvements in model performance were observed when regression coefficients were re-estimated or complete centre-specific prediction models were developed. Exceptions were the cohorts of France and Italy-Naples among which the re-estimated model exhibited considerably higher predictive accuracy than the original overall model (0.63 vs. 0.56 and 0.70 vs. 0.61, respectively).

As the overall risk score performed better among men than among women, it was further investigated whether sex-specific prediction models may yield higher predictive accuracy than the overall model. However, model performance in terms of discrimination did virtually not change in sex-specific models, further underlining the finding of a considerably lower predictability of weight gain among women compared to men. Also, the inclusion of information on menopausal status and use of hormones for menopause at recruitment did not materially alter the predictive accuracy among women.

Table 14. Performance of the overall risk score compared to the re-estimated overall score and centre-specific scores in single cohorts of the training and validation set

	Overall model		Re-estimated overall model *		Centre-specific prediction models †	
	aROC (95% CI)	HLT	aROC (95% CI)	HLT	aROC (95% CI)	HLT
Training sample						
UK-Nor	0.65 (0.62–0.68)	0.50	0.69 (0.65–0.72)	0.14	0.68 (0.65–0.71)	0.15
NL-Doe	0.70 (0.66–0.74)	0.15	0.72 (0.68–0.76)	0.53	0.71 (0.66–0.75)	0.40
NL-AmMa	0.76 (0.72–0.80)	0.89	0.76 (0.72–0.80)	0.94	0.75 (0.71–0.79)	0.97
IT-Flo	0.66 (0.62–0.69)	0.14	0.67 (0.64–0.70)	0.17	0.68 (0.65–0.71)	0.82
GER-Pot	0.69 (0.66–0.71)	0.05	0.68 (0.66–0.71)	0.05	0.68 (0.66–0.70)	0.07
DK-CopAa	0.69 (0.67–0.70)	0.38	0.69 (0.68–0.71)	0.84	0.69 (0.68–0.71)	0.08
Validation Sample						
Total	0.57 (0.57–0.58)	<.0001	0.61 (0.60–0.61)	0.003	-	-
France	0.56 (0.55–0.57)	0.08	0.63 (0.62–0.64)	0.002	0.63 (0.62–0.64)	0.01
It-Nap	0.60 (0.50–0.70)	0.49	0.70 (0.61–0.80)	0.52	0.66 (0.54–0.78)	0.23
Spain	0.62 (0.60–0.63)	0.07	0.65 (0.63–0.66)	0.23	0.65 (0.63–0.66)	0.33
UK-GP	0.65 (0.61–0.68)	0.60	0.66 (0.63–0.69)	0.53	0.65 (0.62–0.69)	0.08
UK-HC	0.58 (0.57–0.59)	0.35	0.60 (0.59–0.61)	0.71	0.61 (0.60–0.62)	0.05
NL-Utr	0.62 (0.61–0.64)	0.58	0.63 (0.62–0.65)	0.37	0.64 (0.62–0.65)	0.49
Ger-Hd	0.66 (0.63–0.68)	0.96	0.67 (0.64–0.70)	0.87	0.67 (0.64–0.70)	0.88
Swe-Mal	0.66 (0.64–0.67)	0.06	0.67 (0.65–0.68)	0.55	0.67 (0.65–0.69)	0.88

* Using centre-specific regression coefficients as weights for all predictor variables included in the overall model.
† Including only centre-specific predictor variables as identified from stepwise Cox regression (for training sample see App. Table 2).
HLT = p-value Hosmer-Lemeshow Test to evaluate calibration of the model.
DK-CopAa = Denmark (Copenhagen and Aarhus) cohort; GER-Pot = Germany (Potsdam) cohort; IT-Flo = Italy (Florence) cohort; NL-AmMa = the Netherlands (Amsterdam and Maastricht) cohort; NL-Doe = Doetinchem cohort; UK-Nor = United Kingdom (Norfolk) cohort; IT-Nap = Italy (Naples) cohort; UK-GP = United Kingdom (General population) cohort; UK-HC = United Kingdom (Health Conscious) cohort; NL-Utr = The Netherlands (Utrecht) cohort; GER-Hd = Germany (Heidelberg) cohort; SWE-Mal = Sweden (Malmö) cohort.

4 Discussion

The research described in this thesis is focused on the development and validation of a risk score predicting substantial weight gain over 5 years among middle-aged European men and women. The risk score was developed in six cohorts of the EPIC study, comprising 47,203 men and women, and subsequently validated in eight independent EPIC cohorts, involving 115,099 men and women. The risk score, which was based on easily assessable information on several socio-demographic, dietary and lifestyle factors, allowed for accurately estimating risk of SWG during the following 5 years in various European populations. The discriminatory accuracy of the risk score was moderate in the six training cohorts and reduced in the external validation cohorts. Importantly, there was considerable variation in discriminatory ability between study populations (aROCs ranging from 0.56–0.76) which mainly persisted when population-specific risk scores were developed. This finding suggests that the present risk score does not predict 5-year risk of SWG equally well across European populations, but it also indicates that the performance of the score is hardly inferior to population-specific risk score models.

4.1 Methods

4.1.1 Study design and population

Major strengths of the present study are its prospective design, the large sample size, and the diverse study cohort. It may be considered, though, that the single study cohorts cannot be assumed to be representative of the underlying source populations. Study populations represent samples of convenience of volunteers [154] and 30% of the original cohort had to be excluded due to lack of follow-up data. Consequently, the descriptive characteristics of the study population in the present study, including the incidence of SWG, the range and distribution of predictors, risk score points and absolute risks of SWG, may not be generalisable to the underlying European population which might be considered a potential weakness of the study. However, this selection of participants should not have affected the internal validity of the present study. According to Rothman et al. bias occurs when non-response or loss to follow-up is associated with both exposure and disease [172], which is rather unlikely in the present study. Therefore, comparisons of incidences between exposure groups among the study population might well reflect the ratios of risks being present in similarly defined subgroups of the source population, independent of differing absolute risks [173, 174]. Valid estimation of relative risks is a fundamental requirement to construct a risk score with high predictive performance and ensures that the score also applies to subjects in the source population. Clearly, the same risk factor profile would result in the same absolute risk.

4.1.2 Data quality of predictors and the outcome

The internal validity and interpretability of the study results depend on the quality of the data at hand. In the present study, information on dietary and lifestyle factors was mainly obtained by means of self-administered questionnaires at baseline. Questionnaires are most feasible for large-scale epidemiologic studies since they are easy to process, inexpensive to be administered and the respondent's burden is lower than for many other assessment methods. Habitual dietary intake of the preceding year was assessed using country-specific FFQs. The FFQ is a widely used dietary assessment instrument in large epidemiologic studies [175, 176]. A general limitation of the FFQ is that participants are required to have a good memory to accurately assess frequency and portion size of food consumption during the previous year. Further inaccuracies may result from incomplete food lists contained in the FFQ [175]. Various pilot studies have been conducted before recruitment of the main study cohorts to assess the relative validity and reproducibility of the FFQs used in EPIC [156, 157, 177-182]. Theses pilot studies showed that the repeatability of the questionnaire was generally good, while the validity ranged from modest to good and varied between food items and according to country [156, 157]. Clearly, the intake measured with highest validity was alcohol (correlation coefficient: 0.63–0.90), however, in general, correlations were not higher than 0.4–0.7 for the main food groups. Exemplarily, correlations varied between 0.38 and 0.70 for meat intake and between 0.30 and 0.60 for vegetable intake across countries. Previous studies on the validity of FFQ measurements obtained similar results, and, from this perspective, the dietary questionnaires developed for the EPIC study can be considered acceptable [156]. Nevertheless, given that the validity varied across food groups and countries, relative risk estimates for some food groups (and in some countries) might have been less precisely estimated which might have lead to an underestimation of the predictive strength of these components in the risk score and might have limited the predictive ability of the score.

Likewise, information on lifestyle factors may suffer from random error and self-reporting bias in the measurement which might also have limited the ability to obtain accurate risk estimates. For example, some behaviours, i.e. smoking and alcohol, are value-laden and prone to underreporting, while others are socially desirable, i.e. physical activity, and likely to be overreported. In particular, it was shown that accuracy in recalling physical activity may vary according to gender, body weight status and exercise type and level [81]. In this respect, several studies have also demonstrated that (structured) high-intensity activities are more accurately and reliably recalled than low-intensity behaviours [81]. Because exercise activities were exclusively considered in the present study, measurement error may be marginal for this predictor.

With the exception of smoking status, information on dietary and lifestyle factors has been assessed only at a single point in time in the present study. Consequently, the meaning of the present findings depends on the stability of dietary and lifestyle habits over time. Although it is not expected that middle-aged men and women, the majority of EPIC participants, dramatically change their lifestyle and diet over time, slight changes are likely to occur over longer periods of follow-up. More importantly, changes in diet or lifestyle, thereby changes in energy balance, are a prerequisite for subsequent changes in weight among individuals not suffering from weight-influencing diseases. Only fairly recently, a study from the US using data of three large prospective cohorts highlighted the impact of changes in dietary and lifestyle behaviours on long-term weight gain [183]. Because individuals achieve a new steady-state of weight within short time periods after they changed their lifestyle, repeated assessments of behavioural factors over time might be important to discern long-term effects and to identify specific lifestyle changes to be translated into obesity-prevention strategies [183]. At this point, however, it should be kept in mind that the purpose of the present risk score was to estimate the 5-year risk of SWG based on the current risk profile. The inclusion of numerous variables indicating dietary and behavioural changes represents a different concept and would have been beyond the scope of this thesis. Future studies might however explore whether the incorporation of lifestyle changes improves the predictability of weight gain.

With respect to the assessment of weight change, differences in the measurement of body weight at baseline and follow-up between study cohorts may be considered a drawback. Although body weight was measured by trained staff in most centres at baseline, local study protocols about the participants' clothing varied between cohorts. At follow-up, body weight was professionally measured in two centres, while it was self-reported by the participants in the follow-up questionnaire in all other centres and was thus most likely underestimated [184, 185]. These methodological limitations were accounted for by correcting body weight for protocol differences in clothing worn during measurement and by applying prediction equations to correct for potential underreporting in self-reported body weight as previously described [158]. Because there were only two measures of body weight available for each individual, case status could be determined only at the end of follow-up. In order to estimate the time point when cases theoretically crossed the threshold of SWG, the extremeness of weight gain was judged through an underlying thought of linear weight gain. This is a strong assumption about the structure of weight gain that might not hold true. Weight gain is reversible, and it is well known that body weight tends to fluctuate over time [186], which may lead to repeated cycles of weight loss and recovery [187, 188] that are not reflected in a two-point-in-time measurement. Thus, some study participants might have experienced periods in which they crossed the threshold of SWG during their follow-up although they appeared

not to have done so as determined from the two points of measurement. Fluctuations or non-linear weight gain in general may have resulted in misclassification of cases and non-cases and additionally in misspecification of the cases' survival time, which might have limited the predictive performance of the obtained risk score model. However, recent findings from the EPIC-Potsdam study using 5 measurements of weight suggest that weight gain can be reasonably well approximated by a straight line over a follow-up period of 8 years on the population-level [189]. Further, the assumption of linearity, although not fully appropriate, appeared to be the most meaningful approach given the structure of the present data and the most self-evident assumption to start from.

4.1.3 Methodological and statistical approaches

Model building

The pre-selection of the set of variables to be considered as candidate predictors for the final model was mainly based on subject matter knowledge, specifically, on reported or hypothesised associations with weight change or risk of overweight/obesity in the literature. A second criterion for inclusion was that candidate predictors should be easy to assess by questionnaire. The questionnaire approach is commonly used as a first step in risk assessment as it does not require blood testing and is thus simple and inexpensive [190]. Although only modifiable risk factors can be addressed by intervention strategies, non-modifiable risk factors like age and current body weight are important components to determine an individual's risk and thus have been previously used in numerous risk prediction models for diabetes [11], CVD [119, 125] and cancer [12, 14].

In order to keep the model parsimonious and easy to use in practice, only the most relevant predictors from the set of candidate variables were retained for the final model. The process of model specification is considered the most difficult part of prediction modelling [100]. In the present study, the selection of the final predictors was based on the variable selection method "stepwise" implemented in the PHREG procedure in SAS [166]. Such an automated variable selection procedure has several advantages and has been applied in previous studies on risk prediction [114, 191]. Such procedures are objective, straightforward to apply in modern statistical packages and usually reach the goal of making a model smaller [100]. Because the number of predictors retained in the final model is determined by the level of significance, the stopping rule for inclusion and exclusion of variables is a central issue [100]. As for many procedures, the standard level of significance (α=0.05) is often used. For inclusion in a Cox PH model, α=0.1 has also been recommended for selecting variables in stepwise regression [192]. Because of different sample sizes across single centres of the training sample, a value of α=0.1 was chosen in the present study in order to be less likely to reject possibly important variables and to obtain a larger overlap in respect of centre-specific

predictors to choose a common set from. Also, in model building, it may be generally better to err on the side of caution, i.e. to include a variable rather than to exclude it, and to decide on a less stringent level of significance [192]. In this regard, simulation studies have shown that noise variables in a model do hardly decrease predictive ability [100]. Despite the advantages of the automated stepwise selection procedure, some concerns have also been raised about this method, including bias in the estimation of regression coefficients, their standard errors and p-values or instability of the selection [100, 172]. Instability of the selection refers to the case that different combinations of predictors may actually produce similar predictive performances. This implies that the selection of predictors may depend on the set of subjects used to develop the model and that the selection is unstable. However, as the sample sizes in the present study were sufficiently large (events per variable ≥10 in all centres and in three of them even >50 which is considered appropriate for reliable selection [100, 101]), it is unlikely, that important predictor variables were missed and that the set of predictors would essentially change if a slightly different selection of participants was considered for the development of the model. Likewise, the problem of biased estimation of coefficients is essentially a problem of small sample size [100]. Also, there was no difference in the final set of predictor variables when the other selection methods available in SAS (forward and backward selection) were used, which supports the stability of the selected model. Ultimately, although many statisticians may have reservations about the use of automated variable selection methods, Austin and Tu concluded that these methods are appropriate in the setting of prediction modelling if the predictive accuracy of the final model is evaluated in an independent population [193], which has been done in the present study.

Apart from the selection of variables to be included in the final model, the way these variables, particularly continuous predictors, are modelled is an important consideration. A crucial assumption of the Cox model is that the log hazard of the outcome relates linearly to a set of covariates [170]. For continuous variables, however, this linearity assumption may be inappropriate and relationships may be more complex. Categorisation of quantitative variables has often been criticised as this approach is tied to loss of power and results in risk step-functions which are biologically less plausible than smooth relationships [194]. Alternatively, (cubic) spline functions, polynomials or fractional polynomials may be used to capture non-linear relationships and increase predictive ability of the model [100, 194, 195]. However, interpretability of these relationships appears to be difficult and in terms of practicability, continuous predictors were modelled using linear terms. Depending on the context in which the risk score will be applied, e.g. in general practices or in medical research, non-linear modelling of continuous variables may be meaningful to consider, though.

Model validation

Performance of the model was evaluated using measures of discrimination and calibration. Although the area under the ROC curve, or c-statistic, is very straightforward to interpret and the most commonly used measure of discrimination, its short-comings should be noted. One of the most obvious concerns is the question of how high discrimination needs to be in order to be "high enough" to justify a model [196]. Further, it has been criticised that the aROC does not account for the stochastic process of the disease outcome in risk prediction modelling [9]. While aROC is a useful method to assess the quality of a test in the diagnostic setting, the random nature of the outcome in the prognostic setting is not taken into account. Also, the c-statistic is based on ranks and does not consider the actual distribution of risk. Thus, minor differences in risk between two individuals, e.g. 1.0% vs. 1.1% would have the same impact on the c-statistic as two individuals who are at moderate versus high risk (e.g. 5% vs. 20%), if their difference in rank is the same [9]. This limitation of the c-statistic may be of particular concern in prospective cohorts that are generally characterised by a preponderance of individuals at low or very low risk. The fact that the c-statistic is rank-based also implies that even in case of perfect discrimination, the predicted risk can be substantially different from the actual observed risk [111]. On account of these limitations, the second component of model performance, calibration, plays a pivotal role in model validation and has been considered in the present study. The Hosmer-Lemeshow Test, which is the most popular measure of calibration, was used to formally test for significant differences between observed and predicted risk across deciles of predicted risk. This test has been criticised because the null hypothesis is that "the model is well calibrated" and the test can only tell that there is insufficient evidence of miscalibration [196]. As with the c-statistic, there is no clear threshold to decide how much miscalibration would be too much implying that a model is not of value. Furthermore, it is very sensitive to the way the groups are formed and large sample sizes [110]. Particularly, the latter property needs to be considered in view of the large sample size of the present study. Therefore, observed incidence of SWG was also compared with predicted risk across intuitively meaningful categories of predicted risk.

A major strength of the present study is that the performance of the risk score was evaluated in independent study populations. A key threat to validity of a prediction model is that the model may not solely explain part of the relationship between the outcome variable and its predictors, but may also explain accidental variability present in that dataset (idiosyncrasies, "random noise") which does not generalise beyond this particular population sample [100]. Collaborative multi-centric studies, such as EPIC, offer the unique opportunity to develop a risk score in a non-randomly selected set of centres and validate it in the remaining, geographically independent centres. This approach has been followed in the present thesis by choosing the cohorts participating in the DiOGenes project to guide model development

and the remaining cohorts to test the model. This approach of geographical validation has been judged more meaningful than standard techniques such as cross-validation to account for overfitting [100]. Further it is noteworthy, that the EPIC study with its multi-centric nature comprising countries from Northern to Southern Europe is characterised by large heterogeneity in dietary and lifestyle habits which allowed for testing the prediction model in several culturally diverse study populations. This aspect can be clearly seen as an advantage of the present study and increased the confidence with respect to general applicability of the model.

4.2 Results

4.2.1 Predictors included in the risk score

The risk score developed in the present study was based on the following factors: age, sex, baseline body weight and height, educational level (technical school, secondary school, university), smoking cessation, sports, non-consumption of alcohol, consumption of alcohol in moderate amounts, intake of fruits and vegetables, red and processed meat, poultry, bread, butter/margarine, cake and cookies, and soft drinks.

In interpreting the associations between predictors and SWG in the risk score model, it has to be kept in mind that the aim of risk prediction is to most accurately predict future disease risk on the basis of a few variables reflecting the structure of the data while these predictors do not necessarily need to be causally related to the outcome [101]. Thus, in some cases, predictors may indeed reflect (established or hypothesised) aetiological associations with weight gain in the present study, while in other circumstances, predictors may essentially be a marker of other lifestyle factors and reflect behavioural patterns that are more/less favourable for the regulation of body weight rather than causal relationships. In the present risk score, for example, education, sex and also alcohol may be considered non-causal predictors. Although numerous epidemiological studies reported on the aetiological role of alcohol consumption in weight gain, results are conflicting [74], alcohol consumption may also represent a measure of other dietary and lifestyle factors. Specifically, in sex-specific analyses, alcohol consumption was predictive of a lower risk for SWG particularly among women for whom it was shown before that higher alcohol consumption is related to higher education [197, 198] which in turn may be accompanied by a greater health and body weight consciousness.

