Original Article

J Prev Med Public Health 2025;58:465-474 • https://doi.org/10.3961/jpmph.25.030

pISSN 1975-8375 eISSN 2233-4521

Journal of Preventive Medicine & Public Health

Population-attributable Fractions of Lifestyle Factors for Prediabetes in Korea: A Regression-based Analysis of National Survey Data

Yeon Woo Oh[1,2], Chung Mo Nam[1,2], Eun-Cheol Park[2,3]

[1]Department of Biostatistics and Computing, Yonsei University College of Medicine, Seoul, Korea; [2]Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Korea; [3]Institute of Health Services Research, Yonsei University, Seoul, Korea

Objectives: Although lifestyle modification programs are widely implemented for diabetes prevention, the contributions of individual lifestyle factors remain unclear. This study investigated lifestyle risk factors for prediabetes and employed a regression-based approach for estimating their population-attributable fractions (PAFs) using nationally representative data.

Methods: We analyzed data from 3104 adults aged ≥30 years without diabetes from the 2022 Korea National Health and Nutrition Examination Survey. Seven lifestyle factors were assessed: body weight, alcohol consumption, smoking, physical activity, sleep duration, vegetable intake, and breakfast consumption. Prediabetes was defined as fasting blood glucose of 100-125 mg/dL or hemoglobin A1c levels of 5.7-6.4%. Complex survey-adjusted logistic regression was used to identify significant lifestyle risk factors, and their PAFs were estimated using a regression-based sequential method.

Results: Five lifestyle factors were significantly associated with prediabetes: abnormal body weight (odds ratio [OR], 2.05; 95% confidence interval [CI], 1.68 to 2.50), excessive alcohol consumption (OR, 1.27; 95% CI, 1.00 to 1.62), smoking (OR, 1.35; 95% CI, 1.07 to 1.71), insufficient exercise (OR, 1.26; 95% CI, 1.05 to 1.51), and irregular breakfast consumption (OR, 1.31; 95% CI, 1.08 to 1.59). In sequential PAF estimation, abnormal body weight had the largest contribution (22.2%; 95% CI, 16.2 to 28.2), followed by smoking (6.4%; 95% CI, 1.1 to 11.6), insufficient exercise (5.8%; 95% CI, 1.2 to 10.5), irregular breakfast consumption (4.9%; 95% CI, 0.5 to 9.2), and excessive alcohol consumption (3.6%; 95% CI, 0.1 to 7.4). These results remained consistent in sensitivity analyses including undiagnosed diabetes cases.

Conclusions: Abnormal body weight emerged as the largest contributor to prediabetes (PAF>20%). Diabetes prevention programs in Korea should prioritize weight management within a comprehensive approach to lifestyle modification.

Key words: Prediabetic state, Lifestyle, Body weight

Received: Jan 11, 2025 Revised: Mar 24, 2025 Accepted: Mar 31, 2025

Corresponding author: Eun-Cheol Park
Department of Preventive Medicine, Yonsei University College of Medicine, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
E-mail: ecpark@yuhs.ac

INTRODUCTION

Diabetes represents a significant global health burden and serves as a major risk factor for cardio-cerebrovascular diseases, chronic kidney disease, and various other complications [1]. As of 2021, the worldwide age-standardized prevalence of diabetes reached 6.1%, accounting for 79.2 million disability-associated life years (DALYs). Notably, type 2 diabetes constitutes the predominant burden, representing 96.0% of the diabetes

prevalence and 95.4% of DALYs attributable to diabetes [2].

Numerous lifestyle modification-based diabetes prevention programs have been implemented internationally. In the United States, the Centers for Disease Control and Prevention (CDC) administers the National Diabetes Prevention Program [3], while the United Kingdom's National Health Service (NHS) operates the "Healthier You" Diabetes Prevention Programme [4]. In the Korea, the Korea Diabetes Prevention Study implemented interventions targeting 10 lifestyle factors, primarily focusing on exercise, diet, and behavioral modifications [5].

To optimize resource allocation in population-based prevention programs, it is essential to assess the population-attributable fractions (PAFs) of various lifestyle risk factors. Studies conducted in various countries, including the United States [6], China [7], and France [8], have estimated the PAF of lifestyle factors for diabetes incidence. However, to our knowledge, no comprehensive assessment of lifestyle factor PAFs for diabetes has been conducted in Korea.

The potential for reverse causality due to lifestyle modifications following diabetes diagnosis presents a methodological challenge [9,10]. Given that prediabetes is a robust predictor of diabetes development, as approximately 70% of individuals with prediabetes progress to diabetes during their lifetime [11], and considering that post-diagnostic lifestyle modifications are likely to be minimal among patients with prediabetes compared to those with diabetes [12], we selected prediabetes as a surrogate endpoint.

This study has 2 aims: first, to identify lifestyle risk factors for prediabetes using nationally representative Korean survey data and estimate their PAFs; and second, to demonstrate the application of regression-based approaches for PAF estimation using cross-sectional data. This investigation is expected not only to elucidate the relative importance of lifestyle factors, considering both their association strength and prevalence, but also to provide a methodological framework for estimating PAFs from prevalence data.

