Survival Analysis Workshop

Jooyoung Lee

Department of Applied Statistics

Chung-Ang University

Seoul Medical Center

January 27, 2023

Week 4: Clinical Study using NHISS II

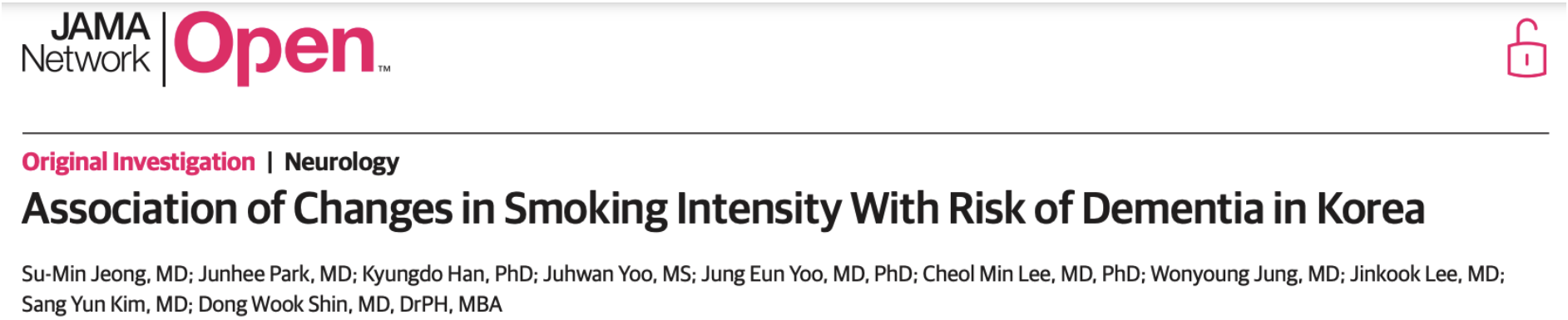
STROBE Checklist

- https://www.strobe-statement.org/

- STROBE stands for an international, collaborative initiative of epidemiologists, methodologists, statisticians, researchers and journal editors involved in the conduct and dissemination of observational studies, with the common aim of STrengthening the Reporting of OBservational studies in Epidemiology.

Example using NHISS

Jeong, S. M., Park, J., Han, K., Yoo, J., Yoo, J. E., Lee, C. M., ... & Shin, D. W. (2023). Association of Changes in Smoking Intensity With Risk of Dementia in Korea. JAMA Network Open, 6(1), e2251506-e2251506.



STROBE: Title and abstract

(a) Indicate the study's design with a commonly used term in the title or the abstract

(b) Provide in the abstract an informative and balanced summary of what was done and what was found

STROBE: Introduction

- Background/rationale: Explain the scientific background and rationale for the investigation being reported

- Objectives: State specific objectives, including any prespecified hypotheses

STROBE: Methods

- Study design: Present key elements of study design early in the paper
- Setting: Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
- Bias: Describe any efforts to address potential sources of bias
- Study size: Explain how the study size was arrived at

Example using NHISS

Methods

Design, Setting, and Population

This population-based, retrospective cohort study used data from the National Health Insurance Service (NHIS), a mandatory universal insurance system that covers the entire population in Korea. This study was approved by the Samsung Medical Center Institutional Review Board, which waived the informed consent requirement because the study used retrospective anonymized data. We followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.

Example using NHISS

The NHIS database contains demographic data and links the data to health care claims. The NHIS database also contains data from the national health checkup program, which covers all employed individuals and beneficiaries 40 years or older, including data from self-administered health questionnaires, anthropometric measurements, and laboratory tests.[15]

Example using NHISS

We identified from the NHIS database participants 40 years or older who underwent biennial health examinations in 2009 and 2011. We selected participants with current smoking status (n = 1 006 855) at the first health examination, according to the definition of the World Health Organization.[16] Participants who were diagnosed with any cancer (n = 15 629), stroke (n = 42 165), myocardial infarction (n = 19 544), or dementia (n = 986) prior to the second health examination period (2011) were excluded. We applied a 1-year lag time to reduce the implications of reverse causality by excluding participants who were diagnosed with cancer (n = 9328), stroke (n = 6426), myocardial infarction (n = 2958), or dementia (n = 714) or participants who died (n = 1834) within 1 year after the second health examination (2011). Those who had missing information on variables (eTable 2 in Supplement 1) used in this study were excluded, resulting in the inclusion of 789 532 individuals in the study cohort.