In contrast, most of the dietary predictors identified in the present study have also emerged as aetiological factors in previous epidemiological studies. For instance, higher intake of fruits and vegetables and bread were associated with lower risks of SWG, while intake of meat and soft drinks showed a positive relation to the outcome in the current analysis and in

previous studies [17, 34, 35, 37, 42, 43, 66]. For these predictors, clear hypotheses about causal associations with weight gain do exist (see chapter 1.2.5) and these modifiable factors may be well addressed by prevention strategies. The inverse association observed with intake of cake and cookies was striking, though. Evidence from aetiological studies is scarce and conflicting [33, 35, 62]. Because cake and cookies tend to be energy dense, higher intake may cause many people to exceed daily energy requirements and would be suspected to increase risk of weight gain. Since cake and cookies are usually perceived as socially undesirable foods, one explanation for the observed inverse association could be selective misreporting of intake in the way that subjects experiencing weight problems underreported consumption. An inverse association might also have occurred if participants experiencing weight gain or weight fluctuations before baseline might have reduced their intake of cake and cookies. This confounding effect of post hoc changes in intake of cakes and cookies was accounted for by excluding cases occurring during the first two years of follow-up; however, findings remained virtually unchanged. Another explanation may refer to the number of eating occasions per day. Because cake and cookies may be mainly consumed between regular meals, intake of these foods may be associated with a higher number of eating occasions and would therefore represent a marker of dietary habits. Higher frequency of eating has been shown to be related to lower insulin concentrations [78] which again may assist in body weight control. A number of cross-sectional studies reported higher eating frequency to be inversely related to body weight [199-201]; data from prospective studies are lacking though. Hence, further prospective (and mechanistic) studies are needed to evaluate a potential causal role of cake and cookies in body weight management.

Exercise activity was predictive of lower risk of SWG which is consistent with the literature on aetiological associations and has a clear causal basis [17]. In accordance with previous observations in epidemiological studies and hypothesised biological mechanisms [96, 98], smoking cessation was related to a higher risk of SWG in the present study and appeared to be one of the strongest predictors in the model. Smoking cessation-induced weight gain may frequently be a source of concern for smokers planning to quit and may discourage many of them from actually trying to do so [98]. Indeed, the rise in prevalence of overweight and obesity in the United States has been partly attributed to smoking cessation, whereby weight gain seems to be greatest within the first few months after quitting and former smokers do not seem to continue weight gain at a higher rate than never smokers [99]. Nevertheless, despite cessation-induced weight gain, it is widely argued that the health benefits associated with smoking cessation far exceed any health risks that may result from (short-term) post-cessation body weight gain [96]. Smoking is one of the strongest risk factors for many types of cancer, for respiratory and cardiovascular diseases and remains the leading cause of preventable mortality worldwide [6]. A total of 9% of all deaths globally are estimated to be

attributable to smoking [6]. Also, despite smoking may facilitate to maintain a lower total body weight, it promotes abdominal body fat distribution which has been increasingly shown to be an independent risk factor for cancer [202], diabetes [203] and overall mortality [204]. Hence, the present risk score underscores the importance of weight control among subjects who attempt to quit smoking. Given the undeniable health benefits associated with cessation, post-cessation weight gain should not hold people off stopping to smoke. To remove barriers to quitting, the promotion of smoking cessation in public health may further be accompanied by offering information on weight management and support to smokers concerned about gaining weight.

4.2.2 Predictive performance of the risk score

The overall discriminatory ability in the training sample was modest and considerably lower than found for most diabetes risk prediction models achieving aROCs of roughly 0.80 [11]. Nevertheless, it lay within the range observed for most cancer risk prediction models that mainly yielded aROCs between 0.60 and 0.75 [102, 137, 138, 140]. Low discriminatory accuracy of a prediction model may be explained by lack of information on important predictors. For instance, television watching [17], sleep duration [94] as well as weight loss attempts [205, 206], weight cycling [207, 208] and large short-term weight changes [186, 208] have been shown to be risk factors for future weight change. Unfortunately, information on these risk factors was not available in the present dataset, but may additionally be considered in future risk prediction models for weight gain. In the field of chronic diseases, hopes have been raised that next to the inclusion of non-invasive factors and biochemical measures, information on common genetic markers may be used to improve discriminatory accuracy and thereby disease prevention programmes [11, 135]. The predictive ability of genetic factors, however, currently appears limited [11, 135]. For example, the addition of seven SNP's to the breast cancer model developed by Gail et al. only modestly improved discriminatory accuracy [135]. Similarly, the additional inclusion of 20 diabetogenic SNP's did barely improve discrimination of incident type 2 diabetes beyond lifestyle factors and metabolic markers in the EPIC-Potsdam cohort [152]. In respect to obesity, the EPIC-Norfolk study reported that 12 obesity-susceptible loci explained 0.9% of variation in BMI, with an aROC of 0.57 for prediction of obesity [209]. Thus, despite overwhelming statistical significances and repeated replications, the explained variance and the predictive value of the currently identified obesity-susceptibility loci is too low to be clinically meaningful [29]. Another promising approach might be the concept of gene-environment interactions describing the differential response to an environmental factor on disease risk dependent on an individual's genotype [29]. Several studies have shown, for instance, that the BMI-increasing effect of the FTO-variant, which is currently the most promising genetic marker for BMI, is more pronounced in individuals who have a sedentary lifestyle and attenuated in

physically active individuals [29]. Overall, only relatively few studies have investigated gene-environment interactions with respect to obesity development so far, which might be an important aspect for better targeting obesity prevention strategies. However, even if this progress in identifying enough genetic variants (or gene-environment interactions) is made, it still needs to be figured out how such information can be communicated to individuals and whether it will motivate people to adopt a healthy lifestyle [210].

At this point, it may be noteworthy though that very large independent relative risks are needed for a single predictor to meaningfully improve discrimination [135]. Hence, although one may think of other possible predictors to be included in a risk prediction model for weight gain, it is questionable whether the predictive ability measured by the c-statistic would meaningfully improve. As the present study is already based on a large number of standard risk factors for weight gain, the findings rather convey the impression that the predictability of weight gain in terms of discrimination may be limited in general. Nonetheless, the addition of new risk predictors might improve the present risk score model in terms of discrimination beyond the ROC curve. Recently, two new measures of discrimination, the net reclassification index and the integrated discrimination improvement, have been suggested to evaluate new variables for their ability to improve upon the original model beyond the ROC curve [211]. These measures are based primarily on stratification into risk categories and attempt to evaluate the ability of an extended model to more accurately reclassify individuals into higher or lower risk categories. These measures might, therefore, be considered in subsequent studies on the improvement of the present risk prediction model for weight gain instead of solely focussing on the c-statistic.

The discriminatory ability of the present risk score was reduced in the external validation sample in the present study, an observation that is also commonly reported for risk prediction models in the field of chronic diseases [10, 11]. Several reasons may be thought of to explain this phenomenon. First, overfitting of the model in the training sample may certainly be an explanation for poorer performance in the validation sample; however, given that the sample size was large in the present study and that the amount of optimism decreases with larger sample size [100], this explanation appears unlikely. Second, lower predictive accuracy in external populations may be due to differences between the training and validation population, especially in methods of data collection, coding of predictors and endpoint, and the availability of all variables used to construct the score [100]. However, given the standardised methodology followed in EPIC, this explanation also seems rather unlikely. Interestingly, apart from the overall difference in predictive ability between training and validation sample, there was considerable variation in discrimination across single cohorts of the training and validation sample, respectively. Specifically, discriminatory power ranged from 0.65 in UK-Nor to 0.76 in NL-AmMa in the training set and varied between 0.56

(France) and 0.66 (GER-Heidelberg and Sweden) in the validation set. In this respect it is also noteworthy, that despite a comparably lower overall predictive accuracy in the validation sample, the score was found to exhibit quite similar discriminatory accuracy among those validation cohorts that may be assumed to be comparable to selected cohorts of the training sample in respect of cultural and lifestyle factors. In particular, the predictive accuracy for GER-Pot and GER-Heidelberg was 0.69 and 0.66, respectively. Likewise, the score's ability to discriminate cases and non-cases differed only slightly between Denmark and Sweden (0.69 vs. 0.66). This comparable predictive accuracy among centres of similar socio-cultural background in addition to the variation in predictive accuracy between culturally (more) diverse populations indicates that the prediction of weight gain may depend on underlying socio-cultural or structural factors that were not similarly represented by the predictors included in the present risk score across study populations and that extrapolation of the model to other populations should be done with caution. This speculation is further supported by the finding, that the discriminatory accuracy among 6.521 individuals stemming from the same source population as the training sample (participant of any cohort included in the DiOGenes project) but whose data were not included in the DiOGenes project, and thus were not used for model development, was very similar to that observed in the training sample. Specifically, the overall c-statistic was 0.68 (0.66–0.71).

It has been suggested that re-estimation of regression coefficients in other populations may lead to more precise risk estimates and thus to better performances of the model in external populations [11, 190]. Nevertheless, in the present study, discriminatory ability did not materially change in most centres when centre-specific regression coefficients were used, which implies that the universal model seems to reasonably well reflect the effects of single predictors on SWG across European countries. Only the cohorts of France and IT-Naples showed a meaningful improvement in discriminatory ability which might be due to differences in the structure of these populations compared to the other populations in the training sample. Specifically, the cohorts of France and IT-Naples were both solely composed of women for whom weighting of predictors differed from weighting in the overall model.

Apart from varying effect sizes of predictors, the set of important predictor variables per se may differ between structurally and culturally diverse populations and it has been suggested that population-specific prediction models may be more useful than a universal risk score working in all populations [11]. In the present study, however, centre-specific risk prediction models did improve predictive accuracy only in the centres of UK-Nor, France and IT-Flo while no improvement was observed in the other centres. Although this observation does not rule out the possibility that important, maybe population-specific, predictors may not have been assessed in the present study, it conveys the impression that the predictive accuracy of the overall model already exhibits good discrimination within the limits of what appears to be

achievable in the prediction of weight gain and that the model is relatively robust against the exchange of single (centre-specific) behavioural predictors. This observation is not surprising given that the added predictive ability of modifiable factors was only modest (**App. Figure 4**). In particular, the addition of smoking cessation to a model containing all non-modifiable predictors strongly improved the aROC from 0.59 (0.58–0.60) to 0.65 (0.64–0.66), while further including various dietary factors and physical activity, those predictors varying between centre-specific models, only improved the aROC to 0.67 (0.66–0.68).

Finally, it may be worthwhile to mention that the predictability of SWG was lower among women than among men in most study centres. Reasons for this difference may include lack of information on important predictor variables or inaccuracies in the reporting of body weight or some predictors in women. Nevertheless, with respect to body weight, body weight was corrected for self-report using prediction equations in the present study [158]. Also, the additional inclusion of information on menopausal status and use of hormones for menopause at recruitment did not materially alter the predictive accuracy among women. Because sex-specific models did not perform better than the overall model, there was no reason for constructing sex-specific risk scores. Future studies, however, might investigate possible differences in the predictability of weight gain between sexes in more detail.

4.3 Implications for public health

In the field of chronic diseases, some risk scores are implemented in national prevention programmes in several Western countries [122, 128, 212, 213] and are recommended for clinical decision-making to decide on prevention intervention, further testing or therapeutic guidance. For example, the Framingham algorithm for predicting coronary heart disease was incorporated into the Third Report of the Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) [122]. Further, the Gail model and the German Diabetes Risk Score are publicly available as interactive Web tools [128, 214]. Likewise, the present risk score to predict SWG among European adults might be implemented in (national) obesity prevention programmes and may be used to identify high-risk individuals who may benefit most from timely intervention. However, test characteristics of the present risk score were only modest. The threshold of 225 points, maximising sensitivity and specificity, implies, that preventive actions will be indicated for a substantial part of the population (34%), which would have important implications for the healthcare system. Of these high-risk individuals, 11% will indeed experience SWG within the timeframe of 5 years. The decision about an appropriate cut-off point to define high-risk individuals requires careful weighing of risks and benefits associated with this threshold [104]. In contrast to serious diseases such as cancer or CVD, however, misclassification-costs may essentially reduce to those costs attached to false-positive findings in the context

of weight gain. Specifically, a false-negative test result may create false reassurance which is most critical when the disease is serious, only curable in early stages and effective intervention is available [104]. False-positives may pose a large economic burden to the healthcare system since a high number of individuals will be unnecessarily advised to undergo intervention programmes of weight gain prevention. Thus, a cut-off point related to higher specificity, i.e. 250 or 275, could substantially reduce the percentage of false-positives in the present study – certainly, at the expense of higher sensitivity.

The comparison of observed and predicted probabilities indicated that the ability of the score to quantify absolute risk of SWG was very good in the cohorts of the training sample and good to adequate in cohorts of the validation sample. Accurate estimation of risk, even from models with modest discriminatory accuracy, may have several important applications [116]. Well-calibrated projections of absolute risk are particularly important for designing intervention trials because the power of those studies depends on the number of incident cases, a reflection of absolute risk [116]. Thus, the present risk prediction model may serve to design more powerful and "smarter" prevention trials for weight gain by enriching the number of observed events. Results from these trials may be helpful to identify successful key strategies for obesity prevention among high-risk individuals. Because effective and sustainable prevention strategies may require action at both the individual and the societal level [215], the strategies identified for high-risk individuals could be translated to population-based programmes in a next step. Accurate risk estimation is also crucial in the context of prevention programmes which focus on the motivation of individuals to change their behaviour by providing information on the expected benefit. Although the present risk score includes several modifiable dietary and lifestyle factors, caution may be warranted with respect to the application of the present risk score in prevention programmes given the associations of some predictor variables with SWG that may be misinterpreted by laymen.

4.4 Conclusion and outlook

The present thesis is the first study to report on the development and validation of a risk score predicting absolute risk of (substantial) weight gain among adults. In this large prospective cohort study of middle-aged European men and women, a model using easily assessable information on several socio-demographic, dietary and lifestyle factors was found to be moderately effective at identifying individuals at high risk (discrimination) and good at quantifying absolute risk (calibration) of experiencing SWG within the next 5 years. The results of the present thesis highlight that the prediction of weight gain is complex and may be limited in general. Within these limits, the findings of the thesis further suggest the appropriateness of one universal, transnational risk score to predict weight gain among European adults.

On the basis of this first work on weight gain prediction among adults, future research may aim to consider additional predictors to attempt to further improve the predictive ability of a risk score model for weight gain. Emphasis may be particularly put on weight gain prediction among women.

Particularly with regard to the motivation of people to adopt a healthy, weight-preventing lifestyle, future studies may also explore the predictive accuracy of a risk score model incorporating changes in modifiable lifestyle factors to predict subsequent weight gain.

In terms of weight gain prevention in general, a supportive social and less obesogenic environment may be needed to facilitate changes in lifestyle and ensuring sustainability of these changes. Because the key to combat the obesity epidemic ultimately lies at the level of individuals, future research may also focus on the identification of factors enabling individuals to achieve a long-term behaviour change.

Summary

Given the rapid rise in obesity prevalence around the world in recent years and the considerable strains involved with the treatment of overweight and obesity, primary prevention of weight gain is of paramount public health importance. Due to limited resources, prevention efforts might be particularly targeted to individuals at high risk of excess weight gain that may result in overweight and obesity. Therefore, the aim of the present thesis was to develop a risk score predicting risk of substantial weight gain (SWG) within the following 5 years. Because this objective was addressed using data of the large-scale, multi-centre European Prospective Investigation into Cancer and Nutrition (EPIC), the present study offered the unique opportunity to simultaneously investigate whether such a risk score is equally well applicable to different European populations and to evaluate the idea of one universal, transnational risk prediction model for weight gain.

Data from six cohorts (47,203 men and women) of the EPIC study were used to develop a risk prediction model for SWG (training sample). Substantial weight gain was defined as gaining ≥10% of baseline weight during follow-up. Significant predictors of SWG were identified using Cox Proportional Hazards regression. Multivariable regression model coefficients were used to assign weights for each predictor and the risk score was calculated as a linear combination of the single predictors. The validity of the risk score was assessed by means of discrimination (area under a receiver operating characteristic curve, aROC) and calibration in the training sample and in external validation samples, comprising eight independent cohorts of the EPIC study (115,099 men and women).

During an average follow-up of 6.2 years (291,748 person-years), a total of 6,471 men and women gained ≥10% of baseline weight in the training sample. The final prediction model was defined on the basis of the following factors: age, sex, baseline body weight and height, technical school, secondary school, university, cessation of smoking, sports, non-consumption of alcohol, moderate consumption of alcohol, intake of fruits and vegetables, red and processed meat, poultry, bread, butter and margarine, cake and cookies, and soft drinks. The probability of experiencing SWG within 5 years increased from 2.0% for 100 to 33.6% for 400 score points. The discriminatory ability of the model measured by the aROC (95% CI) was 0.67 (0.66–0.68) in the training sample and 0.57 (0.566–0.578) in the external validation sample. Variations in discrimination between cohorts were observed, ranging from 0.65 to 0.76 in the training and from 0.56 to 0.66 in the validation sample. Calibration was very good and good in cohorts of the training and validation sample, respectively.

The present thesis could demonstrate that it was possible to develop a risk score predicting SWG over 5 years among European adults using easily assessable information on several socio-demographic, dietary and lifestyle factors. The score was found to exhibit moderate

discriminatory accuracy and to accurately predict risk of experiencing SWG during the following 5 years. The results of the present thesis highlight that the prediction of weight gain is complex and may be limited in general. Within these limits, the findings of the thesis further suggest the appropriateness of one universal, transnational risk score to predict weight gain among European adults.

Zusammenfassung

Die Primärprävention von Übergewicht und Adipositas ist von hoher gesamtgesellschaftlicher Bedeutung. Angesichts limitierter Ressourcen könnten Interventionsmaßnahmen zur Prävention vorrangig an jene Personen gerichtet werden, die ein hohes absolutes Risiko für die Entwicklung übermäßigen Körpergewichts aufweisen. Daher war das Ziel der vorliegenden Dissertation, einen Risikoscore zu entwickeln, der das absolute Risiko für eine starke Gewichtszunahme (SGZ) in den folgenden 5 Jahren vorhersagt. Da die multizentrische European Prospective Investigation into Cancer and Nutrition (EPIC)-Studie die Datengrundlage dieser Arbeit darstellte, konnte überdies der Frage nachgegangen werden, ob ein Risikoscore zur Prädiktion einer SGZ gleichermaßen in unterschiedlichen europäischen Populationen einsetzbar ist und ob es ein universelles, länderübergreifendes Risikoprädiktionsmodell für SGZ gibt.