METHODS

Study Population

This study utilized data from the Korea National Health and Nutrition Examination Survey (KNHANES), conducted between January 2022 and December 2022 by the Korea Disease Control and Prevention Agency (KDCA). The KNHANES employs a nationally representative sampling design to collect comprehensive data on the health status, health behaviors, and nutritional intake of the Korean population [13].

The sampling process employed a 2-stage stratified cluster sampling design. Initially, 192 enumeration districts (EDs) were selected as primary sampling units. Subsequently, 25 households were randomly selected within each ED as secondary sampling units. All eligible household members from these selected households were included as participants, resulting in an initial sample of 6265 individuals.

The analytic sample was restricted to adults aged ≥30 years who did not have diabetes. After excluding participants with missing data on lifestyle variables of interest and relevant covariates, the final analytic sample comprised 3104 participants.

Exposure (Lifestyles)

Lifestyle factors were primarily based on the Life's Essential 8 (LE8) metrics [14], which were established by the American Heart Association as "the key measures for improving and maintaining cardiovascular health" [15]. LE8 consists of 8 components, divided into 2 categories: health behaviors and health factors. Health behaviors include "eat better," "be more active," "quit tobacco," and "get healthy sleep," whereas health factors include "manage weight," "control cholesterol," "manage blood sugar," and "manage blood pressure."

From these 8 LE8 components, we focused on the health behaviors and weight management components, as these factors were more directly related to lifestyle practices rather than laboratory measurements. Additionally, breakfast consumption was included as a lifestyle factor based on its previously documented association with diabetes risk [16,17].

All lifestyle factors were evaluated dichotomously. Dietary assessment consisted of 2 components: alcohol consumption, categorized according to whether participants consumed 5 or more alcoholic drinks per sitting, and vegetable intake, determined by vegetable consumption at every meal. Physical activity was categorized based on weekly aerobic exercise duration, with 2.5 hr/wk serving as the threshold. Smoking status was classified as ever-smoker versus never-smoker. Sleep habits were evaluated based on weekday sleep duration, with 7-8 hr/day considered optimal. Body weight status was assessed using body mass index (BMI), with BMI values between 18.5 kg/m^2 and 23.0 kg/m^2 considered within the target range. Breakfast consumption was determined based on weekly frequency, with regular consumption defined as breakfast intake 5-7 times/wk.

Outcome (Prediabetes)

Initially, the diabetes group was defined as participants who met any of the following criteria: fasting blood glucose (FBG) ≥126 mg/dL, hemoglobin A1c (HbA1c) ≥6.5%, current use of antidiabetic medications (oral or insulin), or self-reported physician diagnosis of diabetes. After excluding all participants with diabetes, prediabetes was defined among the remaining participants as having either an FBG between 100 mg/dL and 125 mg/dL or an HbA1c between 5.7% and 6.4%.

In sensitivity analyses, we expanded our case definition to include undiagnosed diabetes. Undiagnosed diabetes was defined as individuals who reported no previous diabetes diagnosis but had an FBG ≥126 mg/dL or an HbA1c ≥6.5%.

Covariates

Demographic, socioeconomic, and clinical factors were included as covariates. Demographic factors included sex, age, and marital status (ever- vs. never-married). Socioeconomic factors consisted of education level, family income, and occupation. Education level was categorized based on the highest level of education attained: elementary school or lower (education ≤6 years), middle school or lower (7-9 years), high school (10-12 years), and college or higher (≥13 years). Family income was classified into quartiles. Occupation was categorized as no occupation, white-collar, or blue-collar.

Clinical covariates included personal histories of hypertension and dyslipidemia. Family history of chronic diseases was considered positive if any first-degree relative (parents or siblings) had a documented history of hypertension, dyslipidemia, diabetes, or ischemic heart disease.

Statistical Analysis

We employed a complex survey-adjusted logistic regression model within a case-control framework to identify significant lifestyle risk factors for prediabetes. Lifestyle factors whose regression coefficients achieved statistical significance ($p<0.05$) were identified as risk factors. It is important to clarify that throughout our analysis, we utilized a single comprehensive regression model, which simultaneously included all 7 lifestyle factors regardless of their statistical significance. The selection based on statistical significance pertained solely to identifying which factors' PAFs would subsequently be reported, rather than influencing the specification of the model itself. This approach ensured that all potential lifestyle factors were accounted for, while the PAF calculations were limited to those showing significant positive associations with prediabetes.

The PAFs for identified risk factors were estimated using 2 approaches, following the method outlined by Rückinger et al. [18]. First, we applied a regression-based method to calculate individual PAFs. For a given dichotomous risk factor A, the PAF was calculated as follows:

(1) The current population prevalence was estimated by computing the weighted sum of individual predicted probabilities from the logistic regression model, using all observed independent variables.
(2) A hypothetical prevalence was then estimated by recalculating the predicted probabilities after setting the risk factor A to zero (unexposed) for all participants.
(3) The PAF was calculated as: PAF=(observed prevalence−hypothetical prevalence)/observed prevalence.