STROBE: Methods

- Variables: Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable

- Data sources/ measurement: For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group

- Quantitative variables: Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why

Example using NHISS

Definition of Change in Cigarette Smoking Intensity

Information on smoking status and change in smoking intensity was obtained from self-administered questionnaires during the health checkup. Participants who acknowledged having smoked at least 100 cigarettes in their lifetime, based on the World Health Organization definition,[16] were asked about current smoking status. Then, current smokers were questioned on the duration of smoking

and mean number of cigarettes smoked per day (eTable 3 in Supplement 1). According to cigarette smoking intensity at the time of the first health examination (2009), participants were categorized as mild smokers (<10 cigarettes per day), moderate smokers (10-19 cigarettes per day), and heavy smokers (≥20 cigarettes per day).[17]

Example using NHISS

In this study, change in cigarette smoking intensity was identified based on (1) relative change in the number of cigarettes smoked per day (reducer or increaser group) and (2) categorical change in the level of smoking intensity (eg, mild to heavy smoking status). Participants were categorized into 5 groups based on a relative change in smoking intensity between the first (2009) and the second (2011) health examinations: quitter, reducer I, reducer II, sustainer, or increaser, based on definitions used in previous studies.[18,19] Quitters were defined as those who completely stopped

Example using NHISS

Outcomes and Follow-up

The primary end point was newly diagnosed dementia, identified on the basis of prescribed antidementia medications (rivastigmine, galantamine, memantine, or donepezil) along with *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10)* codes for dementia (F00, F01, F02, F03, G30, or G31).[20-24] For valid medical expense claims submitted to the NHIS, physicians need to document the evidence for cognitive dysfunction according to strict criteria. In particular, a Mini-Mental State Examination score of 26 or lower (scale: 0-30, with the lowest score indicating dementia) and either a Clinical Dementia Rating of 1 or higher (scale: 0-5, with the highest rating indicating dementia) or Global Deterioration Scale score of 3 or higher (scale: 1-7, with the highest score indicating dementia) were required.[25] Patients were categorized as having Alzheimer disease (AD; *ICD-10* code F00 or G30), vascular dementia (VaD; *ICD-10* code F01), or another type of dementia.[3] The cohort was followed up after 1 year of lag time from the second health examination (2011) until the end of the study period (December 31, 2018). Information on covariates is provided in the eMethods in Supplement 1.

STROBE: Methods

Statistical methods:

(a) Describe all statistical methods, including those used to control for confounding

(b) Describe any methods used to examine subgroups and interactions

(c) Explain how missing data were addressed

(d) If applicable, explain how loss to follow-up was addressed

(e) Describe any sensitivity analyses

Example using NHISS

Statistical Analysis

Continuous variables were presented as mean (SD), and categorical variables were presented as number (percentage). Hazard ratios (HR) with 95% CIs for all dementia and subtypes (AD or VaD) were calculated using a Cox proportional hazards regression model. The proportional hazards assumption was tested using Schoenfeld residuals. Multivariable models were adjusted for age; sex; household income; alcohol consumption; regular physical activity; area of residence; comorbidities, such as hypertension, diabetes, dyslipidemia, and chronic kidney disease; and body mass index.

Example using NHISS

Analysis was performed using 2 reference groups: (1) sustainers (relative change of <20% in number of cigarettes smoked per day) and (2) those in the same category of smoking intensity (eg, both in the mild smoker category for 2009 and 2011 health examinations) (eTable 4 in Supplement 1). We analyzed the restricted cubic spline curve to assess the association between change in smoking intensity treated as a continuous variable and the incidence of dementia.