Für die Ableitung des Prädiktionsmodells wurden die Daten von 47.203 Männern und Frauen aus sechs Kohorten der EPIC-Studie verwendet (Lernstudienpopulation). Eine starke Gewichtszunahme war definiert als eine Zunahme von ≥10% des Ausgangsgewichts während der Nachbeobachtung. Signifikante Prädiktoren wurden mit Hilfe der Cox Regression identifiziert. Multivariable Regressionskoeffizienten dienten der Gewichtung der einzelnen Prädiktoren in der Berechnung des Risikoscores. Die Validität des Risikoscores wurde anhand von Diskrimination (Fläche unter der Receiver Operating Characteristic-Kurve, aROC) und Kalibrierung bewertet. Die externe Validität wurde mit Hilfe von acht unabhängigen EPIC-Kohorten beurteilt (Teststudienpopulation, 115.099 Männer und Frauen).

Während einer mittleren Nachbeobachtungszeit von 6,2 Jahren (291.748 Personenjahre) verzeichneten 6.471 Männer und Frauen in der Lernstudienpopulation eine SGZ. Das finale Prädiktionsmodell wurde auf Basis der folgenden Prädiktoren definiert: Alter, Geschlecht, Körpergewicht und –höhe, Bildung, Aufgabe des Rauchens, Sport, Alkoholabstinenz, moderater Alkoholkonsum, Verzehr von Obst und Gemüse, rotem und verarbeitetem Fleisch, Geflügel, Brot, Butter und Margarine, Kuchen und Keksen und von Softgetränken. Das Risiko für eine SGZ in den nächsten 5 Jahren stieg von 2.0% bei 100 auf 33.6% bei 400 Score-Punkten. Die Fähigkeit zur Diskrimination des Modells gemessen anhand der aROC (95% Konfidenzintervall, KI) betrug 0,67 (0,66–0,68) in der Lernstudienpopulation und 0,57 (0,566–0,578) in der Teststudienpopulation. Die Diskriminationsfähigkeit variierte dabei zwischen den einzelnen Studienkohorten und schwankte von 0,65 bis 0,76 in der Lern- und von 0,56 bis 0,66 in der Teststudienpopulation. In den Kohorten der Lern- bzw. Teststudienpopulation wies das Modell eine sehr gute bzw. gute Kalibrierung auf.

In der vorliegenden Arbeit wurde erstmalig ein Risikoscore zur Prädiktion von Gewichtszunahmen entwickelt. Dieser Score basiert auf einfach zu erhebenden Informationen zu soziodemographischen, Ernährungs- und Lebensstilfaktoren und wies eine moderate Diskriminationsfähigkeit auf. Die Ergebnisse der vorliegenden Arbeit stellen die Komplexität der Prädiktion von Gewichtszunahmen heraus und weisen darauf hin, dass deren Prädiktion im Allgemeinen begrenzt sein könnte. Innerhalb dieser Grenzen scheint ein universeller Risikoscore zur Prädiktion von Gewichtszunahmen bei europäischen Erwachsenen jedoch geeignet zu sein.

References

1. World Health Organization, *Obesity: Preventing and Managing The Global Epidemic. Report of a WHO Consultation.* Report No.: 894, Geneva, 2000.
2. World Health Organization, *Fact sheet: Obesity and Overweight.* Available online: http://www.who.int/mediacentre/factsheets/fs311/en/index.html (accessed on 28 March 2011).
3. Berghofer, A., et al., *Obesity prevalence from a European perspective: a systematic review.* BMC Public Health, 2008. **8**: p. 200.
4. Branca, F., H. Nikogocian, and T. Lobstein (eds), *The Challenge of Obesity in the WHO European Region and the Strategies for Response*. 2007: WHO Regional Office for Europe: Copenhagen, Denmark.
5. Guh, D.P., et al., *The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis.* BMC Public Health, 2009. **9**: p. 88.
6. World Health Organization, *The World Health Report 2002: Reducing Risks, Promoting Healthy Life*. 2002.
7. Muller-Riemenschneider, F., et al., *Health-economic burden of obesity in Europe.* Eur J Epidemiol, 2008. **23**(8): p. 499-509.
8. Kumanyika, S.K., et al., *Population-based prevention of obesity: the need for comprehensive promotion of healthful eating, physical activity, and energy balance: a scientific statement from American Heart Association Council on Epidemiology and Prevention, Interdisciplinary Committee for Prevention (formerly the expert panel on population and prevention science).* Circulation, 2008. **118**(4): p. 428-64.
9. Cook, N.R., *Use and misuse of the receiver operating characteristic curve in risk prediction.* Circulation, 2007. **115**(7): p. 928-35.
10. Cui, J., *Overview of risk prediction models in cardiovascular disease research.* Ann Epidemiol, 2009. **19**(10): p. 711-7.
11. Buijsse, B., et al., *Risk Assessment Tools for Identifying Individuals at Risk of Developing Type 2 Diabetes.* Epidemiol Rev, 2011.
12. Gail, M.H., et al., *Projecting individualized probabilities of developing breast cancer for white females who are being examined annually.* J Natl Cancer Inst, 1989. **81**(24): p. 1879-86.
13. Freedman, A.N., et al., *Colorectal cancer risk prediction tool for white men and women without known susceptibility.* J Clin Oncol, 2009. **27**(5): p. 686-93.
14. Driver, J.A., et al., *Development of a risk score for colorectal cancer in men.* Am J Med, 2007. **120**(3): p. 257-63.
15. Cassidy, A., et al., *The LLP risk model: an individual risk prediction model for lung cancer.* Br J Cancer, 2008. **98**(2): p. 270-6.
16. Steur, M., et al., *Predicting the risk of newborn children to become overweight later in childhood: The PIAMA birth cohort study.* Int J Pediatr Obes, 2010.
17. World Cancer Research Fund and American Institute for Cancer Research, *Food, Nutrition, Physical Actitivity, and the Prevention of Cancer: A Global Perspective*. 2007, Washington DC: AICR.
18. Stubbs, R.J. and B.J. Tolkamp, *Control of energy balance in relation to energy intake and energy expenditure in animals and man: an ecological perspective.* Br J Nutr, 2006. **95**(4): p. 657-76.
19. Drewnowski, A. and B.M. Popkin, *The nutrition transition: new trends in the global diet.* Nutr Rev, 1997. **55**(2): p. 31-43.
20. Ogden, C.L., et al., *Prevalence of overweight and obesity in the United States, 1999-2004.* Jama, 2006. **295**(13): p. 1549-55.
21. Grundy, S.M., et al., *Assessment of cardiovascular risk by use of multiple-risk-factor assessment equations: a statement for healthcare professionals from the American Heart Association and the American College of Cardiology.* Circulation, 1999. **100**(13): p. 1481-92.
22. Kalaitzidis, R.G. and K.C. Siamopoulos, *The role of obesity in kidney disease: recent findings and potential mechanisms.* Int Urol Nephrol, 2011.

23. Wearing, S.C., et al., *Musculoskeletal disorders associated with obesity: a biomechanical perspective.* Obes Rev, 2006. **7**(3): p. 239-50.
24. Murugan, A.T. and G. Sharma, *Obesity and respiratory diseases.* Chron Respir Dis, 2008. **5**(4): p. 233-42.
25. Yach, D., D. Stuckler, and K.D. Brownell, *Epidemiologic and economic consequences of the global epidemics of obesity and diabetes.* Nat Med, 2006. **12**(1): p. 62-6.
26. *Modern Nutrition in Health and Disease*. Tenth ed, ed. M.E. Shils, et al. 2006, Baltimore: Lippincott Williams & Wilkins.
27. Korner, J. and R.L. Leibel, *To eat or not to eat - how the gut talks to the brain.* N Engl J Med, 2003. **349**(10): p. 926-8.
28. Gale, S.M., V.D. Castracane, and C.S. Mantzoros, *Energy homeostasis, obesity and eating disorders: recent advances in endocrinology.* J Nutr, 2004. **134**(2): p. 295-8.
29. Vimaleswaran, K.S. and R.J. Loos, *Progress in the genetics of common obesity and type 2 diabetes.* Expert Rev Mol Med, 2010. **12**: p. e7.
30. de Graaf, C., et al., *Biomarkers of satiation and satiety.* Am J Clin Nutr, 2004. **79**(6): p. 946-61.
31. Rolls, B.J., J.A. Ello-Martin, and B.C. Tohill, *What can intervention studies tell us about the relationship between fruit and vegetable consumption and weight management?* Nutr Rev, 2004. **62**(1): p. 1-17.
32. He, K., et al., *Changes in intake of fruits and vegetables in relation to risk of obesity and weight gain among middle-aged women.* Int J Obes Relat Metab Disord, 2004. **28**(12): p. 1569-74.
33. Parker, D.R., et al., *Dietary factors in relation to weight change among men and women from two southeastern New England communities.* Int J Obes Relat Metab Disord., 1997. **21**(2): p. 103-9.
34. Kahn, H.S., et al., *Stable behaviors associated with adults' 10-year change in body mass index and likelihood of gain at the waist.* Am J Public Health, 1997. **87**(5): p. 747-54.
35. Schulz, M., et al., *Food groups as predictors for short-term weight changes in men and women of the EPIC-Potsdam cohort.* J Nutr, 2002. **132**(6): p. 1335-40.
36. Buijsse, B., et al., *Fruit and vegetable intakes and subsequent changes in body weight in European populations: results from the project on Diet, Obesity, and Genes (DiOGenes).* Am J Clin Nutr, 2009. **90**(1): p. 202-9.
37. Tohill, B.C., et al., *What epidemiologic studies tell us about the relationship between fruit and vegetable consumption and body weight.* Nutr Rev, 2004. **62**(10): p. 365-74.
38. Alinia, S., O. Hels, and I. Tetens, *The potential association between fruit intake and body weight--a review.* Obes Rev, 2009. **10**(6): p. 639-47.
39. Paddon-Jones, D., et al., *Protein, weight management, and satiety.* Am J Clin Nutr, 2008. **87**(5): p. 1558S-1561S.
40. Summerbell, C.D., et al., *The association between diet and physical activity and subsequent excess weight gain and obesity assessed at 5 years of age or older: a systematic review of the epidemiological evidence.* Int J Obes (Lond), 2009. **33 Suppl 3**: p. S1-92.
41. Halkjaer, J., et al., *Intake of total, animal and plant protein and subsequent changes in weight or waist circumference in European men and women: the Diogenes project.* Int J Obes (Lond), 2010.
42. Bes-Rastrollo, M., et al., *Predictors of weight gain in a Mediterranean cohort: the Seguimiento Universidad de Navarra Study 1.* Am J Clin Nutr., 2006. **83**(2): p. 362-70; quiz 394-5.
43. French, S.A., et al., *Predictors of weight change over two years among a population of working adults: the Healthy Worker Project.* Int J Obes Relat Metab Disord, 1994. **18**(3): p. 145-54.
44. Rosell, M., et al., *Weight gain over 5 years in 21,966 meat-eating, fish-eating, vegetarian, and vegan men and women in EPIC-Oxford.* Int J Obes (Lond), 2006. **30**(9): p. 1389-96.

45. Stamler, J. and T.A. Dolecek, *Relation of food and nutrient intakes to body mass in the special intervention and usual care groups in the Multiple Risk Factor Intervention Trial.* Am J Clin Nutr, 1997. **65**(1 Suppl): p. 366S-373S.
46. Vergnaud, A.C., et al., *Meat consumption and prospective weight change in participants of the EPIC-PANACEA study.* Am J Clin Nutr, 2010. **92**(2): p. 398-407.
47. Turner-McGrievy, G.M., N.D. Barnard, and A.R. Scialli, *A two-year randomized weight loss trial comparing a vegan diet to a more moderate low-fat diet.* Obesity (Silver Spring), 2007. **15**(9): p. 2276-81.
48. Mahon, A.K., et al., *Protein intake during energy restriction: effects on body composition and markers of metabolic and cardiovascular health in postmenopausal women.* J Am Coll Nutr, 2007. **26**(2): p. 182-9.
49. Burke, L.E., et al., *A randomized clinical trial testing treatment preference and two dietary options in behavioral weight management: preliminary results of the impact of diet at 6 months--PREFER study.* Obesity (Silver Spring), 2006. **14**(11): p. 2007-17.
50. Leslie, W.S., et al., *Weight management: a comparison of existing dietary approaches in a work-site setting.* Int J Obes Relat Metab Disord, 2002. **26**(11): p. 1469-75.
51. Hodgson, J.M., et al., *Partial substitution of carbohydrate intake with protein intake from lean red meat lowers blood pressure in hypertensive persons.* Am J Clin Nutr, 2006. **83**(4): p. 780-7.
52. Yamashita, T., et al., *Arterial compliance, blood pressure, plasma leptin, and plasma lipids in women are improved with weight reduction equally with a meat-based diet and a plant-based diet.* Metabolism, 1998. **47**(11): p. 1308-14.
53. Koh-Banerjee, P. and E.B. Rimm, *Whole grain consumption and weight gain: a review of the epidemiological evidence, potential mechanisms and opportunities for future research.* Proc Nutr Soc, 2003. **62**(1): p. 25-9.
54. Fardet, A., *New hypotheses for the health-protective mechanisms of whole-grain cereals: what is beyond fibre?* Nutr Res Rev, 2010. **23**(1): p. 65-134.
55. van de Vijver, L.P., et al., *Whole-grain consumption, dietary fibre intake and body mass index in the Netherlands cohort study.* Eur J Clin Nutr, 2009. **63**(1): p. 31-8.
56. Du, H., et al., *Dietary fiber and subsequent changes in body weight and waist circumference in European men and women.* Am J Clin Nutr, 2010. **91**(2): p. 329-36.
57. Zemel, M.B., *The role of dairy foods in weight management.* J Am Coll Nutr, 2005. **24**(6 Suppl): p. 537S-46S.
58. Rajpathak, S.N., et al., *Calcium and dairy intakes in relation to long-term weight gain in US men.* Am J Clin Nutr, 2006. **83**(3): p. 559-66.
59. Vergnaud, A.C., et al., *Dairy consumption and 6-y changes in body weight and waist circumference in middle-aged French adults.* Am J Clin Nutr, 2008. **88**(5): p. 1248-55.
60. Rosell, M., N.N. Hakansson, and A. Wolk, *Association between dairy food consumption and weight change over 9 y in 19,352 perimenopausal women.* Am J Clin Nutr, 2006. **84**(6): p. 1481-8.
61. Barr, S.I., *Increased dairy product or calcium intake: is body weight or composition affected in humans?* J Nutr, 2003. **133**(1): p. 245S-248S.
62. Sammel, M.D., et al., *Weight gain among women in the late reproductive years.* Fam Pract, 2003. **20**(4): p. 401-9.
63. Mattes, R.D., *Dietary compensation by humans for supplemental energy provided as ethanol or carbohydrate in fluids.* Physiol Behav, 1996. **59**(1): p. 179-87.
64. DiMeglio, D.P. and R.D. Mattes, *Liquid versus solid carbohydrate: effects on food intake and body weight.* Int J Obes Relat Metab Disord, 2000. **24**(6): p. 794-800.
65. St-Onge, M.P., et al., *Added thermogenic and satiety effects of a mixed nutrient vs a sugar-only beverage.* Int J Obes Relat Metab Disord, 2004. **28**(2): p. 248-53.
66. Malik, V.S., M.B. Schulze, and F.B. Hu, *Intake of sugar-sweetened beverages and weight gain: a systematic review.* Am J Clin Nutr, 2006. **84**(2): p. 274-88.
67. Vartanian, L.R., M.B. Schwartz, and K.D. Brownell, *Effects of soft drink consumption on nutrition and health: a systematic review and meta-analysis.* Am J Public Health, 2007. **97**(4): p. 667-75.

68. Suter, P.M., *Is alcohol consumption a risk factor for weight gain and obesity?* Crit Rev Clin Lab Sci., 2005. **42**(3): p. 197-227.
69. Rissanen, A.M., et al., *Determinants of weight gain and overweight in adult Finns.* Eur J Clin Nutr., 1991. **45**(9): p. 419-30.
70. Wannamethee, S.G. and A.G. Shaper, *Alcohol, body weight, and weight gain in middle-aged men.* Am J Clin Nutr, 2003. **77**(5): p. 1312-7.
71. Liu, S., et al., *A prospective study of alcohol intake and change in body weight among US adults.* Am J Epidemiol, 1994. **140**(10): p. 912-20.
72. Wannamethee, S.G., et al., *Alcohol intake and 8-year weight gain in women: a prospective study.* Obes Res, 2004. **12**(9): p. 1386-96.
73. Wang, L., et al., *Alcohol consumption, weight gain, and risk of becoming overweight in middle-aged and older women.* Arch Intern Med, 2010. **170**(5): p. 453-61.
74. Sayon-Orea, C., M.A. Martinez-Gonzalez, and M. Bes-Rastrollo, *Alcohol consumption and body weight: a systematic review.* Nutr Rev, 2011. **69**(8): p. 419-31.
75. Berg, C., et al., *Eating patterns and portion size associated with obesity in a Swedish population.* Appetite, 2009. **52**(1): p. 21-6.
76. Kelly, M.T., et al., *Increased portion size leads to a sustained increase in energy intake over 4 d in normal-weight and overweight men and women.* Br J Nutr, 2009. **102**(3): p. 470-7.
77. Naska, A., et al., *Eating out, weight and weight gain. A cross-sectional and prospective analysis in the context of the EPIC-PANACEA study.* Int J Obes (Lond), 2010. **35**(3): p. 416-26.
78. Jenkins, D.J., et al., *Nibbling versus gorging: metabolic advantages of increased meal frequency.* N Engl J Med, 1989. **321**(14): p. 929-34.
79. Berteus Forslund, H., et al., *Meal patterns and obesity in Swedish women-a simple instrument describing usual meal types, frequency and temporal distribution.* Eur J Clin Nutr, 2002. **56**(8): p. 740-7.
80. Toschke, A.M., et al., *Meal frequency and childhood obesity.* Obes Res, 2005. **13**(11): p. 1932-8.
81. DiPietro, L., *Physical activity, body weight, and adiposity: an epidemiologic perspective.* Exerc Sport Sci Rev, 1995. **23**: p. 275-303.
82. Jakicic, J.M., *The effect of physical activity on body weight.* Obesity (Silver Spring), 2009. **17 Suppl 3**: p. S34-8.
83. Mekary, R.A., et al., *Physical activity patterns and prevention of weight gain in premenopausal women.* Int J Obes (Lond), 2009. **33**(9): p. 1039-47.
84. Hankinson, A.L., et al., *Maintaining a high physical activity level over 20 years and weight gain.* Jama, 2010. **304**(23): p. 2603-10.
85. Saris, W.H., et al., *How much physical activity is enough to prevent unhealthy weight gain? Outcome of the IASO 1st Stock Conference and consensus statement.* Obes Rev, 2003. **4**(2): p. 101-14.
86. Erlichman, J., A.L. Kerbey, and W.P. James, *Physical activity and its impact on health outcomes. Paper 2: Prevention of unhealthy weight gain and obesity by physical activity: an analysis of the evidence.* Obes Rev, 2002. **3**(4): p. 273-87.
87. Pate, R.R., et al., *Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine.* Jama, 1995. **273**(5): p. 402-7.
88. US Department of Health and Human Services, *Physical activity and health: A report of the Surgeon General*. 1996, US Department of Health and Human Services: Atlanta, GA.
89. US Department of Health and Human Services, *A report of the Surgeon General: physical activity and health at-a-glance*. 1996, US Department of Health and Human Services,: Atlanta, GA.
90. Haskell, W.L., et al., *Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association.* Circulation, 2007. **116**(9): p. 1081-93.