While this approach provides valid estimates of individual PAFs, the sum of PAFs for multiple risk factors may exceed 100%. Therefore, we additionally employed the sequential PAF estimation method proposed by Eide and Gefeller [19], ensuring that the sum of individual PAFs remains bounded at 100%.

For the k identified risk factors, we calculated PAFs for all possible sequences ($k!$ permutations) of risk factor removal. For each sequence, we sequentially computed the PAF for each factor by setting it to zero in turn. The final reported PAF for each risk factor was the average of its calculated PAF across all permutations. For comparison, we also calculated PAFs using the formula suggested by Miettinen [20], which is known to be appropriate when adjusted relative risks are utilized [21]. In this method, we replaced the relative risk in the formula with prevalence rate ratios obtained from a log-binomial model that did not account for the survey structure.

For statistical inference, we estimated 95% confidence intervals (CIs) for the PAFs using a bootstrap approach. First, we obtained individual predicted probabilities of prediabetes from the previously fitted logistic regression model. We then performed 1000 bootstrap iterations. In each iteration, we (1) generated new binary outcomes for prediabetes by randomly sampling from a Bernoulli distribution using these predicted probabilities while holding all covariates fixed; and (2) recalculated PAFs using both individual and sequential methods on the bootstrapped dataset. The 2.5th and 97.5th percentiles of the resulting PAF distributions were used to construct the CIs.

As a secondary analysis, we explored potential interaction effects between age or sex and lifestyle factors. For each demographic variable (age and sex), we fitted 7 separate logistic

regression models, each including 1 interaction term between the demographic variable and a specific lifestyle factor. If any of these interaction terms achieved statistical significance ($p<0.05$), we conducted subgroup analyses and calculated PAFs stratified by that demographic variable.

To address potential reverse causality bias, our primary analysis excluded participants with diagnosed diabetes. As an additional sensitivity analysis, we broadened our case definition to include both prediabetes and undiagnosed diabetes (FBG ≥100 mg/dL or HbA1c ≥5.7%), under the assumption of minimal lifestyle modifications among individuals without a diabetes diagnosis. The significance level was set at 0.05 for all analyses. All analyses were performed using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA).

Ethics Statement

The study adhered to the ethical guidelines outlined in the Helsinki Declaration. All protocols were approved by the Institutional Review Board of the KDCA, and the requirement for informed consent was waived (IRB No. 2018-01-03-4C-A).

RESULTS

Baseline Characteristics

Among the 3104 participants, 1344 met the criteria for prediabetes (Table 1). Baseline characteristics differed significantly between prediabetes and control groups across all covariates, except for family history of chronic diseases. The prediabetes group had higher proportions of male, older individuals, and participants with hypertension and dyslipidemia. Among the 7 lifestyle factors evaluated, the prediabetes group exhibited higher proportions of abnormal body weight, excessive alcohol consumption, ever-smoking status, insufficient aerobic exercise, and irregular breakfast consumption. Conversely, the control group had a higher proportion of individuals who did not consume vegetables at every meal.

Lifestyle Risk Factor Identification

Logistic regression analysis accounting for the complex survey structure identified 5 lifestyle factors significantly associated with prediabetes (Table 2): abnormal body weight (odds ratio [OR], 2.05, $p<0.001$), excessive alcohol consumption (OR, 1.27, $p=0.050$), smoking history (OR, 1.35, $p=0.011$), insufficient exercise (OR, 1.26, $p=0.014$), and irregular breakfast consumption (OR, 1.31, $p=0.007$). Sleep duration showed no statistically significant association (OR, 1.02, $p=0.843$), and vegetable consumption demonstrated an inverse association with prediabetes (OR, 0.80, $p=0.031$). Since PAFs are conventionally calculated for risk factors rather than protective factors, vegetable consumption was not included in PAF calculations, despite remaining in the comprehensive regression model. Therefore, the PAF estimates for body weight, alcohol, smoking, exercise, and breakfast were assessed in subsequent analyses.

Population-attributable Fractions

PAFs were estimated using 2 methods (Table 3). Individual PAF estimation identified abnormal body weight as the largest contributor (23.3%; 95% CI, 17.1 to 29.5), followed by smoking (6.9%; 95% CI, 1.2 to 12.8), insufficient exercise (6.2%; 95% CI, 1.3 to 11.0), irregular breakfast consumption (5.7%; 95% CI, 0.6 to 10.5), and excessive alcohol consumption (4.1%; 95% CI, 0.1 to 8.1). The sequential estimation method yielded similar

Table 1. Baseline characteristics of the study population

Characteristics	Total population	Prediabetes[1]	Normal	*p*-value[2]
Total	3104 (100)	1344 (40.8)	1760 (59.2)	
Sex				<0.001
Male	1282 (47.8)	637 (56.2)	645 (42.0)	
Female	1822 (52.2)	707 (43.8)	1115 (58.0)	
Age (y)				<0.001
<45	524 (21.6)	127 (13.3)	397 (27.3)	
45-64	1668 (57.8)	684 (56.3)	984 (58.8)	
≥65	912 (20.6)	533 (30.4)	379 (13.9)	
Marital status				<0.001
Ever married	2819 (88.6)	1271 (92.5)	1548 (85.8)	
Never married	285 (11.4)	73 (7.5)	212 (14.2)	

