Demographic Factors and Smoking Status: Impact on Health Status

John Mariano
July 17th, 2024

Study Objective

This study aims to investigate the relationships between various demographic factors (age, sex, race, ethnicity, and children in the household) and smoking status with health status. We predict that there is a statistically significant relationship between smoking status and health status.

Methods

NHIS data

The NHIS data was collected in 2022 targeting noninstitutionalized civilians of the United States in all 50 states as well as Washington DC. The data is collected in geographic clusters to allow for cost effective sampling. The data was collected from January to December. This study looked at a subset of the NHIS data that only included individuals who have smoked at least 100 cigarettes in their life (National Health Interview Survey Description, 2022).

Eligibility

The population included consist of residential households as well as noninstitutional group quarters. This excludes those living in places like military bases, nursing homes, and prisons or jails. The homeless population with no address were not included, and those in military service or living on a military base, even civilians, were also not included in the target population. The data is collected in geographic clusters to allow for cost effective sampling (National Health Interview Survey Description, 2022).

Variables Used

Age is a continuous variable between 18 and 85 disclosing the age of a participant in the dataset. All individuals above 85 are denoted as 85+. Sex describes a participant in the data as male or female. Two race or ethnicity variables were used. One denoted participants as Hispanic or not Hispanic, and the other contained the categories, White only, Black only, Asian only, AIAN only, AIAN and any other group, or Other single and multiple races. Children in the household describes the number of people under the age of 18 that live in the quarters with the participant. This was denoted as 0, 1, 2, or 3+. The smoke now variable denotes participants of the data as smoking every day, some days, or not at all. The health status variable denotes participants as being in excellent, very good, good, fair, or poor health (*2022 National Health Interview Survey (NHIS) Codebook for Sample Adult File)*.

Statistical Analysis

For each of the variables in this paper, univariate analysis was conducted. For each categorical variable (all except age), counts and percentages for each category were calculated. For the age variable, which was the only continuous variable, the mean, median, minimum, and maximum were calculated. A test of normality was conducted for the age variable as well.

Two different tests were run for the bivariate analysis of each variable pairing. The bivariate analysis of all the categorical variables was done with a chi-square test to test the

relationship of the variables. A Kruscal Wallis test was conducted to test the relationship of the continuous age variable with the Health Status and Smoke Now variables. This test was chosen due to the resulting test of normality done on the age variable, which showed that the distribution was not normal.

Two regression analysis tests were conducted, both using a multinomial logistic regression. The first was a univariable model that used just the Health Status and Smoke Now variables, and the second test included all the covariates (sex, race, ethnicity, age, and children in household). A multinomial logistic regression was used because the data is categorical and contains more than two categories.

Results

Table 1: Variable Statistics

	n	Mean (std)	Median (min, max)
Age	9876	57.20 (16.68)	60.00(18.00,85.00)
	n	**%**	
Sex	9888	100.00	
Male	5243	53.02	
Female	4645	46.98	
Hispanic, Latino, or of Spanish ethnicity	9888	100.00	
Yes	918	9.28	
No	8970	90.72	
Race/ethnicity	9592	100.00	
1: White only	8042	83.84	
2: Black only	972	10.13	
3: Asian only	266	2.77	
4: AIAN only	106	1.11	
5: AIAN and any other group	105	1.09	
6: Other single and multiple races	101	1.05	
Smoke Now	9888	100.00	
1 Every day	2430	24.58	
2 Some days	714	7.22	
3 Not at all	6744	68.20	
Health Status	9888	100.00	

1 Excellent	1396	14.12
2 Very Good	3096	31.31
3 Good	3315	33.53
4 Fair	1488	15.05
5 Poor	593	6.00
Children in Household	9888	100.00
0 children	7802	78.90
1 child	895	9.05
2 children	757	7.66
3+ children	434	4.39
Std: standard deviation, min: minimum, max: maximum		

Table 1: Univariate analysis

Looking at table 1, the study sample consisted of 9,888 participants. The mean age was 57.20 years (SD = 16.68), with a median age of 60 years, ranging from 18 to 85 years. Although not shown in Table 1, the age distribution was found to be not normally distributed. Regarding sex, 53.02% of the participants were male (n = 5,243), and 46.98% were female (n = 4,645). Participants identifying as Hispanic, Latino, or of Spanish ethnicity comprised 9.28% (n = 918) of the sample, while 90.72% (n = 8,970) did not identify as such.