91. Blair, S.N., M.J. LaMonte, and M.Z. Nichaman, *The evolution of physical activity recommendations: how much is enough?* Am J Clin Nutr, 2004. **79**(5): p. 913S-920S.
92. Lee, I.M., et al., *Physical activity and weight gain prevention.* Jama, 2010. **303**(12): p. 1173-9.
93. Swinburn, B. and A. Shelly, *Effects of TV time and other sedentary pursuits.* Int J Obes (Lond), 2008. **32 Suppl 7**: p. S132-6.
94. Patel, S.R. and F.B. Hu, *Short sleep duration and weight gain: a systematic review.* Obesity (Silver Spring), 2008. **16**(3): p. 643-53.
95. Patel, S.R., *Reduced sleep as an obesity risk factor.* Obes Rev, 2009. **10 Suppl 2**: p. 61-8.
96. U.S. Department of Health and Human Services, *The Health Benefits of Smoking Cessation.* 1990.
97. Canoy, D., et al., *Cigarette smoking and fat distribution in 21,828 British men and women: a population-based study.* Obes Res, 2005. **13**(8): p. 1466-75.
98. Filozof, C., M.C. Fernandez Pinilla, and A. Fernandez-Cruz, *Smoking cessation and weight gain.* Obes Rev, 2004. **5**(2): p. 95-103.
99. Flegal, K.M., et al., *The influence of smoking cessation on the prevalence of overweight in the United States.* N Engl J Med, 1995. **333**(18): p. 1165-70.
100. Steyerberg, E.W., *Clinical Prediction Models - A Practical Approach to Development, Validation, and Updating*, ed. M. Gail, et al. 2009, New York: Springer Science + Business Media.
101. Royston, P. and W. Sauerbrei, *Multivariable Model-building*. 2008, West Sussex: John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester.
102. Spitz, M.R., et al., *A risk model for prediction of lung cancer.* J Natl Cancer Inst, 2007. **99**(9): p. 715-26.
103. Cook, N.R., *Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve.* Clin Chem, 2008. **54**(1): p. 17-23.
104. Gordis, L., *Epidemiology*. 4th ed. 2009, Philadelphia: Saunders, Elsevier.
105. Rosner, B., *Fundamentals of Biostatistics*. Sixth ed. 2006, United States of America: Thomson Brooks/Cole.
106. Spijkerman, A.M., et al., *The performance of a risk score as a screening test for undiagnosed hyperglycemia in ethnic minority groups: data from the 1999 health survey for England.* Diabetes Care, 2004. **27**(1): p. 116-22.
107. Youden, W.J., *Index for rating diagnostic tests.* Cancer, 1950. **3**(1): p. 32-5.
108. Bewick, V., L. Cheek, and J. Ball, *Statistics review 13: receiver operating characteristic curves.* Crit Care, 2004. **8**(6): p. 508-12.
109. Hosmer, D.W. and S. Lemeshow, *Applied Logistic Regression*. 1989, New York: John Wiley & Sons Inc.
110. Hosmer, D.W., et al., *A comparison of goodness-of-fit tests for the logistic regression model.* Stat Med, 1997. **16**(9): p. 965-80.
111. Diamond, G.A., *What price perfection? Calibration and discrimination of clinical prediction models.* J Clin Epidemiol, 1992. **45**(1): p. 85-9.
112. Gail, M.H. and R.M. Pfeiffer, *On criteria for evaluating models of absolute risk.* Biostatistics, 2005. **6**(2): p. 227-39.
113. Wilson, P.W., et al., *Prediction of coronary heart disease using risk factor categories.* Circulation, 1998. **97**(18): p. 1837-47.
114. Schulze, M.B., et al., *An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes.* Diabetes Care, 2007. **30**(3): p. 510-5.
115. Conroy, R.M., et al., *Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project.* Eur Heart J, 2003. **24**(11): p. 987-1003.
116. Gail, M.H. and J.P. Costantino, *Validating and improving models for projecting the absolute risk of breast cancer.* J Natl Cancer Inst, 2001. **93**(5): p. 334-5.
117. Kannel, W.B., D. McGee, and T. Gordon, *A general cardiovascular risk profile: the Framingham Study.* Am J Cardiol, 1976. **38**(1): p. 46-51.

118. Anderson, K.M., et al., *An updated coronary risk profile. A statement for health professionals.* Circulation, 1991. **83**(1): p. 356-62.
119. D'Agostino, R.B., Sr., et al., *General cardiovascular risk profile for use in primary care: the Framingham Heart Study.* Circulation, 2008. **117**(6): p. 743-53.
120. Kannel, W.B., et al., *Profile for estimating risk of heart failure.* Arch Intern Med, 1999. **159**(11): p. 1197-204.
121. Wolf, P.A., et al., *Probability of stroke: a risk profile from the Framingham Study.* Stroke, 1991. **22**(3): p. 312-8.
122. *Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III).* Jama, 2001. **285**(19): p. 2486-97.
123. D'Agostino, R.B., Sr., et al., *Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation.* Jama, 2001. **286**(2): p. 180-7.
124. De Backer, G., et al., *European guidelines on cardiovascular disease and prevention in clinical practice.* Atherosclerosis, 2003. **171**(1): p. 145-55.
125. Assmann, G., P. Cullen, and H. Schulte, *Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Munster (PROCAM) study.* Circulation, 2002. **105**(3): p. 310-5.
126. Hippisley-Cox, J., et al., *Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study.* Bmj, 2007. **335**(7611): p. 136.
127. Woodward, M., P. Brindle, and H. Tunstall-Pedoe, *Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC).* Heart, 2007. **93**(2): p. 172-6.
128. National Cancer Institute, *Breast Cancer Risk Assessment Tool, http://www.cancer.gov/bcrisktool/*.
129. Rockhill, B., et al., *Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention.* J Natl Cancer Inst, 2001. **93**(5): p. 358-66.
130. MacKarem, G., C.A. Roche, and K.S. Hughes, *The effectiveness of the Gail model in estimating risk for development of breast cancer in women under 40 years of age.* Breast J, 2001. **7**(1): p. 34-9.
131. Tice, J.A., et al., *Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population.* Breast Cancer Res Treat, 2005. **94**(2): p. 115-22.
132. Bondy, M.L., et al., *Validation of a breast cancer risk assessment model in women with a positive family history.* J Natl Cancer Inst, 1994. **86**(8): p. 620-5.
133. Boyle, P., et al., *Contribution of three components to individual cancer risk predicting breast cancer risk in Italy.* Eur J Cancer Prev, 2004. **13**(3): p. 183-91.
134. Chen, J., et al., *Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density.* J Natl Cancer Inst, 2006. **98**(17): p. 1215-26.
135. Gail, M.H., *Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk.* J Natl Cancer Inst, 2008. **100**(14): p. 1037-41.
136. Gail, M.H., *Value of adding single-nucleotide polymorphism genotypes to a breast cancer risk model.* J Natl Cancer Inst, 2009. **101**(13): p. 959-63.
137. Ma, E., et al., *10-Year risk of colorectal cancer: development and validation of a prediction model in middle-aged Japanese men.* Cancer Epidemiol, 2011. **34**(5): p. 534-41.
138. Rosner, B.A., et al., *Mathematical models of ovarian cancer incidence.* Epidemiology, 2005. **16**(4): p. 508-15.
139. van Vugt, H.A., et al., *Prediction of prostate cancer in unscreened men: external validation of a risk calculator.* Eur J Cancer, 2011. **47**(6): p. 903-9.
140. Bach, P.B., et al., *Variations in lung cancer risk among smokers.* J Natl Cancer Inst, 2003. **95**(6): p. 470-8.

141. Etzel, C.J. and P.B. Bach, *Estimating individual risk for lung cancer.* Semin Respir Crit Care Med, 2011. **32**(1): p. 3-9.
142. Etzel, C.J., et al., *Development and validation of a lung cancer risk prediction model for African-Americans.* Cancer Prev Res (Phila), 2008. **1**(4): p. 255-65.
143. Spitz, M.R., et al., *An expanded risk prediction model for lung cancer.* Cancer Prev Res (Phila), 2008. **1**(4): p. 250-4.
144. Tammemagi, C.M., et al., *Lung Cancer Risk Prediction: Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial Models and Validation.* J Natl Cancer Inst, 2011.
145. Cho, E., et al., *Risk factors and individual probabilities of melanoma for whites.* J Clin Oncol, 2005. **23**(12): p. 2669-75.
146. Fears, T.R., et al., *Identifying individuals at high risk of melanoma: a practical predictor of absolute risk.* J Clin Oncol, 2006. **24**(22): p. 3590-6.
147. Park, Y., et al., *Validation of a colorectal cancer risk prediction model among white patients age 50 years and older.* J Clin Oncol, 2009. **27**(5): p. 694-8.
148. Cronin, K.A., et al., *Validation of a model of lung cancer risk prediction among smokers.* J Natl Cancer Inst, 2006. **98**(9): p. 637-40.
149. Lindstrom, J. and J. Tuomilehto, *The diabetes risk score: a practical tool to predict type 2 diabetes risk.* Diabetes Care, 2003. **26**(3): p. 725-31.
150. Schmidt, M.I., et al., *Identifying individuals at high risk for diabetes: The Atherosclerosis Risk in Communities study.* Diabetes Care, 2005. **28**(8): p. 2013-8.
151. Griffin, S.J., et al., *Diabetes risk score: towards earlier detection of type 2 diabetes in general practice.* Diabetes Metab Res Rev, 2000. **16**(3): p. 164-71.
152. Schulze, M.B., et al., *Use of multiple metabolic and genetic markers to improve the prediction of type 2 diabetes: the EPIC-Potsdam Study.* Diabetes Care, 2009. **32**(11): p. 2116-9.
153. Riboli, E. and R. Kaaks, *The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition.* Int J Epidemiol, 1997. **26 Suppl 1**: p. S6-14.
154. Riboli, E., et al., *European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection.* Public Health Nutr, 2002. **5**(6B): p. 1113-24.
155. Bingham, S. and E. Riboli, *Diet and cancer--the European Prospective Investigation into Cancer and Nutrition.* Nat Rev Cancer, 2004. **4**(3): p. 206-15.
156. Kaaks, R., N. Slimani, and E. Riboli, *Pilot phase studies on the accuracy of dietary intake measurements in the EPIC project: overall evaluation of results. European Prospective Investigation into Cancer and Nutrition 10.1093/ije/26.suppl_1.S26.* Int. J. Epidemiol., 1997. **26**(suppl_1): p. S26-36.
157. Margetts, B.M. and P. Pietinen, *European Prospective Investigation into Cancer and Nutrition: validity studies on dietary assessment methods.* Int J Epidemiol, 1997. **26 Suppl 1**: p. S1-5.
158. Haftenberger, M., et al., *Overweight, obesity and fat distribution in 50- to 64-year-old participants in the European Prospective Investigation into Cancer and Nutrition (EPIC).* Public Health Nutr, 2002. **5**(6B): p. 1147-62.
159. Saris, W.H. and A. Harper, *DiOGenes: a multidisciplinary offensive focused on the obesity epidemic.* Obes Rev, 2005. **6**(2): p. 175-6.
160. Saris, W.H., *DiOGenes: an integrated multidisciplinary approach to the obesity problem in Europe.* Nutrition Bulletin, 2005. **30**(2): p. 188-193.
161. Food and Agriculture Organization, World Health Organization, and U.N. University, *Energy and protein requirements.* World Health Organization, Technical Report Series, 1985. **724**.
162. Snijder, M.B., et al., *What aspects of body fat are particularly hazardous and how do we measure them?* Int J Epidemiol, 2006. **35**(1): p. 83-92.
163. Schwarzer, G., *meta: An R package for meta-analysis.* R News, 2007. **7**(3): p. 40-45.
164. Patel, S.R., et al., *Association between reduced sleep and weight gain in women.* Am J Epidemiol, 2006. **164**(10): p. 947-54.
165. Cox, D.R., *Regression models and life-tables.* J Roy Stat Soc B, 1972. **34**: p. 187-220.

166. SAS Institute Inc., *SAS/STAT User's Guide, Version 9.2*. 2007: Cary, NC.
167. Smith-Warner, S.A., et al., *Methods for pooling results of epidemiologic studies: the Pooling Project of Prospective Studies of Diet and Cancer.* Am J Epidemiol, 2006. **163**(11): p. 1053-64.
168. Borenstein, M., L. Hedges, and H. Rothstein, *Introduction to Meta-Analysis*. 2007, http://www.meta-analysis.com/downloads/Meta%20Analysis%20Fixed%20vs%20Random%20effects.pdf, last access: 12th June 2011.
169. Higgins, J.P., et al., *Measuring inconsistency in meta-analyses.* Bmj, 2003. **327**(7414): p. 557-60.
170. Kleinbaum, D.G. and M. Klein, *Survival Analysis - A Self-Learning Text*. 2 ed, ed. M. Gail, et al. 2005, New York: Springer Science + Business Media, LLC.
171. DeLong, E.R., D.M. DeLong, and D.L. Clarke-Pearson, *Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach.* Biometrics, 1988. **44**(3): p. 837-45.
172. Rothman, K.J., S. Greenland, and T.L. Lash, *Modern Epidemiology*. Third ed. 2008, Philadelphia: Lippincott Williams & Wilkins.
173. Boeing, H., A. Korfmann, and M.M. Bergmann, *Recruitment procedures of EPIC-Germany. European Investigation into Cancer and Nutrition.* Ann Nutr Metab, 1999. **43**(4): p. 205-15.
174. Nohr, E.A., et al., *Does low participation in cohort studies induce bias?* Epidemiology, 2006. **17**(4): p. 413-8.
175. Thompson, F.E. and A.F. Subar, *Dietary assessment methodology*. 2001, http://riskfactor.cancer.gov/diet/adi/thompson_subar_dietary_assessment_methodology.pdf, last access: March 17th, 2010.
176. Drewnowski, A., *Diet image: a new perspective on the food-frequency questionnaire.* Nutr Rev, 2001. **59**(11): p. 370-2.
177. Ocke, M.C., et al., *The Dutch EPIC food frequency questionnaire. I. Description of the questionnaire, and relative validity and reproducibility for food groups.* Int J Epidemiol, 1997. **26 Suppl 1**: p. S37-48.
178. *Relative validity and reproducibility of a diet history questionnaire in Spain. I. Foods. EPIC Group of Spain. European Prospective Investigation into Cancer and Nutrition.* Int J Epidemiol, 1997. **26 Suppl 1**: p. S91-9.
179. *Relative validity and reproducibility of a diet history questionnaire in Spain. II. Nutrients. EPIC Group of Spain. European Prospective Investigation into Cancer and Nutrition.* Int J Epidemiol, 1997. **26 Suppl 1**: p. S100-9.
180. Pisani, P., et al., *Relative validity and reproducibility of a food frequency dietary questionnaire for use in the Italian EPIC centres.* Int J Epidemiol, 1997. **26 Suppl 1**: p. S152-60.
181. van Liere, M.J., et al., *Relative validity and reproducibility of a French dietary history questionnaire.* Int J Epidemiol, 1997. **26 Suppl 1**: p. S128-36.
182. Bohlscheid-Thomas, S., et al., *Reproducibility and relative validity of food group intake in a food frequency questionnaire developed for the German part of the EPIC project. European Prospective Investigation into Cancer and Nutrition.* Int J Epidemiol, 1997. **26 Suppl 1**: p. S59-70.
183. Mozaffarian, D., et al., *Changes in diet and lifestyle and long-term weight gain in women and men.* N Engl J Med, 2011. **364**(25): p. 2392-404.
184. Niedhammer, I., et al., *Validity of self-reported weight and height in the French GAZEL cohort.* Int J Obes Relat Metab Disord, 2000. **24**(9): p. 1111-8.
185. Spencer, E.A., et al., *Validity of self-reported height and weight in 4808 EPIC-Oxford participants.* Public Health Nutr, 2002. **5**(4): p. 561-5.
186. Colditz, G.A., et al., *Patterns of weight change and their relation to diet in a cohort of healthy women.* Am J Clin Nutr., 1990. **51**(6): p. 1100-5.
187. Lahti-Koski, M., et al., *Prevalence of weight cycling and its relation to health indicators in Finland.* Obes Res, 2005. **13**(2): p. 333-41.

188. Vergnaud, A.C., et al., *Weight fluctuations and risk for metabolic syndrome in an adult cohort.* Int J Obes (Lond), 2008. **32**(2): p. 315-21.
189. von Rüsten, A., et al., *Trend in Obesity Prevalence in European Adult Cohort Populations during Follow-up since 1996 and Their Predictions to 2015.* PLoS ONE, 2011. **6**(11).
190. Alssema, M., et al., *The Evaluation of Screening and Early Detection Strategies for Type 2 Diabetes and Impaired Glucose Tolerance (DETECT-2) update of the Finnish diabetes risk score for prediction of incident type 2 diabetes.* Diabetologia, 2010. **54**(5): p. 1004-12.
191. Kanaya, A.M., et al., *Predicting the development of diabetes in older adults: the derivation and validation of a prediction rule.* Diabetes Care, 2005. **28**(2): p. 404-8.
192. Parmar, M.K.B. and D. Machin, *Survival analysis - a practical approach*. 1 ed. 1995, Chichester: John Wiley & Sons.
193. Austin, P.C. and J.V. Tu, *Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality.* J Clin Epidemiol, 2004. **57**(11): p. 1138-46.
194. Royston, P., G. Ambler, and W. Sauerbrei, *The use of fractional polynomials to model continuous risk variables in epidemiology.* Int J Epidemiol, 1999. **28**(5): p. 964-74.
195. Durrleman, S. and R. Simon, *Flexible regression models with cubic splines.* Stat Med, 1989. **8**(5): p. 551-61.
196. Vickers, A.J. and A.M. Cronin, *Traditional statistical methods for evaluating prediction models are uninformative as to clinical value: towards a decision analytic framework.* Semin Oncol, 2010. **37**(1): p. 31-8.
197. Sieri, S., et al., *Alcohol consumption patterns, diet and body weight in 10 European countries.* Eur J Clin Nutr, 2009. **63 Suppl 4**: p. S81-100.
198. Bergmann, M.M., et al., *The association of lifetime alcohol use with measures of abdominal and general adiposity in a large-scale European cohort.* Eur J Clin Nutr, 2011. **65**(10): p. 1079-87.
199. Bellisle, F., R. McDevitt, and A.M. Prentice, *Meal frequency and energy balance.* Br J Nutr, 1997. **77 Suppl 1**: p. S57-70.
200. Ma, Y., et al., *Association between eating patterns and obesity in a free-living US adult population.* Am J Epidemiol, 2003. **158**(1): p. 85-92.
201. Keast, D.R., T.A. Nicklas, and C.E. O'Neil, *Snacking is associated with reduced risk of overweight and reduced abdominal obesity in adolescents: National Health and Nutrition Examination Survey (NHANES) 1999-2004.* Am J Clin Nutr, 2010. **92**(2): p. 428-35.
202. Pischon, T., U. Nothlings, and H. Boeing, *Obesity and cancer.* Proc Nutr Soc, 2008. **67**(2): p. 128-45.
203. Wang, Y., et al., *Comparison of abdominal adiposity and overall obesity in predicting risk of type 2 diabetes among men.* Am J Clin Nutr, 2005. **81**(3): p. 555-63.
204. Pischon, T., et al., *General and abdominal adiposity and risk of death in Europe.* N Engl J Med, 2008. **359**(20): p. 2105-20.
205. Coakley, E.H., et al., *Predictors of weight change in men: results from the Health Professionals Follow-up Study.* Int J Obes Relat Metab Disord, 1998. **22**(2): p. 89-96.
206. Korkeila, M., et al., *Weight-loss attempts and risk of major weight gain: a prospective study in Finnish adults.* Am J Clin Nutr, 1999. **70**(6): p. 965-75.
207. Field, A.E., et al., *Relationship of a large weight loss to long-term weight change among young and middle-aged US women.* Int J Obes Relat Metab Disord, 2001. **25**(8): p. 1113-21.
208. Kroke, A., et al., *Recent weight changes and weight cycling as predictors of subsequent two year weight change in a middle-aged cohort.* Int J Obes Relat Metab Disord, 2002. **26**(3): p. 403-9.
209. Li, S., et al., *Cumulative effects and predictive value of common obesity-susceptibility variants identified by genome-wide association studies.* Am J Clin Nutr, 2010. **91**(1): p. 184-90.

