Why Independent Surveys with the Same Objective Yield Different Estimates

Yumiko Siegfried[1], David Morganstein[1], Andrea Piesse[1], Sharon Lohr[1]
[1]Westat, 1600 Research Blvd, Rockville, MD 20850

Abstract
Independent surveys that measure the same quantity typically produce different estimates. While sampling error is one obvious explanation, there are many non-sampling error sources that could contribute to an even greater variation across different studies. In this paper, the authors compare important tobacco-use estimates from several national surveys. They explore variations in the definitions, timing, instrument wording, interview modes, and other methodological dissimilarities that might contribute to those differences. The authors' objective is to inform users of these tobacco-use estimates about notable differences between the surveys and to show the impact these differences can make on the resulting estimates.

Keywords: Sampling error, non-sampling error, inconsistent estimates

1. Introduction

In this paper, we contrast several national surveys that produce estimates related to tobacco use and explore reasons why they yield different estimates.

All surveys discussed in this paper provide valid national estimates, yet they produce estimates for similar variables that do not agree. There are many potential sources of error, both sampling and non-sampling, that could contribute to observed differences. While the impact of sampling error may be easier to measure, in large surveys, such as the ones included here, major contributors to error may often be from non-sampling sources. Increasing sample size reduces the sampling error and results in narrower confidence intervals but does not reduce non-sampling errors. Thus, differences observed as 'statistically significant' may actually be explainable by a variety of common causes.

The surveys we include are: Population Assessment of Tobacco and Health (PATH) Study; National Adult Tobacco Survey (NATS); National Health Interview Survey (NHIS); National Health Interview Survey Cancer Control Supplement (NHIS-CCS); Tobacco Use Supplements to the Current Population Survey (TUS-CPS); and National Survey on Drug Use and Health (NSDUH). In addition, we mention one widely used estimate from National Health and Nutrition Examination Survey (NHANES).

2. Comparison of Survey Characteristics

2.1 Multiple Objectives

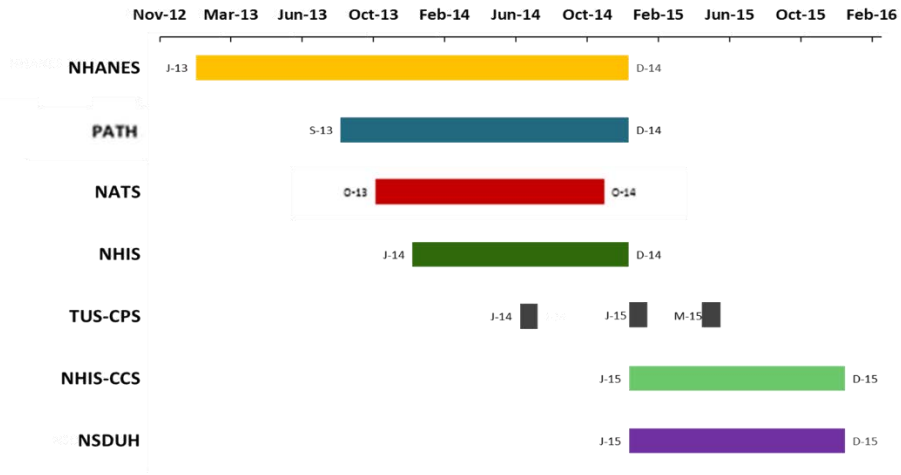
Even though these surveys share a common aim to assess health and tobacco-related behaviors of the population, they have multiple objectives that make them unique. The PATH Study makes longitudinal observations of changes in various aspects of tobacco use and attitudes toward tobacco products. NATS assesses prevalence of tobacco use and the

factors promoting and impeding tobacco use. NHIS collects data on a wide range of health-related topics, not limited to smoking or tobacco. TUS-CPS evaluates and monitors changes in tobacco-related behaviors, environment, and policies. In addition, because TUS is a supplement to CPS, the data from TUS can be linked to labor force data from CPS. NSDUH provides estimates on substance use, including but not limited to tobacco use and mental health issues.

2.2 Fielding Period

Chart 1 describes the fielding period(s) for each survey dataset examined. NHANES covers years 2013 and 2014; Wave 1 of the PATH Study collected data from September 2013 to December 2014; NATS was fielded from October 2013 to October 2014; and TUS-CPS estimates are based on data collected in July 2014, January 2015, and May 2015. NHIS is conducted continuously throughout the year and therefore, NHIS 2014 contains data from January 2014 to December 2014, and NHIS-CCS 2015 and NSDUH contain data from January 2015 to December 2015. In this paper, subsequent references to characteristics and estimates by survey acronym refer to these specific timeframes, unless otherwise noted.

Chart 1: Fielding Periods



2.3 Target Population

Even though there are some small differences, the target population for all these surveys is essentially the noninstitutionalized population residing in the 50 states and the District of Columbia. NATS did not explicitly exclude or include Armed Forces personnel. Some information was collected from Armed Forces personnel in CPS and NHIS, but they were excluded from our analyses so they do not affect the estimates in this paper. Also, some of these surveys collected information from those under 18 years old, but this paper only discusses estimates based on data from adults ages 18 and over.

2.4 Sample and Study Design

Table 1 compares high-level survey characteristics. The PATH Study, NHIS, TUS-CPS, NHIS-CCS, and NSDUH all used stratified multistage household sample designs, and NATS used dual-frame random digit dialing (RDD). In the "Adult respondents" column, the count for TUS-CPS is the total number of self-respondents over 3 months used in analyses and does not include proxy respondents since they were not included in the estimates discussed in this paper. The mode of data collection is not the same for all surveys.

The PATH Study, NHIS, and NSDUH were all done in person; however, NHIS used a computer-assisted personal interviewing (CAPI) system, whereas, the PATH Study and NSDUH used an audio computer-assisted self-interview (ACASI) system for the interviews, which asked about tobacco use. About 64 percent of TUS-CPS interviews were done by phone using a computer-assisted telephone interviewing (CATI) system, and the rest were done in person using CAPI; NATS was conducted entirely by phone using CATI.