The racial/ethnic composition of the sample was as follows: 83.84% White only (n = 8,042), 10.13% Black only (n = 972), 2.77% Asian only (n = 266), 1.11% American Indian or Alaska Native (AIAN) only (n = 106), 1.09% AIAN and any other group (n = 105), and 1.05% other single and multiple races (n = 101). For smoking status, 24.58% of participants (n = 2,430) reported smoking every day, 7.22% (n = 714) smoked some days, and 68.20% (n = 6,744) did not smoke at all.

Health status was reported as follows: 14.12% (n = 1,396) rated their health as excellent, 31.31% (n = 3,096) as very good, 33.53% (n = 3,315) as good, 15.05% (n = 1,488) as fair, and 6.00% (n = 593) as poor. Regarding the presence of children in the household, 78.90% of participants (n = 7,802) reported having no children, 9.05% (n = 895) had one child, 7.66% (n = 757) had two children, and 4.39% (n = 434) had three or more children.

	Dependent variable Health Status (PHSTAT_A)	**Age (integer)**	**Sex (dichotomous)**	**Race (6 categories)**	**Ethnicity (dichotomous)**	Children in Household (PCNTLT18TC)

Independent variable Smoke Now(SMKNOW_A)	Chi-square	Kruskal Wallis	Chi-square	Chi-square	Chi-square	Chi-square
Dependent variable Health Status (PHSTAT_A)		Kruskal Wallis	Chi-square	Chi-square	Chi-square	Chi-square

Table 2: Bivariate analysis tests conducted for each variable pair

Table 2 shows each of the tests run for each variable pairing. The categorical data relationships were determined using a chi-square test while the relationship to age of the health status and smoking status was determined with a Kruskal Wallis test.

	Dependent variable Health Status (PHSTAT_A)	**Age (integer)**	**Sex (dichotomous)**	**Race (6 categories)**	**Ethnicity (dichotomous)**	Children in Household (PCNTLT18TC)
Independent variable Smoke Now(SMKNOW_A)	<0.0001	<0.0001	0.3791	<0.0001	<0.0001	<0.0001

Dependent variable Health Status (PHSTAT_A)	[redacted]	<0.0001	0.2648	<0.0001	0.0024	<0.0001

Table 3: Bivariate analysis p values

In Table 3, the bivariate analysis revealed significant associations between health status and several independent variables. Age, race, ethnicity, and the presence of children in the household were all significantly associated with health status. Also, there was a statistically significant relationship between smoking status and health status. However, there was no statistically significant relationship between sex and health status.

Smoking status also showed significant relationships with age, ethnicity, race, and the presence of children in the household, as well as with health status. Similar to health status, smoking status did not have a statistically significant relationship with sex.

Parameter	PHSTAT_A	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	2 Very Good	1	1.0482	0.0789	176.5381	<.0001
Intercept	3 Good	1	1.4192	0.0757	351.9075	<.0001
Intercept	4 Fair	1	0.8104	0.0816	98.6525	<.0001
Intercept	5 Poor	1	-0.0376	0.0969	0.1502	0.6983
SMKNOW_A	2 Some days	2 Very Good	-0.2986	0.1392	4.6041	0.0319
SMKNOW_A	2 Some days	3 Good	-0.767	0.1389	30.4783	<.0001
SMKNOW_A	2 Some days	4 Fair	-0.7582	0.1551	23.8944	<.0001
SMKNOW_A	2 Some days	5 Poor	-1.2152	0.2226	29.7894	<.0001
SMKNOW_A	3 Not at all	2 Very Good	-0.3066	0.0872	12.3554	0.0004
SMKNOW_A	3 Not at all	3 Good	-0.6942	0.0843	67.7369	<.0001
SMKNOW_A	3 Not at all	4 Fair	-1.0008	0.0934	114.7582	<.0001
SMKNOW_A	3 Not at all	5 Poor	-1.0714	0.1148	87.1546	<.0001