210. Khoury, M.J., R. Valdez, and A. Albright, *Public health genomics approach to type 2 diabetes.* Diabetes, 2008. **57**(11): p. 2911-4.
211. Pencina, M.J., et al., *Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond.* Stat Med, 2008. **27**(2): p. 157-72; discussion 207-12.
212. Saaristo, T., et al., *National type 2 diabetes prevention programme in Finland: FIN-D2D.* Int J Circumpolar Health, 2007. **66**(2): p. 101-12.
213. Schwarz, P.E., et al., *Development of a diabetes prevention management program for clinical practice.* Public Health Rep, 2007. **122**(2): p. 258-63.
214. German Institute of Human Nutrition, *German Diabetes Risk Score, http://drs.dife.de/en*.
215. Kumanyika, S.K., *Minisymposium on obesity: overview and some strategic considerations.* Annu Rev Public Health, 2001. **22**: p. 293-308.
216. Pandya, A., M.C. Weinstein, and T.A. Gaziano, *A comparative assessment of non-laboratory-based versus commonly used laboratory-based cardiovascular disease risk scores in the NHANES III population.* PLoS One, 2011. **6**(5): p. e20416.
217. Voko, Z., et al., *How do American stroke risk functions perform in a Western European population?* Neuroepidemiology, 2004. **23**(5): p. 247-53.
218. Empana, J.P., et al., *Are the Framingham and PROCAM coronary heart disease risk functions applicable to different European populations? The PRIME Study.* Eur Heart J, 2003. **24**(21): p. 1903-11.
219. Hippisley-Cox, J., et al., *Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: a validation study.* Heart, 2008. **94**(1): p. 34-9.
220. Costantino, J.P., et al., *Validation studies for models projecting the risk of invasive and total breast cancer incidence.* J Natl Cancer Inst, 1999. **91**(18): p. 1541-8.
221. D'Amelio, A.M., Jr., et al., *Comparison of discriminatory power and accuracy of three lung cancer risk models.* Br J Cancer, 2010. **103**(3): p. 423-9.
222. Spitz, M.R., et al., *Re: Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk.* J Natl Cancer Inst, 2009. **101**(24): p. 1731-2; author reply 1732.
223. Wilson, P.W., et al., *Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study.* Arch Intern Med, 2007. **167**(10): p. 1068-74.

Appendix

Tables

Figures

App. Table 1. Overview about the most well-known risk scores in the field of major chronic diseases and weight gain*

First author, Year [Ref]	Population, Country	Variables included in the Risk Score	Discrimination (aROC) *	Calibration (p for HLT) †	Validation in external populations (e.g.),aROC
Cardiovascular disease					
Kannel, 1976 [117] (updated in 1991 by Anderson) [118]	Framingham Study, USA	Sex-specific risk function for **CVD**: age, systolic blood pressure, serum cholesterol, cigarette smoking, glucose intolerance, left ventricular hypertrophy (Update: + HDL cholesterol)	Not reported	Not reported	For score from 1991: Men: 0.78 Women: 0.82 [216]
Wolf, 1991 [121]	Framingham Study, USA	Sex-specific risk functions for **stroke** including age, systolic blood pressure, treatment of hypertension, diabetes status, smoking, prior CVD, atrial fibrillation, left ventricular hypertrophy	Not reported	Not reported	0.75 [217]
Wilson, 1998 [113]	Framingham Study, USA	Sex-specific risk function for **CHD** including age, diabetes status, smoking, categories of blood pressure, total cholesterol (or LDL), and HDL	Men: 0.73 Women: 0.76	Not reported	D'Agostino, 2001, Hense, 2003, Liu, 2004, aROC betw. 0.63 – 0.88
Assmann, 2002 [125]	PROCAM study (only among men), Germany	Risk function for **acute coronary event** including age, LDL cholesterol, smoking, HDL cholesterol, systolic blood pressure, family history of premature myocardial infarction, diabetes status, triglycerides	0.82	P > 0.3	0.61 and 0.64 among two cohorts of men; overestimation of risk [218]
Conroy, 2003 [115]	SCORE project (12 European cohort studies), Europe	Sex-specific models for fatal **CVD**: age, smoking status, systolic blood pressure, ratio of total cholesterol and HDL cholesterol stratified by low- and high-risk populations	0.71 to 0.84 across populations	Not reported	Men: 0.79 Women: 0.79 [216]
Hippisley-Cox, 2007 [126]	QRESEARCH database of health records, UK	For **CVD**: Age, ratio of total cholesterol to HDL, systolic blood pressure, BMI, family history of premature CVD, smoking status, deprivation score, use of at least on blood pressure treatment, interaction of systolic blood pressure and blood pressure treatment	Men: 0.77 Women: 0.79	Overall good calibration, overestimation in the first decile	Men: 0.76 Women: 0.79 [219]
D'Agostino, 2008 [119]	Framingham Study, US	For **CVD**: Sex-specific risk functions including age, total and HDL cholesterol, systolic blood pressure, treatment for hypertension, smoking, diabetes status	Men: 0.76 Women: 0.79	Men: 0.14 Women: 0.56	Men: 0.78 Women: 0.83 [216]

App. Table 1. cont.

First author, Year [Ref]	Population, Country	Variables included in the Risk Score	Discrimination (aROC) *	Calibration (p for HLT) †	Validation in external populations (e.g.),aROC
Cancer					
Breast					
Gail, 1989 [12] (Revised in 1999 by Costantino) [220]	Breast Cancer Detection Demonstration Project, US	Age (<50 years vs. ≥50 years), age at menarche, number of previous breast biopsies, age at first live birth, number of first degree relatives with breast cancer, interaction of age category and number of biopsies, interaction of age at first birth and number of affected relatives	Not reported	Not reported	0.58 [129] Gail model and Gail model + mammographic breast density, aROCs: 0.60 and 0.64 [131]
Chen, 2006 [134]	Breast Cancer Detection Demonstration Project	Gail model Gail model + mammographic density	0.60 0.64	Well calibrated	-
Gail, 2008 [135]	No ind. data, distribution of risk factors taken from the 2000 National Health Interview Survey	Gail model + information on 7 SNPs	0.63	Not reported	-
Colorectum					
Driver, 2007 [14]	Physician's Health Study	Age, BMI, history of smoking, alcohol use	0.70	0.91	-
Freedman, 2009 [13]	Population-based case-control study combined with national incidence data	**For men:** sigmoidoscopy/colonoscopy in the last 10 years, colorectal cancer in first degree relatives, aspirin and nonsteroidal anti-inflammatory drug use, smoking, BMI, vigorous activity, vegetable intake **For women:** sigmoidoscopy/colonoscopy, polyp history, colorectal cancer in first degree relatives, aspirin and nonsteroidal anti-inflammatory drug use, BMI, vigorous activity, vegetable intake, hormone replacement therapy, estrogen exposure	Not reported	Not reported	0.61 both for men and women, also well-calibrated [147]
Ma, 2010 [137]	Japan Public Health Center-based Prospective Cohort Study II (only men)	Age, BMI, alcohol consumption, smoking status, daily physical activity level	0.70	0.08	Validation in Cohort I: 0.64

App. Table 1 cont.

First author, Year [Ref]	Population, Country	Variables included in the Risk Score	Discrimination (aROC) *	Calibration (p for HLT) †	Validation in external populations (e.g.),aROC
Cancer					
Lung					
Bach, 2003 [140]	Carotene and Retinol Efficacy Trial (subjects with a presence of smoking history), US	Age, duration of smoking, average number of cigarettes per day, duration of abstinence, history of asbestos, study drug	0.72	Excellent calibration (calibration plot)	0.69 [148] 0.66 [221]
Spitz, 2007 [102] (extended in 2008 and 2009) [143, 222]	Case-control study combined with national incidence data, US	**Never smokers:** exposure to environmental tobacco smoke, family history of cancer	0.59	0.78	Overall: 0.69
		Former smokers: emphysema, dust exposure, family history of cancer, age stopped smoking, no prior hay fever	0.63	0.71	0.70
		Current smokers: emphysema, pack-years, dust exposure, asbestos exposure, family history of cancer, no prior hay fever	0.63	0.69	0.68 [221]
		Extended model among former smokers and smokers (2008): basic model + DNA repair capacity and mutagen sensitivity data	Increase from 0.67 to 070 and 0.68 to 0.73, respectively	0.61 and0.43, respectively	
		Extended model (2009): basic model + three SNPs	Increase from 0.66 to 0.67	Not reported	
Cassidy, 2008 [15]	Population-based case control study combined with age-gender-specific incidence rates, UK	Smoking duration, prior diagnosis of pneumonia, occupational exposure to asbestos, prior diagnosis of malignant tumour, family history of lung cancer	0.71 (0.70 after 10-fold cross-validation)	Not reported	0.69 [221]
Etzel, 2008 [142]	Case-control study among African-Americans combined with national incidence rates, US	Smoking status, pack-years smoked, age at smoking cessation, number of years since smoking cessation, chronic obstructive pulmonary disease or hay fever, exposure to asbestos or wood dusts	0.75	Not reported	0.63 in two other case-control studies on lung cancer in African-Americans [142]

App. Table 1 cont.

First author, Year [Ref]	Population, Country	Variables included in the Risk Score	Discrimination (aROC) *	Calibration (p for HLT) †	Validation in external populations (e.g.),aROC
Cancer					
Tammemagi, 2011 [144]	Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (only among subjects of the control arm), US	**For all:** Age, sex, education, BMI, family history of lung cancer, chronic obstructive pulmonary disease, recent chest x-ray, smoking status, pack-years smoked, smoking duration	0.86	0.27	0.84 in the intervention arm of the study
		For ever smokers: basic model + smoking quit time	0.81	0.42	0.78 in the intervention arm
Ovary					
Rosner, 2005	Nurses Health Study, US	Age at menarche, age at menopause, use of oral contraceptives, tubal ligation	0.60	0.63	-
Melanom					
Cho, 2005	Nurses Health Study, Nurses Health Study II, Health Professionals Follow-up Study, US	Age, male sex, family history of melanoma, higher number of nevi, history of sunburn, light hair colour	0.62	0.41	-
Fortes, 2010	Italy	Nevi, skin and hair colour, freckles, sunburns in childhood	0.79	-	0.79 in Brazilian population

App. Table 1 cont.

First author, Year [Ref]	Population, Country	Variables included in the Risk Score	Discrimination (aROC) *	Calibration (p for HLT) †	Validation in external populations (e.g.),aROC
Diabetes ‡					
Schmidt, 2005 [150]	Atherosclerosis Risk in Communities study, US	**Clinical model**: age, ethnicity, parental history, systolic blood pressure, waist circumference, height Clinical model + fasting plasma glucose Clinical model + fasting plasma glucose, triglycerides, HDL	0.71 0.78 0.80	Not reported	Validation in several external populations with aROCs ranging from 0.70 to 0.87 (see [11])
Griffin, 2000 [151]	Population from general practices, UK	Age, sex, BMI, smoking status, corticosteroid use, antihypertensive use, family history	0.80	Not reported	Validation in external populations aROCs from 0.58 to 0.81 (see [11])
Lindström, 2003 [149]	FINRISK, Finland	**Concise model**: age, BMI, WC, history of antihypertensive use, previous diabetes **Full model**: concise model + physical inactivity, fruit and vegetable intake	0.857 0.860	Not reported	Validation in external populations: aROCs from 0.65 to 0.87 (see [11])
Schulze, 2007 [114]	EPIC-Potsdam, Germany	**Full model**: age, WC; height, hypertension, physical activity, smoking, and consumption of whole-grain bread, red meat, coffee, moderate alcohol	0.84	Well calibrated	Validation in EPIC-Heidelberg, aROC = 0.82 [114]
Wilson, 2007 [223]	Framingham Offspring Study, US	**Personal model**: age, sex, parental history, BMI **Simple clinical model with categorical variables**: age, sex, parental history, BMI, WC, fasting glucose, HDL, triglycerides, hypertension **Simple clinical model with continuous variables**	0.72 0.85 0.88	Not reported	Validation in several external populations, aROCs for simple clinical model ranging from 0.73 to 86 (see [11])
Weight gain/Overweight					
Steur, 2010 [16]	Prevention and Incidence of Asthma and Mite Allergy birth cohort, The Netherlands	Risk of overweight at the age of 8 years among newborns: paternal and maternal BMI, female gender, smoking in the parental house, birth weight, hospital delivery	0.75	0.30	-

* The overview is restricted to a selection of risk score mainly based on publicity of the scores.
† Area under a receiver operating characteristic curve.
‡ HLT, Hosmer-Lemeshow-Test.
¶ Adapted from [11]. For a comprehensive review see [11].

App. Table 2. General characteristics at baseline and follow-up in the single cohorts of the validation population (n=115,099)

	France	IT-Nap	Spain	UK-GP	UK-HC	NL-Utr	GER-Hd	SWE-Mal
N	45,841	359	23,763	2,079	12,684	6,248	14,570	9,557
Men (%)	–	–	37.6	27.4	23.1	–	44.2	37.7
Age at baseline (y)	50.8 (5.1)	49.1 (4.2)	47.0 (6.93)	49.0 (5.3)	45.6 (6.6)	54.0 (3.3)	49.4 (7.5)	52.5 (4.1)
Duration of follow-up (y)	3.42 (0.76)	8.83 (1.27)	3.29 (0.34)	5.58 (0.58)	5.32 (0.47)	4.36 (0.81)	2.06 (0.57)	4.96 (0.51)
Anthropometry								
Weight								
Baseline (kg)	59.9 (7.5)	62.8 (6.9)	68.7 (9.8)	67.4 (11.2)	64.9 (10.2)	65.9 (8.4)	71.1 (11.7)	69.6 (11.4)
Follow-up (kg)	61.4 (8.4)	67.1 (7.9)	70.1 (10.3)	70.6 (11.7)	68.1 (11.3)	69.9 (9.2)	72.1 (12.0)	72.8 (11.8)
Absolute change (kg)	1.55 (3.1)	4.27 (4.5)	1.34 (3.53)	3.13 (4.6)	3.22 (4.5)	4.03 (4.0)	1.04 (2.8)	3.22 (4.12)
Annual change (g/y)	461.9 (978)	492.2 (524)	413.9 (1088)	560.6 (820)	607.1 (844)	951.6 (966)	544.9 (1457)	650.1 (836)
% change	2.58 (5.1)	6.93 (7.1)	2.04 (5.2)	4.88 (6.9)	5.04 (6.8)	6.26 (6.1)	1.52 (4.0)	4.84 (6.0)
BMI								
Baseline (kg/m²)	22.9 (2.5)	25.3 (2.5)	26.0 (2.4)	23.9 (2.7)	23.2 (2.7)	24.1 (2.7)	24.5 (2.8)	24.1 (2.8)
Follow-up (kg/m²)	23.5 (2.8)	27.1 (3.0)	26.5 (2.6)	25.1 (3.1)	24.3 (3.1)	25.6 (3.0)	24.9 (2.9)	25.2 (2.9)
Obese at follow-up (%)	2.4	1.2	8.7	6.5	5.1	8.0	3.3	5.6
Physical activity								
At Work (%)								
Sedentary	23.4	31.8	23.5	59.2	46.9	24.0	52.2	43.0
Standing	53.6	7.5	33.0	17.7	27.0	21.3	23.4	34.8
Manual	2.1	12.0	11.7	5.7	6.6	21.3	4.8	9.5
Non-workers	21.0	48.8	31.9	17.5	19.5	33.3	19.6	12.6
Sports (hours/week)	1.25 (1.6)	0.28 (0.93)	1.00 (2.3)	1.71 (2.5)	2.14	1.56 (2.2)	1.80 (2.5)	1.27 (2.3)
Education (%)								
No school / primary school	10.0	21.2	66.6	–	–	12.9	23.3	31.2
Techn. / profess. school	0	5.0	10.3	39.0	28.9	33.4	33.7	28.0
Secondary school	51.1	52.7	8.0	15.2	16.6	32.2	7.9	10.7
University degree	38.9	21.2	15.1	45.8	54.6	21.5	35.1	30.1

App. Table 2 cont.

