1. INTRODUCTION

1.1. Background

1.1.1. Cancer Burden

As one of the leading causes of death worldwide, cancer is a public health problem well known for its impact on society. Characterized by mutations of DNA (the genetic blueprint of health and homeostasis), an estimated 1,658,370 new cases of cancer were diagnosed in the USA in 2015; and over half a million died from the disease that same year[1]. Furthermore, in addition to the human toll of cancer, the financial implications, both direct (treatment and rehabilitation) and indirect (morbidity and mortality), are profound. The direct costs of cancer in the USA in 2011 was estimated to be nearly $89 billion[2].

1.1.2. Carcinogens

The genomic alterations that allow cancer to manifest are identified to be caused by both inherited tendencies and external/environmental elements. Non-hereditary influences include lifestyle and behavioral factors, such as smoking, along with naturally occurring exposures such as, sunlight and other household or workplace exposures. Though dependent on dose, potency, and length of exposure- any substance that can cause damage to DNA, directly or indirectly, are said to be carcinogenic[1].

1.1.3. Occupations

It is no surprise that some occupations inherently have varying levels of exposures to potential carcinogens[3-4]. Most staggeringly may be miners; exposed to elements such as uranium, dust, chemicals, and dangerous conditions, mining is one of the most dangerous professions. Firefighters and medical professionals also consent to potentially higher levels of occupational exposures, from which we take measures to safeguard against. But one of the most common occupations, serving in our nation’s military, has largely lacked attention, and research (on a broad scale), to investigate what occupational exposures may excessively render them more susceptible to certain diseases, including cancer.

1.1.4. Veterans, Carcinogens & Cancer

1.1.4.1. *Coast Guard*

In a study conducted assessing mortality related to specific assignments of marine inspectors within the branch of the U.S. Coast Guard, this population was compared to non-inspector officers of the same branch, as well as to national mortality standards for civilians. Though when inspectors were compared to non-inspectors, no statistically significant figures were found; inspectors and non-inspector officers were shown to have decreased all-cause life expectancy, as well as discrepant deficits for all malignant neoplasms[5].

1.1.4.2. *Agent Orange*

Between 1962 – 1971 US military forces sprayed herbicides over the thick jungle landscape of Vietnam in hopes to: eradicate the leafy green canopy that served to conceal opposition forces, destroy crops that enemy forces might depend on, and clear tall grasses and bushes from the perimeters of US bases. Since then, numerous studies have been conducted on the herbicides used: Agent Orange and several other tetrachlorodibenzo-p-dioxins (TCDDs). Once data had been stratified to statistically account for time served in Southeast Asia, it was discovered that the unit responsible for spraying, known as Operation Ranch Hand, had increased risks for prostate cancer and melanoma[6-7].

1.1.4.3. *Testicular Cancer*

Although testicular cancer only comprises 1% of all male cancers, it is the most commonly diagnosed malignancy in young men. In a French study looking at rising incidence of testicular diagnosis within the military, they found the mean patient age to be 30.8 years when compared to the general public, 37.4 years[8]. Similar findings were reported for U.S. military service men, with mean age at diagnosis approximately 29.8 (data from 1990-2003)[9]. Though age is certainly a confounding demographic, in which active military personnel have younger male populations than the general public, it is still believed that environmental exposures, particularly for technicians exposed to aviation fuels, is the cause of a rising incidence of stage 1 testicular germ cell tumors (TGCT)[8]. This was certainly discovered to be the case for another study looking at military deployment to the Persian Gulf during the 1991 conflicts. Using national cancer registry data and information from 621,902 Gulf War veterans, testicular cancer was found to be the malignancy that was most significantly increased when compared to non-Gulf War veterans (n=726,248). Though research into the prominence of cancer among veterans has gotten more attention since the Vietnam War and the infamous Agent Orange, this study clearly linked not only a risk between veteran status, but an increased risk linked to specific deployments[10].

1.1.4.4. *Camp Lejeune*

One particular disaster to have taken place in US history, on US soil, that seems to avoid the limelight, was the continual pollution of water systems supplying a Marine Corp base in North Carolina; a base covering nearly 200 square miles of land and currently home to 54,000 people. The discovery of nearly a dozen volatile compounds (VOCs) in late 1980 was far too late to address the contamination that had been occurring since the early 1950s. It was estimated to have exposed an average of 10,000 individuals annually in areas serviced by the water systems discovered to be highly carcinogenic[11-14].

There certainly has been a flurry of efforts since the true unveiling of this public health problem, but they regrettably fall just short of being enough. For example, the Agency for Toxic Substances and Disease Registry (ATSDR) succeeded in publishing a critical Health Assessment in 1997, which found a statistically significant association between the exposure to toxic water at Camp Lejeune and adverse pregnancy outcomes (such as small gestational age), and are still conducting valuable research on in utero assaults, however, the larger population of adults that have been exposed have had no additional support, through the form of research, to compare and evaluate their potential increased risk for poor health outcomes or malignancies[11-14].

1.2. Research Question

Stirred by the hundreds of anecdotal stories and legal cases against the government for premature and atypical cancer clusters in veterans stationed at the specific bases known to have been supplying hundreds of thousands of service individuals with water polluted with carcinogenic toxins, along with the plethora of published research on higher incidences of specific cancers being linked with particular military assignments (either by time frame, locale, or chemical exposure), my aim with this paper is to capture a glimpse of what insight the data from the Behavioral Risk Factor Surveillance System (BRFSS) survey of 2013 could afford on the issue of veterans and their association with all cancers (skin cancers and solid organ cancers included). In looking at veteran status as the exposure and cancer as the outcome, it was my objective to contribute to current literature through answering the following umbrella question: Is there an association between a history of serving in any branch of the military and any cancer diagnosis?

1.3 Hypothesis

With previous significant findings from research conducted on similar topics, most commonly specific cancer diagnosis with unambiguous branches of military, deployments, service time, locations, or duties, I hypothesize that overall, veterans would have a higher prevalence of cancer, than civilians in the US population.

2. METHODS

2.1. Data Source

This study used data from the 2013 BRFSS; a Centers for Disease Control (CDC) sponsored population-based health telephone survey, initiated in 1984, that collects cross-sectional information on the health behaviors and protective characteristics of the non-institutionalized US adult population, aged 18 years or older. Conducted year-round in all 50 states, the District of Columba, and three territories, survey respondents are identified through random-digit-dialing, which includes listed and non-listed numbers. In addition to landlines, cellular telephones have also been surveyed since 2011 and comprise approximately 20% of respondents. In 2013 the BRFSS obtained data on 491,773 individuals. And by means of complex design involving the aggregation of data by state, post-stratification weighing, ranking methodology, clustering and multistage sampling the BRFSS yields nationally representative estimates[15-17].

2.2. <u>Measures</u>

It should also be noted that unless otherwise mentioned, for all measures, responses of *don't know/not sure* and *refused* were recoded as *missing*.

2.2.1. <u>Exposure</u>

The conceptual definition of the exposure (independent/predictor variable) is having been a member of any branch of the US armed forces. The operational definition for this variable used the BRFSS demographic question, "Have you ever served on active duty in the US Armed forces, either in the regular military or in a National Guard or military reserve unit?" It should be noted that active duty does not include training for the Reserves or National Guard, but DOES include activation, for example, for the Persian Gulf War. This question was asked in every state, as it is a part of a core section. Veterans are "exposed" and non-veterans/civilians are "unexposed." My exposure is categorically binary/dichotomous. Subjects can either be veterans or not be veterans[18].

2.2.2. <u>Outcome</u>

Although aforementioned examples focus on solid organ cancers, since there is a documented link of sun exposure associated with the occupational exposure of being a in the armed forces, as seen in a study conducted by the Department of Defense (DoD) in which active duty military personnel were seen to have higher rates of melanoma between 2000 and 2007[19]; I felt it best to combine the skin cancer data with "all/any other cancer" for the purposes of this research. Thus, the conceptual definition of the outcome (dependent/response variable) is having had any cancer (including any skin and/or non-skin /solid organ cancer diagnosis). The operational definition is devised from the questions that originally asked, "Have you ever been told you have skin cancer?" and the follow up question, "Have you ever been told you have any other type of cancer?" My outcome variable is also categorical, and binary/dichotomous. Subjects can either have (or had) cancer or not have (or had) cancer. This variable was also asked of all respondents being that is also a core module (Chronic Health Conditions, Section 7, Questions 6 & 7)[18].

2.3. Analytic Sample

The US population of non-institutionalized adults, >18 years of age is the target population for the BRFSS. The 491,773 respondents in 2013, compose the sample population[16]. Of these observations a total of 741 were removed for invalid responses for the exposure of veteran status (166 answered don't know/not sure, 330 refused to answer and 245 had missing data for this variable). From the remaining 491,032 observations- an additional 2,176 were removed for invalid data for the outcome variables to the two questions pertaining to skin and any/all other cancer types. There was no missing data for the question pertaining skin cancer, but 1,121 answered don't know/not sure and another 183 refused to answer. As for the data for any/all other cancers, again there was no missing data, but 992 answered don't know/not sure and another 230 refused to answer. The resulting 488,856 essentially comprise our final analytic sample, with < 1% of the eligible population excluded, see Chart 1.

Chart 1: Initial Population & Final Analytic Sample

Target Population	Non-Institutionalized Adults (18+)
Sample Population	491,773 (BRFSS 2013 Respondents)
Observations Removed for Invalid Data* for Exposure and/or Outcome variables	2,917
Analytic Sample	488,856

**Invalid consisted of recorded answers of Don't Know/Not Sure, Refused or Missing for questions pertaining to veteran status (exposure) and both skin and non-skin cancer questions (outcome).*

2.4. Covariates / Confounders

The BRFSS provides information on a wide range of demographic and background characteristics. Given their relevance and presence in previous literature pertaining to the exposure and/or outcome, a total of 11 potential confounders were initially considered for this research. However, through bivariate analysis, comparisons of crude and adjusted odds ratios (ORs), as well as calculations of percent change between the two, plus model building and estimations through forward and backward stepwise estimation and deletion– six were eliminated from the final regression model: health insurance coverage, heavy drinking habits[20], BMI>25 Overweight/Obese[21-23], fruit intake of at least 1 serving daily, vegetable intake of at least one serving daily, and any exercise in past 30 days; leaving: sex, age, ethnicity, doctor, and smoking.

2.4.1. Sex

The variable for sex had no missing information and was a binary variable in which respondents were identified as male or female. Important to note here is that sex certainly plays a role larger that delineating gender as a demographic covariate. In addition to recognizing that cancer prevalence varies by gender, specific research on Male Breast Cancer in the veteran population, has shown discrepant care and deficits in overall survival rates when compared to their female counterparts for identical diagnosis and stage of cancer diagnosis[11].

2.4.2. Age

Investigating previously published research on the veteran population[23], including the VetPop Initiative[36, 24], allowed for the assessment if particular stratifications were more appropriate for this sample. Due to the nature of the outcome, and considering that 77% of cancer cases are diagnosed after 55 years of age[1], I felt it most important to make sure we had disaggregated data before and after that critical marker and thus, did not precisely mirror the prominent data reporting trends, that often reported data for the age group 44-59[25]. In the end I used four bins for age: 18-34, 35-54, 55-64, and 65+. There was no missing data.

2.4.3. Ethnic Background

In commonly published research on the veteran population, as well as research conducted by the CDC, the racial background groupings are near identical to that provided by the BRFSS[24]. However, three of the ethnic categories had less than 1% proportion of the sample, and thus I combined *American Indian or Alaskan Native* (1%), *Asian* (4.5%), *Native Hawaii or Pacific Islander* (<1%), *Multiracial* (1.3%), and *Other* (<1%). In the end, the three largest ethnic groups were represented: *White* (n=375,265, 63%), *Black* (n=39,000, 11%), and *Hispanic* (n=36,909, 16%), plus *Other* (containing the aforementioned groups). This four group arrangement consistently matches nicely with other research conducted on the veteran population[25]. According to U.S. CDC cancer statistics, the risk of developing cancers of diverse types is not comparable across diverse ethnic groups[24], making this demographic covariate critical to be included in our statistical analysis. In fact, as of 2012 the American Cancer Society has presented statistics revealing Hispanic populations experience cancer as the leading cause of death, when compared to non-Hispanic populations, for which heart disease is the leading cause of death[26].