Table 4: Multinomial linear regression

Effect	PHSTAT_A	Point Estimate	95% Wald Confidence Limits

SMKNOW_A 2 Some days vs 1 Every day	2 Very Good	0.742	0.565-0.974
SMKNOW_A 2 Some days vs 1 Every day	3 Good	0.464	0.354-0.610
SMKNOW_A 2 Some days vs 1 Every day	4 Fair	0.468	0.346-0.635
SMKNOW_A 2 Some days vs 1 Every day	5 Poor	0.297	0.192-0.459
SMKNOW_A 3 Not at all vs 1 Every day	2 Very Good	0.736	0.620-0.873
SMKNOW_A 3 Not at all vs 1 Every day	3 Good	0.499	0.423-0.589
SMKNOW_A 3 Not at all vs 1 Every day	4 Fair	0.368	0.306-0.441
SMKNOW_A 3 Not at all vs 1 Every day	5 Poor	0.343	0.274-0.429

Table 5: Multinomial linear regression odds ratios

The results from the multinomial linear regression analysis, as shown in Tables 4 and 5, indicate significant relationships between smoking status and health status. Table 4 shows the parameter estimates for the regression model. The estimates for smoking "Some days" versus "Every day" were significant across all health status categories, with p-values ranging from 0.0319 to <0.0001. Similarly, the estimates for smoking "Not at all" versus "Every day" were also significant across all health status categories, with p-values of 0.004 to <0.0001.

Table 5 shows the odds ratios for the health status categories based on smoking status. Participants who smoked "Some days" had lower odds of reporting "Very Good," "Good," "Fair," and "Poor" health compared to those who smoked "Every day," with odds ratios of 0.742 (95% CI: 0.565-0.974), 0.464 (95% CI: 0.354-0.610), 0.468 (95% CI: 0.346-0.635), and 0.297 (95% CI: 0.192-0.459), respectively. Similarly, participants who did not smoke at all had lower odds of reporting "Very Good," "Good," "Fair," and "Poor" health compared to those who smoked "Every day," with odds ratios of 0.736 (95% CI: 0.620-0.873), 0.499 (95% CI: 0.423-0.589), 0.368 (95% CI: 0.306-0.441), and 0.343 (95% CI: 0.274-0.429), respectively.