(*Continued to the next page*)

Table 1. Continued from the previous page

Characteristics	Total population	Prediabetes[1]	Normal	*p*-value[2]
Education (y)				<0.001
Elementary school or lower (≤6)	541 (11.5)	313 (16.6)	228 (8.0)	
Middle school (7-9)	276 (7.4)	152 (9.9)	124 (5.7)	
High school (10-12)	961 (31.9)	430 (33.6)	531 (30.7)	
College or higher (≥13)	1326 (49.2)	449 (39.9)	877 (55.6)	
Family income				0.025
Low	554 (13.3)	287 (15.9)	267 (11.6)	
Middle-low	753 (21.7)	336 (22.1)	417 (21.5)	
Middle-high	847 (30.1)	345 (29.3)	502 (30.7)	
High	950 (34.8)	376 (32.8)	574 (36.1)	
Occupation				<0.001
No occupation	1124 (32.4)	506 (32.8)	618 (32.1)	
White-collar	1215 (44.6)	458 (40.4)	757 (47.5)	
Blue-collar	765 (23.0)	380 (26.9)	385 (20.3)	
Hypertension				<0.001
Undiagnosed	2303 (77.8)	859 (67.1)	1444 (85.2)	
Diagnosed	801 (22.2)	485 (32.9)	316 (14.8)	
Dyslipidemia				<0.001
Undiagnosed	2369 (79.5)	893 (69.9)	1476 (86.1)	
Diagnosed	735 (20.5)	451 (30.1)	284 (13.9)	
Family history of chronic diseases				0.165
Undiagnosed	802 (25.0)	343 (23.4)	459 (26.0)	
Diagnosed	2302 (75.0)	1001 (76.6)	1301 (74.0)	
Body weight (kg/m^2)				<0.001
Normal (18.5≤BMI<23.0)	1193 (36.9)	388 (25.5)	805 (44.8)	
Abnormal (other)	1911 (63.1)	956 (74.5)	955 (55.2)	
Sleep (hr/day)				0.389
7-8	1669 (53.6)	706 (52.6)	963 (54.3)	
Otherwise	1435 (46.4)	638 (47.4)	797 (45.7)	
Alcohol (drinks/sitting)				0.008
≤4 or non-drinker	2253 (67.7)	954 (64.3)	1299 (70.1)	
≥5	851 (32.3)	390 (35.7)	461 (29.9)	
Smoking				<0.001
Never-smoker	1906 (57.0)	751 (48.7)	1155 (62.8)	
Ever-smoker	1198 (43.0)	593 (51.3)	605 (37.2)	
Exercise (hr/wk)				<0.001
Aerobic exercise ≥2.5	1378 (47.1)	532 (41.7)	846 (50.9)	
Aerobic exercise <2.5	1726 (52.9)	812 (58.3)	914 (49.1)	
Breakfast (times/wk)				<0.001
5-7	1920 (55.9)	903 (60.8)	1017 (52.4)	
0-4	1184 (44.1)	441 (39.2)	743 (47.6)	
Diet				<0.001
Vegetables at every meal	1222 (34.6)	615 (41.1)	607 (30.1)	
Otherwise	1882 (65.4)	729 (58.9)	1153 (69.9)	

Values are presented as number (%).
[1]Prediabetes was defined as either hemoglobin A1c 5.7-6.4% or fasting blood glucose 100-125 mg/dL.
[2]Using χ^2-test.

results: abnormal body weight (22.2%; 95% CI, 16.2 to 28.2), smoking (6.4%; 95% CI, 1.1 to 11.6), insufficient exercise (5.8%; 95% CI, 1.2 to 10.5), irregular breakfast consumption (4.9%; 95% CI, 0.5 to 9.2), and excessive alcohol consumption (3.6%; 95% CI, 0.1 to 7.4).

All lifestyle risk factors demonstrated statistically significant PAFs, with CIs excluding zero. The PAF for abnormal body weight was significantly higher than for other lifestyle risk factors, as indicated by non-overlapping CIs. Absolute differences between the 2 methods were minimal, approximately 1 percentage point for all lifestyle factors. Estimates using Miettinen [20]'s formula were approximately 60% of those obtained using regression-based methods.

In secondary analyses exploring interaction effects, no significant interactions between sex and any lifestyle factor were observed. However, for age, significant interactions were identified with abnormal weight ($p=0.027$) and excessive alcohol consumption ($p=0.015$). Consequently, we stratified the population into younger (≤50 years) and older (≥51 years) subgroups based on the weighted median age (50.6 years) and calculated subgroup-specific PAFs.