2.4.4. Access to Health Care

Conceptually I wanted to include an element of "access" to health/medical care in this research to understand if there was a relationship with cancer diagnosis and veteran status. In early stages of bivariate analysis and developing the final regression model both insurance and primary care provider/doctor (PCP) data were considered to serve as a proxy for "access;" using the following two questions from the Health Care Access section of the BRFSS survey: (1) "Do you have any kind of health care coverage, including health insurance, prepaid plans such as HMOs, government plans such as Medicare, or Indian Health Service?; and (2) "Do you have one person you think of as your personal doctor or health care provider?"[16, 18]. With the established Veteran Affairs (VA) system and associated medical care and system afforded to veterans, you would expect that the former question pertaining to insurance coverage would result in 100% of respondents indicating they have coverage, however in bivariate analysis we found that greater than 10% of veteran respondents did not indicate they had access to health insurance coverage (n=55,242 – 11.28%). Nevertheless, in the end, insurance was excluded while the variable elaborating primary care doctor access remained in the final regression model.

2.4.5. Smoking Status

Since smokers use approximately 25% of health care spending nationally[27] and since smoking status is considered to contribute to disproportions of prevalence of cancer, assessing such for our analytic sample seemed indispensable. Although smoking prevalence overall has been on a slow, though statistically significant, decline- decreasing from 21% to 19% from 2005 to 2011[27-29], military veterans have been shown to be at high risk for nicotine dependence. Smoking rates have been found to increase with deployment, and so much as a 9% escalation has been cited[30-32]. In one particular study of personnel serving in the first Gulf War, 7% reported starting smoking for the first time during deployment. In another study focused on American military personnel on active duty in Iraq and Afghanistan, smoking rates of >50% were revealed[30-32]. Moreover, when compared to the British, of which 29% of their military population of preexisting smokers increased cigarette consumption on deployment, the US military has witnessed an 56% increase[31, 33].

Previous studies indicate that not only are veterans more likely to be current smokers, but the VA system found in a study, of three 500 veteran cohorts, that 43% of current smokers had an interest in clinical programs to help with smoking cessation, (77% of whom participated)[34].

Tobacco use was assessed primarily through one question: (1) "Have you smoked at least 100 cigarettes in your entire life?"[16, 18] Those that answered no to the first question were considered non-smokers, where as those you answered yes to the first question were then promoted to elaborate as to whether they are a former or current smoker (either every day or some days). For the purpose of my research current and former smokers were combined and a binary variable was generated separating never/non-smokers from former or current smokers.

2.5. Statistical Analysis

2.5.1. Software

The statistical software, STATA (version 14.0)[35], was used for all statistical analyses. In order to account for the complex survey design and report weighted and nationally representative data, survey commands were used in the statistical software.

2.5.2. Bivariate Analysis

Bivariate analysis was performed to compare demographic characteristics and potential confounders among the BRFSS sample with valid data for exposure and outcome to test for covariance. Using Pearson's chi-squared tests of independence, veteran status was compared against cancer and all covariates mentioned. Weighted data was used to generate data as presented in Table 1. P-values <0.05

were considered statistically significant, though it should be noted that all reported findings had p-values <0.001.

Table 1: Cancer Outcome and Covariates of Interest for Veterans & Civilians: BRFSS 2013
Though 11 covariates were initially assessed only those included in the final regression model have been presented below.

Veteran Status			No n=429,527 (87.47%)	Yes n=61,505 (12.53%)
Cancer	No	406,581 (89%)	361,440 (90%)	45,141 (80%)
	Yes	82,275 (11%)	66,309 (10%)	15,966 (20%)
Age	18-34	77,033 (30%)	73,005 (32%)	4,028 (14%)
	35-54	143,375 (35%)	132,066 (36%)	11,306 (27%)
	55-64	108,705 (16%)	96,721 (16%)	11,984 (18%)
	65+	159,746 (19%)	125,957 (16%)	33,789 (41%)
Sex	Male	199,865 (49%)	144,249 (44%)	55,616 (91%)
	Female	288,991 (51%)	283,500 (56%)	5,491 (8.9%)
Ethnic Background	White	374,367 (64%)	324,441 (63%)	49,926 (75%)
	Black	38,966 (12%)	34,711 (12%)	4,255 (12%)
	Hispanic	36,868 (17%)	34,598 (18%)	2,270 (7.1%)
	Other	30,326 (7.6%)	26,939 (7.9%)	3,387 (5.2%)
At least 1 Primary Care Doctor	No	79,074 (24%)	70,506 (24%)	8,568 (18%)
	Yes	408,025 (76%)	355,727 (76%)	52,298 (82%)
Smoker Status	Never	260,511 (57%)	238,844 (59%)	22,067 (39%)
	Yes (Is &/or Was)	213,477 (43%)	176,170 (41%)	37,307 (61%)

**All P-Values were statistically significant and <0.001*

2.5.3. Model Building & Excluded Covariates

All variables eliminated in either forward or backward stepwise estimations were removed, along with all covariates with non-significant regression coefficients and those with less than a ten percent change[i] between crude and adjusted odds ratio, (with the exception of sex which was forced into the final model despite a 3% change in odds ratio), were eliminated.

2.5.4. Logistic Regression

I ran separate multiple logistic regression models to assess the independent association between cancer and each demographic and potentially confounding covariate. Then multivariate logistic regression analyses were conducted to assess whether veteran status predicted the odds of cancer in the presence of three demographic confounders and two other covariates. Crude ORs, adjusted ORs and 95% confidence intervals (CIs) were calculated from these logistic regressions and are presented in Table 2. P-values <0.05 were considered statistically significant, though it should be noted that all reported findings had p-values <0.001.

3. RESULTS

Bivariate analysis revealed that 12.5% of respondents were veterans (n=61,505), and 11% had cancer (n=82,275). Among veterans 20% had cancer compared to 10% of non-veterans, (P<0.001). Veterans were more likely to be older when compared to non-veterans. 41% of veterans were older than 65 years, compared to 16% of non-veterans (P<0.001). Veterans were also more likely to be male. 91% of veterans were male, compared to 44% of non-veterans, (P<0.001). There were proportionally more Hispanics among non-veterans (18%) than among veterans (7.1%). However, this was not true for Whites, which were proportionally more among veterans (75%) than among non-veterans (63%), (P<0.001). Veterans were less likely to not have a PCP when compared to non-veterans, respectively 18% compared to 24% (P<0.001). And there were more current and/or former smokers among veterans (61%), when compared to non-veterans 41%, (P<0.001).

[i] Investigators determine whether there is confounding by estimating the measure of association before and after adjusting for a potential confounding variable. A change in the estimated measure of association of 10% or more would be evidence that confounding was present, but if the measure of association changes by <10%, there is likely to be little, if any, confounding by that variable.

Table 2: Unadjusted and Adjusted Odds Ratio of Cancer Among Veterans: BRFSS 2013
Though 11 covariates were initially assessed only those included in the final regression model have been presented below.

Characteristics		**Crude Odds Ratio (95% CI)***	**Adjusted Odds Ratio (95% CI)****
Veteran	No	1.00 (ref)	1.00 (ref)
	Yes	2.29 (2.20, 2.38)	1.45 (1.38, 1.52)
Age	18-34	0.04 (0.038, 0.045)	0.061 (0.056, 0.066)
	35-54	0.16 (0.15, 0.17)	0.205 (0.196, 0.215)
	55-64	0.41 (0.40, 0.43)	0.47 (0.45, 0.49)
	65+	1.00 (ref)	1.00 (ref)
Sex	Male	1.00 (ref)	1.00 (ref)
	Female	1.31 (1.27, 1.35)	1.38 (1.33, 1.44)
Ethnic Background	White	1.00 (ref)	1.00 (ref)
	Black	0.34 (0.31, 0.36)	0.40 (0.37, 0.43)
	Hispanic	0.22 (0.20, 0.24)	0.38 (0.35, 0.42)
	Other	0.29 (0.26, 0.33)	0.44 (0.39, 0.49)
At Least 1 Primary Care Doctor	No	1.00 (ref)	1.00 (ref)
	Yes	4.14 (3.89, 4.40)	1.72 (1.61, 1.84)
Smoker Status	No, Never	1.00 (ref)	1.00 (ref)
	Yes, Former or Current	1.63 (1.58, 1.68)	1.25 (1.21, 1.30)

**All Crude OR have P-values <0.001*
***All Adjusted OR have P-values <0.001*

Unadjusted, non-institutionalized adults in the USA who are veterans have 2.29 the odds of any/all cancer compared to those who are non-veterans (P<0.001). Also, among the same sample population, those who are aged 18-34 years, 35-54 years, and 55-64 years have, respectively, 0.04, 0.16, and 0.41 the odds of cancer than those who are 65 years and older (P<0.001). Further, women of the same sample population, have 1.31 the odds of cancer when compared to males (P<0.001). Additionally, blacks, Hispanics, and individuals placed in the other category respectively have 0.34, 0.22 and 0.29 the odds when compared to whites (P<0.001). Those with access to at least one PCP have 4.14 times the odds of a cancer diagnosis when compared to those who do have at least one PCP (P<0.001). And former and/or current smokers have 1.63 times the odds of a history of or current diagnosis when compared to non-smokers (P<0.001). However, when adjusted the relationship between exposure/predictive variable veteran status and outcome/response variable, cancer, changes by nearly 37% when accounting for the five selected covariates: age, sex, ethnicity, doctor, and smoking. Among non-institutionalized adults in the USA those who are veterans have 1.45 the odds of cancer compared to those who are non-veterans, independent of age, sex, ethnicity, PCP and smoking status (P<0.001).

4. DISCUSSION

A more thorough understanding of the occupational risks to which we expose our military is more than just an ethical concern. The costs to society and burden on quality of life is impacted, and is particularly exorbitant considering the premature and uncharacteristic manifestations of malignancies in the veteran population[5, 8-14]. This study demonstrates that, consistent with our hypothesis, veterans are associated with a higher incidence of cancer and to my knowledge, this is the first study to look at the broad association of this occupational exposure as a risk factor for any/all cancer. Though cancer-specific or branch-specific research is valuable, so too is the comprehensive understanding of the association of these two variables. I believe these findings are novel and substantively add to the literature on veterans and cancer for a myriad of reasons, including: the political and medical leverage it affords for cancer clusters[ii] within the veteran population that have gained little saliency[11-14], the generalizability of the data gleamed, the impact on policy and public programming such as providing backing to cessation interventions within the veteran community[37] and the implications of the

[ii] Since cancer is a common disease (approximately one in two men and one in three women, over their lifetime, will develop or die from cancer) it can be difficult to discover when true cancer clusters arise from workplace, or occupational, exposure. What constitutes a true cancer cluster, which would then demand further investigation, is the homogeneity of the cancer type within a given workplace, along with whether they are primary or metastasized cancers[3-4].

effectiveness of the VA system, of which more than 10% of the veteran population is unaware or unsure of their available benefits.

4.1. Population Comparison

Of our analytic sample, approximately 12.52% of them indicated they are a veteran (n=61,322). The remaining 428,591 observations are non-veterans/civilians and comprise 87.48% of the sample. On Sept 30th of 2013 the US census indicated the resident population to be around 317,133,991[36], and the Veterans Affairs (VA), through an initiative called VetPop2014, indicated that the estimated population of living veterans residing in the US to be 22,299,350 on Sept 30th of 2013[37]. These statistics indicate that our target population allegedly has a near 14% veteran population, which is less than a 2% margin from what the BRFSS portrays in their data for 2013, which translates to fairly good representation for a potentially hard to reach population due to high rates of homelessness[23, 25].