Table 1: High-Level Survey Characteristics

Survey	*Sample design*	*Adult respondents*	*Adult response rate*	*Adult interview mode*	*Proxy*
PATH	stratified, multistage, household	32,320	40%*	in-person ACASI	N
NHIS		36,697	59%*	in-person CAPI	Y
TUS-CPS		163,920	48%*+	64% phone CATI, 36% in-person CAPI	Y
NHIS-CCS		33,672	55%*	in-person CAPI	Y
NSDUH		51,118	55%*+	in-person ACASI	N
NATS	dual-frame RDD	75,233	36%	phone CATI	N

*Calculated as household or family response rate times conditional adult response rate.
+Approximate

2.5 Raking and Post-Stratification Dimensions

In general, weighting in all surveys started with base weights, which were the inverse of probabilities of selection. These base weights were adjusted for nonresponse, then post-stratified or raked to known population control totals. Though the general steps are similar, the weighting processes in these adjustment steps varied considerably across the surveys, and this can affect estimates. We will focus on the differences in the raking and post-stratification steps in this paper.

Table 2 shows characteristics used to form the raking or post-stratification adjustment cells, as well as the data sources for the control totals. Note that these characteristics and data sources have been simplified in the table to illustrate major differences, and there may be more differences in details. Some surveys used different sets of characteristics to define household-level and person-level weighting adjustments, so "H" in the table below indicates a characteristic used at the household level and "P" at the person level. "R" marks characteristics used for ratio adjustments prior to post-stratification in applicable surveys. These surveys were benchmarked to different control totals using different dimensions, which may be another factor contributing to observed differences among estimates.

2.6 Other Characteristics

Other differences among surveys are likely to affect the estimates, including nonresponse follow-up protocol, survey context, questionnaires, and definitions. Though this paper does not cover details in these aspects, some of these will be discussed in the next section.

Table 2: Raking and Post-Stratification Dimensions

Characteristics	*PATH*	*NATS*	*TUS-CPS*	*NHIS, NHIS-CCS*	*NSDUH*
Census region and/or division	H, P			R	H, P
CBSA				R	
State		P	R, P		H, P
# Adults in HH	H				
# Non-adults in HH	H				
Race and ethnicity	P	P	R, P	R, P	H, P
Age group	P	P	P	P	H, P
Sex	P	P	P	P	H, P
Education	P	P			
Marital status		P			
House type		P			
Phone category		R, P			
Control Totals	2013 1-year ACS	2013 NHIS, 2013 1-year ACS	Totals estimated based on Census 2010 count		

3. Tobacco-Use Variables

This section compares and contrasts the tobacco-use variables to illustrate specific differences that could be influencing the estimates. The estimates for the tobacco-use variables listed in the following tables were computed from public-use files. In the comparison charts, the surveys are organized in temporal order, and the confidence intervals provided are Wilson intervals.

3.1 Source of Variation in Questionnaires and Definitions

Each survey used its own questionnaire, which varied in length, wording, order of questions, and content of the questions. When looking at tobacco-use questions, there were differences such as the grouping of tobacco products, whether or not there was a filtering question prior to asking about ever use, how the questions were phrased, and who was asked these questions. Because the questions and answer options differed across surveys, it was not always possible to construct variables with exactly the same meaning. Another difference is the use of visual aids in the PATH Study, which might have helped respondents recognize some products like pipes and e-cigarettes.

3.2 Ever Cigarette Use

Ever use of cigarettes was defined in two ways, either “ever in lifetime” or “100 cigarettes in lifetime” (see Table 4). NHIS and TUS did not have “ever in lifetime” questions, so ever use could only be defined using the “100 cigarettes in lifetime” definition. On the other hand, the PATH Study, NATS, and NSDUH defined ever use of cigarettes using “ever one or two puffs” or “part or all of a cigarette.” These three surveys also collected enough information to define the “100 cigarettes in lifetime” version of ever use, and therefore, both estimates are provided in Chart 2 and Table 3. It is obvious that estimates using the

"100 cigarettes in lifetime" definition are all noticeably smaller than estimates using the "ever in lifetime" definition. This is an extreme case, but it is a good example to illustrate the importance of understanding what "ever" means when looking at survey estimates, such as ever use of tobacco.

Chart 2: Ever Cigarette Use (*Estimates for ever in lifetime)

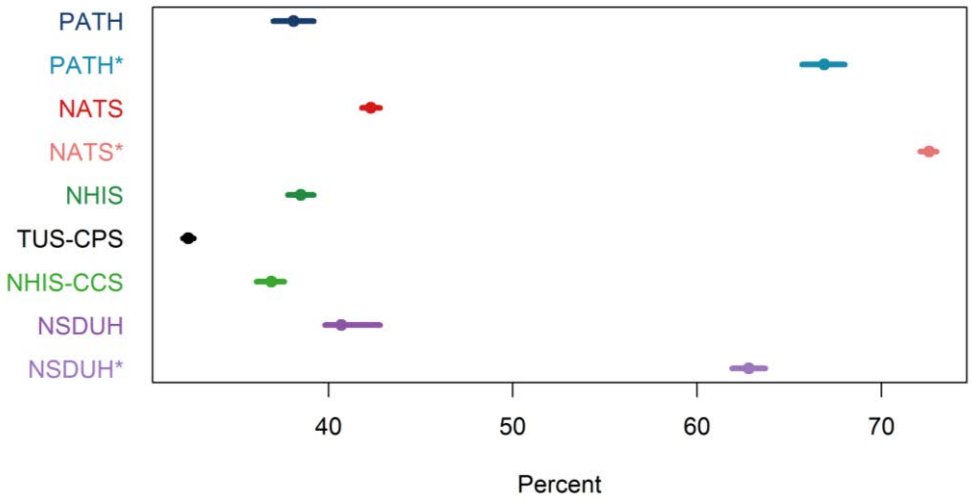


Table 3: Ever Cigarette Use (*Estimates For Ever in Lifetime)

