Predicting risk of substantial weight gain in German adults—a multi-center cohort approach

Ursula Bachlechner[1], Heiner Boeing[1], Marjolein Haftenberger[2], Anja Schienkiewitz[2],
Christa Scheidt-Nave[2], Susanne Vogt[3], Barbara Thorand[3], Annette Peters[3], Sabine Schipf[4,5],
Till Ittermann[4], Henry Völzke,[4,5,6] Ute Nöthlings[7], Jasmine Neamat-Allah[8], Karin-Halina Greiser[8],
Rudolf Kaaks[8], Annika Steffen[1]

1 Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany
2 Department of Epidemiology and Health Monitoring, Robert Koch Institute, Berlin, Germany
3 Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health,
　Neuherberg, Germany
4 Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
5 German Centre for Diabetes Research, Site Greifswald, Greifswald, Germany
6 German Centre for Cardiovascular Research, Partner Site Greifswald, Greifswald, Germany
7 Department of Nutrition and Food Science, Institute for Nutrition and Food Science, Rheinische Friedrich-Wilhelms-
　University Bonn, Bonn, Germany
8 Department of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany

Correspondence: Annika Steffen, German Institute of Human Nutrition (DIfE), Arthur-Scheunert-Allee 114-116, 14558
Nuthetal, Germany, Tel: +49 33200 882723, Fax: +49-33200-882721, e-mail: annika.steffen@dife.de

Background: A risk-targeted prevention strategy may efficiently utilize limited resources available for prevention of
overweight and obesity. Likewise, more efficient intervention trials could be designed if selection of subjects was
based on risk. The aim of the study was to develop a risk score predicting substantial weight gain among German
adults. **Methods:** We developed the risk score using information on 15 socio-demographic, dietary and lifestyle
factors from 32 204 participants of five population-based German cohort studies. Substantial weight gain was
defined as gaining $\geq$10% of weight between baseline and follow-up (>6 years apart). The cases were censored
according to the theoretical point in time when the threshold of 10% baseline-based weight gain was crossed
assuming linearity of weight gain. Beta coefficients derived from proportional hazards regression were used as
weights to compute the risk score as a linear combination of the predictors. Cross-validation was used to evaluate
the score's discriminatory accuracy. **Results:** The cross-validated c index (95% CI) was 0.71 (0.67–0.75). A cutoff value
of $\geq$475 score points yielded a sensitivity of 71% and a specificity of 63%. The corresponding positive and negative
predictive values were 10.4% and 97.6%, respectively. **Conclusions:** The proposed risk score may support healthcare
providers in decision making and referral and facilitate an efficient selection of subjects into intervention trials.

Introduction

Obesity increases morbidity and, consequently, entails a large
societal and economic burden.[1–3] Considering the limited
resources, a risk-targeted prevention strategy may efficiently
complement a population-based approach.[4] Prevention efforts may
be specifically targeted at those individuals with high risk of
developing overweight or obesity. High-risk individuals can be
identified using risk prediction models that combine information
on multiple characteristics to estimate the individual's risk of a
future outcome. While there are numerous prediction models to
estimate the risk of cardiovascular disease, type 2 diabetes and
cancer,[5–8] models predicting weight gain are scarce.

We recently presented a risk score to predict substantial weight
gain (SWG) among European adults based on 13 socio-demo-
graphic, dietary and lifestyle factors.[9] This European obesity risk
score exhibited a rather low discriminatory accuracy [c index
(95% CI) of 0.64 (0.63–0.65) and 0.57 (0.56–0.58) in the develop-
ment and validation sample, respectively]. Interestingly, we observed
remarkable variation in performance across different socio-cultural
backgrounds, i.e. across countries, while study centers assumed to
have a similar underlying socio-cultural structure showed
comparable performance. Hence, the approach to first develop

country-specific prediction models and subsequently compare
them across socio-cultural backgrounds could be a successful
strategy.[5]

Based on data from five population-based German,cohort studies,
we developed and validated a risk score predicting SWG, the
German obesity risk score (GORS). Given the importance of
abdominal fat in relation to metabolic diseases[10], we additionally
evaluated a risk score predicting gain in waist circumference
beyond gain in BMI.