Odds Ratio Estimates				
Effect	PHSTAT_A	Point Estimate	95% Wald Confidence Limits	
SMKNOW_A 2 Some days vs 1 Every day	2 Very Good	0.803	0.605	1.065
SMKNOW_A 2 Some days vs 1 Every day	3 Good	0.491	0.370	0.652
SMKNOW_A 2 Some days vs 1 Every day	4 Fair	0.506	0.368	0.695
SMKNOW_A 2 Some days vs 1 Every day	5 Poor	0.316	0.199	0.502
SMKNOW_A 3 Not at all vs 1 Every day	2 Very Good	0.709	0.595	0.845
SMKNOW_A 3 Not at all vs 1 Every day	3 Good	0.444	0.374	0.526
SMKNOW_A 3 Not at all vs 1 Every day	4 Fair	0.304	0.251	0.369
SMKNOW_A 3 Not at all vs 1 Every day	5 Poor	0.252	0.198	0.321
SEX_A 1 Male vs 2 Female	2 Very Good	0.997	0.876	1.136
SEX_A 1 Male vs 2 Female	3 Good	1.040	0.913	1.185
SEX_A 1 Male vs 2 Female	4 Fair	0.937	0.804	1.091
SEX_A 1 Male vs 2 Female	5 Poor	1.107	0.906	1.354
RACEALLP_A 2 Black/African American only vs 1 White only	2 Very Good	0.914	0.711	1.175
RACEALLP_A 2 Black/African American only vs 1 White only	3 Good	1.549	1.222	1.965
RACEALLP_A 2 Black/African American only vs 1 White only	4 Fair	2.127	1.644	2.752
RACEALLP_A 2 Black/African American only vs 1 White only	5 Poor	1.749	1.256	2.436
RACEALLP_A 3 Asian only vs 1 White only	2 Very Good	0.637	0.454	0.893
RACEALLP_A 3 Asian only vs 1 White only	3 Good	0.743	0.529	1.043
RACEALLP_A 3 Asian only vs 1 White only	4 Fair	0.431	0.256	0.726
RACEALLP_A 3 Asian only vs 1 White only	5 Poor	0.584	0.293	1.165
RACEALLP_A 4 AIAN only vs 1 White only	2 Very Good	2.104	0.972	4.557
RACEALLP_A 4 AIAN only vs 1 White only	3 Good	1.647	0.744	3.642
RACEALLP_A 4 AIAN only vs 1 White only	4 Fair	3.090	1.356	7.039
RACEALLP_A 4 AIAN only vs 1 White only	5 Poor	3.815	1.463	9.950
RACEALLP_A 5 AIAN and any other group vs 1 White only	2 Very Good	0.471	0.254	0.874
RACEALLP_A 5 AIAN and any other group vs 1 White only	3 Good	0.705	0.397	1.252
RACEALLP_A 5 AIAN and any other group vs 1 White only	4 Fair	1.077	0.568	2.044
RACEALLP_A 5 AIAN and any other group vs 1 White only	5 Poor	1.885	0.908	3.916
RACEALLP_A 6 Other single and multiple races vs 1 White only	2 Very Good	1.204	0.654	2.215
RACEALLP_A 6 Other single and multiple races vs 1 White only	3 Good	1.215	0.650	2.271
RACEALLP_A 6 Other single and multiple races vs 1 White only	4 Fair	1.368	0.638	2.931
RACEALLP_A 6 Other single and multiple races vs 1 White only	5 Poor	1.563	0.552	4.423
HISP_A 2 No vs 1 Yes	2 Very Good	1.254	0.965	1.629
HISP_A 2 No vs 1 Yes	3 Good	0.850	0.659	1.095

HISP_A 2 No vs 1 Yes	4 Fair	0.776	0.574	1.048
HISP_A 2 No vs 1 Yes	5 Poor	0.974	0.626	1.515
PCNTLT18TC 1 child vs 0 children	2 Very Good	0.894	0.719	1.112
PCNTLT18TC 1 child vs 0 children	3 Good	0.845	0.676	1.057
PCNTLT18TC 1 child vs 0 children	4 Fair	0.680	0.510	0.907
PCNTLT18TC 1 child vs 0 children	5 Poor	0.571	0.369	0.884
PCNTLT18TC 2 children vs 0 children	2 Very Good	0.832	0.664	1.043
PCNTLT18TC 2 children vs 0 children	3 Good	0.682	0.538	0.865
PCNTLT18TC 2 children vs 0 children	4 Fair	0.628	0.460	0.858
PCNTLT18TC 2 children vs 0 children	5 Poor	0.292	0.157	0.542
PCNTLT18TC 3 3+ children vs 0 children	2 Very Good	0.901	0.663	1.225
PCNTLT18TC 3 3+ children vs 0 children	3 Good	0.971	0.713	1.324
PCNTLT18TC 3 3+ children vs 0 children	4 Fair	0.842	0.565	1.254
PCNTLT18TC 3 3+ children vs 0 children	5 Poor	0.315	0.134	0.743
AGEP_A	2 Very Good	1.008	1.004	1.013
AGEP_A	3 Good	1.020	1.015	1.024
AGEP_A	4 Fair	1.031	1.025	1.036
AGEP_A	5 Poor	1.042	1.034	1.050