PATH	*PATH**	*NATS*	*NATS**	*NHIS*
38.1%	66.9%	42.3%	72.6%	38.5%
[37.0%, 39.2%]	[65.7%, 68.0%]	[41.8%, 42.8%]	[72.1%, 73.0%]	[37.8%, 39.2%]

Table 3: Ever Cigarette Use (*Estimates For Ever in Lifetime) (continued)

TUS-CPS	*NHIS-CCS*	*NSDUH*	*NSDUH**
32.4%	36.9%	40.7%	62.8%
[32.1%, 32.7%]	[36.1%,37.6%]	[39.8%, 42.8%]	[61.9%, 63.7%]

Table 4: Ever Cigarette Use Definition

	Ever (one puff)	*Ever (100 cigarettes)*
PATH	"Have you ever smoked a cigarette, even one or two puffs?" = "Yes"	"Have you ever smoked a cigarette, even one or two puffs?" = "Yes" and "How many cigarettes have you smoked in your entire life? A pack usually has 20 cigarettes in it." = "At least 100 or more cigarettes (5 packs or more)"

Table 4: Ever Cigarette Use Definition (continued)

NATS	"Have you ever tried cigarette smoking, even one or two puffs?" = "Yes" or "Have you smoked at least 100 cigarettes in your entire life? HELP: 100 CIGARETTES=5 PACKS." = "Yes" or "Do you now smoke cigarettes every day, some days, or not at all?" = "Every day" or "Some days"	"Have you smoked at least 100 cigarettes in your entire life? HELP: 100 CIGARETTES=5 PACKS." = "Yes"
NHIS	NA	"Have you smoked at least 100 cigarettes in your ENTIRE LIFE?" = "Yes"
TUS-CPS	NA	"(Have/Has) (you/name) smoked at least 100 cigarettes in (your/his/her) entire life? 100 cigarettes = approximately 5 packs." = "Yes"
NHIS-CCS	NA	"Have you smoked at least 100 cigarettes in your ENTIRE LIFE?" = "Yes"
NSDUH	"Have you ever smoked part or all of a cigarette?" = "Yes"	("Now think about the past 30 days, that is, from [DATEFILL] up to and including today. During the past 30 days, have you smoked part or all of a cigarette?" = "Yes" and a logical assignment from responses to ("During the past 30 days, that is, since [DATEFILL], on how many days did you smoke part or all of a cigarette?," "What is your best estimate of the number of days you smoked part or all of a cigarette during the past 30 days?," and "On the [CG07 days / CIGDKRE] you smoked cigarettes during the past 30 days, how many cigarettes did you smoke per day, on average?") indicates that a respondent smoked at least 100 cigarettes in the past 30 days) or ("Have you smoked at least 100 cigarettes in your entire life?" = "Yes")

3.3 Current Cigarette Use

Current cigarette use is defined in all but NSDUH as having smoked at least 100 cigarettes and currently smoking cigarettes every day or some days (see Table 6). Chart 3 and Table 5 show that generally these estimates are close, especially in comparison to the sizes of their respective confidence intervals. However, TUS-CPS and NHIS-CCS produced lower estimates than the PATH Study, NATS, NHIS, and NHANES. Noticing that TUS-CPS and NHIS-CCS are more recent than the others, it could be that the decreasing trend in current cigarette smoking is playing a role (CDC, 2016). Note that it is difficult to compare against NSDUH because current use of cigarettes is defined as being use in the past month due to

NSDUH not including the "do you now smoke cigarettes every day, some days, or not at all" question.

Chart 3: Current Cigarette Use (*Estimate uses past 30-day use)

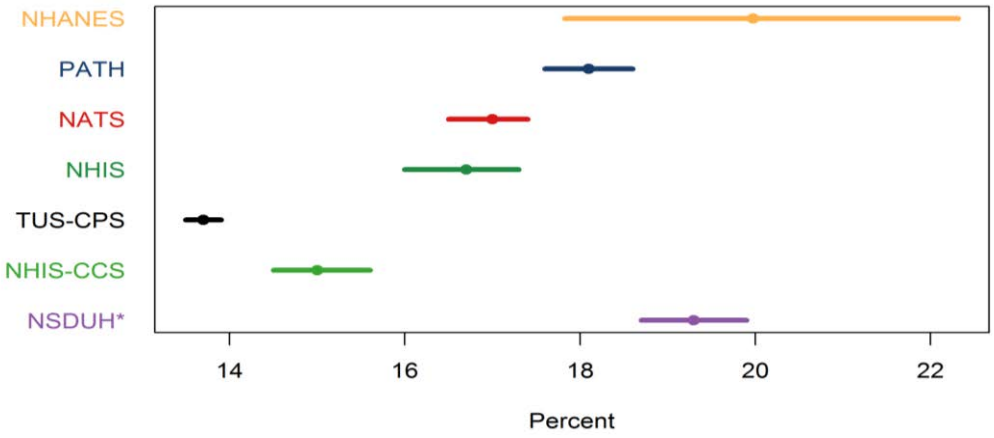


Table 5: Current Cigarette Use (*Estimate uses past 30-day use)

NHANES	*PATH*	*NATS*	*NHIS*	*TUS-CPS*	*NHIS-CCS*	*NSDUH**
20.0%	18.1%	17.0%	16.7%	13.7%	15.0%	19.3%
[17.8%, 22.3%]	[17.6%, 18.6%]	[16.5%, 17.4%]	[16.0%, 17.3%]	[13.5%, 13.9%]	[14.5%, 15.6%]	[18.7%, 19.9%]