4.2. Limitations

As with much research, limitations are a natural element in which components such as study design, sampling, measurement tools, and the measures themselves, e.g. self-reported verse clinically obtained biometrics, have the potential to be sources of confounding and could contribute to an attenuated magnitude of effect.

4.2.1. Cross-Sectional Data

The cross-sectional study design poses challenges to establishing causality considering the missing element of temporal sequence between military service and health outcomes later in life, during veteran-hood. Further, the trajectory of and nature of malignancy and cancer development and diagnosis represents a complex interplay of risk and protective factors operating at the individual-, familial-, and community-levels, and these can conspire in ways that may moderate the effect of exposures as a veteran.

4.2.2. Confounders

Although through the use of a conventional approach of multivariate analysis, confounders may originate from the study design, such as not collecting data on a potential confounder. And though the BRFSS does collect information on a great deal of health-related variables, I speculate that richness of data could be improved with specifying whether respondents were currently serving or resigned from active military duty. The phrasing of the current question/codebook allows for ambiguity. The conceptual definition of a veteran is prior servitude, not active. Thus, if an individual were state-side and were surveyed they could theoretically answer the question “Have you ever served on active duty in the US Armed forces, either in the regular military or in a National Guard or military reserve unit?” – Yes, but technically not yet be a veteran. Instead of a binary variable, value could be gained from delineating civilian, actively serving and veteran-status post resignation/discharge/retirement.

4.2.3. Language

Though the CDC provides a Spanish translation for the core questionnaire and optional modules, no further language support or translations are provide. Instead the BRFSS indicates that if any particular state has a significant population of non-English speakers, the state has the option to translate the questionnaire[17]. This leaves a great deal of potential for poor quality data collection, especially since no data is available on which states chose to do so and what their process was to conduct such translations for which languages, which begs me to question the consistency and administration of the survey in languages other than English or Spanish. Without language translations for major non-English speaking groups, or with poor quality translations, we are potentially either missing millions of respondents and thus creating a sampling bias and generating less generalizable data or simply collecting poor data. The US Census reported that in 2013 61.6 million spoke a language other than English at home and of that 41% (25.1 million) were of Limited English Proficiency (LEP). Of LEPs 64% were Spanish-speakers (16.2 million) but other major languages of LEPs included: Chinese with 6% (1.6 million), Vietnamese 3% (847,000), Korean 2% (599,000) and Tagalog also approximately 2% (509,000)[38].

4.3. Conclusion

Despite the aforementioned limitations, not all of which were thoroughly discussed (such as the potential for recall bias) the BRFSS telephone survey has been shown to establish true validity and reliable measures[39-40] from which we can confidently trust that the findings presented offer some statistically sound insight to a connection between veteran status and any/all cancers. However, further studies are needed to better understand the mechanisms through which these two variables interact. The cross sectional nature of the data does not allow for extrapolating a causal or temporal direction to their relationship, but the results do indicate significant insight that can impact future studies, policy setting, funding trends, and even physician interaction with veteran patients.

5. REFERENCES

[1] American Cancer Society. *Cancer Facts & Figures 2013*. Atlanta: American Cancer Society; 2013.

[2] American Cancer Society. *Global Cancer Facts & Figures* 3rd Edition. Atlanta: American Cancer Society; 2015.

[3] National Institute for Occupational Safety and Health (NIOSH). Division of Surveillance, Hazard Evaluations, and Field Studies. (2014, May)

[4] Straif K [2008]. The burden of occupational cancer. Occupational and Environmental Medicine. 65(12):787-788.

[5] Rusiecki, J. Thomas, D., & Blair, A. (2009, Aug) Mortality Among United States Coast Guard Marine Inspectors: Follow Up. Military Medicine, 174, 843-851.

[6] National Research Council. (2004). *Veterans and Agent Orange: Length of Presumptive Period for Associaton Between Exposure and Respiratory Cancer*. Washington, D.C.: Institute of Medicine of The National Academies Press.

[7] Michalek, J. E. and M. Pavuk (2008). "*Diabetes and cancer in veterans of Operation Ranch Hand after adjustment for calendar period, days of spraying, and time spent in Southeast Asia*." J Occup Environ Med 50(3): 330-340.

[8] Dusaud, M., et al. (2015). "*A 20-Year Epidemiological Review of Testis Cancer at a French Military Hospital*." Mil Med 180(11): 1184-1188.

[9] Enewold, L., Zhou, J., Devesa, S., Erickson, R., Zhu, K., and McGlynn, K. (2011, Oct). *Trends in Testicular Germ Cell Tumors Among U.S. Military Servicemen, 1990-2003*. Military Medicine, 176, 1184-1187.

[10] Levine, P., et al (2005), "*Is Testicular Cancer Related to Gulf War Deployment? Evidence from a Pilot Population-Based Study of Gulf War Era Veteran and Cancer Registries*." Military Medicine, 170, 149-153.

[11] Nahleh, Z. A., et al. "*Male breast cancer in the veterans affairs population: a comparative analysis*." Cancer 109 (8): 1471-1477. (2007)

[12] Boudreau, Abbie. "*Male breast cancer patients blame a Marine base*." CNN. (2009, Sept).

[13] Chiu, W. A., et al. (2013). "*Human health effects of trichloroethylene: key findings and scientific issues*." Environ Health Perspect 121(3): 303-311.

[14] United States Government Accountability Office. "*Activities Related to Past Drinking Water Contamination at Marine Corps Base Camp Lejeune*." May 2007.

[15] Centers for Disease Control and Prevention (CDC). *The BRFSS Data User Guide* 08.15.2013.

[16] Centers for Disease Control and Prevention (CDC). *Behavioral Risk Factor Surveillance System (BRFSS) 2013 Codebook Report Land-line and Cell-Phone data*. Aug 18, 2014.

[17] Centers for Disease Control and Prevention (CDC). *Behavioral Risk Factor Surveillance System (BRFSS) 2013 Overview*. Aug 15, 2014.

[18] Centers for Disease Control and Prevention (CDC). *Behavioral Risk Factor Surveillance System Questionnaire (BRFSS)*. Final 12.28.2012.

[19] Lea, S., Efird, J., Toland, A., Lewis, D., & Phillips, C. (2014, Mar) *Melanoma incidence Rates in Active Duty Military Personnel*. *Military Medicine, 179*, 247-253.

[20] Kennet J GJ, eds. *Evaluating and improving methods used in the National Survey on Drug Use and Health*. Rockville, MD: US Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies; 2006.

[21] Das SR, Kinsinger LS, Yancy WS, et al: *Obesity prevalence among veterans at Veterans Affairs medical facilities*. American Journal of Preventative Medicine 2005; 28: 291.

[22] Ko, L., Allicok, M., Campbell, M., et al: *An Examination of Sociodemographic, Health, Psychological Factors, and Fruit and Vegetable Consumption Among Overweight and Obese U.S. Veterans.* Military Medicine. 176 (2011, Nov): 1281-1286.

[23] US Department of Veterans Affairs Office of Inspector General. Homeless incidence and risk factors for becoming homeless in veterans. http://www.va.gov/oig/pubs/VAOIG-11-03428-173.pdf.

[24] U.S. Cancer Statistics Working Group. *United States Cancer Statistics: 1999–2012 Incidence and Mortality Web-based Report.* Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; 2015. Available at: www.cdc.gov/uscs.

[25] Gundlapallt, A., et al. *Military Misconduct and Homelessness Among US Veterans Separated From Active Duty, 2001-2012.* Journal of American Medical Association 314 (8) 832-834. (2015, Aug).

[26] American Cancer Society. Cancer Facts & Figures for Hispanics/Latinos 2015-2017. Atlanta: American Cancer Society; 2015.

[27] Centers for Disease Control and Prevention. *Current Cigarette Smoking Among Adults Aged ≥18 Years – United States, 2005-2010*. *MMWR Morb Mortal Wkly Rep*. 2011;60(35):1207-1212.

[28] Centers for Disease Control and Prevention. *Current Cigarette Smoking Among Adults Aged ≥18 Years – United States, 2011*. *MMWR Morb Mortal Wkly Rep*. 2012;61(44):889-894.

[29] Jean C. Beckham, PhD*tt; Mary E. Becker, PhD*; Kim W. Hamlett-Berry. *Preliminary*

Findings from a Clinical Demonstration Project for Veterans Returning from Iraq or Afghanistan. Military Medicine. 173, 5: 448-451, (2008, May).

[30] Department of Veterans Aifairs, Veterans Health Administration, Office of the Assistant Under the Deputy for Health for Policy and Planning, *2005 Survey of Veteran Enrollees' Health and Reliance upon VA*. Washington, DC, September 2006.

[31] Centers for Disease Control and Prevention Coordinating Center for Health Promotion. *Targeting Tobacco Use—the Nation's Leading Cause of Preventable Death*, January 2007. Available at http://www.cdc.gov/tobacco/basic_information/FastFacts.htm.

[32] Miller DR, Kaiman D, Ren XS, et al: *Health Behaviors of Veterans in the VHA: Tobacco Abuse: 1999 Large Health Survey of VHA Enrollees*. Washington, DC, Office of Quality and Performance, Veterans Health Administration, Department of Veterans Affairs, 2001.

[33] Boos CJ, Croft AM: *Smoking rates in the staff of a military field hospital before and after wartime deployment*. J R Soc Med 2004; 97: 20-2.

[34] Forgas L, Meyer D, Cohen M: *Tobacco use habits of naval personnel during Desert Storm*. Milit Med 1996; 161: 165-8.

[35] StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX:StataCorp LP.

[36] U.S. Census Bureau, Population Division. *Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2014*. (2015, May).

[37] Department of Veteran Affairs. *Veteran Population Projection Model VetPop2014*. Office of the Assistant Secretary for Policy and Planning. (2015, Sept).

[38] U.S. Census Bureau. Census 2013.

[39] Bowlin SJ, Morrill BD, Nafziger AN, Lewis C, Pearson TA. Reliability and changes in validity of self-reported cardiovascular disease risk factors using dual response: the behavioral risk factor survey. J Clin Epidemiol 1996;49(5):511–7.

[40] Shea S, Stein AD, Lantigua R, Basch CE. Reliability of the behavioral risk factor survey in a triethnic population. Am J Epidemiol 1991;133 (5):489–500.