Methods

Study population

The study is based on five population-based longitudinal German
cohorts comprising participants from the national cohort from the
German Health Interview and Examination Survey for Adults
(DEGS) and participants from four regional cohorts including the
two German cohorts of the EPIC study based in Potsdam and
Heidelberg, the Study of Health In Pomerania (SHIP) and the
Cooperative Health Research in the Region of Augsburg (KORA)
S4 Study. Detailed information on these studies has already been
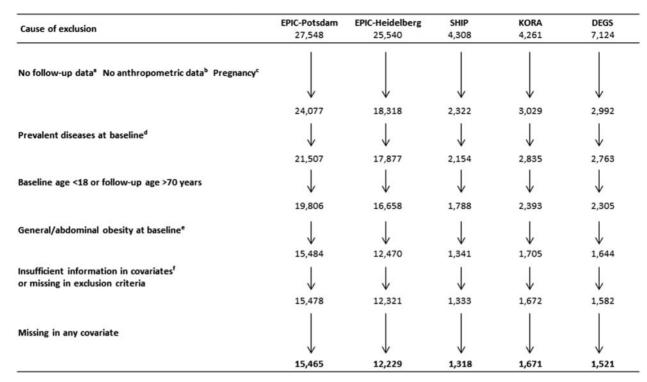described.[11–15] In brief, participants for all studies were recruited

Cause of exclusion	EPIC-Potsdam 27,548	EPIC-Heidelberg 25,540	SHIP 4,308	KORA 4,261	DEGS 7,124
No follow-up data[a] No anthropometric data[b] Pregnancy[c]	↓ 24,077	↓ 18,318	↓ 2,322	↓ 3,029	↓ 2,992
Prevalent diseases at baseline[d]	↓ 21,507	↓ 17,877	↓ 2,154	↓ 2,835	↓ 2,763
Baseline age <18 or follow-up age >70 years	↓ 19,806	↓ 16,658	↓ 1,788	↓ 2,393	↓ 2,305
General/abdominal obesity at baseline[e]	↓ 15,484	↓ 12,470	↓ 1,341	↓ 1,705	↓ 1,644
Insufficient information in covariates[f] or missing in exclusion criteria	↓ 15,478	↓ 12,321	↓ 1,333	↓ 1,672	↓ 1,582
Missing in any covariate	↓ **15,465**	↓ **12,229**	↓ **1,318**	↓ **1,671**	↓ **1,521**

Figure 1 Flow diagram of participants excluded from the present study. [a]No follow-up questionnaire (e.g. due to death before follow-up, emigration or non-response to invitation). [b]Missing data on baseline or follow-up body weight, waist circumference or body height. [c]Pregnancy at baseline or follow-up; [d]Baseline diabetes, cardiovascular disease or cancer. [e]Definition of general obesity: BMI $\geq 30 \, kg/m^2$; definition of abdominal obesity: $WC_{men} \geq 102 \, cm$, $WC_{women} \geq 88 \, cm$. [f]Insufficient information (e.g. "don't know") in assessment of pregnancy or sleep disorders

from the general population as a random sample from the respective population registries. The study population from DEGS virtually covered the whole adult life span (18–79 years). The age ranges for SHIP, KORA and EPIC-Germany were 20–79, 25–74 and 35–65 years at baseline, respectively. From the 68 781 men and women recruited between 1994 and 2001, a total of 55 689 were followed up between 2004 and 2012 (participation rate varied between 43% in DEGS and 89% in EPIC-Potsdam). Of these, 32 204 individuals were finally used in the present study (for exclusions, see figure 1). Briefly, individuals were excluded if follow-up information was not available, anthropometric data were missing, if they were pregnant at baseline or follow-up or if they reported diabetes, cardiovascular diseases or cancer at baseline. To minimize confounding from changes in body composition and shape occurring in older age, our study was restricted to participants aged <70 years at follow-up. Further exclusions refer to participants with general or abdominal obesity at baseline (BMI $\geq 30 \, kg/m^2$; waist circumference of $\geq 102 \, cm$ and $\geq 88 \, cm$ among men and women, respectively) and to participants with missing values in any exclusion criteria or covariate (<2% except for DEGS).

Ethics statement

All studies were conducted conforming to the principles of the Declaration of Helsinki. The regional studies were approved by the local ethics committees. The study protocol of DEGS was consented with the Federal and State Commissioners for Data Protection and approved by the Charité-Universitätsmedizin Berlin ethics committee. All participants provided written informed consent.