Age stratification revealed distinct patterns in lifestyle factor contributions to prediabetes (Table 4). Generally, PAFs were higher in the younger subgroup and lower in the older subgroup, irrespective of the estimation method used. Abnormal body weight had a notably higher sequential PAF in the younger subgroup (29.2%; 95% CI, 18.6 to 40.8) compared to the older subgroup (17.5%; 95% CI, 11.3 to 24.3). In the older subgroup, excessive alcohol consumption (0.4%; 95% CI, -2.8 to 3.5), smoking (3.9%; 95% CI, -2.6 to 10.1), and irregular breakfast consumption (1.1%; 95% CI, -2.1 to 4.1) yielded non-significant PAFs, with CIs including 0. Interestingly, insufficient exercise demonstrated the opposite trend, with a lower PAF in the younger subgroup (2.9%; 95% CI, -4.1 to 10.3) compared to the older subgroup (7.3%; 95% CI, 1.8 to 13.0).

Table 2. Associations between lifestyle factors and prediabetes

Factors	aOR (95% CI)
Body weight (kg/m^2)	
Normal (18.5≤BMI<23.0)	1.00 (reference)
Abnormal (other)	2.05 (1.68, 2.50)*
Sleep (hr/day)	
7-8	1.00 (reference)
Otherwise	1.02 (0.86, 1.21)
Alcohol (drinks/sitting)	
≤4 or non-drinker	1.00 (reference)
≥5	1.27 (1.00, 1.62)*
Smoking	
Never-smoker	1.00 (reference)
Ever-smoker	1.35 (1.07, 1.71)*
Exercise (hr/wk)	
Aerobic exercise ≥2.5	1.00 (reference)
Aerobic exercise <2.5	1.26 (1.05, 1.51)*
Breakfast (times/wk)	
5-7	1.00 (reference)
0-4	1.31 (1.08, 1.59)*
Diet	
Vegetables at every meal	1.00 (reference)
Otherwise	0.80 (0.66, 0.98)

aOR, adjusted odds ratio; CI, confidence interval.
*$p<0.05$.

Sensitivity Analysis

A sensitivity analysis was conducted by expanding the case definition to include undiagnosed diabetes, increasing the number of cases from 1344 to 1463. Logistic regression analysis identified the same set of significant lifestyle factors: abnormal body weight (OR, 2.13, $p<0.001$), excessive alcohol consumption (OR, 1.30, $p=0.031$), smoking (OR, 1.32, $p=0.013$), insufficient exercise (OR, 1.26, $p=0.011$), and irregular breakfast consumption (OR, 1.28, $p=0.014$), associated with the combined outcome of prediabetes or undiagnosed diabetes.

Table 3. PAF of lifestyle factors for prediabetes

Lifestyle factors	aOR	Prevalence (%)	Individual PAF, % (95% CI)	Sequential PAF, % (95% CI)	Miettinen's formula (%)[1]
Abnormal weight	2.05	63.1	23.3 (17.1, 29.5)	22.2 (16.2, 28.2)	14.2
Excessive alcohol consumption	1.27	32.3	4.1 (0.1, 8.1)	3.6 (0.1, 7.4)	2.2
Smoking	1.35	43.0	6.9 (1.2, 12.8)	6.4 (1.1, 11.6)	4.1
Insufficient exercise	1.26	52.9	6.2 (1.3, 11.0)	5.8 (1.2, 10.5)	3.7
Irregular breakfast	1.31	44.1	5.7 (0.6, 10.5)	4.9 (0.5, 9.2)	2.6

PAF, population-attributable fractions; aOR, adjusted odds ratio; CI, confidence interval.
[1]Miettinen's formula estimates were derived by substituting prevalence rate ratio for relative risk.

Table 4. PAF of lifestyle factors for prediabetes for age subgroups

Lifestyle factors	aOR	Prevalence (%)	Individual PAF, % (95% CI)	Sequential PAF, % (95% CI)	Miettinen's formula (%)[1]
Age ≤ 50					
Abnormal weight	2.37	61.6	34.1 (21.8, 47.2)	29.2 (18.6, 40.8)	18.0
Excessive alcohol consumption	1.47	42.1	11.4 (1.0, 21.7)	9.0 (0.7, 17.1)	7.9
Smoking	1.48	45.0	12.7 (0.2, 25.4)	10.1 (0.1, 20.1)	5.7
Insufficient exercise	1.13	47.8	3.8 (-5.3, 12.8)	2.9 (-4.1, 10.3)	1.1
Irregular breakfast	1.53	64.6	16.7 (0.5, 32.2)	13.2 (0.4, 24.9)	10.0
Age ≥ 51					
Abnormal weight	1.85	64.5	17.5 (11.3, 24.4)	17.5 (11.3, 24.3)	15.1
Excessive alcohol consumption	1.04	23.1	0.4 (-2.8, 3.5)	0.4 (-2.8, 3.5)	-0.2
Smoking	1.24	41.0	3.8 (-2.5, 10.2)	3.9 (-2.6, 0.1)	4.5
Insufficient exercise	1.34	57.7	7.3 (1.8, 13.1)	7.3 (1.8, 13.0)	6.7
Irregular breakfast	1.11	25.0	1.2 (-2.2, 4.3)	1.1 (-2.1, 4.1)	0.2

PAF, population-attributable fractions; aOR, adjusted odds ratio; CI, confidence interval.
[1]Miettinen's formula estimates were derived by substituting prevalence rate ratio for relative risk.