Table 6: Current Cigarette Use Definition

NHANES	"These next questions are about cigarette smoking and other tobacco use. {Have you/Has SP} smoked at least 100 cigarettes in {your/his/her} entire life?" = "Yes" and "{Do you/Does SP} now smoke cigarettes?" = "Every day" or "Some days"
PATH	"Have you ever smoked a cigarette, even one or two puffs?" = "Yes" and "How many cigarettes have you smoked in your entire life? A pack usually has 20 cigarettes in it." = "At least 100 or more cigarettes (5 packs or more)" and "Do you now smoke cigarettes…" = "Every day" or "Some days"
NATS	"Have you smoked at least 100 cigarettes in your entire life? HELP: 100 CIGARETTES=5 PACKS." = "Yes" and "Do you now smoke cigarettes every day, some days, or not at all?" = "Every day" or "Some days"
NHIS	"Have you smoked at least 100 cigarettes in your ENTIRE LIFE?" = "Yes" and "Do you NOW smoke cigarettes every day, some days or not at all?" = "Every day" or "Some days"
TUS-CPS	"(Have/Has) (you/name) smoked at least 100 cigarettes in (your/his/her) entire life? 100 cigarettes = approximately 5 packs." = "Yes" and "(Do/Does) (you/name) now smoke cigarettes every day, some days, or not at all?" = "Every day" or "Some days"

Table 6: Current Cigarette Use Definition (continued)

NHIS-CCS	"Have you smoked at least 100 cigarettes in your ENTIRE LIFE?" = "Yes" and "Do you NOW smoke cigarettes every day, some days or not at all?" = "Every day" or "Some days"
NSDUH	("Now think about the past 30 days, that is, from [DATEFILL] up to and including today. During the past 30 days, have you smoked part or all of a cigarette?" = "Yes") and (("Have you smoked at least 100 cigarettes in your entire life?" = "Yes") or (a logical assignment from responses to ("During the past 30 days, that is, since [DATEFILL], on how many days did you smoke part or all of a cigarette?," "What is your best estimate of the number of days you smoked part or all of a cigarette during the past 30 days?," and "On the [CG07 days / CIGDKRE] you smoked cigarettes during the past 30 days, how many cigarettes did you smoke per day, on average?") indicates that a respondent smoked at least 100 cigarettes in the past 30 days))

3.4 Current Pipe Use

Current pipe use is a good example where some surveys combine tobacco products that are asked separately in others, as well as where answer options differ (see Table 8). In Chart 4 and Table 7 below, there is no estimate for NHIS because it did not ask about pipe use specifically. NHIS-CCS did include a question about pipes, but it also included water pipe and hookah use in the same question. NATS included in its current use definition those who reported use of pipes rarely, which was not offered as a response option in the other surveys. Like the item on current cigarette use, NSDUH asked about past month use of pipes instead of use now, every day, or some days. These various differences could be contributing to higher estimates in NATS, NHIS-CCS, and NSDUH.

There were a few differences in the way the PATH Study asked about pipe use. For example, a filtering question (ever seen or heard of a pipe) was asked prior to asking about ever use, and consequently current use, and the study used fairly regular use as a condition in defining the current pipe use variable. In addition to that, visual aids were also used, which might have helped respondents recognize pipe products.

Chart 4: Current Pipe Use

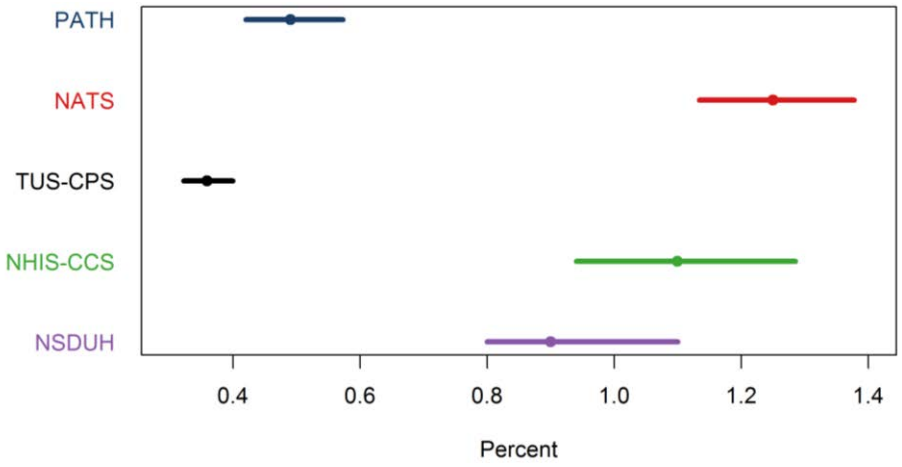


Table 7: Current Pipe Use

PATH	*NATS*	*NHIS*	*TUS-CPS*	*NHIS-CCS*	*NSDUH*
0.5%	1.2%	NA	0.4%	1.1%	0.9%
[0.4%, 0.6%]	[1.1%, 1.4%]		[0.3%, 0.4%]	[0.9%, 1.3%]	[0.8%, 1.1%]