Assessment of anthropometry

At baseline, body weight and height were measured by trained staff following standardized procedures in all cohorts. Measurements were obtained without shoes and in underwear (EPIC) or light clothing (SHIP, KORA and DEGS). At follow-up, body weight was measured according to the same measurement procedures in SHIP and KORA. In follow-up of DEGS, participants were only wearing underwear for anthropometric measurements. In EPIC, body weight was self-reported at follow-up, and we applied prediction equations to correct for potential misreporting.[16]

Assessment of diet and lifestyle

In KORA, SHIP and DEGS usual frequency of intake over the past 12 months was assessed by means of short questionnaires including around 40 food items covering the main food groups. In EPIC, habitual diet in the past 12 months was assessed using a validated food frequency questionnaire.[17] Dietary information was harmonized across the five cohorts and expressed as frequency of consumption using three categories per food item. Across cohorts, socio-demographic, lifestyle and further health-related characteristics were assessed in a comparable manner by means of extensive questionnaires, face-to-face interviews and/or computer-assisted interviews.

Statistical analysis

Definition of cases and time-to-event

SWG was defined as gaining $\geq 10\%$ of baseline weight during follow-up, as described previously.[9] Each participant was followed up for incidence of SWG from study entry to the second assessment of body weight (end of follow-up). Those individuals who did not experience SWG during follow-up were censored at time of their second weight assessment and participants experiencing SWG constituted the set of cases. We estimated the time theoretically needed for the cases to cross the threshold of $\geq 10\%$ baseline-based weight gain by assuming linear weight gain.

Definition of predictors

In addition to all predictors forming the European obesity risk score,[9] we included intake of fruits and vegetables, chocolate and sleep disorders as predictors in the GORS. Because intake of meat types was not available in DEGS, we used a single variable representing total meat intake. Similarly, information on whole-grain bread was not available in EPIC-Heidelberg and SHIP, and we used intake of non-white bread in these cohorts.

Construction of the GORS

The GORS was derived using a proportional hazards model to account for diverse follow-up times between participants and individual velocities of weight gain. To best reflect the association of predictors with SWG in the German population, we used a meta-analytical approach to develop the risk score function. Specifically, cohort-specific β coefficients of the 15 predictors were estimated and combined using random-effects meta-analysis.[18] Score points (weights) for each predictor were assigned based on the value of the corresponding pooled β coefficients multiplied by 100 and rounded to two decimal places. For each individual, a risk score (RS_i) was computed as a linear combination of the weighted predictors. The score was rescaled by adding 800. The individual probability of gaining $\geq 10\%$ of weight in the following 5 years ($P_{(SWG, 5y)i}$) was computed by inserting the individual risk score (RS_i) into the survival function of the proportional hazards model. For this, the cohort-specific background survival probabilities at 5 years were estimated based on the average value of each predictor over all individuals of the five cohorts and subsequently meta-analytically combined.

Evaluation of the GORS

Discrimination was assessed by the c index for time-to-event analysis.[19] We applied 10-fold cross-validation by estimating the risk score function in (all combinations of) two cohorts and testing it in the remaining three, respectively. Calibration was evaluated by comparing observed and predicted risk across categories of predicted risk estimated over all cohorts. To propose a threshold value of the continuous GORS to define high-risk individuals, we used the Youden's index.[20] Following Menke,[21, 22] we applied bivariate random-effects meta-analysis to pool cohort-specific sensitivities and specificities for a range of possible cut-off points of the GORS.

Statistical analyses were performed in SAS, version 9.4 (SAS Institute, Cary, NC). Meta-analyses were conducted using the package "meta" developed by Schwarzer[23] for R software (version 3.2.3).

Risk score predicting substantial gain in waist circumference beyond gain in BMI

We additionally constructed a risk score predicting substantial gain in waist circumference beyond gain in BMI defined on the basis of the residuals of waist circumference regressed on BMI (WC_{BMI}), thus specifically reflecting gain in abdominal fat. Substantial gain in WC_{BMI} was defined as gaining ≥ 3.5 cm of baseline WC_{BMI} during follow-up. Height and baseline waist circumference were included in this model additionally to the 15 predictors from the GORS.

Results

The present study included 32 204 individuals and the average follow-up ranged from 6.9 years (KORA) to 10.9 years (DEGS) (table 1). A total of 1661 individuals (5.1%) gained $\geq 10\%$ of their baseline weight within 5 years after baseline, with 66 (5.0%), 67 (4.4%), 89 (5.3%), 504 (4.1%) and 935 (6.0%) in SHIP, DEGS,

KORA, EPIC-Heidelberg and EPIC-Potsdam, respectively. The overall incidence rate amounted to 241 per 10 000 person-years (PY). Across cohorts, incidence rate was highest in SHIP and EPIC-Potsdam (295 and 277 per 10 000 PY, respectively) while it was lowest in KORA and EPIC-Heidelberg (196 and 198 per 10 000 PY, respectively) (data not shown). On average, study participants gained 299 g (KORA) to 429 g (EPIC-Potsdam) weight per year (table 1). Individuals from SHIP and DEGS were on average younger than participants in the other cohorts, proportion of men was highest in KORA and DEGS (almost 50%), while in EPIC-Potsdam, it was lowest (37%).

