Research Brief

Factors Associated With Changes in Fruit Intake During Young Adulthood: A Classification and Regression Tree Analysis of Longitudinal Data

David Lohrmann, PhD; Wasantha Jayawardene, MD, PhD; Mohammad Torabi, PhD

ABSTRACT

Objective: To examine whether distinct participant groupings for changes in fruit intake (FI) levels between ages 23 and 31 years are identifiable based on both time-varying and time-invariant sociodemographic and behavioral variables.

Methods: Data were derived from the National Longitudinal Survey of Youth–1997, US. Change in FI frequency constituted the dependent variable. For 21 variables, changes and averages in 2007–2011 were calculated. Classification and regression tree analysis was conducted using Generalized, Unbiased, Interaction Detection, and Estimation software.

Results: Analysis isolated 5 variables (changes in smoking, drinking alcohol, and television viewing, plus 5-year mean of income-to-poverty ratio and computer use) and associated cutoff values to identify 7 groups of participants with differing degrees of FI change.

Conclusions and Implications: Multiple groupings existed within upper social strata; a majority maintained healthy behaviors whereas some adopted substance use stress-coping mechanisms. Some low-income individuals demonstrated a capacity to adopt healthy behaviors. Dietary interventions could identify behavioral clustering, with emphasis on drinking, smoking, and screen time.

Key Words: fruit intake, classification and regression tree, health-related behaviors, alcohol, smoking, screen time (*J Nutr Educ Behav*. 2017;■:1-8.)

Accepted January 31, 2017.

INTRODUCTION

Adequate fruit and vegetable intake (FVI) is negatively correlated with the risk of developing chronic diseases and conditions[1]; however, recent surveys and public health guidelines indicated that US adults fail to consume adequate fruits and vegetables.[2] Therefore, increases in mean daily FVI by 80% for fruits and 37.5% for vegetables are important public health objectives for 2020.[3] The 2010 Dietary Guidelines for Americans (the guidance at the time of the study) stipulated mean daily vegetable intake (VI) for adults at the 2,000-cal level of at least 2.5 cup-equivalents, allowing 0.7 cups of starchy vegetables, including french fries and potato chips.[4] For fruit intake (FI), including fruit juice, cup-equivalents for adults at the 2,000-cal level is ≥2.0.[4]

Fruit intake or VI and overall diet quality of both young and older adults are determined by many socioeconomic, environmental, and behavioral factors. A study revealed that compared with non-students, young adult students and college graduates had greater FVI.[5] Among adults, FVI was associated with higher income and education levels[6-8] as well as being female and, for both males and females, being married.[7-9] Less than high school education and being obese were associated with decreased FVI.[10] Food insecurity was associated with lower-quality diets[11]; and among young adults, FVI was positively affected by perceptions of neighborhood access.[12] For college students, dieting to lose weight, higher meal and snack frequency, not skipping breakfast, not eating red meat once per day, avoiding fat, trying to eat fiber, not smoking, and not binge drinking were associated with FI.[13]

Several non-dietary, health-related behaviors were associated with FVI. Comparisons of college students and graduates with comparable non-students found that non-students of both sexes were more likely to smoke and to have less healthy diets.[5] Among Minnesota adolescents, inadequate FVI was correlated with tobacco, alcohol, and marijuana use.[14] Another study of US adults found increased FVI over time among nonsmokers.[15] Weight gain among college students was associated with alcohol consumption rather than FVI, physical activity, or sleeping.[16] Although not directly related to FVI, television viewing was associated with young adults' energy-dense snack consumption.[17]

Although fruits and vegetables are often grouped together, fruits are

Applied Health Science, School of Public Health–Bloomington, Indiana University, Bloomington, IN

Conflict of Interest Disclosure: The authors' conflict of interest disclosures can be found online with this article on www.jneb.org.

Address for correspondence: Wasantha Jayawardene, MD, PhD, Applied Health Science, School of Public Health–Bloomington, SPH 116, Indiana University, 1025 E 7th St, Bloomington, IN 47405; Phone: (812) 272-9136; Fax: (812) 855-3936; E-mail: wajayawa@indiana.edu

http://dx.doi.org/10.1016/j.jneb.2017.01.017

generally sweeter tasting and often consumed as part of dessert or snacks, whereas vegetables are typically more savory and consumed within meals. Furthermore, economic barriers, intrinsic motivations for eating a healthy diet, psychosocial factors, and other health-related behaviors were more strongly associated with FI than VI.[18,19] Therefore, prior studies suggested that interventions to increase FVI should treat fruit and vegetables as separate groups, considering differences in stages of change and sources of motivation for FI vs VI.[19,20]

Consequently, only FI was analyzed for this study. The researchers employed classification and regression tree analysis of prospective data from a nationally representative sample[4] to test 2 hypotheses: (1) time-varying and time-invariant variables, ie, sociodemographics, body composition, and behaviors (eg, change in smoking), are associated with age-related FI-changes; and (2) through follow-up assessments over time, the study population could be segmented into multiple distinct groupings.