6. APPENDIX

```
. use "D:\[[ COURSES ]]\11. [2508 Biostats & Data Analysis II]\[Data]\2013 BRFSS data and
documentation\LLCP2013.start\LLCP2013.start.dta", clear

. save "D:\[[ COURSES ]]\11. [2508 Biostats & Data Analysis II]\[Final Project]\Final Project
DataSet.dta"

file D:\[[ COURSES ]]\11. [2508 Biostats & Data Analysis II]\[Final Project]\Final Project
DataSet.dta saved

Generated new variables

Tab to look at original variable from dataset

Generate to create new variables to work with

Replace or recode to recode data as desired

Tab to check recoding of new variable

. tab veteran3

  ARE YOU A |

    VETERAN |      Freq.     Percent        Cum.

------------+-----------------------------------

          1 |     61,505       12.51       12.51

          2 |    429,527       87.39       99.90

          7 |        166        0.03       99.93

          9 |        330        0.07      100.00

------------+-----------------------------------

      Total |    491,528      100.00

. gen vet=veteran3

(245 missing values generated)

. recode vet (1=1) (2=0) (7=.) (9=.)

(vet: 430023 changes made)

. tab vet

        vet |      Freq.     Percent        Cum.

------------+-----------------------------------

          0 |    429,527       87.47       87.47

          1 |     61,505       12.53      100.00

------------+-----------------------------------

      Total |    491,032      100.00
```

```
. tab chcocncr

(EVER TOLD) |
YOU HAD ANY |
OTHER TYPES |
       OF C |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |     47,139        9.59        9.59
          2 |    443,482       90.18       99.77
          7 |        922        0.19       99.95
          9 |        230        0.05      100.00
------------+-----------------------------------
      Total |    491,773      100.00

. gen can=chcocncr

. recode can (1=1) (2=0) (7=.) (9=.)
(can: 444634 changes made)

. tab can

        can |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |    443,482       90.39       90.39
          1 |     47,139        9.61      100.00
------------+-----------------------------------
      Total |    490,621      100.00

. tab chcscncr
(EVER TOLD) |
    YOU HAD |
       SKIN |
    CANCER? |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |     45,529        9.26        9.26
          2 |    444,940       90.48       99.73
          7 |      1,121        0.23       99.96
          9 |        183        0.04      100.00
------------+-----------------------------------
      Total |    491,773      100.00
. gen skin=chcscncr
. recode skin (1=1) (2=0) (7=.) (9=.)
(skin: 446244 changes made)
```

```
. tab skin

       skin |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |    444,940       90.72       90.72
          1 |     45,529        9.28      100.00
------------+-----------------------------------
      Total |    490,469      100.00

. tab _age_g

IMPUTED AGE |

     IN SIX |

     GROUPS |      Freq.     Percent        Cum.

------------+-----------------------------------

          1 |     27,188        5.53        5.53

          2 |     50,154       10.20       15.73

          3 |     60,371       12.28       28.00

          4 |     83,734       17.03       45.03

          5 |    109,444       22.25       67.29

          6 |    160,882       32.71      100.00

------------+-----------------------------------

      Total |    491,773      100.00


. gen age=_age_g


. recode age (1=1) (2=1) (3=2) (4=2) (5=3) (6=4)

(age: 464585 changes made)


. tab age


        age |      Freq.     Percent        Cum.

------------+-----------------------------------

          1 |     77,342       15.73       15.73

          2 |    144,105       29.30       45.03

          3 |    109,444       22.25       67.29

          4 |    160,882       32.71      100.00

------------+-----------------------------------

      Total |    491,773      100.00


. tab sex


RESPONDENTS |
```

```
        SEX |      Freq.     Percent        Cum.
------------+-----------------------------------
     1.Male |    201,275       40.93       40.93
   2.Female |    290,498       59.07      100.00
------------+-----------------------------------
      Total |    491,773      100.00

. gen sexo=sex

. recode sexo (1=0) (2=1)
(sexo: 491773 changes made)

. tab sexo

       sexo |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |    201,275       40.93       40.93
          1 |    290,498       59.07      100.00
------------+-----------------------------------
      Total |    491,773      100.00

. tab _race

         COMPUTED |
   RACE-ETHNICITY |
         GROUPING |      Freq.     Percent        Cum.
------------------+-----------------------------------
      1.White, nh |    376,451       76.55       76.55
      2.Black, nh |     39,151        7.96       84.52
       3.AIAN, nh |      7,683        1.56       86.08
      4.Asian, nh |      9,510        1.93       88.01
       5.NHPI, nh |      1,546        0.31       88.33
      6.other, nh |      2,693        0.55       88.87
7.multiracial, nh |      9,130        1.86       90.73
       8.Hispanic |     37,054        7.54       98.27
      9.dk/ns/ref |      8,530        1.73      100.00
------------------+-----------------------------------
            Total |    491,748      100.00
```

```
. gen eth=_race
(25 missing values generated)

. recode eth (1=1) (2=2) (3=4) (4=4) (5=4) (6=4) (7=4) (8=3) (9=.)
(eth: 66636 changes made)

. tab eth

        eth |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |    376,451       77.91       77.91
          2 |     39,151        8.10       86.01
          3 |     37,054        7.67       93.68
          4 |     30,562        6.32      100.00
------------+-----------------------------------
      Total |    483,218      100.00

. tab _smoker3

   COMPUTED |
    SMOKING |
     STATUS |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |     55,157       11.22       11.22
          2 |     21,455        4.36       15.58
          3 |    138,218       28.11       43.68
          4 |    261,621       53.20       96.88
          9 |     15,322        3.12      100.00
------------+-----------------------------------
      Total |    491,773      100.00

. gen smoke=_smoker3

. recode smoke (1=1) (2=1) (3=1) (4=0) (9=.)
(smoke: 436616 changes made)

. tab smoke
```

```
      smoke |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |    261,621       54.91       54.91
          1 |    214,830       45.09      100.00
------------+-----------------------------------
      Total |    476,451      100.00

. tab persdoc2

   MULTIPLE |
HEALTH CARE |
PROFESSIONA |
         LS |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |    369,084       75.05       75.05
          2 |     41,306        8.40       83.45
          3 |     79,587       16.18       99.63
          7 |      1,176        0.24       99.87
          9 |        620        0.13      100.00
------------+-----------------------------------
      Total |    491,773      100.00

. gen doc=persdoc2

. recode doc (1=1) (2=1) (3=0) (7=.) (9=.)
(doc: 122689 changes made)

. tab doc

        doc |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |     79,587       16.24       16.24
          1 |    410,390       83.76      100.00
------------+-----------------------------------
      Total |    489,977      100.00

. tab hlthpln1
```

```
   HAVE ANY |
HEALTH CARE |
   COVERAGE |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |    434,627       88.38       88.38
          2 |     55,242       11.23       99.61
          7 |      1,023        0.21       99.82
          9 |        881        0.18      100.00
------------+-----------------------------------
      Total |    491,773      100.00

. gen ins=hlthpln1

. recode ins (1=1) (2=0) (7=.) (9=.)
(ins: 57146 changes made)

. tab ins

        ins |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |     55,242       11.28       11.28
          1 |    434,627       88.72      100.00
------------+-----------------------------------
      Total |    489,869      100.00

. tab _rfbmi5

 OVERWEIGHT |
   OR OBESE |
 CALCULATED |
   VARIABLE |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |    163,257       33.20       33.20
          2 |    301,795       61.37       94.57
          9 |     26,721        5.43      100.00
------------+-----------------------------------
      Total |    491,773      100.00
```

```
. gen owob =_rfbmi5

. recode owob (1=0) (2=1) (9=.)
(owob: 491773 changes made)

. tab owob

       owob |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |    163,257       35.11       35.11
          1 |    301,795       64.89      100.00
------------+-----------------------------------
      Total |    465,052      100.00

. tab _rfdrhv4

      HEAVY |
    ALCOHOL |
CONSUMPTION |
 CALCULATED |
         VA |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |    442,353       89.95       89.95
          2 |     25,546        5.19       95.15
          9 |     23,874        4.85      100.00
------------+-----------------------------------
      Total |    491,773      100.00

. gen drink=_rfdrhv4

. recode drink (1=0) (2=1) (9=.)
(drink: 491773 changes made)

. tab drink

      drink |      Freq.     Percent        Cum.
------------+-----------------------------------
```

```
          0 |    442,353       94.54       94.54
          1 |     25,546        5.46      100.00
------------+-----------------------------------
      Total |    467,899      100.00

. tab _frtlt1

    CONSUME |
 FRUIT 1 OR |
 MORE TIMES |
    PER DAY |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |    291,757       59.33       59.33
          2 |    171,326       34.84       94.17
          9 |     28,690        5.83      100.00
------------+-----------------------------------
      Total |    491,773      100.00

. gen fruit=_frtlt1

. recode fruit (1=1) (2=0) (9=.)
(fruit: 200016 changes made)

. tab fruit

      fruit |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |    171,326       37.00       37.00
          1 |    291,757       63.00      100.00
------------+-----------------------------------
      Total |    463,083      100.00

. tab _veglt1

    CONSUME |
 VEGETABLES |
  1 OR MORE |
TIMES PER D |      Freq.     Percent        Cum.
```

```
------------+-----------------------------------
          1 |    359,902       73.18       73.18
          2 |    101,722       20.68       93.87
          9 |     30,149        6.13      100.00
------------+-----------------------------------
      Total |    491,773      100.00

. gen veg=_veglt1

. recode veg (1=1) (2=0) (9=.)
(veg: 131871 changes made)

. tab veg

        veg |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |    101,722       22.04       22.04
          1 |    359,902       77.96      100.00
------------+-----------------------------------
      Total |    461,624      100.00

. tab exerany2

EXERCISE IN |
    PAST 30 |
       DAYS |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |    332,429       72.20       72.20
          2 |    125,314       27.22       99.41
          7 |        561        0.12       99.53
          9 |      2,154        0.47      100.00
------------+-----------------------------------
      Total |    460,458      100.00

. gen exercise=exerany2
(31,315 missing values generated)

. recode exercise (1=1) (2=0) (7=.) (9=.)
```

```
(exercise: 128029 changes made)

. tab exercise

   exercise |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |    125,314       27.38       27.38
          1 |    332,429       72.62      100.00
------------+-----------------------------------
      Total |    457,743      100.00
```

Generated 1-Combined Outcome Variable Using skin and can variables

Egen to create new variable from two variables

Replace or recode to recode data as desired

Tab to check recoding of new variable

```
. egen allcan = group (can skin), label
(2217 missing values generated)

. tab allcan, missing

    group(can |
        skin) |      Freq.     Percent        Cum.
--------------+-----------------------------------
  0. No 0. No |    407,191       82.80       82.80
 0. No 1. Yes |     35,436        7.21       90.01
 1. Yes 0. No |     36,975        7.52       97.53
1. Yes 1. Yes |      9,954        2.02       99.55
            . |      2,217        0.45      100.00
--------------+-----------------------------------
        Total |    491,773      100.00

. tab allcan, nolabel

  group(can |
```

```
      skin) |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |    407,191       83.18       83.18
          2 |     35,436        7.24       90.41
          3 |     36,975        7.55       97.97
          4 |      9,954        2.03      100.00
------------+-----------------------------------
      Total |    489,556      100.00
```

```
. gen allcan4=allcan
(2,217 missing values generated)
```

***Even though I may not use allcan4 in my project, it will be my combined data (skin and non-skin cancers) that is not going to be binary, but still categorical**

```
. recode allcan (1=0) (2=1) (3=1) (4=1)
(allcan: 489556 changes made)
```

```
. tab allcan, missing

    group(can |
        skin) |      Freq.     Percent        Cum.
--------------+-----------------------------------
            0 |    407,191       82.80       82.80
            1 |     82,365       16.75       99.55
            . |      2,217        0.45      100.00
--------------+-----------------------------------
        Total |    491,773      100.00
```

```
. tab allcan, nolabel

  group(can |
      skin) |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |    407,191       83.18       83.18
          1 |     82,365       16.82      100.00
------------+-----------------------------------
      Total |    489,556      100.00
```

```
. recode allcan4 (1=0) (2=1) (3=2) (4=3)
(allcan4: 489556 changes made)

. tab allcan4

   group(can |
       skin) |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |    407,191       83.18       83.18
          1 |     35,436        7.24       90.41
          2 |     36,975        7.55       97.97
          3 |      9,954        2.03      100.00
------------+-----------------------------------
      Total |    489,556      100.00
```