Table 5. PAF of lifestyle factors for either prediabetes or undiagnosed diabetes

Lifestyle factors	aOR	Prevalence (%)	Individual PAF, % (95% CI)	Sequential PAF, % (95% CI)	Miettinen's formula (%)[1]
Abnormal weight	2.13	63.7	24.1 (18.5, 30.1)	23.1 (18.3, 29.8)	13.2
Excessive alcohol consumption	1.30	32.5	4.0 (-3.8, 7.6)	3.8 (0.7, 7.1)	2.4
Smoking	1.32	43.1	5.3 (-3.6, 10.2)	5.7 (0.7, 11.1)	3.3
Insufficient exercise	1.26	53.1	6.3 (-2.6, 10.1)	5.6 (1.3, 10.1)	2.9
Irregular breakfast	1.28	43.8	4.9 (-3.6, 9.1)	4.4 (-0.2, 8.8)	2.1

PAF, population-attributable fractions; aOR, adjusted odds ratio; CI, confidence interval.
[1]Miettinen's formula estimates were derived by substituting the prevalence rate ratio for the relative risk.

The PAF estimates for these 5 factors yielded results consistent with the main analysis in terms of order of importance (Table 5). Individual PAF estimation revealed abnormal body weight as the largest contributor (24.1%; 95% CI, 18.5 to 30.1), followed by insufficient exercise (6.3%; 95% CI, -2.6 to 10.1), smoking (5.3%; 95% CI, -3.6 to 10.2), irregular breakfast consumption (4.9%; 95% CI, -3.6 to 9.1), and excessive alcohol consumption (4.0%; 95% CI, -3.8 to 7.6). Sequential PAF estimation showed a similar pattern: abnormal body weight (23.1%; 95% CI, 18.3 to 29.8), smoking (5.7%; 95% CI, 0.7 to 11.1), insufficient exercise (5.6%; 95% CI, 1.3 to 10.1), irregular breakfast consumption (4.4%; 95% CI, -0.2 to 8.8), and excessive alcohol consumption (3.8%; 95% CI, 0.7 to 7.1). While all lifestyle risk factors showed significant PAFs in the primary analysis, the sensitivity analysis revealed statistical significance only for abnormal body weight in individual PAF estimation, with irregular breakfast consumption losing significance in sequential PAF estimation.

DISCUSSION

In this nationally representative study of the Korean population, we identified lifestyle risk factors for prediabetes and estimated their PAFs using 2 different regression-based approaches. Five modifiable lifestyle factors were significantly associated with prediabetes: abnormal body weight, excessive alcohol consumption, smoking, insufficient exercise, and irregular breakfast consumption. Both PAF estimation methods consistently identified abnormal body weight as the largest contributor, yielding PAFs exceeding 20%, with 95% CIs that did not overlap those of other lifestyle factors. Each of the remaining lifestyle factors showed PAFs of approximately 5%, all with CIs excluding zero. Both individual and sequential PAF estimation methods produced similar results, with minimal differences between the 2 approaches. Age exhibited significant interaction effects, with higher PAFs observed in the younger subgroup and lower PAFs in the older subgroup. Consistent findings were

observed when the case cohort was expanded to include individuals with undiagnosed diabetes.

The concept of the PAF was first proposed by Levin [22] and defined as "the proportion of disease that could be eliminated by eliminating the exposure" [23]. Numerous formulas have since been developed to handle more complex situations, but most require incidence relative risk in the estimation process [21]. Bruzzi et al. [24] approximated the PAF from case-control data by substituting adjusted ORs, derived from logistic regression, for relative risks. Greenland and Drescher [25] further derived a maximum likelihood estimate of PAF suitable for case-control studies. However, these approaches apply only under the rarity assumption, wherein approximating relative risks with ORs is plausible.

While our method also employs logistic regression, we directly derived each participant's probability of prediabetes under various hypothetical scenarios and calculated the expected number of cases, rather than substituting ORs for relative risks. PAFs were then computed based on these expected case numbers. Thus, our regression-based method effectively avoids the problems associated with approximating relative risks with ORs, such as overestimation of PAFs. Nonetheless, we acknowledge that cross-sectional designs inherently limit causal inference; therefore, our PAF estimates should be interpreted as the proportion of prevalent prediabetes cases attributable to each risk factor, rather than as proportions of incident cases preventable through elimination of exposure.

This method was suggested by Kooperberg and Petitti [26] and illustrated by Rückinger et al. [18], using data from the National Health and Nutrition Examination Survey focused on cardiovascular diseases. Notably, the rarity assumption is unnecessary in this method, as it does not approximate relative risks using ORs. Therefore, this approach is particularly useful for many non-communicable diseases that do not satisfy the rarity assumption.