Example using NHISS

In the sensitivity analysis, a subdistribution hazard model regression using the Fine-Gray methods was performed to estimate the subdistribution HR for dementia incidence, accounting for death as a competing event.[26] Multiple imputation was also used to account for missing covariate data.[27] Stratification analyses by smoking intensity at the first health examination, with age, sex, and alcohol consumption as confounding factors, were performed to assess the association between change in smoking intensity and incidence of dementia.

STROBE: Results

Participants:

(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed

(b) Give reasons for non-participation at each stage

(c) Consider use of a flow diagram

STROBE: Results

Descriptive data:

(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders

(b) Indicate number of participants with missing data for each variable of interest

(c) Summarise follow-up time (eg, average and total amount)

Example using NHISS

Table 1. Baseline Characteristics of the Study Population in 2011

Variable	All participants (n = 789 532)	Participant group[a]					*P* value
		Quitter (n = 114 959)	Reducer I (n = 60 767)	Reducer II (n = 111 890)	Sustainer (n = 376 393)	Increaser (n = 125 523)	
Age, mean (SD), y	52.2 (8.5)	53.3 (8.8)	53.9 (9.4)	51.9 (8.3)	52.0 (8.3)	52.0 (8.6)	<.001
Sex, No. (%)							
Female	33 063 (4.2)	4811 (4.2)	4072 (6.7)	4501 (4.0)	12 765 (3.4)	6914 (5.5)	<.001
Male	756 469 (95.8)	110 148 (95.8)	56 695 (93.3)	107 389 (96.0)	363 628 (96.6)	118 609 (94.5)	<.001
Alcohol consumption, No. (%)							
None	203 562 (25.8)	32 916 (28.6)	17 800 (29.3)	28 038 (25.1)	93 169 (24.8)	31 639 (25.2)	<.001
Mild	278 957 (35.3)	43 483 (37.8)	23 939 (39.4)	41 800 (37.4)	128 661 (34.2)	41 074 (32.7)	NA
Moderate	178 391 (22.6)	22 745 (19.8)	11 900 (19.6)	25 469 (22.8)	89 670 (23.8)	28 607 (22.8)	NA
Heavy	128 622 (16.3)	15 815 (13.8)	7128 (11.7)	16 583 (14.8)	64 893 (17.2)	24 203 (19.3)	NA
Regular physical activity, No. (%)							
None	632 538 (80.1)	85 519 (74.4)	47 665 (78.4)	89 733 (80.2)	307 234 (81.6)	102 387 (81.6)	<.001
Regular	156 994 (19.9)	29 440 (25.6)	13 102 (21.6)	22 157 (19.8)	69 159 (18.4)	23 136 (18.4)	<.001

Example using NHISS

Variables in Table 1: Age, Sex, Alcohol consumption, Regular physical activity, Height, Weight, Waist circumference, BMI, Blood pressure, Comorbidities (Hypertension, Diabetes, Dyslipidemia, CKD), Laboratory findings (Glucose, Total cholesterol, HDL, LDL, GFR), Urban residency, Household income, Smoking status...

STROBE: Results

Outcome data: Report numbers of outcome events or summary measures over time Main results:

(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included

(b) Report category boundaries when continuous variables were categorized

Example using NHISS

Table 2. Association Between Relative Changes in Cigarette Smoking Intensity and Risk of Dementia