Table 8: Current Pipe Use Definition

PATH	"Have you ever seen or heard of a pipe before this study?" = "Yes" and "Have you ever smoked a pipe filled with tobacco, even one or two puffs?" = "Yes" and "Have you ever smoked a pipe filled with tobacco fairly regularly?" = "Yes" and "Do you now smoke a pipe filled with tobacco…" = "Every day" or "Some days"
NATS	"Have you smoked a regular pipe filled with tobacco at least 50 times in your entire life?" = "Yes"* and "Do you now smoke a regular pipe filled with tobacco every day, some days, rarely, or not at all?" = "Every day" or "Some days" or "Rarely" *Question/criterion only applies to 30+ years old
NHIS	NA
TUS-CPS	"(Have/Has)(you/name) EVER used any of the following EVEN ONE TIME? ... A regular pipe filled with tobacco?" = "Yes" and "(Do you /Does name) NOW smoke a regular pipe every day, some days or not at all?" = "Every day" or "Some days"
NHIS-CCS	"Have you EVER smoked a pipe filled with tobacco-either a regular pipe, water pipe, or hookah EVEN ONE TIME?" = "Yes" and "Do you now smoke pipes filled with tobacco – either regular pipes, water pipes, or hookahs, every day, some days, or not at all?" = "Every day" or "Some days" *Read if necessary: Do not include pipes filled with substances other than tobacco.
NSDUH	"The last questions on tobacco products are about using a pipe to smoke tobacco. Have you ever smoked tobacco in a pipe, even once?" = "Yes" and "During the past 30 days, that is, since [DATEFILL], have you smoked tobacco in a pipe, even once?" = "Yes"

3.5 Cigarette Smoking Cessation Attempt

Cigarette cessation estimates were not available from NSDUH. In all surveys where cessation attempt estimates were available, the numerator of this estimate was defined as "Current smokers who quit for 1 day or longer in the past 12 months PLUS former smokers who completely quit in the past year." The denominator was "Current cigarette smokers PLUS former smokers who completely quit in the past year." The questions used in NHIS and NHIS-CCS were essentially the same, and the estimates are quite close. However, the other surveys showed some variation in how the length of time since quitting was collected and how quit attempt questions were phrased (see Table 10).

Besides the question wording, interview mode differences might be another factor contributing to inconsistent estimates for this outcome variable (see Chart 5 and Table 9). Respondents may be more likely to provide more socially desirable responses in some modes. This may be one of the reasons for a higher estimate in NATS, a telephone survey,

and a lower estimate in the PATH Study, which used in-person ACASI, allowing the respondent more privacy when entering answers (Tourangeau and Yan, 2007).

Chart 5: Cigarette Smoking Cessation Attempt

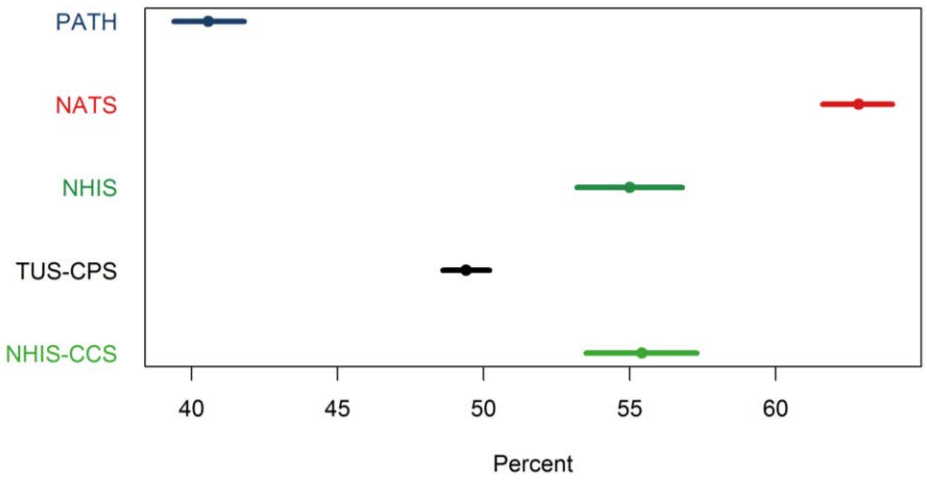


Table 9: Cigarette Smoking Cessation Attempt

PATH	*NATS*	*NHIS*	*TUS-CPS*	*NHIS-CCS*
40.6%	62.8%	55.0%	49.4%	55.4%
[39.4%, 41.8%]	[61.6%, 64.0%]	[53.2%, 56.8%]	[48.6%, 50.2%]	[53.5%, 57.3%]

Table 10: Cigarette Smoking Cessation Attempt Definition

	Numerator	*Denominator*
PATH	("How many cigarettes have you smoked in your entire life? A pack usually has 20 cigarettes in it." = "At least 100 or more cigarettes (5 packs or more)" and "Do you now smoke cigarettes…" = "Every day" or "Some days" and "In the past 12 months, have you stopped [ND FILL3 | tobacco products] for one day or longer because you were trying to quit?" = "Yes") or ("How many cigarettes have you smoked in your entire life? A pack usually has 20 cigarettes in it." = "At least 100 or more cigarettes (5 packs or more)" and "Do you now smoke cigarettes…" = "Not at all" and "About how long has it been since you completely quit smoking cigarettes? (If it was earlier today, enter 1 day)" = (less than or equal to 365 days or less than or equal to 12 months or 1 year))	Numerator + ("How many cigarettes have you smoked in your entire life? A pack usually has 20 cigarettes in it." = "At least 100 or more cigarettes (5 packs or more)" and "Do you now smoke cigarettes…" = "Every day" or "Some days" and "In the past 12 months, have you stopped [ND FILL3 | tobacco products] for one day or longer because you were trying to quit?" not equal to "Yes")

Table 10: Cigarette Smoking Cessation Attempt Definition (continued)