Although this regression-based method was initially developed in 1995, it has been underutilized in practical applications. Beyond the advantage of not requiring the rarity assumption, the sequential estimation method offers a significant benefit by partially accounting for correlations between risk factors, ensuring that the total PAF for multiple risk factors remains bounded at 100%. This methodological strength makes the sequential method highly relevant for future PAF estimation studies.

In our study, abnormal body weight demonstrated the highest OR (2.05) and prevalence (63.1%), leading to PAFs exceeding 20% across both estimation methods. No other lifestyle factor had a PAF exceeding 10%. Although our cross-sectional study design limits causal inference between body weight and prediabetes, previous research has robustly demonstrated causal relationships between body weight and diabetes risk. For instance, a randomized controlled trial involving overweight Finnish participants found significantly lower diabetes incidence among individuals assigned to a weight loss intervention compared to controls [27]. Combined with our results, such findings highlight the critical importance of weight management among lifestyle factors in diabetes prevention.

Among the remaining lifestyle factors, insufficient exercise warrants particular attention. Despite having the lowest OR (1.26) among the 5 significant lifestyle factors, its high prevalence (52.9%) resulted in a larger PAF compared to excessive alcohol consumption and irregular breakfast consumption. This finding underscores the substantial influence of risk factor prevalence when evaluating the potential impact of population-level interventions.

Several limitations of this study should be considered. First, the cross-sectional nature of our data prevented us from establishing causal relationships between identified lifestyle factors and prediabetes. Moreover, because our regression model was based on prevalence rather than incidence data, the estimated PAFs reflect prevalence risk rather than incidence risk attributable to these lifestyle factors. Thus, our findings should be interpreted specifically in the context of prevalence. Prospective studies using incident cases would be valuable to validate PAF estimates derived from this method.

Second, the validity of regression-based PAF estimates depends upon the appropriateness of the regression model used. If the model inadequately captures outcome variation due to omitted relevant variables or if underlying assumptions (e.g., linearity of parameters, choice of link function) are violated, PAF estimates may be biased. Although we controlled for known confounders, residual confounding and model misspecification remain potential limitations.

Third, while our use of prediabetes as a surrogate endpoint minimized potential reverse causality bias associated with diagnosed diabetes, this choice implies our PAF estimates pertain specifically to prediabetes rather than diabetes itself. Nevertheless, reverse causality risk remains even with prediabetes. In our study, insufficient vegetable consumption exhibited a protective effect, contrary to existing literature, suggesting

potential reverse causality bias. Future studies employing this method with cross-sectional data should carefully consider study design and interpretation.

Fourth, our study could not include impaired glucose tolerance in the case definition due to data limitations, potentially causing misclassification of some prediabetes cases as controls. Because such misclassification biases associations toward the null, our reported associations and PAF estimates are likely conservative.

Fifth, the potential reversibility of prediabetes risk via lifestyle modification may vary across factors, as may the feasibility of achieving these modifications. Thus, policymakers should consider not only the magnitude of PAFs but also the practicality and effectiveness of interventions targeting each lifestyle factor when designing prevention programs.

In this nationally representative study of a Korean cohort, we identified 5 lifestyle factors significantly associated with prediabetes and estimated their PAFs using regression-based sequential methods. Abnormal body weight had the largest contribution, with a PAF exceeding 20%, while other lifestyle factors each demonstrated PAFs of approximately 5%. Based on these findings, we recommend that diabetes prevention programs in Korea prioritize weight management interventions while maintaining comprehensive lifestyle modification strategies.

NOTES

Conflict of Interest

The authors have no conflicts of interest associated with the material presented in this paper.

Funding

None.

Acknowledgements

None.

Author Contributions

Conceptualization: Oh YW, Nam CM, Park EC. Data curation: Oh YW. Formal analysis: Oh YW. Funding acquisition: None. Methodology: Oh YW, Nam CM. Project administration: Park EC. Visualization: Oh YW. Writing – original draft: Oh YW. Writing – review & editing: Park EC, Nam CM.

ORCID

Yeon Woo Oh *https://orcid.org/0009-0002-2485-1400*
Chung Mo Nam *https://orcid.org/0000-0003-0985-0928*
Eun-Cheol Park *https://orcid.org/0000-0002-2306-5398*