Smoking status in 2009[a]	Smoking status in 2011[a]	Participants, No. (%)	Dementia cases, No.	Duration, person-year[b]	IR	Crude model HR (95% CI)	Age-adjusted model aHR (95% CI)[c]	Multivariate model aHR (95% CI)[d]	Multivariate competing risk SHR (95% CI)
All dementia (total)									
All current smokers (n = 789 532)	Quitter	114 959 (14.6)	1730	722 407	2.4	1.13 (1.07-1.20)	0.87 (0.83-0.92)	0.92 (0.87-0.97)	0.92 (0.87-0.97)
	Reducer I	60 767 (7.7)	1585	376 954	4.2	1.99 (1.88-2.11)	1.26 (1.19-1.33)	1.25 (1.18-1.33)	1.24 (1.17-1.32)
	Reducer II	111 890 (14.2)	1570	699 937	2.2	1.07 (1.01-1.13)	1.05 (0.99-1.11)	1.06 (1.00-1.12)	1.05 (0.99-1.12)
	Sustainer	376 393 (47.7)	4958	2 355 282	2.1	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
	Increaser	125 523 (15.9)	2069	783 225	2.6	1.26 (1.19-1.32)	1.13 (1.07-1.20)	1.12 (1.06-1.18)	1.11 (1.05-1.17)
Mild smokers: <10 CPD (n = 69 292)	Quitter	18 417 (26.6)	435	115 189	3.8	0.84 (0.73-0.96)	0.97 (0.84-1.11)	1.00 (0.87-1.15)	1.02 (0.89-1.18)
	Reducer I	2694 (3.9)	116	16 587	7.0	1.55 (1.26-1.91)	1.24 (1.01-1.52)	1.24 (1.01-1.52)	1.29 (1.04-1.61)
	Reducer II	6280 (9.1)	282	38 662	7.3	1.62 (1.39-1.89)	1.18 (1.02-1.38)	1.19 (1.02-1.38)	1.19 (1.01-1.40)
	Sustainer	14 693 (21.2)	411	91 292	4.5	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
	Increaser	27 208 (39.3)	838	168 476	5.0	1.11 (0.98-1.24)	1.17 (1.04-1.32)	1.17 (1.04-1.32)	1.20 (1.06-1.36)
Moderate smokers: 10-19 CPD (n = 295 770)	Quitter	47 357 (16.0)	637	298 035	2.1	0.97 (0.89-1.06)	0.85 (0.77-0.93)	0.88 (0.81-0.97)	0.88 (0.80-0.97)
	Reducer I	13 062 (4.4)	424	80 973	5.2	2.38 (2.14-2.64)	1.35 (1.22-1.51)	1.30 (1.17-1.45)	1.28 (1.14-1.44)
	Reducer II	38 056 (12.9)	654	237 973	2.7	1.25 (1.14-1.37)	1.10 (1.01-1.12)	1.12 (1.02-1.22)	1.13 (1.03-1.24)
	Sustainer	125 764 (42.5)	1731	788 212	2.2	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
	Increaser	71 531 (24.2)	970	447 233	2.2	0.99 (0.91-1.07)	1.16 (1.08-1.26)	1.15 (1.06-1.24)	1.14 (1.05-1.23)
Heavy smokers: ≥20 CPD (n = 424 470)	Quitter	49 185 (11.6)	658	309 183	2.1	1.11 (1.02-1.21)	0.86 (0.79-0.94)	0.92 (0.85-1.01)	0.92 (0.84-1.00)
	Reducer I	45 011 (10.6)	1045	279 395	3.7	1.95 (1.82-2.10)	1.21 (1.12-1.30)	1.21 (1.13-1.30)	1.19 (1.11-1.29)
	Reducer II	67 554 (15.9)	634	423 302	1.5	0.79 (0.72-0.86)	0.96 (0.88-1.05)	0.98 (0.90-1.07)	0.96 (0.88-1.05)
	Sustainer	235 936 (55.6)	2816	1 475 778	1.9	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
	Increaser	26 784 (6.3)	261	167 517	1.6	0.82 (0.72-0.93)	1.08 (0.95-1.23)	1.06 (0.93-1.20)	1.05 (0.93-1.20)