The meta-analytically combined estimates of relative risk for the association of the predictors with SWG and corresponding score points are shown in table 2. As for single predictors, baseline age and weight most strongly determined risk, indicating a lower risk for future weight gain with increasing age and weight.

Under average conditions 96% of the population will not experience SWG within 5 years (average survival probability at 5 years = 0.9634). The individual risk of gaining $\geq 10\%$ of baseline weight within 5 years ($P_{(SWG, 5y)i}$) was estimated by inserting the individual risk score points into the following survival function:

$$P_{(SWG,5y)i} = 1 - 0.9634 + \exp^{((RS_i - 440.94)/100)}$$

The median of the GORS across the five cohorts was 439, ranging from 269 to 684. The probability of experiencing SWG within the following 5 years for 350, 400, 450, 500 and 550 score points was 1.5, 2.4, 4.0, 6.5 and 10.5%, respectively. The cross-validated c index (95% CI) was 0.71 (0.67–0.75). While discriminatory ability was quite similar among four of the five cohorts, ranging from 0.68 (0.67–0.70) in EPIC-Potsdam to 0.72 (0.66–0.78) in SHIP, discrimination was strikingly higher in DEGS [0.79 (0.74–0.84)] (data not shown). Table 3 displays pooled sensitivity and specificity across a range of possible cut-off points of the GORS. Sensitivity and specificity were maximized at a score value of ≥ 475 (Youden's index of 0.34). This threshold captured 71% of the cases who experienced SWG. Likewise, 63% of individuals who did not experience SWG had a score below this threshold. The corresponding positive and negative predictive values were 10.4% and 97.6%, respectively, implying that around 10% of individuals with a score value of ≥ 475 in fact experience SWG, while almost 98% of individuals having a score below 475 remain free of SWG.

Observed incidence was largely comparable with predicted risk across the total study population (data not shown). Within cohorts, we observed very good calibration in KORA and SHIP, good to acceptable calibration in the EPIC cohorts with underestimation of risk in the higher risk groups and overestimation in DEGS.

A total of 3095 individuals showed substantial gain in WC_{BMI} within 5 years after baseline (256 cases per 10 000 PY). The cross-validated c index (95% CI) of the model predicting substantial gain in WC_{BMI} was 0.70 (0.65–0.74). Within cohorts, discrimination varied between 0.65 (0.61–0.69) in KORA to 0.78 (0.70–0.86) in SHIP. Calibration was heterogeneous across cohorts, ranging from very good calibration in EPIC-Heidelberg to poor calibration in EPIC-Potsdam, SHIP and DEGS. Importantly, while dietary and lifestyle predictors mostly showed associations of similar direction compared with the GORS, the relationship of the most influential predictors age, sex and weight was reversed when predicting substantial gain in WC_{BMI} [HR (95% CI) of 1.04 (1.03–1.04), 0.53 (0.32–0.90) and 1.17 (1.15–1.20), respectively].

Discussion

In this multi-center prospective study, we presented a risk score to predict SWG within the next 5 years. The GORS is based on data from five population-based German cohort studies, representing