Label Everything

Label Define to set new label names

Label Value to assign label to variable output options

Tab to check labeling

Label Variable to give title or label to variable

Tab to check labeling

```
. label define yn 0 "0. No" 1 "1. Yes"

. label value vet can skin allcan smoke doc owob ins drink veg fruit exercise

. label define sexo 0 "0. Male" 1 "1. Female"

. label value sexo sexo

. label define age_groups 1 "1. 18-34" 2 "2. 35-54" 3 "3. 55-64" 4 "4. 65+"
```

```
. label value age age_groups

. label define eth_groups 1 "1. White" 2 "2. Black" 3 "3. Hispanic" 4 "4. Other"

. label value eth eth_groups

*Again even though I may not use allcan4 in my project, it will be my combined data (skin and
non-skin cancers) that is not going to be binary, but still categorical

. label define allcan4 0 "0. No Solid Organ & No Skin Cancer" 1 "1. No Solid Organ But Yes Skin
Cancer" 2 "2. Yes Solid Organ But No Skin Cancer" 3 "3. Yes Solid Organ and Yes Skin Cancer"

. label value allcan4 allcan4

. tab vet

        vet |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    429,527       87.47       87.47
     1. Yes |     61,505       12.53      100.00
------------+-----------------------------------
      Total |    491,032      100.00

. tab can

        can |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    443,482       90.39       90.39
     1. Yes |     47,139        9.61      100.00
------------+-----------------------------------
      Total |    490,621      100.00

. tab skin

       skin |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    444,940       90.72       90.72
     1. Yes |     45,529        9.28      100.00
------------+-----------------------------------
```

```
      Total |    490,469      100.00
```

```
. tab allcan

     allcan |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    407,191       83.18       83.18
     1. Yes |     82,365       16.82      100.00
------------+-----------------------------------
      Total |    489,556      100.00
```

```
. tab smoke

      smoke |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    261,621       54.91       54.91
     1. Yes |    214,830       45.09      100.00
------------+-----------------------------------
      Total |    476,451      100.00
```

```
. tab doc

        doc |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |     79,587       16.24       16.24
     1. Yes |    410,390       83.76      100.00
------------+-----------------------------------
      Total |    489,977      100.00
```

```
. tab ins

        ins |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |     55,242       11.28       11.28
     1. Yes |    434,627       88.72      100.00
------------+-----------------------------------
      Total |    489,869      100.00
```

```
. tab owob

       owob |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    163,257       35.11       35.11
     1. Yes |    301,795       64.89      100.00
------------+-----------------------------------
      Total |    465,052      100.00

. tab drink

      drink |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    442,353       94.54       94.54
     1. Yes |     25,546        5.46      100.00
------------+-----------------------------------
      Total |    467,899      100.00

. tab fruit

      fruit |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    171,326       37.00       37.00
     1. Yes |    291,757       63.00      100.00
------------+-----------------------------------
      Total |    463,083      100.00

. tab veg

        veg |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    101,722       22.04       22.04
     1. Yes |    359,902       77.96      100.00
------------+-----------------------------------
      Total |    461,624      100.00

. tab exercise
```

```
   exercise |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    125,314       27.38       27.38
     1. Yes |    332,429       72.62      100.00
------------+-----------------------------------
      Total |    457,743      100.00
```

```
. tab sexo
```

```
       sexo |      Freq.     Percent        Cum.
------------+-----------------------------------
    0. Male |    201,275       40.93       40.93
  1. Female |    290,498       59.07      100.00
------------+-----------------------------------
      Total |    491,773      100.00
```

```
. tab age
```

```
        age |      Freq.     Percent        Cum.
------------+-----------------------------------
   1. 18-34 |     77,342       15.73       15.73
   2. 35-54 |    144,105       29.30       45.03
   3. 55-64 |    109,444       22.25       67.29
     4. 65+ |    160,882       32.71      100.00
------------+-----------------------------------
      Total |    491,773      100.00
```

```
. tab eth
```

```
        eth |      Freq.     Percent        Cum.
------------+-----------------------------------
   1. White |    376,451       77.91       77.91
   2. Black |     39,151        8.10       86.01
3. Hispanic |     37,054        7.67       93.68
   4. Other |     30,562        6.32      100.00
------------+-----------------------------------
      Total |    483,218      100.00
```

```
. tab allcan4

                               allcan4 |      Freq.     Percent        Cum.
---------------------------------------+-----------------------------------
    0. No Solid Organ & No Skin Cancer |    407,191       83.18       83.18
   1. No Solid Organ But Yes Skin Cancer |     35,436        7.24       90.41
   2. Yes Solid Organ But No Skin Cancer |     36,975        7.55       97.97
 3. Yes Solid Organ and Yes Skin Cancer |      9,954        2.03      100.00
---------------------------------------+-----------------------------------
                                 Total |    489,556      100.00

. label variable vet "Veteran?"

. label variable can "Has/Had Non-Skin Cancer?"

. label variable skin "Have/Had Skin Cancer?"

. label variable allcan "Has/Had ANY Cancer?"

. label variable age "Age Groups"

. label variable sexo "Sex"

. label variable eth "Ethnic Background"

. label variable doc "Has at least 1 doctor?"

. label variable smoke "Has been or is a smoker?"

. label variable allcan4 "Skin and Organ Cancers?"

. label variable ins "Has Health Insurance?"

. label variable owob "Overweight or Obese: BMI > 25?"

. label variable drink "Current Heavy Drinker?"

. label variable fruit "Consumes at least 1 Fruit per day?"
```

```
. label variable veg "Consumes at least 1 Vegetable per day?"

. label variable exercise "During the past month, participated in physical activities or
exercise?"

. tab vet

   Veteran? |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    429,527       87.47       87.47
     1. Yes |     61,505       12.53      100.00
------------+-----------------------------------
      Total |    491,032      100.00

. tab can

    Has/Had |
   Non-Skin |
    Cancer? |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    443,482       90.39       90.39
     1. Yes |     47,139        9.61      100.00
------------+-----------------------------------
      Total |    490,621      100.00

. tab age

 Age Groups |      Freq.     Percent        Cum.
------------+-----------------------------------
   1. 18-34 |     77,342       15.73       15.73
   2. 35-54 |    144,105       29.30       45.03
   3. 55-64 |    109,444       22.25       67.29
     4. 65+ |    160,882       32.71      100.00
------------+-----------------------------------
      Total |    491,773      100.00

. tab sexo
```

```
        Sex |      Freq.     Percent        Cum.
------------+-----------------------------------
    0. Male |    201,275       40.93       40.93
  1. Female |    290,498       59.07      100.00
------------+-----------------------------------
      Total |    491,773      100.00

. tab eth

     Ethnic |
 Background |      Freq.     Percent        Cum.
------------+-----------------------------------
   1. White |    376,451       77.91       77.91
   2. Black |     39,151        8.10       86.01
3. Hispanic |     37,054        7.67       93.68
   4. Other |     30,562        6.32      100.00
------------+-----------------------------------
      Total |    483,218      100.00

. tab doc

     Has at |
    least 1 |
    doctor? |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |     79,587       16.24       16.24
     1. Yes |    410,390       83.76      100.00
------------+-----------------------------------
      Total |    489,977      100.00

. tab smoke

Has been or |
       is a |
    smoker? |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    261,621       54.91       54.91
```

```
     1. Yes |    214,830       45.09      100.00
------------+-----------------------------------
      Total |    476,451      100.00
```

```
. tab allcan4
```

```
               Skin and Organ Cancers? |      Freq.     Percent        Cum.
---------------------------------------+-----------------------------------
    0. No Solid Organ & No Skin Cancer |    407,191       83.18       83.18
 1. No Solid Organ But Yes Skin Cancer |     35,436        7.24       90.41
 2. Yes Solid Organ But No Skin Cancer |     36,975        7.55       97.97
3. Yes Solid Organ and Yes Skin Cancer |      9,954        2.03      100.00
---------------------------------------+-----------------------------------
                                 Total |    489,556      100.00
```

```
. tab ins
```

```
 Has Health |
Insurance? |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |     55,242       11.28       11.28
     1. Yes |    434,627       88.72      100.00
------------+-----------------------------------
      Total |    489,869      100.00
```

```
. tab owob
```

```
 Overweight |
  or Obese: |
  BMI > 25? |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    163,257       35.11       35.11
     1. Yes |    301,795       64.89      100.00
------------+-----------------------------------
      Total |    465,052      100.00
```

```
. tab fruit
```

```
Consumes at |
    least 1 |
  Fruit per |
       day? |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    171,326       37.00       37.00
     1. Yes |    291,757       63.00      100.00
------------+-----------------------------------
      Total |    463,083      100.00

. tab veg

Consumes at |
    least 1 |
  Vegetable |
   per day? |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    101,722       22.04       22.04
     1. Yes |    359,902       77.96      100.00
------------+-----------------------------------
      Total |    461,624      100.00

. tab exercise

 During the |
past month, |
participate |
       d in |
   physical |
 activities |
         or |
  exercise? |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    125,314       27.38       27.38
     1. Yes |    332,429       72.62      100.00
------------+-----------------------------------
      Total |    457,743      100.00
```

```
. tab drink

    Current |
      Heavy |
   Drinker? |      Freq.     Percent        Cum.
------------+-----------------------------------
      0. No |    442,353       94.54       94.54
     1. Yes |     25,546        5.46      100.00
------------+-----------------------------------
      Total |    467,899      100.00
```

Generate Instudy(instudies)

Generate Instudy = 0 first

Replace to set Instudy to include only valid data

Tab to Check Instudy

***I did not end up using instudy in the end, it was my original idea to only look at solid organ cancers.**

```
. gen instudy=0

. replace instudy=1 if vet!=. & can!=.
(489,913 real changes made)

. tab instudy

    instudy |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      1,860        0.38        0.38
          1 |    489,913       99.62      100.00
------------+-----------------------------------
      Total |    491,773      100.00

. gen instudy2=0
```

```
. replace instudy2=1 if vet!=. & allcan!=.
(488,856 real changes made)

. tab instudy2

   instudy2 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      2,917        0.59        0.59
          1 |    488,856       99.41      100.00
------------+-----------------------------------
      Total |    491,773      100.00
```

*I did not use instudy3 in my project I was just curious only skin cancer compared.

```
. gen instudy3=0

. replace instudy3=1 if vet!=. & skin!=.
(489,759 real changes made)

. tab instudy3

   instudy3 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      2,014        0.41        0.41
          1 |    489,759       99.59      100.00
------------+-----------------------------------
      Total |    491,773      100.00
```

*If I had thought to consider cancer data stratified by non-skin vs. skin I would have used this instudy, much more interesting this way.

```
. gen instudy4=0

. replace instudy4=1 if vet!=. & allcan4!=.
(488,856 real changes made)

. tab instudy4

   instudy4 |      Freq.     Percent        Cum.
```

```
------------+-----------------------------------
          0 |      2,917        0.59        0.59
          1 |    488,856       99.41      100.00
------------+-----------------------------------
      Total |    491,773      100.00
```

Obtaining survey estimates (and p-values of Pearson chi-squared tests for independence) for covariates of interest through bivariate analysis between outcome and covariates against exposure, veteran status. Output data is used for Table 1.