REFERENCES

1. Farmaki P, Damaskos C, Garmpis N, Garmpi A, Savvanis S, Diamantis E. Complications of the type 2 diabetes mellitus. Curr Cardiol Rev 2020;16(4):249-251. https://doi.org/10.2174/1573403X1604201229115531
2. GBD 2021 Diabetes Collaborators. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2023;402(10397):203-234. https://doi.org/10.1016/S0140-6736(23)01301-6
3. Centers for Disease Control and Prevention (CDC). National diabetes prevention program [cited 2024 Dec 10]. Available from: https://www.cdc.gov/diabetes-prevention/index.html
4. National Health Service. Healthier you [cited 2024 Dec 10]. Available from: https://healthieryou.org.uk/
5. Woo JT. Follow-up study of Korea Diabetes Prevention Study (KDPS). Seoul: Kyung Hee University; 2023 (Korean).
6. Hu FB, Manson JE, Stampfer MJ, Colditz G, Liu S, Solomon CG, et al. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. N Engl J Med 2001;345(11):790-797. https://doi.org/10.1056/NEJMoa010492
7. Gu F, Zhou S, Lou K, Deng R, Li X, Hu J, et al. Lifestyle risk factors and the population attributable fractions for overweight and obesity in Chinese students of Zhejiang Province. Front Pediatr 2021;9:734013. https://doi.org/10.3389/fped.2021.734013
8. Rajaobelina K, Dow C, Romana Mancini F, Dartois L, Boutron-Ruault MC, Balkau B, et al. Population attributable fractions of the main type 2 diabetes mellitus risk factors in women: findings from the French E3N cohort. J Diabetes 2019;11(3):242-253. https://doi.org/10.1111/1753-0407.12839
9. Celentano DD, Szklo M, Gordis L. Gordis epidemiology. 6th ed. Philadelphia: Elsevier; 201, p. 164-165.
10. Szklo M, Nieto FJ. Epidemiology: beyond the basics. 4th ed. Burlington: Jones & Bartlett Learning; 2019, p. 157-158.
11. Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. Prediabetes: a high-risk state for diabetes development. Lancet 2012;379(9833):2279-2290. https://doi.org/10.1016/S0140-6736(12)60283-9

12. Barry E, Greenhalgh T, Fahy N. How are health-related behaviours influenced by a diagnosis of pre-diabetes? A meta-narrative review. BMC Med 2018;16(1):121. https://doi.org/10.1186/s12916-018-1107-6
13. Kweon S, Kim Y, Jang MJ, Kim Y, Kim K, Choi S, et al. Data resource profile: the Korea National Health and Nutrition Examination Survey (KNHANES). Int J Epidemiol 2014;43(1):69-77. https://doi.org/10.1093/ije/dyt228
14. Lloyd-Jones DM, Allen NB, Anderson CA, Black T, Brewer LC, Foraker RE, et al. Life's essential 8: updating and enhancing the American Heart Association's construct of cardiovascular health: a presidential advisory from the American Heart Association. Circulation 2022;146(5):e18-e43. https://doi.org/10.1161/CIR.0000000000001078
15. American Heart Association. Life's Essential 8 [cited 2024 Dec 5]. Available from: https://www.heart.org/en/healthy-living/healthy-lifestyle/lifes-essential-8
16. Mekary RA, Giovannucci E, Willett WC, van Dam RM, Hu FB. Eating patterns and type 2 diabetes risk in men: breakfast omission, eating frequency, and snacking. Am J Clin Nutr 2012;95(5):1182-1189. https://doi.org/10.3945/ajcn.111.028209
17. Seo JS, Kim SY, Choi SH, Park HG, Lim KY. Relationship between breakfast skipping and prediabetes risk in non-diabetic adults. Korean J Fam Pract 2016;6(5):464-469 (Korean). https://doi.org/10.21215/kjfp.2016.6.5.464
18. Rückinger S, von Kries R, Toschke AM. An illustration of and programs estimating attributable fractions in large scale surveys considering multiple risk factors. BMC Med Res Methodol 2009;9:7. https://doi.org/10.1186/1471-2288-9-7
19. Eide GE, Gefeller O. Sequential and average attributable fractions as aids in the selection of preventive strategies. J Clin Epidemiol 1995;48(5):645-655. https://doi.org/10.1016/0895-4356(94)00161-i
20. Miettinen OS. Proportion of disease caused or prevented by a given exposure, trait or intervention. Am J Epidemiol 1974;99(5):325-332. https://doi.org/10.1093/oxfordjournals.aje.a121617
21. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. Am J Public Health 1998;88(1):15-19. https://doi.org/10.2105/ajph.88.1.15
22. Levin ML. The occurrence of lung cancer in man. Acta Unio Int Contra Cancrum 1953;9(3):531-541.
23. Rothman KJ, Lash TL, VanderWeele TJ, Haneuse S. Modern epidemiology. 4th ed. Philadelphia: Wolters Kluwer; 2021, p. 93.
24. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. Am J Epidemiol 1985;122(5):904-914. https://doi.org/10.1093/oxfordjournals.aje.a114174
25. Greenland S, Drescher K. Maximum likelihood estimation of the attributable fraction from logistic models. Biometrics 1993;49(3):865-872. https://doi.org/10.2307/2532206
26. Kooperberg C, Petitti DB. Using logistic regression to estimate the adjusted attributable risk of low birthweight in an unmatched case-control study. Epidemiology 1991;2(5):363-366. https://doi.org/10.1097/00001648-199109000-00009
27. Tuomilehto J, Lindström J, Eriksson JG, Valle TT, Hämäläinen H, Ilanne-Parikka P, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. N Engl J Med 2001;344(18):1343-1350. https://doi.org/ 10.1056/NEJM200105033441801