NATS	("Have you smoked at least 100 cigarettes in your entire life? HELP: 100 CIGARETTES=5 PACKS." = "Yes" and "Do you now smoke cigarettes every day, some days, or not at all?" = "Every day" or "Some days" and "During the past 12 months, have you stopped smoking for 24 hours or more because you were trying to quit?" = "Yes") or ("Have you smoked at least 100 cigarettes in your entire life? HELP: 100 CIGARETTES=5 PACKS." = "Yes" and "Do you now smoke cigarettes every day, some days, or not at all?" = "Not at all" and "How long has it been since you completely stopped smoking cigarettes?" = less than or equal to 365 days*) *Question allowed for response in days, weeks, months, and date. Variable used in derivation was converted to days.	Numerator + ("Have you smoked at least 100 cigarettes in your entire life? HELP: 100 CIGARETTES=5 PACKS." = "Yes" and "Do you now smoke cigarettes every day, some days, or not at all?" = "Every day" or "Some days" and "During the past 12 months, have you stopped smoking for 24 hours or more because you were trying to quit?" not equal to "Yes")
NHIS	("Have you smoked at least 100 cigarettes in your ENTIRE LIFE?" = "Yes" and "Do you NOW smoke cigarettes every day, some days or not at all?" = "Every day" or "Some days" and "During the PAST 12 MONTHS, have you stopped smoking for more than one day BECAUSE YOU WERE TRYING TO QUIT SMOKING?" = "Yes") or ("Have you smoked at least 100 cigarettes in your ENTIRE LIFE?" = "Yes" and "Do you NOW smoke cigarettes every day, some days or not at all?" = "Not at all" and ("How long has it been since you quit smoking cigarettes?" = less than or equal to 365 days, less than or equal to 52 weeks, less than or equal to 12 months, or 1 year))	Numerator + ("Have you smoked at least 100 cigarettes in your ENTIRE LIFE?" = "Yes" and "Do you NOW smoke cigarettes every day, some days or not at all?" = "Every day" or "Some days") and "During the PAST 12 MONTHS, have you stopped smoking for more than one day BECAUSE YOU WERE TRYING TO QUIT SMOKING?" not equal to "Yes"
TUS-CPS	Current cigarette smokers who quit for 1 day or longer in the past 12 months and former smokers who completely quit smoking in the past year.	Current cigarette smokers or former smokers who completely quit smoking in the past year.

Table 10: Cigarette Smoking Cessation Attempt Definition (continued)

NHIS-CCS	("Have you smoked at least 100 cigarettes in your ENTIRE LIFE?" = "Yes" and "Do you NOW smoke cigarettes every day, some days or not at all?" = "Every day" or "Some days" and "During the PAST 12 MONTHS, have you stopped smoking for more than one day BECAUSE YOU WERE TRYING TO QUIT SMOKING?" = "Yes") or ("Have you smoked at least 100 cigarettes in your ENTIRE LIFE?" = "Yes" and "Do you NOW smoke cigarettes every day, some days or not at all?" = "Not at all" and ("How long has it been since you quit smoking cigarettes?" = less than or equal to 365 days, less than or equal to 52 weeks, less than or equal to 12 months, or 1 year))	Numerator + ("Have you smoked at least 100 cigarettes in your ENTIRE LIFE?" = "Yes" and "Do you NOW smoke cigarettes every day, some days or not at all?" = "Every day" or "Some days") and "During the PAST 12 MONTHS, have you stopped smoking for more than one day BECAUSE YOU WERE TRYING TO QUIT SMOKING?" not equal to "Yes"

3.6 Cigarette smoking cessation success

The last comparison is for cigarette cessation success. This is one example where all surveys produced very similar estimates (see Chart 6 and Table 11). In general, the cessation success estimate is defined as the number of "former smokers who last smoked 6 months to 1 year ago" divided by the number of "former smokers who last smoked less than or equal to 1 year ago, or current smokers who initiated smoking 2 or more years ago" (see Table 12). There was some variation in the questions, which did not seem to affect the results greatly, except that two sets of NATS estimates were needed to improve comparability due to a skip pattern in one of the questions used in the definition of the denominator. In short, the denominator for "v1" estimates in Chart 6 and Table 11 below is smaller than that produced by the definition used in the other surveys, and the denominator for "v2" estimates is larger. Therefore, the truly comparable denominator would lie somewhere in-between. Either way, the NATS cessation success estimates are in line with those from the other surveys.

Chart 6: Cigarette Smoking Cessation Success

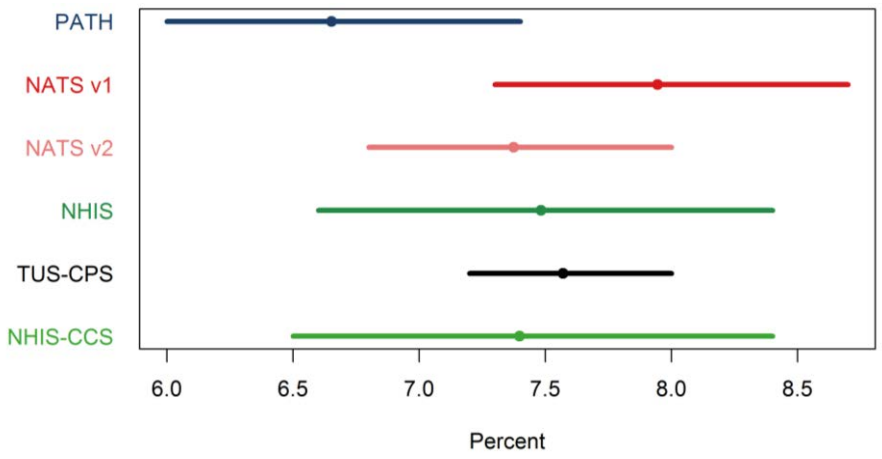


Table 11: Cigarette Smoking Cessation Success

PATH	*NATS (v1)*	*NATS (v2)*	*NHIS*	*TUS-CPS*	*NHIS-CCS*
6.7%	7.9%	7.4%	7.5%	7.6%	7.4%
[6.0%, 7.4%]	[7.3%, 8.7%]	[6.8%, 8.0%]	[6.6%, 8.4%]	[7.2%, 8.0%]	[6.5%, 8.4%]