	France	IT-Nap	Spain	UK-GP	UK-HC	NL-Utr	GER-Hd	SWE-Mal
Smoking habits (%)								
Non-smokers	86.5	60.9	67.6	87.3	89.6	75.5	75.4	68.8
Stable smokers	8.1	12.3	25.2	8.2	5.8	17.4	17.3	22.4
Started smokers	3.9	17.0	2.9	1.5	2.3	1.5	2.7	1.9
Quitters	1.5	9.8	4.3	3.0	2.3	5.6	4.6	6.9
Alcohol use (%)								
No alcohol	12.9	45.1	34.6	0.1	0.1	11.6	4.3	12.7
> 0 - ≤ 6g/d	37.4	14.8	21.5	44.5	47.9	39.5	31.1	22.9
> 6 - ≤ 18g/d	29.4	17.3	16.4	39.0	37.2	26.5	28.4	40.6
> 18 - ≤ 30g/d	11.9	16.7	9.0	6.2	5.5	13.3	16.9	16.2
> 30 - ≤ 60g/d	7.3	5.3	12.0	9.5	8.6	8.6	14.5	6.8
> 60g/d	1.2	0.8	6.5	0.7	0.8	0.6	4.8	0.8
Dietary factors (g/d)								
Fruits and vegetable	555.5 (236)	687.5 (247)	614.4 (295)	496.7 (239)	588.4 (284)	377.3 (157)	243.4 (115)	355.2 (181)
Red and processed meat	80.48 (45)	62.5 (23)	81.8 (51)	63.6 (38)	24.8 (34)	77.0 (40)	78.3 (56)	85.8 (48)
Poultry	21.0 (19)	23.9 (17)	34.9 (26)	29.2 (20)	12.5 (19)	11.5 (11)	12.2 (14)	12.4 (17)
Fish	36.7 (26)	52.5 (28)	62.4 (41)	39.4 (25)	27.3 (32)	10.5 (11)	18.7 (19)	38.1 (30.4)
Milk and yogurt	202.8 (186)	108.7 (98)	265.6 (178)	384.6 (189)	326.1 (212)	359.5 (239)	173.8 (198)	329.9 (247)
Pasta and rice	73.8 (50)	164.0 (66)	60.4 (34)	64.6 (41)	84.0 (54)	37.3 (36)	37.7 (26)	37.8 (42)
Bread	126.7 (80)	173.3 (86)	145.3 (87)	87.9 (57)	89.9 (61)	115.8 (44)	134.5 (71.9)	122.1 (72)
Vegetable oil	5.8 (4)	30.2 (7.7)	27.9 (13)	3.7 (3)	2.9 (2.7)	3.6 (3)	6.6 (4)	1.8 (3)
Butter and margarine	10.5 (10)	2.9 (2)	2.67 (6)	20.2 (15)	18.5 (15)	17.0 (11)	14.7 (11)	35.8 (24)
Chocolate	9.6 (17)	3.1 (5)	3.2 (12)	13.1 (17)	10.8 (15)	8.0 (9)	11.9 (20)	8.0 (11)
Cake and cookies	38.9 (35)	50.7 (32)	36.9 (47)	60.8 (50)	50.8 (42)	33.3 (23)	63.0 (59)	43.2 (36)
Soft drinks	9.4 (44)	9.8 (40)	26.3 (89)	113.3 (178)	97.2 (178)	76.4 (101)	62.6 (180)	84.5 (147)

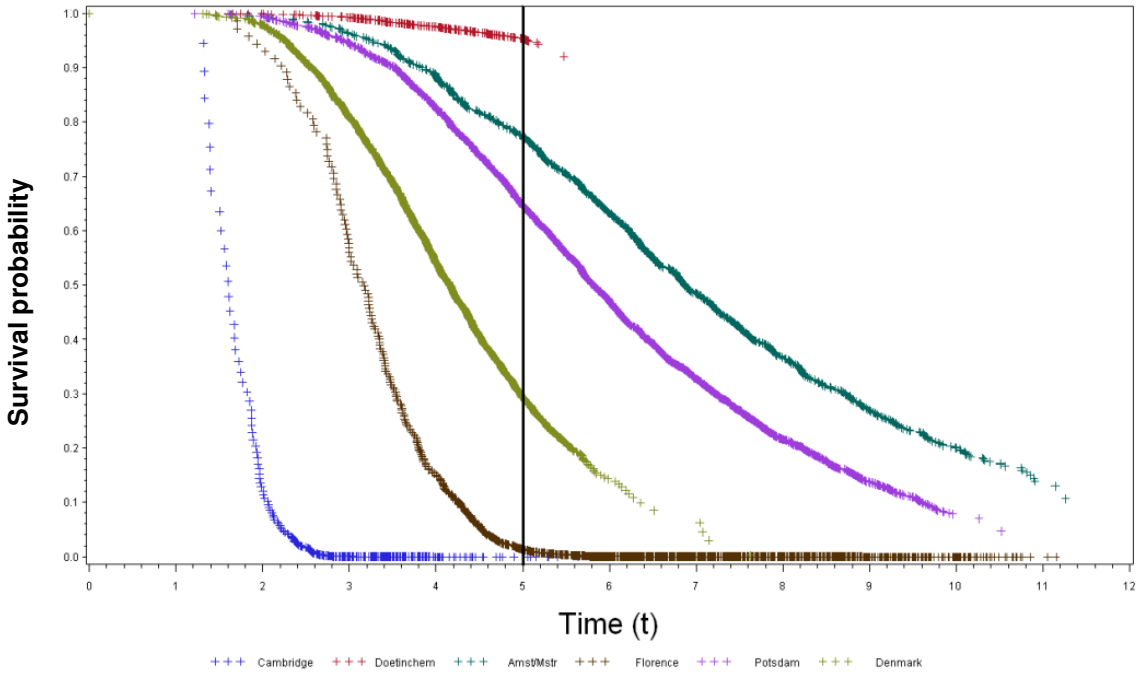
IT-Nap = Italy (Naples) cohort; UK-GP = United Kingdom (General population) cohort; UK-HC = United Kingdom (Health Conscious) cohort; NL-Utr = The Netherlands (Utrecht) cohort; GER-Hd = Germany (Heidelberg) cohort; SWE-Mal = Sweden (Malmö) cohort; Data are means (SD) or percentages.
† Defined as the sum of hours spent on cycling and doing sports.

App. Table 3. Significant* predictors of substantial weight gain separately by cohort in the training sample (n=47,203)

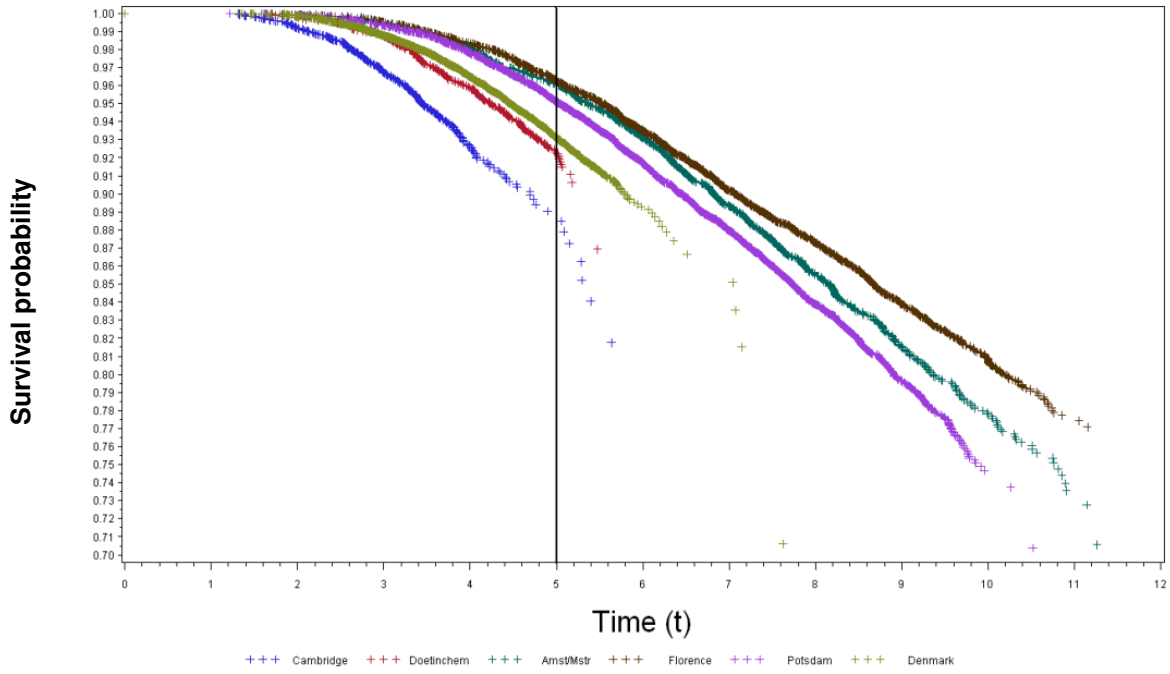
	UK-Nor	NL-Doe	NL-AmMa	IT-Flo	GER-Pot	DK-CopAa
Socio-demographic and anthropo-metric factors		Age ↓	Age ↓	Age ↓	Age ↓	Age ↓
	Sex (F vs. M) ↑		Sex (F vs. M) ↑		Sex (F vs. M) ↑	
		Baseline weight ↓	Baseline weight ↓	Baseline weight ↓	Baseline weight ↓	Baseline weight ↓
	Baseline height ↓			Baseline height ↓		Baseline height ↑
					Techn. school ↓	Techn. school ↓
		Sec. school ↓			Sec. school ↓	Sec. school ↓
	University ↓	University ↓	University ↓		University ↓	University ↓
Lifestyle factors					Start smoking ↓	
	Smoking cessation ↑	Smoking cessation↑	Smoking cessation ↑	Smoking cessation ↑	Smoking cessation ↑	Smoking cessation ↑
						Non-working ↑
		Sports (h/week) ↓		Sports (h/week) ↓	Sports (h/week) ↓	Sports (h/week) ↓
		No alcohol ↑				No alcohol ↑
			Alcohol >6 - ≤18g/d ↓	Alcohol >6 - ≤18g/d ↓	Alcohol >6 - ≤18g/d ↓	Alcohol >6 - ≤18g/d ↓
			Alc. >18 to ≤30g/d ↓	Alc. >18 to ≤30g/d ↓		Alc. >18 to ≤30g/d ↓
	Alc. >30 to ≤60g/d ↓		Alc. >30 to ≤60g/d ↓			Alc. >30 to ≤60g/d ↓
						Alcohol >60g/d ↓
Dietary factors					Fruits/vegetables ↓	Fruits/vegetables ↓
		Red/proc. meat ↑	Red/proc. meat ↑	Red/proc. meat ↑	Red/proc. meat ↑	Red/proc. meat ↑
			Poultry ↑	Poultry ↑		Poultry ↑
				Fish ↑		
				Pasta/rice ↓		
			Bread ↓	Bread ↓		Bread ↓
	Cake and cookies ↓			Cake and cookies ↓		
	Vegetable oil ↑			Vegetable oil ↓		
		Butter/marg. ↓				Butter/marg. ↓
						Milk/yogurt ↓
					Chocolate ↑	
	Soft drinks ↑				Soft drinks ↑	Soft drinks ↑

* Significance at the 10% level.
F vs. M = female vs. male, tech. school = technical school, sec. school = secondary school, red/proc meat = red and processed meat, butter/marg. = butter and margarine.

a)



b)



App. Figure 1. Baseline survival functions for centres of the training sample with a) all covariates set to zero and b) all covariates set to mean values

Let $X = (x_1,\ldots,x_m)$ be a vector of predictors. The survival function can be written as

$$S(t, X=0) = S_0(t)^{\exp(\sum_{i=1}^{m}\beta_i * x_i)} = S_0(t)$$

with the baseline survival function $S_0(t)$ estimated for an individual with all predictors equal to 0 (standard method) [170].
Alternatively, the "baseline" survival function can be expressed based on the "average individual" ($S_M(t)$), where all predictor values equal their respective mean value:

$$S_M(t) = S_0(t)^{\exp(\sum_{i=1}^{m}\beta_i * \bar{x}_i)}$$

The corresponding survival function, corrected for the means of participants' values is then:

$$S(t, X) = S_M(t)^{\exp(\sum_{i=1}^{m}\beta_i * (x_i - \bar{x}_i))}$$

Hence, it has to be proven that:

$$S(t, X) = S_M(t)^{\exp(\sum_{i=1}^{m}\beta_i * (x_i - \bar{x}_i))} = S_0(t)^{\exp(\sum_{i=1}^{m}\beta_i * x_i)}$$

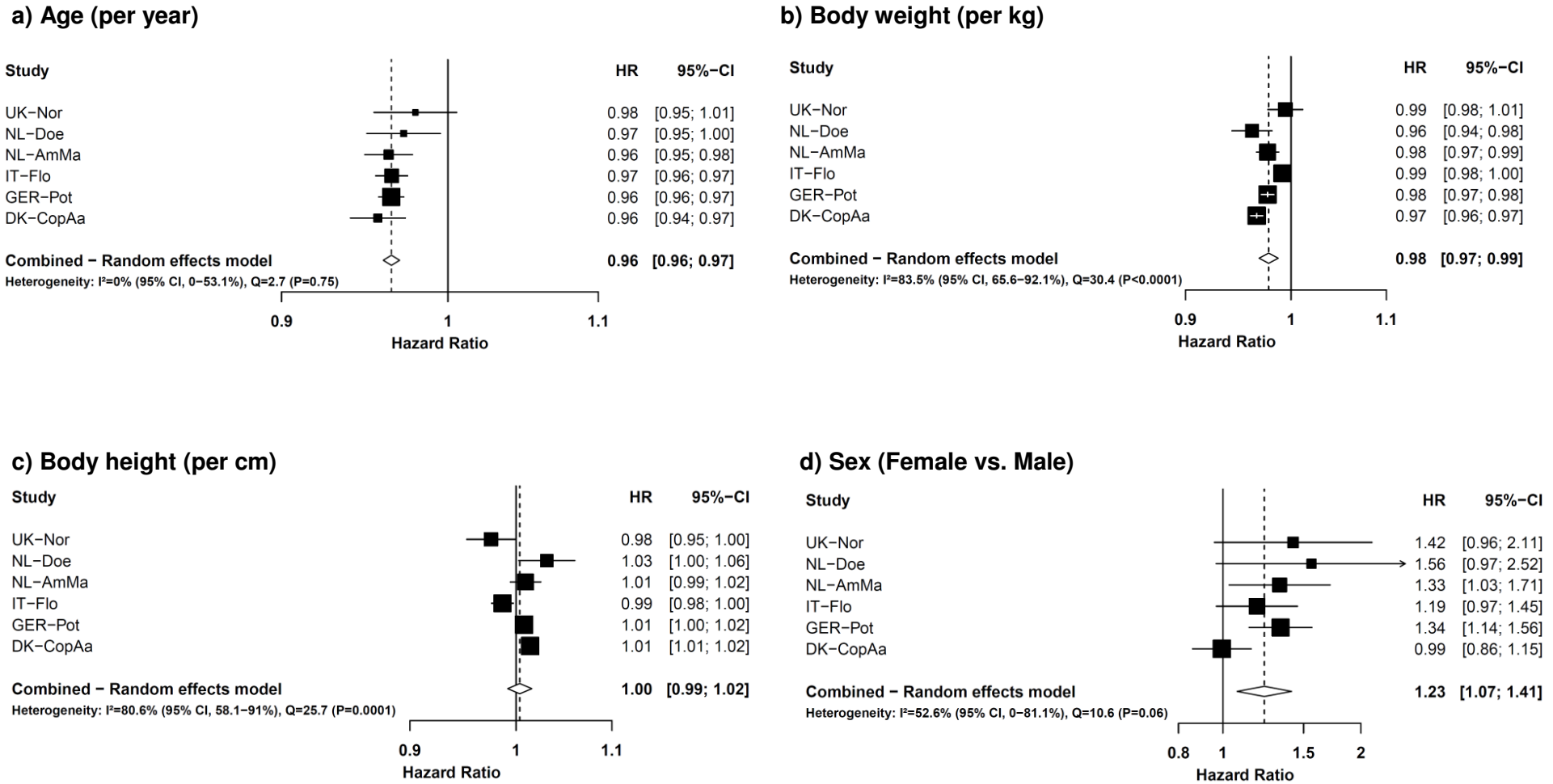
To prove this, the

$$\begin{aligned} S(t, X) &= S_M(t)^{\exp(\sum_{i=1}^{m}\beta_i * (x_i - \bar{x}_i))} \\ &= S_M(t)^{\exp(\sum_{i=1}^{m}\beta_i * x_i - \sum_{i=1}^{m}\beta_i * \bar{x}_i)} \\ &= S_M(t)^{\exp(\sum_{i=1}^{m}\beta_i * x_i) * \exp(-\sum_{i=1}^{m}\beta_i * \bar{x}_i)} \\ &= \left(S_M(t)^{\exp(-\sum_{i=1}^{m}\beta_i * \bar{x}_i)}\right)^{\exp(\sum_{i=1}^{m}\beta_i * x_i)} \end{aligned}$$

Finally, it has to be shown that:

$$\begin{aligned} S_0(t) &= S_M(t)^{\exp(-\sum_{i=1}^{m}\beta_i * \bar{x}_i)} \\ &= \left(S_0(t)^{\exp(\sum_{i=1}^{m}\beta_i * \bar{x}_i)}\right)^{\exp(-\sum_{i=1}^{m}\beta_i * \bar{x}_i)} \\ &= S_0(t)^{\exp(\sum_{i=1}^{m}\beta_i * \bar{x}_i - \sum_{i=1}^{m}\beta_i * \bar{x}_i)} \\ &= S_0(t)^{\exp(0)} \\ S_0(t) &= S_0(t) \end{aligned}$$

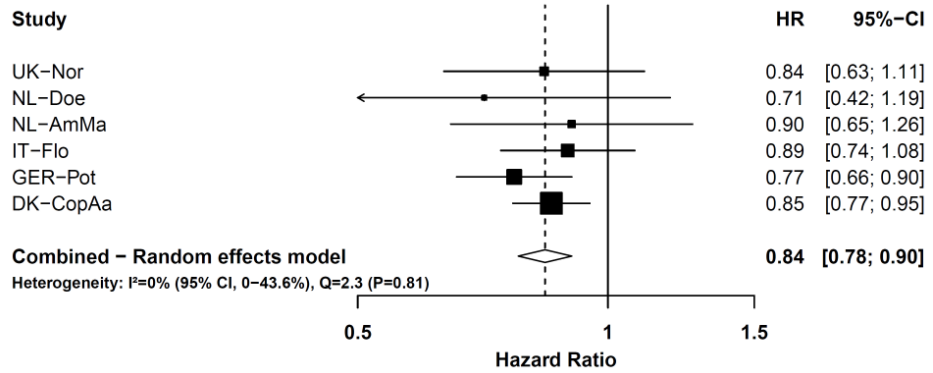
App. Figure 2. Proof that the same estimates of survival probability are obtained irrespective of setting all covariates to zero or to the average of all participants' values and correcting afterwards for these average values



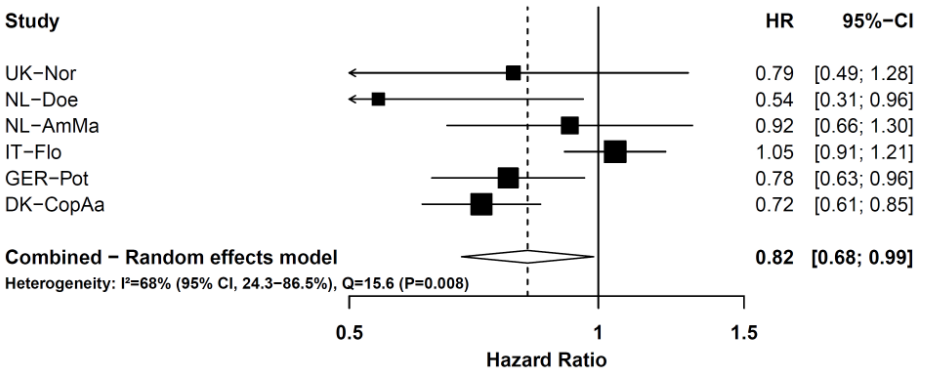
App. Figure 3. Associations of predictors with substantial weight gain across the six cohorts of the training sample and meta-analysis

Substantial weight gain was defined as gaining ≥10% of baseline weight during follow-up.