Table 6: Odds ratio estimates of multinomial linear regression with all coviariates

In table 6, we see a large amount of data from the multinomial linear regression with all coviariates. Looking at the sex rows of Table 6, we see that sex odds ratios are all around 1 indicating a non likely significant effect on health status. The ethnicity variable had odds ratios that all included 1, indicating a non significant effect on health status. Each other variable, race, smoke status, and children in houshold, had at least one category with an odds ratio limit above or below 1 indicating a significant effect on health status.

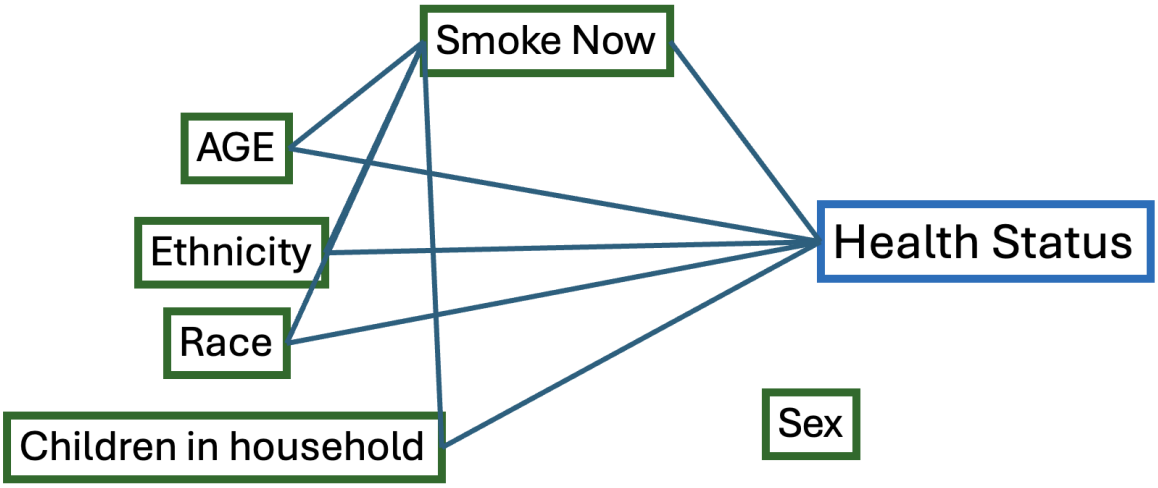

Figure 1: Concept model of variable relationships

In Figure 1, we see the statistically significant relationships shown through the bivariate analysis done. Health status and smoke now have a statistically significant relationship with age, ethnicity, race, and children in household, as well as with each other. Neither smoke now or health status have a statistically significant relationship with sex.

References

National Center for Health Statistics. "National Health Interview Survey Description, 2022." *U.S. Department of Health and Human Services*, Centers for Disease Control and Prevention, 2022, https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2022/srvydesc-508.pdf

National Center for Health Statistics. "*2022 National Health Interview Survey (NHIS) Codebook for Sample Adult File."* Document Version: 05 June 2023, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2023, https://learn-us-east-1-prod-fleet02-xythos.content.blackboardcdn.com/5f9c84db7d38a/23638184?X-Blackboard-S3-Bucket=learn-us-east-1-prod-fleet01-xythos&X-Blackboard-Expiration=1717923600000&X-Blackboard-Signature=Snr1PRh3mPcuq%2BtYFK6%2BQVfZn5o1fcQj5vH6FdvpTYM%3D&X-Blackboard-Client-Id=301799&X-Blackboard-S3-Region=us-east-1&response-cache-control=private%2C%20max-age%3D21600&response-content-disposition=inline%3B%20filename%2A%3DUTF-8%27%27NHIS%2520adult-codebook%25202022.pdf&response-content-type=application%2Fpdf&X-Amz-Security-Token=IQoJb3JpZ2luX2VjEIz%2F%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJGMEQCIDK0UqpFL4jET1wLJsckmoN%2Fz08GViukQQyDPHA63d5ZAiBmX5a9H9y4l%2F8utnJQwby9TY96sGXkQE2JJ1%2