Table 1 General characteristics of the study population

	EPIC-Potsdam	EPIC-Heidelberg	SHIP	KORA	DEGS
n	15 465	12 229	1318	1671	1521
Duration of follow-up (y)	8.0 (1.5)	8.2 (1.3)	10.1 (2.2)	6.9 (0.8)	10.9 (2.3)
Cases within 5 years (*N*)	935	504	66	89	67
Age at baseline (y)	47.3 (8.0)	48.6 (7.4)	39.7 (10.8)	43.0 (10.5)	39.2 (10.6)
Age range at baseline, median (IQR)	46.0 (41.0–55.0)	48.0 (42.0–55.0)	39.0 (31.0–49.0)	42.0 (34.0–51.0)	39.0 (31.0–48.0)
Men (%)	37.2	44.6	45.1	49.9	47.6
University degree (%)	41.6	36.5	13.3	20.1	19.5
Anthropometry					
BMI baseline (kg/m^2)	24.3 (2.6)	24.1 (2.7)	24.4 (2.8)	24.6 (2.7)	24.1 (2.7)
BMI follow-up (kg/m^2)	25.6 (3.1)	25.2 (3.4)	26.1 (3.4)	25.2 (3.0)	25.5 (3.4)
Obese at follow-up (%)	7.6	5.9	14	5.9	9.3
Annual weight change (g/year)	429 (588)	359 (723)	388 (560)	299 (661)	310 (536)
Lifestyle factors and health-related factors					
Sports (%)					
0 h/wk	56.3	37.5	41.1	25.4	33.7
>0 to < 2 h/wk	18.9	20.9	35.3	49.8	41.6
≥2 h/wk	24.8	41.6	23.6	24.8	24.7
Current smoking (%)	20.8	22.4	33.9	29.1	33.8
Alcohol consumption (%)					
0 g/day	2.4	4.2	28.7	23.2	12.5
>0 to ≤ 6 g/day	37.5	31.8	22	23.2	47.7
>6 to ≤ 18 g/day	32.9	29.3	22.3	19.6	22.2
>18 g/day	27.2	34.6	27	34	17.6
Sleep disorders (%)	13.4	11.9	16.8	17.2	16
Dietary factors					
Fruits and vegetables (%)					
<1 time/day	14.3	34.5	56.4	52.4	51.2
≥1 time to < 2 times/day	17.5	26.4	25.8	23.5	19
≥2 times/day	68.2	39.1	17.8	24.1	29.9
Meat (%)					
<4–6 times/week	12.2	58.3	20.9	39	27.8
≥4–6 times/week to < 1 time/day	12	20.6	19.2	27.4	31.6
≥1 time/day	75.8	21.1	59.9	33.7	40.6
Fish (%)					
<2–3 times/month	17.9	38.6	34.8	32.7	22.8
≥2–3 times/month to < 1 time/week	24	23.2	30.1	28.1	25.2
≥1 time/week	58.1	38.2	35.1	39.1	52.1
Whole grain bread (%)[a]					
<2–3 times/week	49.6	9.4	6.4	30.2	33.4
≥2–3 to < 4–6 times/week	17.1	15.4	7.5	33.2	26
≥4–6 times/week	33.3	75.3	86.1	36.6	40.6
Cake and cookies (%)					
<1 time/week	15.6	19.3	32.1	32	21.1
≥1 to < 2–3 times/week	33.3	31.1	29.6	30	33.5
≥2–3 times/week	51.2	49.7	38.3	38	45.3
Chocolate (%)					
<1 time/week	22.4	29.5	42.6	37.7	27.4
≥1 to < 2–3 times/week	23.8	19.6	17.6	22.2	24.1
≥2–3 times/week	53.8	50.9	39.8	40.1	48.6
Soft drinks (%)					
<2–3 times/month	70.8	69.9	56.6	56.3	40.6
≥2–3 times/month to 4–6 times/week	23.1	20.2	28.5	30.9	38.1
≥4–6 times/week	6.2	9.9	15	12.9	21.3

Notes. For continuous variables mean (SD) are shown. [a]In EPIC-Heidelberg and SHIP non-white bread was used instead of whole grain bread.

four different regions and Germany as a whole, thereby including a diverse structure of predictors. Discriminatory accuracy and calibration of the GORS encourage its application in practice. Specifically, the GORS may be used to identify subgroups that would benefit most from risk-reducing interventions or might help to design more efficient intervention studies by specifically selecting high(er)-risk individuals.

Major strengths of the present study are its prospective, multicenter design, the large sample size and the availability of information on a variety of predictors. Given the meta-analytical approach followed by cross-validation with non-random splits, our results can be considered generalizable, i.e. externally validated.

As a limitation, we relied on only two assessments of body weight and assumed weight gain to be linear. On the population-level, however, weight gain can be reasonably well approximated by a straight line over a follow-up period of 8 years.[24] In SHIP, KORA and DEGS, body weight was assessed by trained staff at both assessments, while in EPIC body weight was self-reported at follow-up. These methodological differences were accounted for by correcting self-reported body weight using prediction equations.[16] As in most epidemiological studies, random error and self-reporting bias in dietary and lifestyle factors might have affected our results. Specifically, we lacked information on objectively measured physical activity and therefore restricted information to sports which has been demonstrated to be more accurately recalled than low-intensity behaviors.[25] Also, because information on portion size was not available in all cohorts, information on dietary predictors was restricted to frequency of consumption. It is well known though

Table 2 Meta-analytically derived estimates for the association of predictors with SWG and associated score points