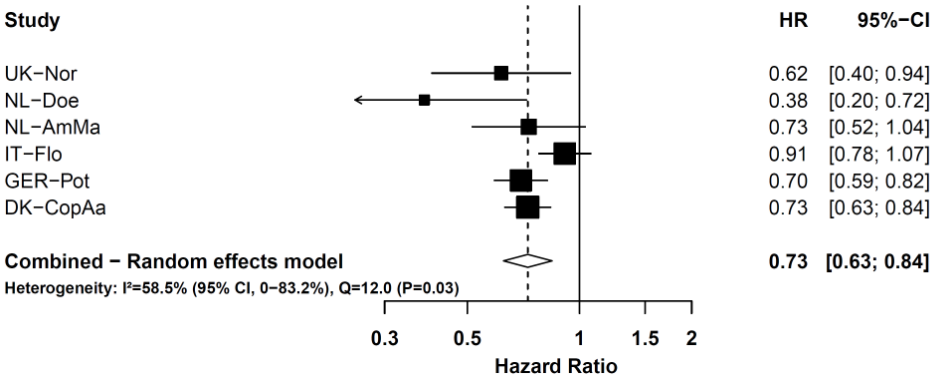
e) Technical school degree vs. no/primary school



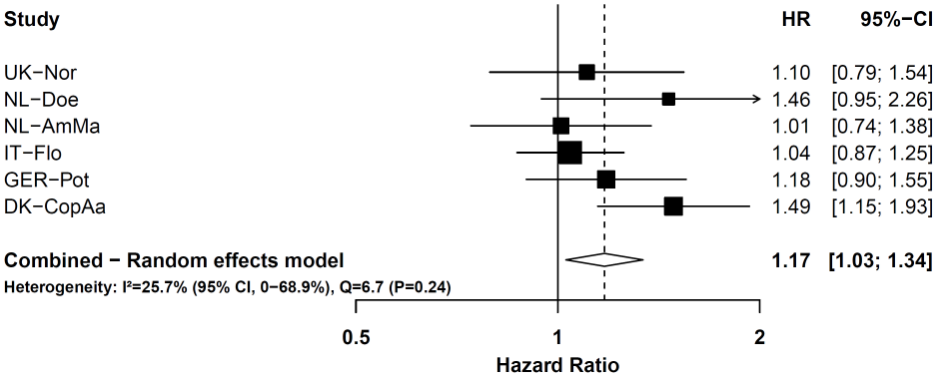
f) Secondary school vs. no/primary school



g) University vs. no/primary school

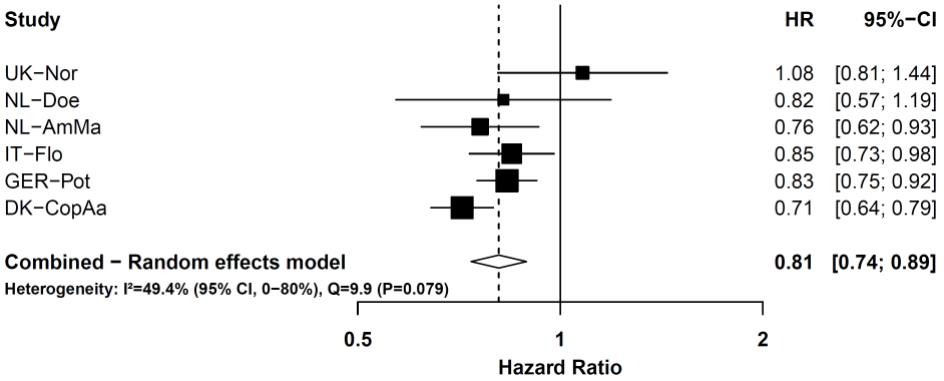


h) No alcohol vs. low alcohol (>0 - ≤ 6g/d)

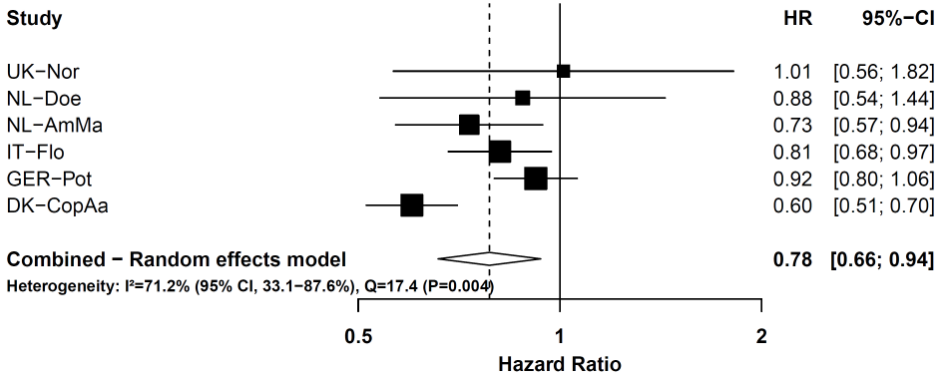


App. Figure 3 cont.

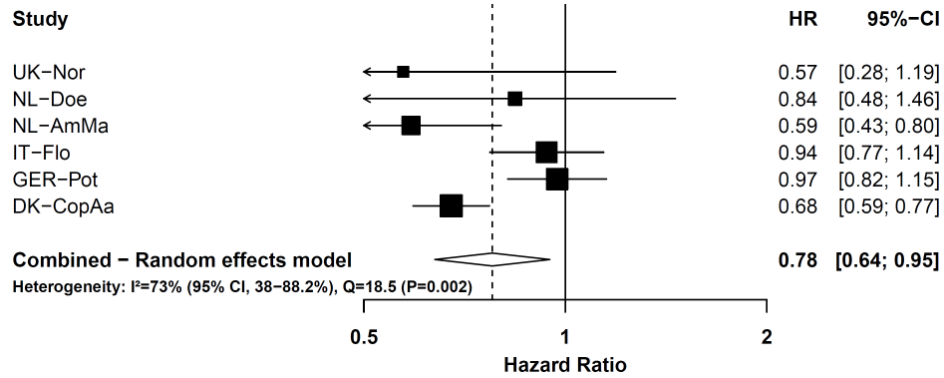
i) Alcohol >6 - ≤18g/d vs. low alcohol (>0 - ≤ 6g/d)



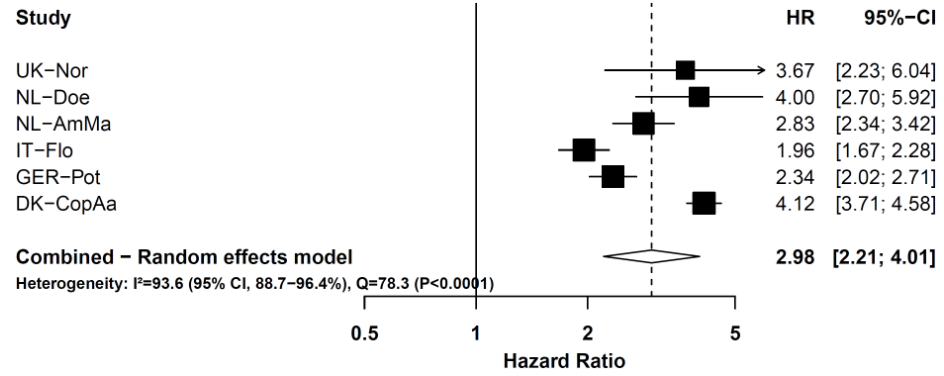
j) Alcohol >18 - ≤30g/d vs. low alcohol (>0 - ≤ 6g/d)



k) Alcohol >30 - ≤60g/d vs. low alcohol (>0 - ≤ 6g/d)

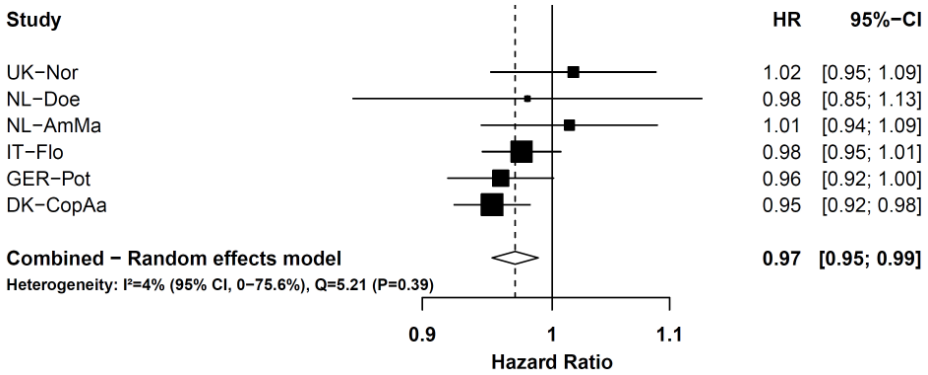


l) Smoking cessation vs. other smoking behaviours

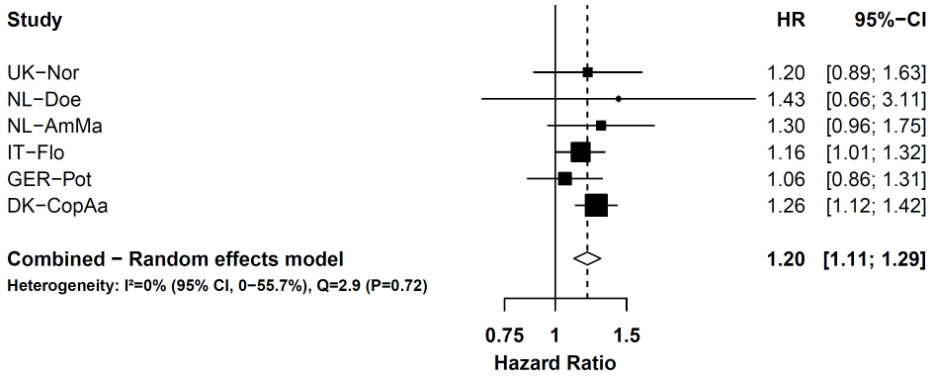


App. Figure 3 cont.

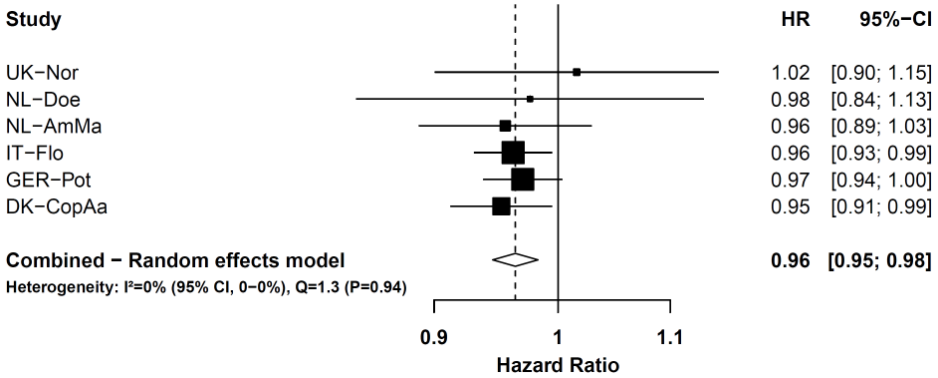
m) Intake of fruits and vegetables (per 125g/d)



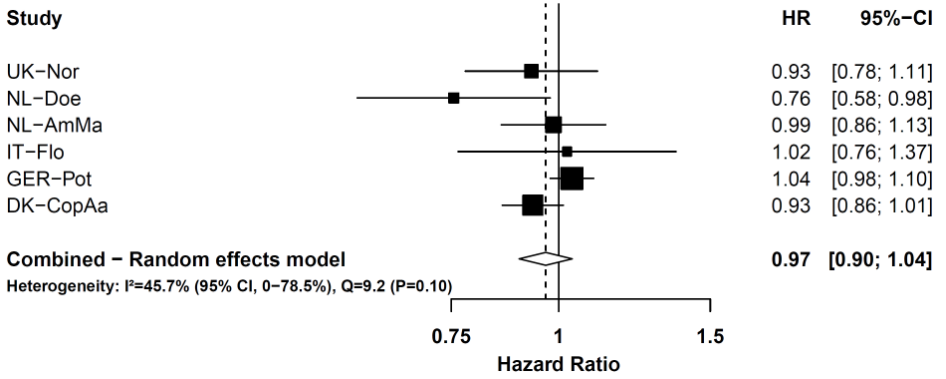
n) Intake of poultry (per 50g/d)



o) Intake of bread (per 50g/d)

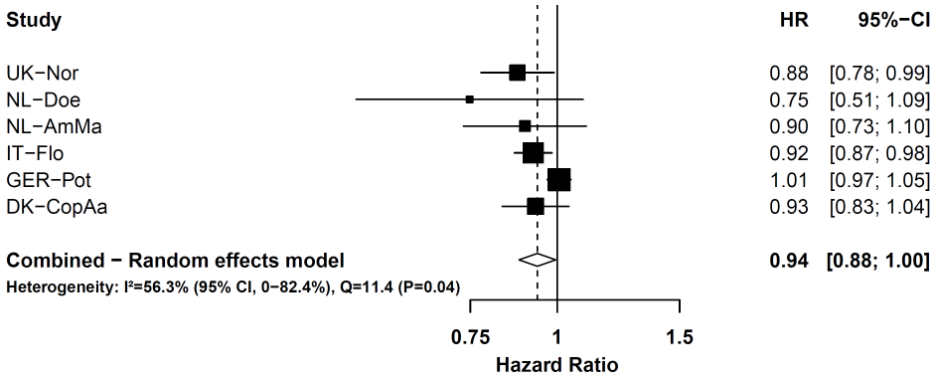


p) Intake of butter and margarine (per 20g/d)

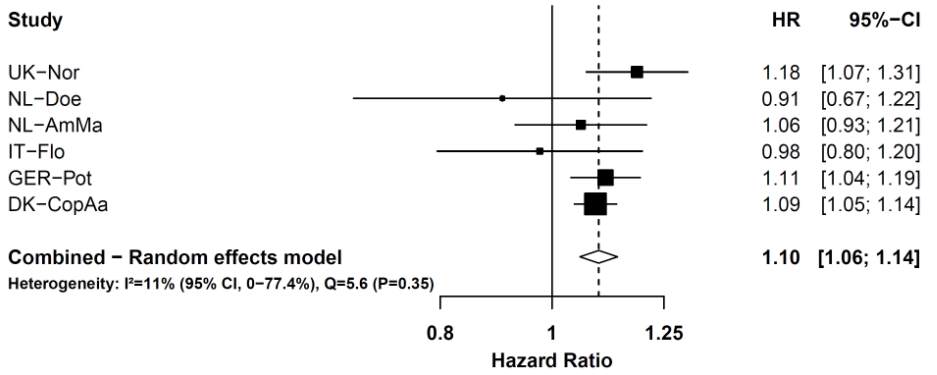


App. Figure 3 cont.

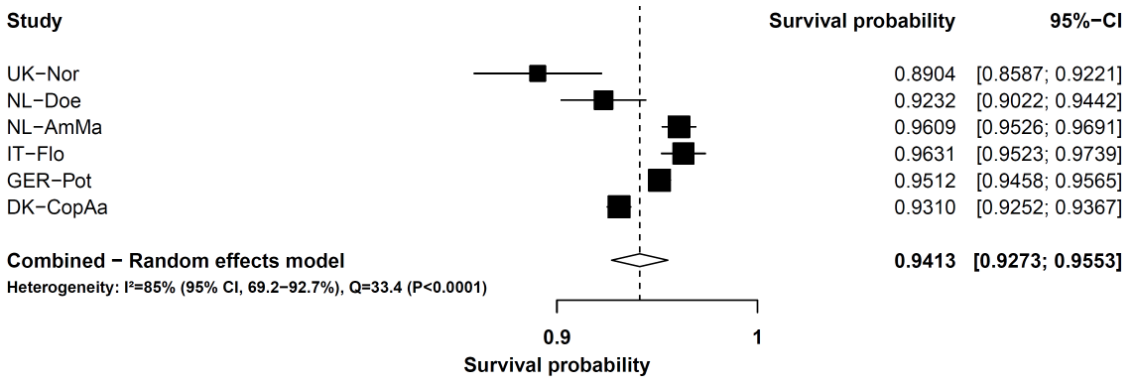
q) Intake of cakes and cookies (per 50g/d)



r) Intake of soft drinks (per 250g/d)