First, svyset to obtain weighted results

Svy= conducts bivariate analysis of analytic sample

```
. svyset _psu [pweight=_llcpwt], strata(_ststr) vce(linearized) singleunit (missing)

      pweight: _llcpwt
          VCE: linearized
  Single unit: missing
     Strata 1: _ststr
         SU 1: _psu
        FPC 1: <zero>

. svy, subpop (if instudy2==1): tab allcan vet, col obs cellwidth(20) format(%15.2g)
(running tabulate on estimation sample)

Number of strata   =     1,303               Number of obs     =      491,773
Number of PSUs     =   491,773               Population size   =  246,024,416
                                             Subpop. no. obs   =      488,856
                                             Subpop. size      =  244,739,321
                                             Design df         =      490,470

-----------------------------------------------------------------------------
Has/Had   |
ANY       |                             Veteran?
Cancer?   |               0, No                1, Yes                 Total
----------+------------------------------------------------------------------
    0, No |                  .9                    .8                   .89
          |              361440                 45141                406581
```

```
          |
   1, Yes |                 .1                  .2                 .11
          |              66309               15966               82275
          |
    Total |                  1                   1                   1
          |             427749               61107              488856
----------------------------------------------------------------------------
  Key:  column proportion
        number of observations

  Pearson:
    Uncorrected   chi2(1)         = 4992.2554
    Design-based  F(1, 490470)    = 1825.1922     P = 0.0000

. svy, subpop (if instudy2==1): tab age vet, col obs cellwidth(20) format(%15.2g)
(running tabulate on estimation sample)

Number of strata   =     1,303              Number of obs     =      491,773
Number of PSUs     =   491,773              Population size   =  246,024,416
                                            Subpop. no. obs   =      488,856
                                            Subpop. size      =  244,739,321
                                            Design df         =      490,470

----------------------------------------------------------------------------
Age       |                            Veteran?
Groups    |               0, No               1, Yes               Total
----------+-----------------------------------------------------------------
 1, 18-34 |                .32                 .14                  .3
          |              73005                4028               77033
          |
 2, 35-54 |                .36                 .27                 .35
          |             132066               11306              143372
          |
 3, 55-64 |                .16                 .18                 .16
          |              96721               11984              108705
          |
   4, 65+ |                .16                 .41                 .19
          |             125957               33789              159746
```

```
          |
    Total |                    1                    1                    1
          |               427749                61107               488856
----------------------------------------------------------------------------
  Key:  column proportion
        number of observations

  Pearson:
    Uncorrected   chi2(3)         =  2.13e+04
    Design-based  F(2.88,  1.4e+06)= 1901.9470    P = 0.0000

. svy, subpop (if instudy2==1): tab sexo vet, col obs cellwidth(20) format(%15.2g)
(running tabulate on estimation sample)

Number of strata   =     1,303              Number of obs     =      491,773
Number of PSUs     =   491,773              Population size   =  246,024,416
                                            Subpop. no. obs   =      488,856
                                            Subpop. size      =  244,739,321
                                            Design df         =      490,470

----------------------------------------------------------------------------
          |                              Veteran?
      Sex |                 0, No               1, Yes                Total
----------+-----------------------------------------------------------------
  0, Male |                  .44                  .91                  .49
          |               144249                55616               199865
          |
 1, Femal |                  .56                 .089                  .51
          |               283500                 5491               288991
          |
    Total |                    1                    1                    1
          |               427749                61107               488856
----------------------------------------------------------------------------
  Key:  column proportion
        number of observations

  Pearson:
    Uncorrected   chi2(1)         =  4.21e+04
```

```
    Design-based  F(1, 490470)    =  1.22e+04     P = 0.0000

. svy, subpop (if instudy2==1): tab eth vet, col obs cellwidth(20) format(%15.2g)
(running tabulate on estimation sample)

Number of strata   =     1,303              Number of obs     =      483,444
Number of PSUs     =   483,444              Population size   =  241,049,359
                                            Subpop. no. obs   =      480,527
                                            Subpop. size      =  239,764,265
                                            Design df         =      482,141

----------------------------------------------------------------------------
Ethnic    |
Backgroun |                            Veteran?
d         |               0, No                1, Yes                Total
----------+-----------------------------------------------------------------
 1, White |                 .63                   .75                  .64
          |              324441                 49926               374367
          |
 2, Black |                 .12                   .12                  .12
          |               34711                  4255                38966
          |
 3, Hispa |                 .18                  .071                  .17
          |               34598                  2270                36868
          |
 4, Other |                .079                  .052                 .076
          |               26939                  3387                30326
          |
    Total |                   1                     1                    1
          |              420689                 59838               480527
----------------------------------------------------------------------------
  Key:  column proportion
        number of observations

  Pearson:
    Uncorrected   chi2(3)         = 4607.3238
    Design-based  F(2.99,  1.4e+06)=  296.0494    P = 0.0000
```

```
. svy, subpop (if instudy2==1): tab smoke vet, col obs cellwidth(20) format(%15.2g)
(running tabulate on estimation sample)

Number of strata   =     1,303              Number of obs     =      476,905
Number of PSUs     =   476,905              Population size   =  236,245,612
                                            Subpop. no. obs   =      473,988
                                            Subpop. size      =  234,960,517
                                            Design df         =      475,602

----------------------------------------------------------------------------
Has been  |
or is a   |                            Veteran?
smoker?   |                0, No                1, Yes                 Total
----------+-----------------------------------------------------------------
    0, No |                  .59                   .39                   .57
          |               238444                 22067                260511
          |
   1, Yes |                  .41                   .61                   .43
          |               176170                 37307                213477
          |
    Total |                    1                     1                     1
          |               414614                 59374                473988
----------------------------------------------------------------------------
  Key:  column proportion
        number of observations

  Pearson:
    Uncorrected   chi2(1)         = 7712.0623
    Design-based  F(1, 475602)    = 2180.5726     P = 0.0000

. svy, subpop (if instudy2==1): tab ins vet, col obs cellwidth(20) format(%15.2g)
(running tabulate on estimation sample)

Number of strata   =     1,303              Number of obs     =      489,929
Number of PSUs     =   489,929              Population size   =  244,681,387
                                            Subpop. no. obs   =      487,012
                                            Subpop. size      =  243,396,292
                                            Design df         =      488,626
```

```
--------------------------------------------------------------------------
Has       |
Health    |
Insurance |                             Veteran?
?         |                0, No                 1, Yes                Total
----------+---------------------------------------------------------------
    0, No |                  .18                   .077                  .17
          |                51307                   3513                54820
          |
   1, Yes |                  .82                    .92                  .83
          |               374752                  57440               432192
          |
    Total |                    1                      1                    1
          |               426059                  60953               487012
--------------------------------------------------------------------------
  Key:  column proportion
        number of observations

  Pearson:
    Uncorrected   chi2(1)         = 3793.2991
    Design-based  F(1, 488626)    =  974.4702     P = 0.0000

. svy, subpop (if instudy2==1): tab doc vet, col obs cellwidth(20) format(%15.2g)
(running tabulate on estimation sample)

Number of strata   =     1,303              Number of obs     =      490,016
Number of PSUs     =   490,016              Population size   =  244,994,897
                                            Subpop. no. obs   =      487,099
                                            Subpop. size      =  243,709,803
                                            Design df         =      488,713

--------------------------------------------------------------------------
          |                               vet
      doc |                    0                      1                Total
----------+---------------------------------------------------------------
        0 |                  .24                    .18                  .24
          |                70506                   8568                79074
```

```
          |
        1 |                .76                 .82                 .76
          |             355727               52298              408025
          |
    Total |                  1                   1                   1
          |             426233               60866              487099
----------------------------------------------------------------------------
  Key:  column proportion
        number of observations

  Pearson:
    Uncorrected   chi2(1)         = 1055.9791
    Design-based  F(1, 488713)    =  265.8488     P = 0.0000


. svy, subpop (if instudy2==1): tab drink vet, col obs cellwidth(20) format(%15.2g)
(running tabulate on estimation sample)

Number of strata   =     1,303              Number of obs     =      468,473
Number of PSUs     =   468,473              Population size   =  229,905,733
                                            Subpop. no. obs   =      465,556
                                            Subpop. size      =  228,620,638
                                            Design df         =      467,170

----------------------------------------------------------------------------
Current   |
Heavy     |                           Veteran?
Drinker?  |              0, No              1, Yes               Total
----------+-----------------------------------------------------------------
    0, No |                .94                 .94                 .94
          |             385112               55024              440136
          |
   1, Yes |                .06                .059                 .06
          |              22196                3224               25420
          |
    Total |                  1                   1                   1
          |             407308               58248              465556
----------------------------------------------------------------------------
```

```
  Key:  column proportion
        number of observations

  Pearson:
    Uncorrected   chi2(1)         =    0.2182
    Design-based  F(1, 467170)    =    0.0610     P = 0.8050

. svy, subpop (if instudy2==1): tab owob vet, col obs cellwidth(20) format(%15.2g)
(running tabulate on estimation sample)

Number of strata   =     1,303              Number of obs     =      465,624
Number of PSUs     =   465,624              Population size   =  231,471,350
                                            Subpop. no. obs   =      462,707
                                            Subpop. size      =  230,186,255
                                            Design df         =      464,321

----------------------------------------------------------------------------
Overweigh |
t or      |
Obese:    |                             Veteran?
BMI > 25? |               0, No                1, Yes                 Total
----------+-----------------------------------------------------------------
    0, No |                 .37                   .26                   .36
          |              146203                 16169                162372
          |
   1, Yes |                 .63                   .74                   .64
          |              256490                 43845                300335
          |
    Total |                   1                     1                     1
          |              402693                 60014                462707
----------------------------------------------------------------------------
  Key:  column proportion
        number of observations

  Pearson:
    Uncorrected   chi2(1)         = 2709.2104
    Design-based  F(1, 464321)    =  826.5618     P = 0.0000
```

```
. svy, subpop (if instudy2==1): tab veg vet, col obs cellwidth(20) format(%15.2g)
(running tabulate on estimation sample)

Number of strata   =     1,303              Number of obs     =      462,137
Number of PSUs     =   462,137              Population size   =  226,617,701
                                            Subpop. no. obs   =      459,220
                                            Subpop. size      =  225,332,606
                                            Design df         =      460,834

----------------------------------------------------------------------------
Consumes  |
at least  |
1         |
Vegetable |                            Veteran?
per day?  |                0, No                1, Yes                 Total
----------+-----------------------------------------------------------------
    0, No |                  .23                   .24                   .24
          |                87497                 13603                101100
          |
   1, Yes |                  .77                   .76                   .76
          |               314224                 43896                358120
          |
    Total |                    1                     1                     1
          |               401721                 57499                459220
----------------------------------------------------------------------------
  Key:  column proportion
        number of observations

  Pearson:
    Uncorrected   chi2(1)         =    5.8702
    Design-based  F(1, 460834)    =    1.7198     P = 0.1897

. svy, subpop (if instudy2==1): tab fruit vet, col obs cellwidth(20) format(%15.2g)
(running tabulate on estimation sample)

Number of strata   =     1,303              Number of obs     =      463,597
Number of PSUs     =   463,597              Population size   =  227,509,101
                                            Subpop. no. obs   =      460,680
```

```
                                              Subpop. size      =  226,224,007
                                              Design df         =      462,294

-----------------------------------------------------------------------------
Consumes  |
at least  |
1 Fruit   |                          Veteran?
per day?  |               0, No                1, Yes                 Total
----------+------------------------------------------------------------------
    0, No |                 .39                    .4                   .39
          |              147905                 22492                170397
          |
   1, Yes |                 .61                    .6                   .61
          |              255066                 35217                290283
          |
    Total |                   1                     1                     1
          |              402971                 57709                460680
-----------------------------------------------------------------------------
  Key:  column proportion
        number of observations

  Pearson:
    Uncorrected   chi2(1)         =   30.8771
    Design-based  F(1, 462294)    =    8.8410     P = 0.0029

. svy, subpop (if instudy2==1): tab exercise vet, col obs cellwidth(20) format(%15.2g)
(running tabulate on estimation sample)

Number of strata   =     1,303              Number of obs     =      458,283
Number of PSUs     =   458,283              Population size   =  224,951,779
                                              Subpop. no. obs   =      455,366
                                              Subpop. size      =  223,666,684
                                              Design df         =      456,980

-----------------------------------------------------------------------------
During    |
the past  |
month,    |
```

```
participa |
ted in    |
physical  |
activitie |
s or      |                      Veteran?
exercise? |             0, No                1, Yes                 Total
----------+--------------------------------------------------------------
    0, No |               .27                   .26                   .27
          |            109253                 15294                124547
          |
   1, Yes |               .73                   .74                   .73
          |            288951                 41868                330819
          |
    Total |                 1                     1                     1
          |            398204                 57162                455366
-------------------------------------------------------------------------
  Key:  column proportion
        number of observations

  Pearson:
    Uncorrected   chi2(1)         =   21.1918
    Design-based  F(1, 456980)    =    6.0540     P = 0.0139
```

```
. svy, subpop (if instudy2==1): tab can vet, col obs cellwidth(20) format(%15.2g)
(running tabulate on estimation sample)

Number of strata   =     1,303              Number of obs     =       491,773
Number of PSUs     =   491,773              Population size   =   246,024,416
                                            Subpop. no. obs   =       488,856
                                            Subpop. size      =   244,739,321
                                            Design df         =       490,470

-------------------------------------------------------------------------
Has/Had   |
Non-Skin  |                      Veteran?
Cancer?   |             0, No                1, Yes                 Total
----------+--------------------------------------------------------------
    0, No |               .94                   .89                   .94
```

```
          |             389302             52675            441977
          |
   1, Yes |                .06               .11              .065
          |              38447              8432             46879
          |
    Total |                  1                 1                 1
          |             427749             61107            488856
----------------------------------------------------------------------------
  Key:  column proportion
        number of observations

  Pearson:
    Uncorrected   chi2(1)         = 1922.5103
    Design-based  F(1, 490470)    =  698.9736     P = 0.0000

. svy, subpop (if instudy2==1): tab skin vet, col obs cellwidth(20) format(%15.2g)
(running tabulate on estimation sample)

Number of strata   =     1,303              Number of obs     =      491,773
Number of PSUs     =   491,773              Population size   =  246,024,416
                                            Subpop. no. obs   =      488,856
                                            Subpop. size      =  244,739,321
                                            Design df         =      490,470

----------------------------------------------------------------------------
Have/Had  |
Skin      |                            Veteran?
Cancer?   |               0, No              1, Yes             Total
----------+-----------------------------------------------------------------
    0, No |                .95               .88               .94
          |             392537             50978            443515
          |
   1, Yes |                .05               .12              .058
          |              35212             10129             45341
          |
    Total |                  1                 1                 1
          |             427749             61107            488856
----------------------------------------------------------------------------
```

```
  Key:  column proportion
        number of observations

  Pearson:
    Uncorrected   chi2(1)         = 4713.8733
    Design-based  F(1, 490470)    = 1930.9062     P = 0.0000

. svy, subpop (if instudy2==1): tab allcan4 vet, col obs cellwidth(20) format(%15.2g)
(running tabulate on estimation sample)

Number of strata   =     1,303              Number of obs     =      491,773
Number of PSUs     =   491,773              Population size   =  246,024,416
                                            Subpop. no. obs   =      488,856
                                            Subpop. size      =  244,739,321
                                            Design df         =      490,470

----------------------------------------------------------------------------
Skin and  |
Organ     |                            Veteran?
Cancers?  |                0, No                1, Yes                Total
----------+-----------------------------------------------------------------
 0, No So |                   .9                    .8                  .89
          |               361440                 45141               406581
          |
 1, No So |                  .04                  .093                 .046
          |                27862                  7534                35396
          |
 2 Yes So |                  .05                  .079                 .053
          |                31097                  5837                36934
          |
 3, Yes S |                .0096                  .031                 .012
          |                 7350                  2595                 9945
          |
    Total |                    1                     1                    1
          |               427749                 61107               488856
----------------------------------------------------------------------------
  Key:  column proportion
        number of observations
```

```
  Pearson:
    Uncorrected   chi2(3)         = 5904.4489
    Design-based  F(2.96,  1.5e+06)=  783.7048    P = 0.0000
```