METHODS

Study Design

Multiple behaviors of young people aged 23–27 years in 2007 were assessed repeatedly until 2011 (aged 27–31 years). Data were derived from the US Department of Labor's National Longitudinal Survey of Youth-1997 (NLSY97), which used stratified, 3-state sampling[21,22] to follow a nationally representative sample of persons born during 1980–1984 (n = 8,984),[4] including those who were away at school or college, in a hospital, correctional facility, etc. In addition to the main sample, supplemental samples of black and Hispanic individuals were obtained to measure changes in these populations more accurately. Consistent clustering effects existed across domains, because design effects were well below 1.5.

Data Collection

In each round, interviews were conducted via computer-assisted personal interview administered by an interviewer using a laptop computer.[4] Phone interviews were conducted only for respondents who refused to be interviewed in person or for location-related reasons. In all years, response rates for most survey questions exceeded 80%. Data were downloaded from the NLSY97 database for 2002, plus each year of the 5-year period 2007–2011, and imported into SAS for Windows statistical software (version 9.4, SAS Institute Inc, Cary, NC, 2002–2012). Variables were obtained or computed separately for each year. Because this was a secondary analysis of public-use data, the Indiana University Bloomington Institutional Review Board confirmed that approval was not required for this study.

Measures

Behavioral variables. Health behavior data were initially collected in 2002 and then annually from 2007 to 2011.[4] Fruit intake constituted the dependent variable analyzed across the survey rounds; NLSY97 respondents were asked, In a typical week, how many times do you eat fruit? Do not count fruit juice. Responses were measured via a 7-level ordinal scale from 0 times/wk to ≥4 times/d.[4] Response categories demonstrated a clear intrinsic ordering as an ordinal variable with 7 levels convertible to a continuous variable ranging from 0 to 28 times/wk: I do not typically eat fruit = 0; 1–3 times/wk = 2; 4–6 times/wk = 5; 1 time/d = 7; 2 times/d = 14; 3 times/d = 21; and ≥4 times/d = 28.

Cigarette smoking and alcohol drinking were measured separately as the number of days of smoking or drinking in the past 30 days (range, 0–30). Sleeping was recorded as the number of hours of sleep per night: 1–9 or ≥10 (entered as 10). The NLSY97 respondents were also asked, In a typical week, how many hours do you watch television? Responses were measured on a 6-level ordinal scale (from <2 to >40 h/wk).[4] These response categories were converted into a 6-level continuous variable ranging from 0 to 45 h/wk: <2 h/wk = 1; 3–10 h/wk = 5; 11–20 h/wk = 15; 21–30 h/wk = 25; 31–40 h/wk = 35; and >40 h/wk = 45. Likewise, computer use response categories were converted into a continuous variable ranging from 0 to 12 h/wk: no use = 0; <1 h/wk = 0.5; 1–3 h/wk = 2; 4–6 h/wk = 5; 7–9 h/wk = 8; and ≥10 h/wk = 12.

Sociodemographic variables and body composition. Annually since 1997, respondents self-reported their age, sex, race/ethnicity (ie, black, Hispanic, non-Hispanic mixed race, or non-black non-Hispanic), marital status, and place of residence, along with household size and income data used to compute the ratio of family income to poverty threshold.[4] For this analysis, non-Hispanic mixed race (n = 17) was combined with blacks into 1 category. The US Census region (ie, Northeast, Midwest, South, and West) and area (ie, rural or urban) for the residence were determined by interviewers based on state and year 2000 census, respectively. Results from 2007–2011 constituted the primary focus for analysis; data from 2002 served as initial reference points. Body mass index (BMI) categories were based on Centers for Disease Control and Prevention criteria.[23]

Statistical analysis. The researchers conducted classification and regression tree analysis using Generalized, Unbiased, Interaction Detection and Estimation (GUIDE) software (version 23.2, Wei-Yin Loh, University of Wisconsin-Madison, 1997–2016). This method of analysis identifies the best combination of various predictor variables and generates specific cut points for each variable, thus providing subgroups most relevant to the outcome of interest. The regression tree method is preferable because the alternative, stepwise multiple regression, requires complicated model selection procedures with multiple higher-order interactions and lacks the convenience of identifying cut points for subgroup analysis.

Observations of height, weight, and BMI in which values were >5 SDs from the mean were removed as outliers (n = 176). For 71% of participants (n = 6,379), data were available for all analyzed variables in 2007–2011 and were included in classification and regression tree analysis. Sampling weights were not used because analysis was longitudinal and confined to subsamples of participants who provided valid responses to certain questions.[4] To analyze changes over time (eg, income, BMI, smoking, FI), individual regression slopes were calculated for each variable and for each participant