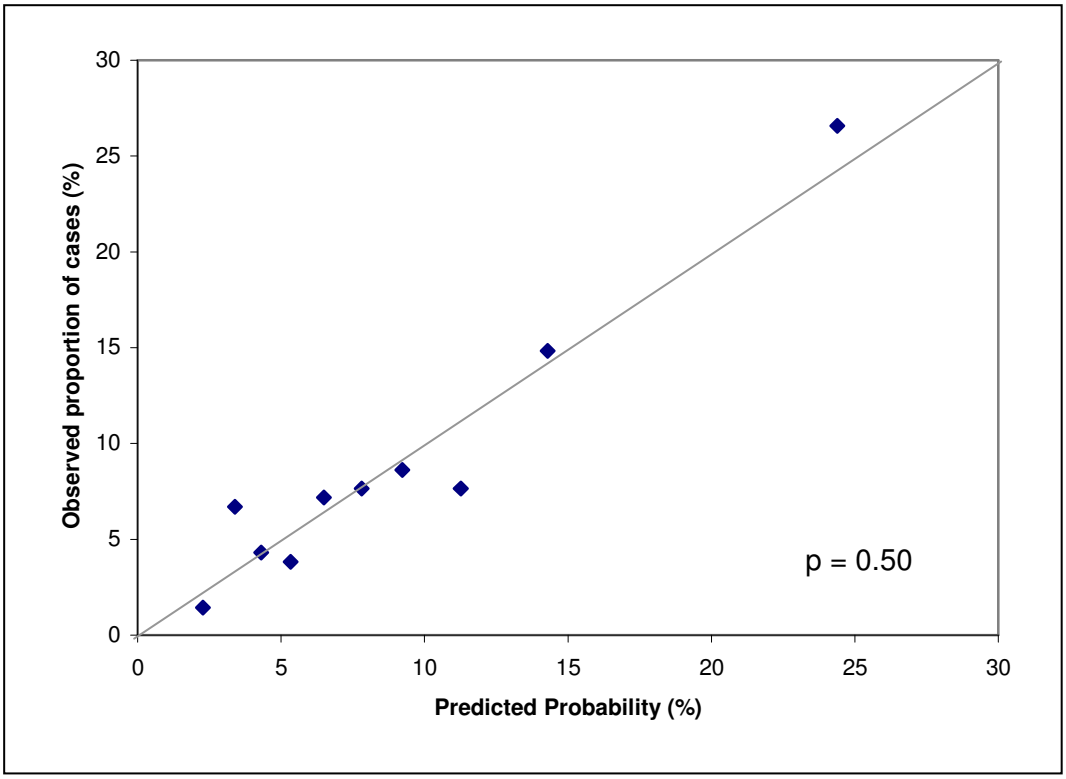


s) Baseline survival probability at t=5years

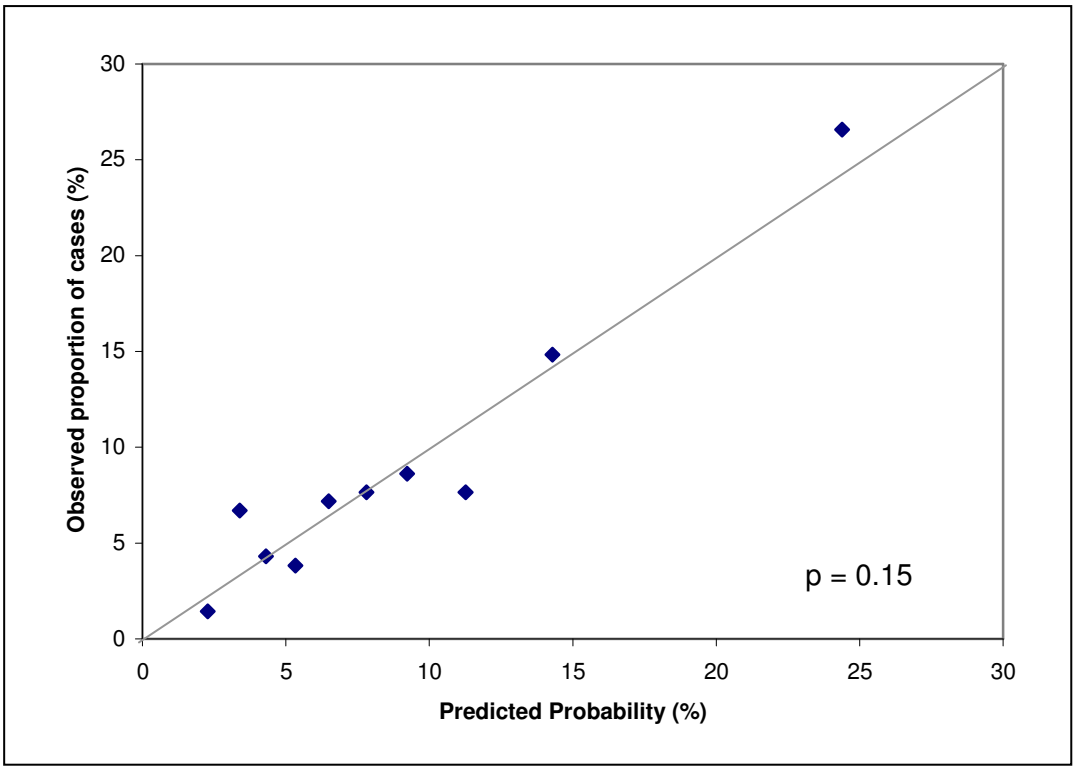


App. Figure 3 cont.

a) UK-Norfolk

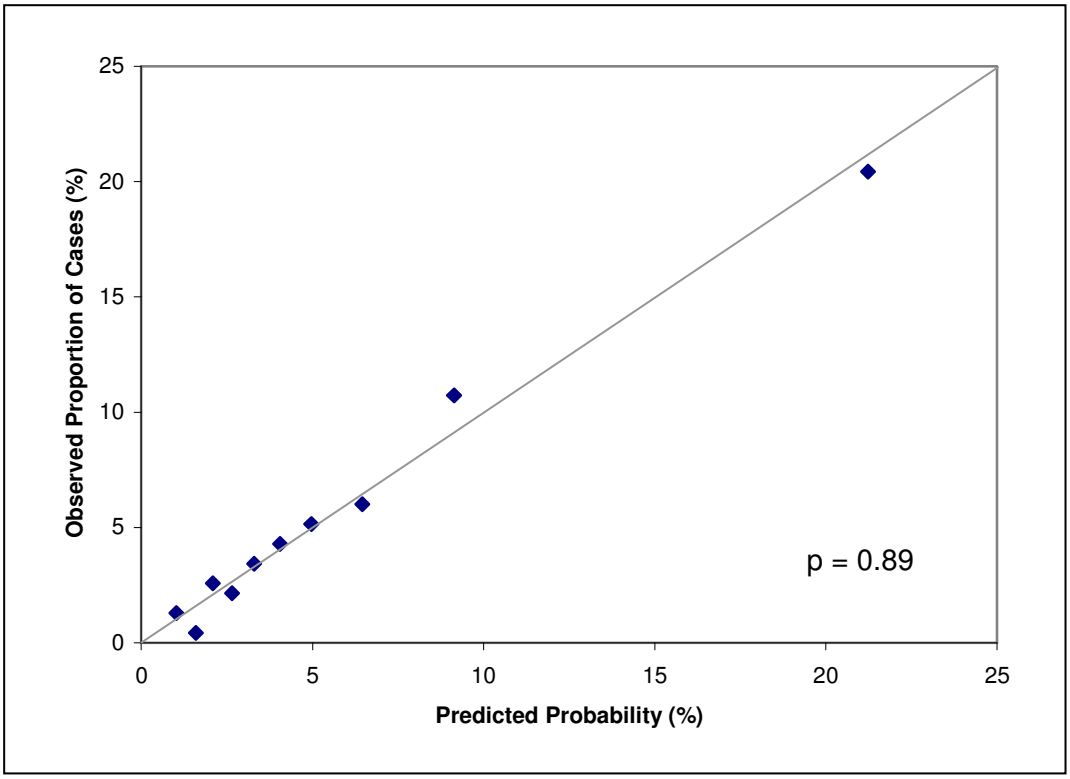


b) NL-Doetinchem

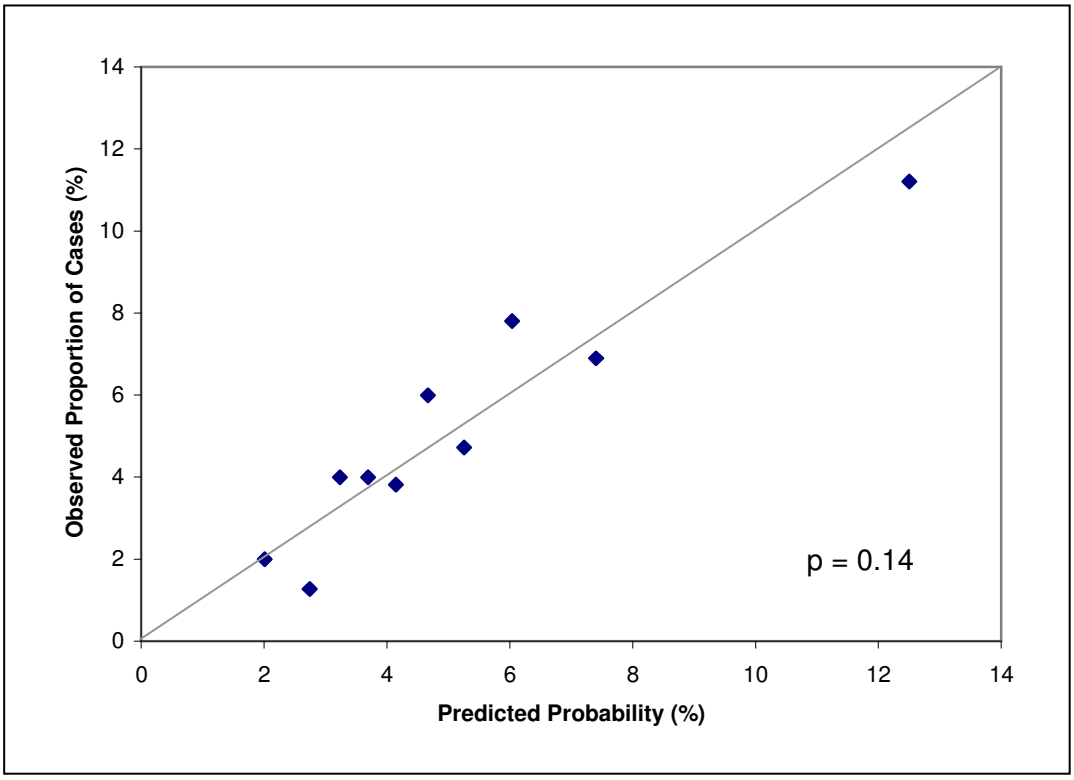


App. Figure 4. Calibration plots (a-f) for the risk score across cohorts of the training sample

c) NL-Amsterdam/Maastricht

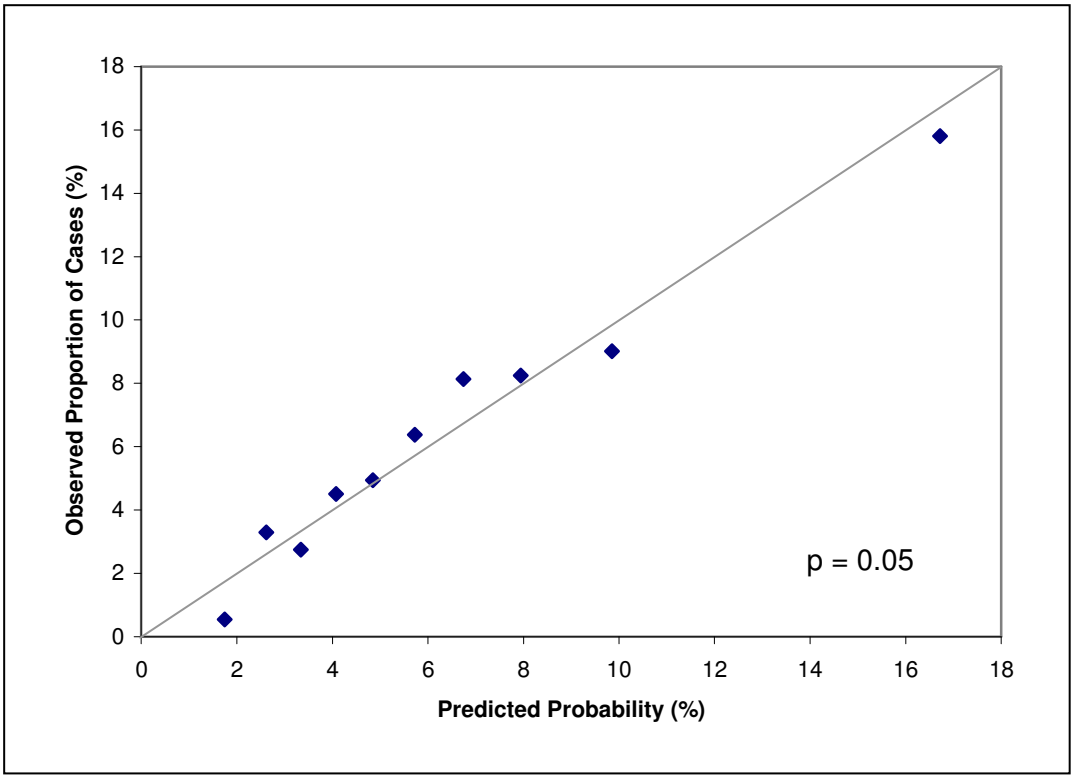


d) IT-Florence

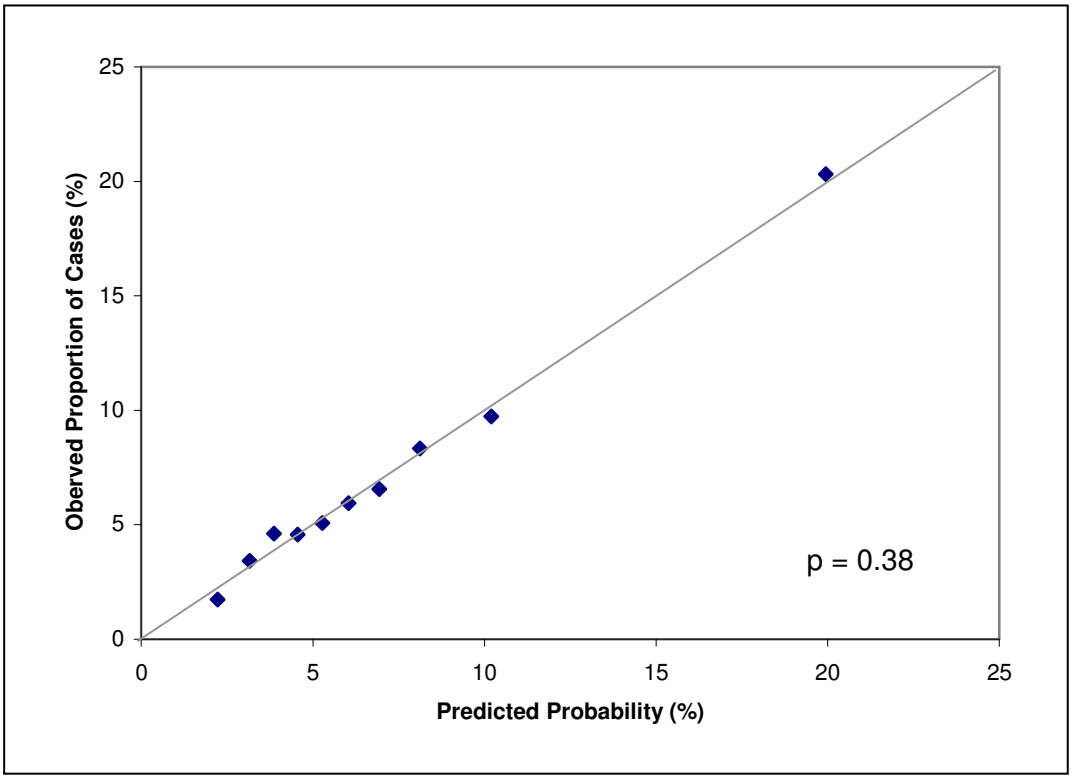


App. Figure 4 cont.

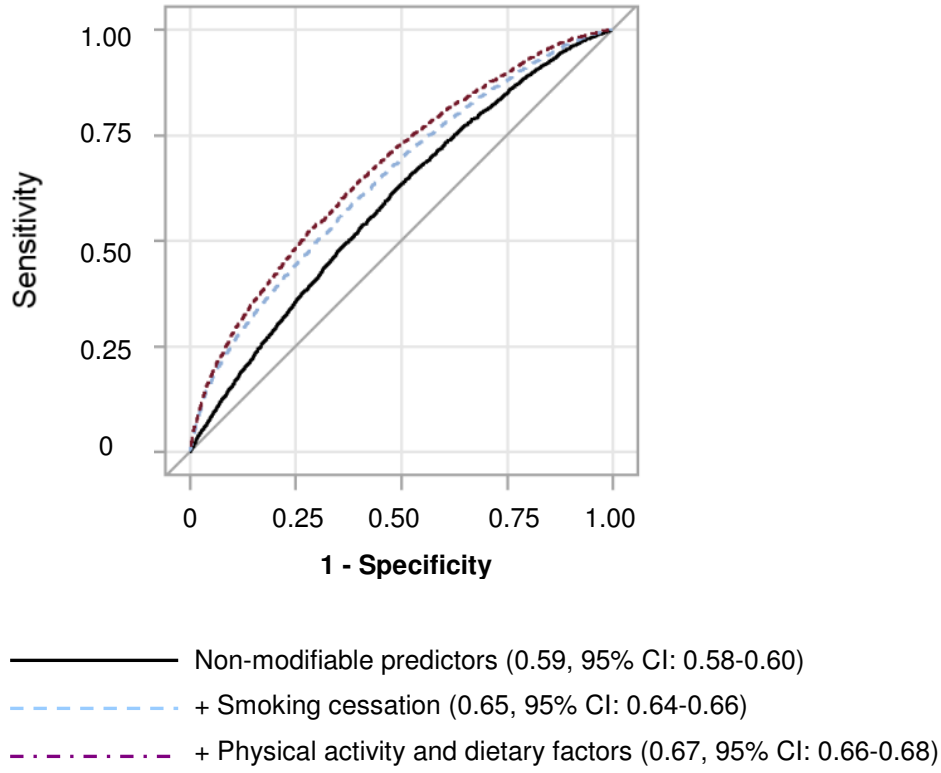
e) GER-Potsdam



f) DK-Aarhus/Copenhagen



App. Figure 4 cont.



App. Figure 5. ROC curves for comparison of models containing non-modifiable and modifiable predictors in the training sample

Danksagung

Die vorliegende Dissertation wurde in der Abteilung Epidemiologie des Deutschen Instituts für Ernährungsforschung (DIfE) angefertigt. An dieser Stelle möchte ich all denen meinen Dank aussprechen, die mich auf beruflicher und privater Ebene unterstützt und damit die Fertigstellung dieser Dissertation möglich gemacht haben.

Mein großer Dank geht an Herrn Prof. Dr. Heiner Boeing, der mich in den letzten Jahren in meiner wissenschaftlichen Entwicklung begleitet, unterstützt und motiviert hat. Herrn Prof. Dr. Reinhard Busse möchte ich danken für die Betreuung dieser Arbeit an der Technischen Universität Berlin und die hilfreichen inhaltlichen Anmerkungen zum Manuskript.

Mein ausdrücklicher Dank gilt Sven Knüppel für seine statistische Expertise, seine unermüdliche Bereitschaft, methodische Ansätze und Ergebnisse zu diskutieren sowie die gewissenhafte Durchsicht der Arbeit.

Ein ganz besonders herzliches Dankeschön gilt Madlen Schütze für ihre freundschaftliche Unterstützung und Hilfsbereitschaft während meiner gesamten Promotionszeit und ihre wertvollen Anmerkungen zu meinem Manuskript.

Weiterhin danke ich meinen lieben KollegInnen der Abteilung Epidemiologie für die freundschaftliche Arbeitsatmosphäre während meiner Zeit in der Abteilung, insbesondere dem „A-Team extended“, Diewertje Sluik und Maria Arregui.

Ganz besonders herzlich danke ich meinen Eltern, die mich in meiner Entwicklung immer nach Kräften unterstützt haben, meinem Bruder Paul-Martin sowie meiner Schwester Christina für ihre vielen offenen Ohren und unterstützenden Worte. Meiner Freundin Sabine danke ich herzlichst für ihre stete moralische Unterstützung und Aufmunterung in frustrierenden und zweifelnden Phasen.

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die am Fachbereich Management im Gesundheitswesen der Technischen Universität Berlin eingereichte Dissertation mit dem Titel „Development and Validation of a Risk Score predicting substantial Weight Gain over 5 Years among middle-aged European Men and Women“ selbständig angefertigt habe und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet wurden. Alle Teile, die wörtlich oder sinngemäß einer Veröffentlichung entstammen sind als solche gekennzeichnet. Weder die gesamte Dissertation noch Teile davon sind bislang veröffentlicht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen.

Berlin, 28. November 2011 Annika Steffen