Example using NHISS

Alzheimer dementia									
All current smokers (n = 789 532)	Quitter	114 959 (14.6)	1314	722 407	1.8	1.17 (1.10-1.25)	0.89 (0.84-0.95)	0.94 (0.88-1.00)	0.94 (0.88-1.01)
	Reducer I	60 767 (7.7)	1191	376 954	3.2	2.04 (1.88-2.11)	1.25 (1.17-1.33)	1.24 (1.16-1.32)	1.23 (1.14-1.32)
	Reducer II	111 890 (14.2)	1162	699 937	1.7	1.08 (1.01-1.15)	1.05 (0.98-1.12)	1.06 (0.99-1.13)	1.05 (0.98-1.13)
	Sustainer	376 393 (47.7)	3636	2 355 282	1.5	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
	Increaser	125 523 (15.9)	1497	783 225	1.9	1.24 (1.17-1.32)	1.10 (1.03-1.17)	1.08 (1.02-1.15)	1.08 (1.01-1.15)
Mild smokers: <10 CPD (n = 69 292)	Quitter	18 417 (26.6)	357	115 189	3.1	0.89 (0.76-1.03)	1.03 (0.89-1.20)	1.08 (0.92-1.25)	1.13 (0.96-1.32)
	Reducer I	2694 (3.9)	93	16 587	5.6	1.60 (1.27-2.02)	1.26 (1.00-1.59)	1.26 (1.00-1.59)	1.34 (1.05-1.71)
	Reducer II	6280 (9.1)	221	38 662	5.7	1.64 (1.38-1.94)	1.18 (0.99-1.40)	1.18 (0.99-1.40)	1.19 (0.99-1.44)
	Sustainer	14 693 (21.2)	319	91 292	3.5	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
	Increaser	27 208 (39.3)	636	168 476	3.8	1.08 (0.95-1.24)	1.15 (1.01-1.32)	1.15 (1.00-1.31)	1.19 (1.03-1.37)
Moderate smokers: 10-19 CPD (n = 295 770)	Quitter	47 357 (16.0)	469	298 035	1.6	0.96 (0.86-1.07)	0.84 (0.75-0.93)	0.88 (0.79-0.98)	0.89 (0.79-0.99)
	Reducer I	13 062 (4.4)	337	80 973	4.2	2.54 (2.25-2.86)	1.40 (1.24-1.58)	1.33 (1.18-1.51)	1.32 (1.15-1.51)
	Reducer II	38 056 (12.9)	492	237 973	2.1	1.27 (1.14-1.41)	1.11 (1.00-1.23)	1.13 (1.01-1.25)	1.15 (1.03-1.28)
	Sustainer	125 764 (42.5)	1286	788 212	1.6	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
	Increaser	71 531 (24.2)	681	447 233	1.5	0.94 (0.85-1.03)	1.12 (1.02-1.23)	1.10 (1.01-1.21)	1.10 (1.00-1.21)
Heavy smokers: ≥20 CPD (n = 424 470)	Quitter	49 185 (11.6)	488	309 183	1.6	1.14 (1.03-1.26)	0.87 (0.79-0.96)	0.93 (0.84-1.03)	0.92 (0.83-1.02)
	Reducer I	45 011 (10.6)	761	279 395	2.7	1.97 (1.81-2.14)	1.16 (1.07-1.26)	1.17 (1.07-1.27)	1.14 (1.05-1.25)
	Reducer II	67 554 (15.9)	449	423 302	1.1	0.77 (0.70-0.86)	0.96 (0.87-1.07)	0.98 (0.89-1.09)	0.95 (0.85-1.06)
	Sustainer	235 936 (55.6)	2031	1 475 778	1.4	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
	Increaser	26 784 (6.3)	180	167 517	1.1	0.78 (0.67-0.91)	1.07 (0.92-1.24)	1.05 (0.90-1.22)	1.04 (0.89-1.22)