Predictors	HR	(95% CI)		Score points
Age (years)	0.96	(0.96–0.97)	-0.0408	-4.1
Sex (female vs. male)	1.10	(0.99–1.24)	0.0953	9.5
Baseline weight (kg)	0.98	(0.98–0.99)	-0.0202	-2.0
Education				
Secondary/Professional school	0.88	(0.79–0.98)	-0.1278	-12.8
University	0.76	(0.70–0.82)	-0.2744	-27.4
Lifestyle and health-related factors				
Sports				
>0 to <2h/wk	0.93	(0.87–0.99)	-0.0726	-7.3
≥2h/wk	0.89	(0.77–1.02)	-0.1165	-11.7
Smoking	1.48	(1.40–1.56)	0.392	39.2
Alcohol				
No alcohol	1.08	(0.90–1.29)	0.077	7.7
Alcohol >6 to ≤18g/d	0.88	(0.83–0.94)	-0.1278	-12.8
Alcohol >18g/d	0.89	(0.80–0.99)	-0.1165	-11.7
Sleep disorders	1.28	(1.20–1.38)	0.2469	24.7
Dietary factors				
Fruits and vegetable				
≥1 time to <2 times/day	0.94	(0.87–1.01)	-0.0619	-6.2
≥2 times/day	0.96	(0.86–1.06)	-0.0408	-4.1
Meat				
≥4–6 times/week to <1 time/day	1.07	(0.97–1.18)	0.0677	6.8
≥1 time/day	1.11	(1.01–1.22)	0.1044	10.4
Fish				
≥2–3 times/month to <1 time/week	0.98	(0.90–1.06)	-0.0202	-2.0
≥1 time/week	0.99	(0.90–1.09)	-0.0101	-1.0
Whole-grain bread				
≥2–3 to <4–6 times/week	0.94	(0.87–1.02)	-0.0619	-6.2
≥4–6 times/week	0.87	(0.81–0.93)	-0.1393	-13.9
Cake and cookies				
≥1 to <2–3 times/week	0.88	(0.75–1.02)	-0.1278	-12.8
≥2–3 times/week	0.76	(0.61–0.96)	-0.2744	-27.4
Chocolate				
≥1 to <2–3 times/week	1.05	(0.97–1.13)	0.0488	4.9
≥2–3 times/week	1.03	(0.96–1.12)	0.0296	3.0
Soft drinks				
≥2–3 times/month to 4–6 times/week	1.11	(1.04–1.18)	0.1044	10.4
≥4–6 times/week	1.18	(1.03–1.34)	0.1655	16.6

Table 3 Pooled sensitivity and specificity and associated test characteristics across a range of possible cut-off points of the German Risk Score to predict SWG

Score points	Absolute risk (%)	% of population	Sensitivity (95% CI), in %	Specificity (95% CI), in %	Youden's index	PPV (%)	NPV (%)	LR +	LR−to
≥400	2.4	74.8	96 (90–99)	19 (9–30)	0.15	6.4	98.9	1.2	0.2
≥425	3.1	59.4	91 (78–98)	32 (16–51)	0.23	7.3	98.5	1.3	0.3
≥450	4.0	42.7	82 (61–95)	47 (28–67)	0.29	8.4	98.0	1.5	0.4
≥475	5.1	27.8	71 (42–91)	63 (41–81)	0.34	10.4	97.6	1.9	0.5
≥500	6.5	15.5	52 (24–80)	77 (56–90)	0.29	12.3	96.7	2.3	0.6
≥525	8.3	7.6	34 (12–64)	87 (71–96)	0.21	14.2	96.0	2.6	0.8

Youden's Index = (sensitivity (%) + specificity (%) − 100)/100. PPV = positive predictive value. NPV = negative predictive value. LR + =likelihood ratio positive. LR-=likelihood ratio negative.

that frequency of consumption explains a larger part of inter-individual variation in food intake than inter-individual variation in portion size.[26]

In a prediction model, associations of predictors with the outcome may not necessarily reflect etiological relationships.[27] In the present study, this may particularly apply to education, smoking and intake of cake and cookies. Educational level is an indicator of acquisition of knowledge and competences that enables individuals to integrate a health-promoting lifestyle.[28] Thus, education may reflect further aspects of dietary and lifestyle behavior in addition to the behavioral factors in the model. For smoking, which is generally related to lower BMI compared with

non-smoking,[29] the higher risk of SWG among baseline smokers can be attributed to the higher risk of weight gain among those individuals who quit smoking during follow-up compared with those who continued smoking, an observation discussed previously.[9,29] In EPIC-Potsdam, for instance, 35% of baseline smokers stopped smoking during follow-up and showed a substantially higher annual weight gain than continuing smokers (770 g/y vs. 422 g/y). Accordingly, smoking cessation during follow-up was related to a significantly higher risk of SWG than continuing smoking (HRs (95% CI) compared with constant non-smoking of [2.53 (2.28–2.80) and 0.97 (0.87–1.08), respectively]. Lastly, the inverse association of cake and cookies with SWG might reflect selective

underreporting of intake. In a sub-study of EPIC-Potsdam, for instance, the inverse association of cake and cookies with BMI was reversed when underreporting was accounted for.[30]