Generate Dummy /Indicator Variables for Categorical Variables Age & Ethnic Groups to allow logistic regression for categorical variables

Tab variable, gen (new dummy variable name)

```
. tab age, gen(i_age)

 Age Groups |      Freq.     Percent        Cum.
------------+-----------------------------------
   1. 18-34 |     77,342       15.73       15.73
   2. 35-54 |    144,105       29.30       45.03
   3. 55-64 |    109,444       22.25       67.29
     4. 65+ |    160,882       32.71      100.00
------------+-----------------------------------
      Total |    491,773      100.00

. tab eth, gen(i_eth)

     Ethnic |
 Background |      Freq.     Percent        Cum.
------------+-----------------------------------
   1. White |    376,451       77.91       77.91
   2. Black |     39,151        8.10       86.01
3. Hispanic |     37,054        7.67       93.68
   4. Other |     30,562        6.32      100.00
------------+-----------------------------------
      Total |    483,218      100.00
```

Checking Created Dummy Variables

Describe new dummy variable name* - to check

```
. describe i_age*
```

```
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------------------------
---------------
i_age1          byte    %8.0g                 age==1. 18-34
i_age2          byte    %8.0g                 age==2. 35-54
i_age3          byte    %8.0g                 age==3. 55-64
i_age4          byte    %8.0g                 age==4. 65+
```

```
. describe i_eth*
```

```
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------------------------
---------------
i_eth1          byte    %8.0g                 eth==1. White
i_eth2          byte    %8.0g                 eth==2. Black
i_eth3          byte    %8.0g                 eth==3. Hispanic
i_eth4          byte    %8.0g                 eth==4. Other
```

Conduct Crude Statistical Analysis- using dummy variables when necessary

Survey Set so you can conduct statistical analysis and run regressions with weighted data (if you haven't already)

```
. svyset _psu [pweight=_llcpwt], strata(_ststr) vce(linearized) singleunit(missing)

      pweight: _llcpwt
          VCE: linearized
  Single unit: missing
     Strata 1: _ststr
         SU 1: _psu
        FPC 1: <zero>
```

Logistic Regression of Exposure and covariates of interest individually against outcome to obtain Crude Odds Ratios - See Unadjusted Column of Table 2

```
. svy, subpop(if instudy2==1): logistic allcan vet
(running logistic on estimation sample)
```

```
Survey: Logistic regression

Number of strata   =     1,303                Number of obs     =      491,773
Number of PSUs     =   491,773                Population size   =  246,024,416
                                              Subpop. no. obs   =      488,856
                                              Subpop. size      =  244,739,321
                                              Design df         =      490,470
                                              F(   1, 490470)   =      1741.94
                                              Prob > F          =       0.0000

------------------------------------------------------------------------------
             |             Linearized
      allcan | Odds Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         vet |   2.290125   .0454665    41.74   0.000     2.202724    2.380994
       _cons |   .1109011   .0009441  -258.33   0.000     .1090662     .112767
------------------------------------------------------------------------------

. svy, subpop(if instudy2==1): logistic allcan i_age1 i_age2 i_age3
(running logistic on estimation sample)

Survey: Logistic regression

Number of strata   =     1,303                Number of obs     =      491,773
Number of PSUs     =   491,773                Population size   =  246,024,416
                                              Subpop. no. obs   =      488,856
                                              Subpop. size      =  244,739,321
                                              Design df         =      490,470
                                              F(   3, 490468)   =      3835.17
                                              Prob > F          =       0.0000

------------------------------------------------------------------------------
             |             Linearized
      allcan | Odds Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      i_age1 |   .0411986   .0017749   -74.03   0.000     .0378627    .0448284
      i_age2 |    .159429   .0034428   -85.03   0.000      .152822    .1663217
```

```
      i_age3 |   .4140488   .0080981   -45.08   0.000      .398477     .430229
       _cons |   .4424886   .0048344   -74.63   0.000     .4331141     .452066
------------------------------------------------------------------------------
```

```
. svy, subpop(if instudy2==1): logistic allcan sexo
(running logistic on estimation sample)

Survey: Logistic regression

Number of strata   =     1,303              Number of obs     =      491,773
Number of PSUs     =   491,773              Population size   =  246,024,416
                                            Subpop. no. obs   =      488,856
                                            Subpop. size      =  244,739,321
                                            Design df         =      490,470
                                            F(   1, 490470)   =       297.04
                                            Prob > F          =       0.0000

------------------------------------------------------------------------------
             |             Linearized
      allcan | Odds Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sexo |   1.309906   .0205174    17.23   0.000     1.270303    1.350743
       _cons |   .1076016   .0012753  -188.10   0.000     .1051309    .1101304
------------------------------------------------------------------------------
```

```
. svy, subpop(if instudy2==1): logistic allcan i_eth2 i_eth3 i_eth4
(running logistic on estimation sample)

Survey: Logistic regression

Number of strata   =     1,303              Number of obs     =      483,444
Number of PSUs     =   483,444              Population size   =  241,049,359
                                            Subpop. no. obs   =      480,527
                                            Subpop. size      =  239,764,265
                                            Design df         =      482,141
                                            F(   3, 482139)   =       765.74
                                            Prob > F          =       0.0000
```

```
------------------------------------------------------------------------------
             |             Linearized
      allcan | Odds Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      i_eth2 |   .3373969    .013249   -27.67   0.000     .3124033      .36439
      i_eth3 |   .2179195   .0096989   -34.23   0.000     .1997155    .2377828
      i_eth4 |     .29361   .0162083   -22.20   0.000     .2635005    .3271601
       _cons |   .1734219   .0013636  -222.83   0.000     .1707699    .1761152
------------------------------------------------------------------------------

. svy, subpop(if instudy2==1): logistic allcan doc
(running logistic on estimation sample)

Survey: Logistic regression

Number of strata   =     1,303               Number of obs     =      490,016
Number of PSUs     =   490,016               Population size   =  244,994,897
                                             Subpop. no. obs   =      487,099
                                             Subpop. size      =  243,709,803
                                             Design df         =      488,713
                                             F(   1, 488713)   =      2048.79
                                             Prob > F          =       0.0000

------------------------------------------------------------------------------
             |             Linearized
      allcan | Odds Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         doc |   4.142069   .1300535    45.26   0.000     3.894853    4.404976
       _cons |   .0373748   .0011336  -108.36   0.000     .0352177     .039664
------------------------------------------------------------------------------

. svy, subpop(if instudy2==1): logistic allcan smoke
(running logistic on estimation sample)

Survey: Logistic regression

Number of strata   =     1,303               Number of obs     =      476,905
Number of PSUs     =   476,905               Population size   =  236,245,612
```

```
                                                Subpop. no. obs   =      473,988
                                                Subpop. size      =  234,960,517
                                                Design df         =      475,602
                                                F(   1, 475602)   =       974.32
                                                Prob > F          =       0.0000

------------------------------------------------------------------------------
             |             Linearized
      allcan | Odds Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       smoke |   1.626969   .0253693    31.21   0.000     1.577998     1.67746
       _cons |   .0999561   .0010941  -210.40   0.000     .0978345    .1021237
------------------------------------------------------------------------------
```

```
. svy, subpop(if instudy2==1): logistic allcan drink
(running logistic on estimation sample)

Survey: Logistic regression

Number of strata   =     1,303              Number of obs     =      468,473
Number of PSUs     =   468,473              Population size   =  229,905,733
                                                Subpop. no. obs   =      465,556
                                                Subpop. size      =  228,620,638
                                                Design df         =      467,170
                                                F(   1, 467170)   =        10.47
                                                Prob > F          =       0.0012

------------------------------------------------------------------------------
             |             Linearized
      allcan | Odds Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       drink |   .8886472   .0324275    -3.24   0.001     .8273099     .954532
       _cons |   .1276873   .0010173  -258.34   0.000     .1257089    .1296967
------------------------------------------------------------------------------
```

```
. svy, subpop(if instudy2==1): logistic allcan ins
(running logistic on estimation sample)
```

```
Survey: Logistic regression

Number of strata   =     1,303                Number of obs     =    489,929
Number of PSUs     =   489,929                Population size   = 244,681,387
                                              Subpop. no. obs   =    487,012
                                              Subpop. size      = 243,396,292
                                              Design df         =    488,626
                                              F(   1, 488626)   =    1129.06
                                              Prob > F          =     0.0000

------------------------------------------------------------------------------
             |             Linearized
      allcan | Odds Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ins |   3.276813   .1157438    33.60   0.000     3.057634    3.511704
       _cons |   .0438063   .0015073   -90.91   0.000     .0409495    .0468624
------------------------------------------------------------------------------

. svy, subpop(if instudy2==1): logistic allcan owob
(running logistic on estimation sample)

Survey: Logistic regression

Number of strata   =     1,303                Number of obs     =    465,624
Number of PSUs     =   465,624                Population size   = 231,471,350
                                              Subpop. no. obs   =    462,707
                                              Subpop. size      = 230,186,255
                                              Design df         =    464,321
                                              F(   1, 464321)   =      18.10
                                              Prob > F          =     0.0000

------------------------------------------------------------------------------
             |             Linearized
      allcan | Odds Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        owob |   1.072264   .0175833     4.25   0.000     1.038349    1.107286
       _cons |   .1220595   .0015808  -162.40   0.000     .1190001    .1251975
------------------------------------------------------------------------------
```

```
. svy, subpop(if instudy2==1): logistic allcan veg
(running logistic on estimation sample)

Survey: Logistic regression

Number of strata   =     1,303              Number of obs     =      462,137
Number of PSUs     =   462,137              Population size   =  226,617,701
                                            Subpop. no. obs   =      459,220
                                            Subpop. size      =  225,332,606
                                            Design df         =      460,834
                                            F(   1, 460834)   =       125.15
                                            Prob > F          =       0.0000

------------------------------------------------------------------------------
             |             Linearized
      allcan | Odds Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         veg |   1.241438   .0239998    11.19   0.000      1.19528     1.28938
       _cons |   .1079103   .0018446  -130.25   0.000     .1043548     .111587
------------------------------------------------------------------------------

. svy, subpop(if instudy2==1): logistic allcan fruit
(running logistic on estimation sample)

Survey: Logistic regression

Number of strata   =     1,303              Number of obs     =      463,597
Number of PSUs     =   463,597              Population size   =  227,509,101
                                            Subpop. no. obs   =      460,680
                                            Subpop. size      =  226,224,007
                                            Design df         =      462,294
                                            F(   1, 462294)   =       220.28
                                            Prob > F          =       0.0000

------------------------------------------------------------------------------
             |             Linearized
      allcan | Odds Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
```

```
-------------+----------------------------------------------------------------
       fruit |   1.278157   .0211351    14.84   0.000     1.237397     1.32026
       _cons |   .1092814   .0014406  -167.93   0.000      .106494    .1121418
------------------------------------------------------------------------------
```

```
. svy, subpop(if instudy2==1): logistic allcan exercise
(running logistic on estimation sample)

Survey: Logistic regression

Number of strata   =     1,303                Number of obs     =      458,283
Number of PSUs     =   458,283                Population size   =  224,951,779
                                              Subpop. no. obs   =      455,366
                                              Subpop. size      =  223,666,684
                                              Design df         =      456,980
                                              F(   1, 456980)   =       118.58
                                              Prob > F          =       0.0000

------------------------------------------------------------------------------
             |             Linearized
      allcan | Odds Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    exercise |   .8270357   .0144232   -10.89   0.000     .7992444    .8557934
       _cons |   .1466261   .0021386  -131.63   0.000     .1424938    .1508783
------------------------------------------------------------------------------
```