Table 12: Cigarette Smoking Cessation Success Definition

	Numerator	*Denominator*
PATH	("How many cigarettes have you smoked in your entire life? A pack usually has 20 cigarettes in it." = "At least 100 or more cigarettes (5 packs or more)" and "Do you now smoke cigarettes…" = "Not at all" and "About how long has it been since you completely quit smoking cigarettes? (If it was earlier today, enter 1 day)" is between 182.5 and 365 days, or between 6 and 12 months, or between 0.5 and 1 year (inclusive))	("How many cigarettes have you smoked in your entire life? A pack usually has 20 cigarettes in it." = "At least 100 or more cigarettes (5 packs or more)" and ("Do you now smoke cigarettes…" = "Every day" or "Some days" and ("About how long [have you been smoking | did you smoke] fairly regularly? Do not count any time you may have stayed off cigarettes." is greater than or equal to 2 years, or is 1 year and 12 or more months, or greater than or equal to 24 months) or ("Do you now smoke cigarettes…" = "Not at all" and "About how long has it been since you completely quit smoking cigarettes? (If it was earlier today, enter 1 day)" is less than or equal to 365 days, or less than or equal to 12 months, or less than or equal to 1 year)

Table 12: Cigarette Smoking Cessation Success Definition (continued)

NATS	"Have you smoked at least 100 cigarettes in your entire life? HELP: 100 CIGARETTES=5 PACKS." = "Yes" and "Do you now smoke cigarettes every day, some days, or not at all?" = "Not at all" and "How long has it been since you completely stopped smoking cigarettes?" is between 182.5 and 365 days (inclusive)* *The question allowed for response in days, weeks, months, and date. The variable used in derivation was converted to days.	v1: ("Have you smoked at least 100 cigarettes in your entire life? HELP: 100 CIGARETTES=5 PACKS." = "Yes" and (("Do you now smoke cigarettes every day, some days, or not at all?" = "Every day" or "Some days" and age minus "How old were you when you first started smoking cigarettes every day?" is greater than or equal to 2) or ("Do you now smoke cigarettes every day, some days, or not at all?" = "Not at all" and "How long has it been since you completely stopped smoking cigarettes?" = less than or equal to 365 days*))) *The question allowed for response in days, weeks, months, and date. The variable used in derivation was converted to days. v2: v1 + those who smoked 100 or more cigarettes and smoke every day or some days and has "How old were you when you first started smoking cigarettes every day?" = inapplicable
NHIS	("Have you smoked at least 100 cigarettes in your ENTIRE LIFE?" = "Yes" and ("Do you NOW smoke cigarettes every day, some days or not at all?" = "Not at all" and "How long has it been since you quit smoking cigarettes?" is between 26 to 52 weeks, or between 6 and 12 months, or between 0.5 and 1 year (inclusive)) *Only integer values were possible so the condition is equivalent to 1 year. **There were no cases with the value between 182 and 365 days because only 1-94 days are possible in SMKQTTP = 1.	("Have you smoked at least 100 cigarettes in your ENTIRE LIFE?" = "Yes" and (("Do you NOW smoke cigarettes every day, some days or not at all?" = "Every day" or "Some days" and age minus "How old were you when you FIRST started to smoke fairly regularly?" is greater than or equal to 2) or ("Do you NOW smoke cigarettes every day, some days or not at all?" = "Not at all" and ("How long has it been since you quit smoking cigarettes?" is less than or equal to 365 days, or less than or equal to 52 weeks, or less than or equal to 12 months, or less than or equal to 1 year)))

Table 12: Cigarette Smoking Cessation Success Definition (continued)

TUS-CPS	Number of persons aged 18 years and over who ever smoked 100 cigarettes, who do not smoke now, and last smoked 6 months to 1 year ago.	Number of persons aged 18 years and over who have ever smoked 100 cigarettes, who do not smoke now, and last smoked less than or equal to 1 year ago (quit less than or equal to 1 year/52 weeks/12 months/365 days ago) + current smokers who initiated smoking at least 2 years ago.
NHIS-CCS	("Have you smoked at least 100 cigarettes in your ENTIRE LIFE?" = "Yes" and ("Do you NOW smoke cigarettes every day, some days or not at all?" = "Not at all" and "How long has it been since you quit smoking cigarettes?" is between 26 to 52 weeks, or between 6 and 12 months, or between 0.5 and 1 year (inclusive)) *Only integer values were possible so the condition is equivalent to 1 year. **There were no cases with the value between 182 and 365 days because only 1-94 days are possible in SMKQTTP = 1.	("Have you smoked at least 100 cigarettes in your ENTIRE LIFE?" = "Yes" and (("Do you NOW smoke cigarettes every day, some days or not at all?" = "Every day" or "Some days" and age minus "How old were you when you FIRST started to smoke fairly regularly?" is greater than or equal to 2) or ("Do you NOW smoke cigarettes every day, some days or not at all?" = "Not at all" and ("How long has it been since you quit smoking cigarettes?" is less than or equal to 365 days, or less than or equal to 52 weeks, or less than or equal to 12 months, or less than or equal to 1 year)))

4. Future Research

Through comparing estimates of various tobacco product use, we found several potential sources for differences in these estimates and surveys. However, more research is needed to understand the impact of individual differences on each of the tobacco-related outcomes. For example, it might be informative to examine response patterns across subgroups in more details. It would also be interesting to apply a standardized weighting approach in all surveys to assess the impact of nonresponse adjustment and raking/post-stratification on various estimates. An experimental design such as a split questionnaire could help understand how individual factors in questionnaire design influence responses and, therefore, the resulting estimates.

Another consideration is to develop an error profile to attempt to quantify the impact of all sources of error. Confidence intervals show us the sampling error but do not account for non-sampling error, which likely contributes to differences in estimates across these large surveys.

We also plan to look further at NHANES and at newer releases of data from the surveys examined in this paper to add to the collection of estimates under comparison.