The observed variation in calibration between cohorts may be explained by the different background risks. For example, in those cohorts where the cohort-specific estimate of background risk resembled the pooled estimate, calibration was very good (KORA and SHIP), while in cohorts where the cohort-specific background risk was higher than the pooled one (EPIC cohorts), the GORS generally underestimated risk. Differences in outcome frequency are often observed between diverse populations, and adjusting the baseline risk has been suggested as a simple method of improving calibration.[31]

The GORS exhibited a higher discrimination than its European counterpart[9] though predictive performance was still moderate. Given the wide range of predictors in our model, it is unlikely that further (strong) predictors have been missed. Also, additional predictors would have to show very large "independent" associations to meaningfully increase discrimination of an already reasonably good model.[32] The nature of weight gain per se may challenge its prediction. Weight gain is reversible, and it is well-known that body weight tends to fluctuate over time, leading to repeated cycles of weight loss and recovery.[33, 34] In some instances, these fluctuations may exacerbate accurate classification of cases, non-cases and the cases' time-to-event, limiting the predictability.

We suggested a cut-off value of $\geq$475 for practical application of the GORS, i.e. individuals exceeding this threshold may be advised to undergo weight management programs. The threshold value was based on the maximization of both sensitivity and specificity while both characteristics were regarded as equally important. In practice, however, the cut-off value should be chosen according to the importance attached to false-positives and false-negatives. In contrast to more severe diseases such as cancer or CVD, misclassification costs may essentially reduce to the costs attached to false-positives in the context of weight gain prediction. False-positives may pose a large economic burden to the healthcare system since a considerable number of individuals will be unnecessarily advised to undergo prevention programs. Thus, for application of the GORS, a cut-off point related to higher specificity may be preferred.

Interestingly, a different group of individuals is defined as "high-risk" when predicting substantial gain in WC_{BMI}, as a measure of abdominal obesity, compared with SWG. While the GORS is more likely to identify individuals who are younger, female and have lower body weights to be at high risk, the WC_{BMI} model assigns higher risks to older, male and heavier individuals. This finding is in line with the observation of sex-specific patterns of body fat gain and fat-redistribution associated with aging.[35, 36] In terms of metabolic disease risk and longevity[10, 37], body fat distribution, i.e. higher visceral fat vs. subcutaneous fat, and the relationship of body compartments to each other, i.e. muscle vs. fat mass, may be important. Thus, future studies may direct research efforts toward the prediction of particular body composition phenotypes to enable a more specific identification of individuals at high metabolic risk.

Our proposed risk score predicting SWG based on easily obtainable information was shown to have adequate discriminatory accuracy and calibration. The GORS may be used to support healthcare providers in decision making and referral. Future intervention studies may apply the GORS to facilitate an efficient selection of study participants.

Funding

Conflicts of interest: None declared.

Key points

- Prevention of overweight and obesity is a major public health challenge.
- A risk-targeted prevention strategy may more efficiently utilize limited resources available for prevention than a population-based strategy.
- Currently, there are no tools to accurately identify high-risk individuals for targeted prevention.
- The present study presents a simple, but adequate tool by which individuals at high risk of experiencing SWG within the next 5 years can be distinguished from individuals at low risk.
- The tool may support healthcare providers in decision making and referral and facilitate an efficient selection of study participants into clinical trials.

References

1 Lehnert T, Streltchenia P, Konnopka A, et al. Health burden and costs of obesity and overweight in Germany: an update. *Eur J Health Econ* 2015;16:957–67.

2 World Health Organization. Obesity: Preventing and Managing The Global Epidemic. Report of a WHO Consultation. Report No: 894 Geneva, 2000.

3 Guh DP, Zhang W, Bansback N, et al. The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC Public Health* 2009;9:88.

4 Zulman DM, Vijan S, Omenn GS, Hayward RA. The relative merits of population-based and targeted prevention strategies. *Milbank Q* 2008;86:557–80.

5 Buijsse B, Simmons RK, Griffin SJ, Schulze MB. Risk assessment tools for identifying individuals at risk of developing type 2 diabetes. *Epidemiol Rev* 2011;33:46–62.