Running a logistic regression model with all covariates of initial interest to calculate if the %change between the crude adjusted odds is greater than 10%

```
. svy, subpop(if instudy2==1): logistic allcan vet i_age1 i_age2 i_age3 sexo i_eth2 i_eth3 i_eth4
owob smoke ins doc drink veg fruit exercise
(running logistic on estimation sample)

Survey: Logistic regression

Number of strata   =     1,303                Number of obs     =      418,007
Number of PSUs     =   418,007                Population size   =  202,111,896
                                              Subpop. no. obs   =      415,090
                                              Subpop. size      =  200,826,801
```

```
                                              Design df         =    416,704
                                              F(  16, 416689)   =     710.42
                                              Prob > F          =     0.0000

------------------------------------------------------------------------------
             |             Linearized
      allcan | Odds Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         vet |   1.424528   .0363574    13.86   0.000     1.355021    1.497599
      i_age1 |   .0597963   .0029207   -57.67   0.000     .0543374    .0658037
      i_age2 |   .2080847   .0049883   -65.48   0.000      .198534     .218095
      i_age3 |    .474873   .0103306   -34.23   0.000     .4550509    .4955586
        sexo |   1.352575   .0289584    14.11   0.000     1.296992     1.41054
      i_eth2 |   .4091847   .0181529   -20.14   0.000     .3751084    .4463565
      i_eth3 |    .391097     .02006   -18.30   0.000     .3536916    .4324582
      i_eth4 |   .4453362    .028107   -12.82   0.000     .3935184    .5039773
        owob |   .9338124   .0173434    -3.69   0.000     .9004311    .9684312
       smoke |   1.239947   .0228827    11.65   0.000     1.195899    1.285617
         ins |   1.079739   .0453487     1.83   0.068     .9944173    1.172382
         doc |    1.70028   .0624917    14.44   0.000     1.582105    1.827281
       drink |   .9753562   .0402781    -0.60   0.546     .8995227    1.057583
         veg |   1.075746   .0242983     3.23   0.001     1.029161     1.12444
       fruit |    1.05386   .0204133     2.71   0.007     1.014601    1.094639
    exercise |   .9551843    .019292    -2.27   0.023     .9181113    .9937543
       _cons |   .2022448   .0120539   -26.82   0.000     .1799473    .2273052
------------------------------------------------------------------------------
```

Running forward and backwards regressions to see what variables are indicated to stay in the model

```
. sw, pr(0.05): regress allcan vet i_age1 i_age2 i_age3 sexo i_eth2 i_eth3 i_eth4 smoke doc ins fruit veg exercise drink owob
                      begin with full model
p = 0.6179 >= 0.0500  removing drink

      Source |       SS           df       MS      Number of obs   =   415,090
-------------+----------------------------------   F(15, 415074)   =   3268.81
       Model |  6289.68964        15  419.312642   Prob > F        =    0.0000
    Residual |  53244.3076   415,074  .128276663   R-squared       =    0.1056
```

```
-------------+----------------------------------   Adj R-squared   =    0.1056
       Total |  59533.9973   415,089  .143424657   Root MSE        =    .35816


------------------------------------------------------------------------------
      allcan |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         vet |   .0475669   .0018591    25.59   0.000      .043923    .0512108
      i_age1 |  -.2629614   .0018948  -138.78   0.000    -.2666751   -.2592477
      i_age2 |  -.2199299   .0014825  -148.35   0.000    -.2228356   -.2170243
      i_age3 |   -.137405   .0015524   -88.51   0.000    -.1404477   -.1343623
        sexo |   .0162458   .0012657    12.84   0.000     .0137651    .0187266
      i_eth2 |   -.089958   .0021213   -42.41   0.000    -.0941156   -.0858003
      i_eth3 |  -.0719267   .0022211   -32.38   0.000    -.0762801   -.0675734
      i_eth4 |  -.0571924   .0023469   -24.37   0.000    -.0617921   -.0525926
       smoke |   .0194199   .0011459    16.95   0.000     .0171739    .0216659
         doc |   .0377014    .001687    22.35   0.000     .0343949    .0410079
         ins |   .0064125   .0019511     3.29   0.001     .0025883    .0102366
       fruit |   .0078187    .001217     6.42   0.000     .0054333    .0102041
         veg |   .0069759   .0014246     4.90   0.000     .0041837    .0097681
    exercise |  -.0064677   .0012967    -4.99   0.000    -.0090092   -.0039261
        owob |  -.0110992   .0011952    -9.29   0.000    -.0134419   -.0087566
       _cons |   .2635219   .0030384    86.73   0.000     .2575668    .2694769
------------------------------------------------------------------------------


. sw, pe(0.05): regress allcan vet i_age1 i_age2 i_age3 sexo i_eth2 i_eth3 i_eth4 smoke doc ins
fruit veg exercise drink owob

                      begin with empty model
p = 0.0000 <  0.0500  adding  vet
p = 0.0000 <  0.0500  adding  i_age1
p = 0.0000 <  0.0500  adding  i_age2
p = 0.0000 <  0.0500  adding  i_age3
p = 0.0000 <  0.0500  adding  i_eth2
p = 0.0000 <  0.0500  adding  i_eth3
p = 0.0000 <  0.0500  adding  doc
p = 0.0000 <  0.0500  adding  i_eth4
p = 0.0000 <  0.0500  adding  smoke
p = 0.0000 <  0.0500  adding  sexo
p = 0.0000 <  0.0500  adding  owob
```

```
p = 0.0000 <  0.0500  adding  fruit

p = 0.0000 <  0.0500  adding  veg

p = 0.0000 <  0.0500  adding  exercise

p = 0.0010 <  0.0500  adding  ins

      Source |       SS           df       MS      Number of obs   =   415,090
-------------+----------------------------------   F(15, 415074)   =   3268.81
       Model |  6289.68964        15  419.312642   Prob > F        =    0.0000
    Residual |  53244.3076   415,074  .128276663   R-squared       =    0.1056
-------------+----------------------------------   Adj R-squared   =    0.1056
       Total |  59533.9973   415,089  .143424657   Root MSE        =    .35816

------------------------------------------------------------------------------
      allcan |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         vet |   .0475669   .0018591    25.59   0.000      .043923    .0512108
      i_age1 |  -.2629614   .0018948  -138.78   0.000    -.2666751   -.2592477
      i_age2 |  -.2199299   .0014825  -148.35   0.000    -.2228356   -.2170243
      i_age3 |   -.137405   .0015524   -88.51   0.000    -.1404477   -.1343623
      i_eth2 |   -.089958   .0021213   -42.41   0.000    -.0941156   -.0858003
      i_eth3 |  -.0719267   .0022211   -32.38   0.000    -.0762801   -.0675734
         doc |   .0377014    .001687    22.35   0.000     .0343949    .0410079
      i_eth4 |  -.0571924   .0023469   -24.37   0.000    -.0617921   -.0525926
       smoke |   .0194199   .0011459    16.95   0.000     .0171739    .0216659
        sexo |   .0162458   .0012657    12.84   0.000     .0137651    .0187266
        owob |  -.0110992   .0011952    -9.29   0.000    -.0134419   -.0087566
       fruit |   .0078187    .001217     6.42   0.000     .0054333    .0102041
         veg |   .0069759   .0014246     4.90   0.000     .0041837    .0097681
    exercise |  -.0064677   .0012967    -4.99   0.000    -.0090092   -.0039261
         ins |   .0064125   .0019511     3.29   0.001     .0025883    .0102366
       _cons |   .2635219   .0030384    86.73   0.000     .2575668    .2694769
------------------------------------------------------------------------------
```

Obtaining Adjusted Odds Ratios for all can- adjusting for all covariates of interest (age, sex, eth, doc and smoke) - Model 1!

Svy, subpop(if instudy==1):logistic outcome variable can, followed by exposure variable vet, followed by **ALL** covariates of interest to be included in Model

```
. svy, subpop(if instudy2==1): logistic allcan vet i_age1 i_age2 i_age3 sexo i_eth2 i_eth3 i_eth4
smoke doc

(running logistic on estimation sample)



Survey: Logistic regression

Number of strata   =     1,303              Number of obs     =      467,610
Number of PSUs     =   467,610              Population size   =  230,682,516
                                            Subpop. no. obs   =      464,693
                                            Subpop. size      =  229,397,421
                                            Design df         =      466,307
                                            F(  10, 466298)   =      1292.05
                                            Prob > F          =       0.0000

------------------------------------------------------------------------------
             |             Linearized
      allcan | Odds Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         vet |   1.452672     .03575    15.17   0.000     1.384266    1.524458
      i_age1 |   .0608758   .0026998   -63.11   0.000     .0558077    .0664041
      i_age2 |   .2053946   .0046395   -70.07   0.000     .1964997    .2146921
      i_age3 |   .4672019    .009476   -37.52   0.000     .4489935    .4861487
        sexo |   1.382182   .0278869    16.04   0.000     1.328591    1.437935
      i_eth2 |   .3993755   .0165246   -22.18   0.000     .3682662    .4331126
      i_eth3 |   .3803597   .0177594   -20.70   0.000      .347097    .4168101
      i_eth4 |   .4375514   .0257611   -14.04   0.000     .3898648    .4910708
       smoke |   1.253415   .0217118    13.04   0.000     1.211575    1.296701
         doc |   1.724657   .0585102    16.07   0.000     1.613708    1.843233
       _cons |   .2114668   .0080409   -40.86   0.000     .1962799    .2278289
------------------------------------------------------------------------------
```