BRM1Lq2iq1BQgkEAQaDDYzNTU2NzkyNDE4MyIMbXkwleJ5aCDN7p35KpIFGwfO9hQlc
Okovdnh1%2FsjXlBx%2F0%2FKqMcVoh8Kngii9C6sL6Ky9e1Q2Y4iIM343rdLxBfRLT%2FEi
%2FBcifKlpn2FWGk3yblV7T9sDt7Tn85%2FbV2z35zbRSbBitQqu53wfACXJlQvzvVM2D6Ie
pXKZDSnp0mjIe8htnHnHK9fcTYBeKSI%2BTY0Gv24%2FtLrFHJSRe9IlzNSQe1HkcmP0TV
0BNkkJyrkPB6Hsp8CUNbHlcWTJn4YlQto0jvEhpruXMlffymxjgNBlcrvB%2BTFBtTD39f4ET
MqpxWOrVq1%2Fb8qM4kZQGjTE0Pz8hxNZLQ0nECcDjBxVOoByqrIx67c9Ko5JH3SRDYD
QRtohFTXnoplFnROZmjS9jXJDZZT3oFQmUVkIlVKOEY27bIiGOEdkQ1RB4uN9HpOpgDd
chPwHdBXeYhzvMavfZrFBz6qMVP0K263Gxvev2euPyjw9HWyHGlG5MOPzv4B9pTBY0Y
mFHA8rHLgaO7MXDEh3NgtaCOv6HAyhRMETk5y4eZM2c2g1smD%2FM9hNoz9sQI8xdfJv
XmH6BgdqszL3NHsb1IdwZXL7ygNQdpt9tIK56jsTvL1D7P3%2F4jKEIK1wVQPmusV4vjIO
Q3hFsmobaF6t%2F7BEwJTTKaH%2BuSgru6qu2kFeHbQ82Zh4BdqWEMw%2FZqEo7OVXy
waSTIwIvGmIfvM5wNfQIIjlHXkzGkxqqoXP%2BFWteguttbI2YgRfe1k%2BaEseT6DdtHj0rR
SrePcjSx5YYiMMIgM%2FJfSqnIjbZiZlohoT%2BdpjPnTNEcpuYTpvXAX2k3dtBEg4XAmFE
kU%2BPt3r7rTubuJkelqRl5m7rt7XTlUykp%2FKiQGpShkCyozRfgk%2B7SqmRt06nupdVj%2
FNDC%2FwJSzBjqyAdPzRZBCiKXsfRcEjdEuc%2FRskab07cELavExPDdjwpZh9OmKJ%2B
gSxFMIhAWk%2B1tIJoEZWrYo%2FooCrMis90nrUtuJhAyrdM6b5mKeH9YAfq2u5tPMbs2mQ
deZPElYyKHXJrKkaFtedT0Lhl61XyD%2Feip4PP%2FeeByEMx438mWfpvi%2BGSti9VQGO
buWY0iNGUAeFG1M9MGbUWRTq3jVkTrOaXRqN3R5xg%2BOIkzGtaS6WvZ1Fto%3D&X
-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20240609T030000Z&X-Amz-Signed
Headers=host&X-Amz-Expires=21600&X-Amz-Credential=ASIAZH6WM4PL7HTIWBWU%
2F20240609%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Signature=b47ba09547a0618c635
1d304b38017a855b4b41d7217993cd2b000cf88289213