6 Cui J. Overview of risk prediction models in cardiovascular disease research. *Ann Epidemiol* 2009;19:711–7.

7 Freedman AN, Slattery ML, Ballard-Barbash R, et al. Colorectal cancer risk prediction tool for white men and women without known susceptibility. *J Clin Oncol* 2009;27:686–93.

8   Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989;81:1879–86.

9   Steffen A, Sorensen TI, Knuppel S, et al. Development and validation of a risk score predicting substantial weight gain over 5 years in middle-aged European men and women. *PloS one* 2013;8:e67429.

10   Finelli C, Sommella L, Gioia S, et al. Should visceral fat be reduced to increase longevity? *Ageing Res Rev* 2013;12:996–1004.

11   Boeing H, Korfmann A, Bergmann MM. Recruitment procedures of EPIC-Germany. European Investigation into Cancer and Nutrition. *Ann Nutr Metab* 1999;43:205–15.

12   Holle R, Happich M, Lowel H, et al. KORA–a research platform for population based health research. *Gesundheitswesen* 2005;67 Suppl 1:S19–25.

13   John U, Greiner B, Hensel E, et al. Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. *Soz Praventivmed* 2001;46:186–94.

14   Scheidt-Nave C, Kamtsiuris P, Gosswald A, et al. German health interview and examination survey for adults (DEGS) - design, objectives and implementation of the first data collection wave. *BMC Public Health* 2012;12:730.

15   Bellach BM, Knopf H, Thefeld W. Der Bundes-Gesundheitssurvey1997/98. *Gesundheitswesen* 1998;60:59–68.

16   Haftenberger M, Lahmann PH, Panico S, 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: 1147–62.

17   Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002;5:1113–24.

18   Smith-Warner SA, Spiegelman D, Ritz J, et al. Methods for pooling results of epidemiologic studies: the Pooling Project of Prospective Studies of Diet and Cancer. *Am J Epidemiol* 2006;163:1053–64.

19   Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med* 2004;23:2109–23.

20   Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32–5.

21   Menke J. Bivariate random-effects meta-analysis of sensitivity and specificity with the Bayesian SAS PROC MCMC: methodology and empirical evaluation in 50 meta-analyses. *Med Decis Making* 2013;33:692–701.

22   Menke J. Bayesian bivariate meta-analysis of sensitivity and specificity: summary of quantitative findings in 50 meta-analyses. *J Eval Clin Pract* 2014;20:844–52.

23   Schwarzer G. meta: An R package for meta-analysis. *R News* 2007;7:40–5.

24   von Ruesten A, Steffen A, Floegel 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:e27455.

25   DiPietro L. Physical activity, body weight, and adiposity: an epidemiologic perspective. *Exerc Sport Sci Rev* 1995;23:275–303.

26   Noethlings U, Hoffmann K, Bergmann MM, Boeing H. European Investigation into C, Nutrition. Portion size adds limited information on variance in food intake of participants in the EPIC-Potsdam study. *J Nutr* 2003;133:510–5.

27   Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009;338:b606.

28   Grossmann M, Kaestner R. *Effects of Education on Health. The Social Benefit of Education*. Ann Arbor: University of Michigan Press, 1997: 69–123.

29   Tian J, Venn A, Otahal P, Gall S. The association between quitting smoking and weight gain: a systemic review and meta-analysis of prospective cohort studies. *Obes Rev* 2015;16:883–901.

30   Gottschald M, Knuppel S, Boeing H, Buijsse B. The influence of adjustment for energy misreporting on relations of cake and cookie intake with cardiometabolic disease risk factors. *Eur J Clin Nutr* 2016;70(11):1318–1324.

31   Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012;98:691–8.

32   Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157–72. discussion 207–112.

33   Lahti-Koski M, Mannisto S, Pietinen P, Vartiainen E. Prevalence of weight cycling and its relation to health indicators in Finland. *Obes Res* 2005;13:333–41.

34   Vergnaud AC, Bertrais S, Oppert JM, et al. Weight fluctuations and risk for metabolic syndrome in an adult cohort. *Int J Obes (Lond)* 2008;32:315–21.

35   St-Onge MP, Gallagher D. Body composition changes with aging: the cause or the result of alterations in metabolic rate and macronutrient oxidation? *Nutrition* 2010;26:152–5.

36   Tchernof A, Despres JP. Pathophysiology of human visceral obesity: an update. *Physiol Rev* 2013;93:359–404.

37   Srikanthan P, Karlamangla AS. Muscle mass index as a predictor of longevity in older adults. *Am J Med* 2014;127:547–53.