5. Summary and discussion

Due to a variety of factors, it is often the case that different surveys produce estimates that disagree even though the estimates are of "comparable" variables. There are sampling and non-sampling errors in estimates created from survey data, but with large surveys, such as those discussed in this paper, non-sampling error often dominates and may be the main reason for differences. It is difficult to determine the exact cause of difference for each variable, however, because some outcomes may be affected by certain survey characteristics that do not affect other outcomes. For example, in the five tobacco-related outcome comparisons in this paper, the possible reasons for differences varied from temporal trend to mode or question wording differences, to name a few of the major and clear differences. In addition, the effect and the degree of the effect these factors and others less obvious may have depends on the outcome variable.

Based on what we saw from these surveys, there is clearly between-survey variability that needs to be accounted for in general inference. Having multiple and repeated measures of a similar outcome will enable future research to gain better understanding of how survey characteristics affect specific outcomes of interest.

Acknowledgements

The authors and reviewers of this paper consisted of Yumiko Siegfried, David Morganstein, Andrea Piesse, and Sharon Lohr. We would like to extend our gratitude and appreciation to the National Cancer Institute staff involved in the TUS-CPS project.

References

Center for Behavioral Health Statistics and Quality. (2016). *2015 National Survey on Drug Use and Health: Methodological summary and definitions*. Rockville, MD: Substance Abuse and Mental Health Services Administration. https://www.samhsa.gov/data/sites/default/files/NSDUH-MethodSummDefsHTML-2015/NSDUH-MethodSummDefsHTML-2015/NSDUH-MethodSummDefs-2015.htm#intro. Accessed July 18, 2017.

Center for Behavioral Health Statistics and Quality. (2016). *2015 National Survey on Drug Use and Health Public Use File Codebook.* Substance Abuse and Mental Health Services Administration, Rockville, MD. https://www.datafiles.samhsa.gov/study-dataset/national-survey-drug-use-and-health-2015-nsduh-2015-ds0001-nid16894. Accessed July 18, 2017. Public-use data file also retrieved from the same source.

Center for Behavioral Health Statistics and Quality. (2017*). 2015 National Survey on Drug Use and Health: Methodological Resource Book (Section 11, Person-Level Sampling Weight Calibration)*. Substance Abuse and Mental Health Services Administration, Rockville MD. https://www.samhsa.gov/data/sites/default/files/NSDUHmrbSamplingWgt2015.pdf. Accessed July 18, 2017.

Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). National Adult Tobacco Survey (NATS), 2013-2014. Public-use data file and documentation. https://www.cdc.gov/tobacco/data_statistics/surveys/nats/. Downloaded on September 22, 2016.

Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data and Questionnaire. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2013-2014. http://wwwn.cdc.gov/Nchs/Nhanes/Search/DataPage.aspx?Component=Questionnaire&CycleBeginYear=2013. Accessed March 15, 2016.

Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). National Health Interview Survey, 2014. Public-use data file and documentation. https://www.cdc.gov/nchs/nhis/nhis_2014_data_release.htm. Downloaded on September 22, 2016.

Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). National Health Interview Survey Cancer Control Supplement, 2015. Public-use data file and documentation. https://www.cdc.gov/nchs/nhis/nhis_2015_data_release.htm. Downloaded on September 22, 2016.

Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). User guide to National Health Interview Survey, 2014 public use file. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2014/srvydesc.pdf. Accessed September 22, 2016.

Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). User guide to National Health Interview Survey, 2015 public use file. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2015/srvydesc.pdf. Accessed September 22, 2016.

Centers for Disease Control and Prevention (CDC). (2016, March). *Trends in Current Cigarette Smoking Among High School Students and Adults, United States, 1965-2014.* https://www.cdc.gov/tobacco/data_statistics/tables/trends/cig_smoking/index.htm. Accessed September 4, 2017.

National Cancer Institute, Division of Cancer Control and Population Sciences. (2017). "What Is the TUS-CPS?" https://cancercontrol.cancer.gov/brp/tcrb/tus-cps/. Accessed June 29, 2017.

Parsons, V. L., Moriarity, C., Jonas, K., et al. (2014). Design and estimation for the National Health Interview Survey, 2006–2015. National Center for Health Statistics. *Vital Health Stat, 2*(165). https://www.cdc.gov/nchs/data/series/sr_02/sr02_165.pdf. Accessed March 13, 2017.

Piscopo, K., Lipari, R.N., Cooney, J., and Glasheen, C. (2016, September). *Suicidal thoughts and behavior among adults: Results from the 2015 National Survey on Drug Use and Health*. NSDUH Data Review. Retrieved from http://www.samhsa.gov/data/. Accessed July 27, 2017.

Tourangeau, R., and Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859-883. http://www.learnlab.org/research/wiki/images/a/a8/Tourangeau_SensitiveQuestions.pdf. Accessed September 7, 2017.

United States Department of Health and Human Services. National Institutes of Health. National Institute on Drug Abuse, and United States Department of Health and Human Services. Food and Drug Administration. Center for Tobacco Products. *Population Assessment of Tobacco and Health (PATH) Study [United States], Wave 1, Public-Use Files and User_guuide.pdf*. ICPSR36498-v6. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2017-06-14. http://www.icpsr.umich.edu/icpsrweb/NAHDAP/studies/36498. Retrieved on September 26, 2016. Persistent URL: https://doi.org/10.3886/ICPSR36498.v6.

U.S. Census Bureau. (2006, October). *Current Population Survey Design and Methodology Technical Paper 66.* https://www.census.gov/prod/2006pubs/tp-66.pdf. Accessed March 16, 2017.

U.S. Department of Commerce, Census Bureau. (2016). National Cancer Institute and Food and Drug Administration co-sponsored Tobacco Use Supplement to the Current Population Survey. 2014-15. Technical Documentation and Data Files. http://thedataweb.rm.census.gov/ftp/cps_ftp.html#cpssupps. Estimates from TUS-CPS provided by Information Management Services, Inc (IMS).