STROBE: Results

Other analyses: Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses

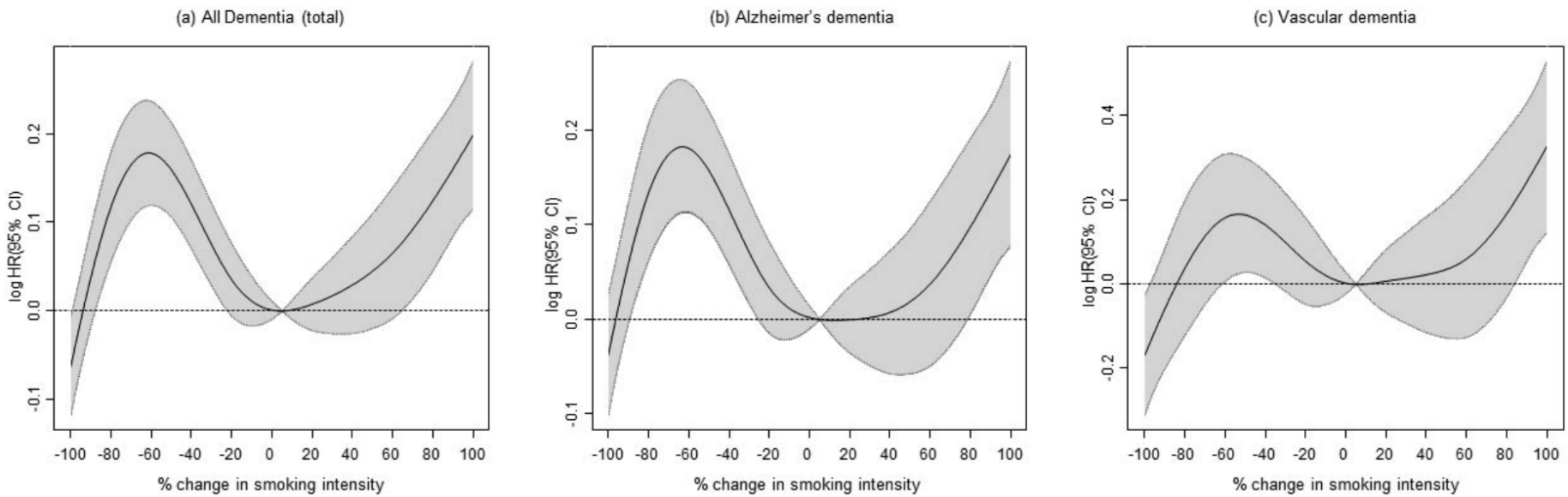
Example using NHISS

eTable 9. Association Between Relative Changes in Cigarette Smoking Intensity and Risk of Dementia Stratified by Alcohol Drinking Status

Subgroup	Smoking status				Sustained smoking at same level as a reference				
	2009	2011	No. (%)		Case No.	Duration	IR	aHR (95% CI)	p for interaction
All Dementia (Total)									
Alcohol drinker	All Smokers (N=203,562)	Quitter	32,916 (16.2)		881	204,776	4.3	0.84 (0.77-0.90)	<0.001
		Reducer I	17,800 (8.7)		753	109,090	6.9	1.27 (1.17-1.37)	
		Reducer II	28,038 (13.8)		636	174,356	3.6	1.10 (1.02-1.18)	
		Sustainer	93,169 (45.8)		2,000	579,717	3.4	1 (Ref.)	
		Increaser	31,639 (15.5)		897	196,038	4.6	1.10 (1.03-1.18)	
Non-drinker	All Smokers (N=585,970)	Quitter	82,043 (14.0)		849	517,631	1.6	1.02 (0.95-1.11)	
		Reducer I	42,967 (7.3)		832	267,864	3.1	1.24 (1.14-1.35)	
		Reducer II	83,852 (14.3)		934	525,581	1.8	1.01 (0.92-1.10)	
		Sustainer	283,224 (48.3)		2,958	1,775,565	1.7	1 (Ref.)	
		Increaser	93,884 (16.0)		1,172	587,187	2.0	1.13 (1.05-1.22)	

Example using NHISS

eFigure 2. Restricted Cubic Spline Curve for Association Between Changes in Smoking Intensity With Risk of Dementia



Thank You!