Prevention Science (2019) 20:246–256
https://doi.org/10.1007/s11121-018-0874-9

Does Marijuana Use at Ages 16–18 Predict Initiation of Daily Cigarette Smoking in Late Adolescence and Early Adulthood? A Propensity Score Analysis of Add Health Data

Trang Quynh Nguyen[1,2] · Cyrus Ebnesajjad[1] · Elizabeth A. Stuart[1,2,3] · Ryan David Kennedy[4] · Renee M. Johnson[1]

Published online: 1 February 2018

Abstract
Given the declining trend in adolescent cigarette smoking and increase in general access to marijuana, it is important to examine whether marijuana use in adolescence is a risk factor for subsequent cigarette smoking in late adolescence and early adulthood. Preliminary evidence from a very small number of studies suggests that marijuana use during adolescence is associated with later smoking; however, to control confounding, previously published studies used regression adjustment, which is susceptible to extrapolation when the confounder distributions differ between adolescent marijuana users and non-users. The current study uses propensity score weighting, a causal inference method not previously used in this area of research, to weight participants based on their estimated probability of exposure given confounders (the propensity score) to balance observed confounders between marijuana users and non-users. The sample consists of participants of Add Health (a nationally representative dataset of youth followed into adulthood) who were 16–18, with no history of daily cigarette smoking at baseline ($n = 2928$ for female and 2731 for male sub-samples). We assessed the effect of adolescent marijuana use (exposure, ascertained at wave 1) on any daily cigarette smoking during the subsequent 13 years (outcome, ascertained at wave 4). Analyses suggest that for females (but not males) who used marijuana in adolescence, marijuana use increased the risk for subsequent daily smoking: OR = 1.71, 95% CI = (1.13, 2.59). We recommend that adolescent marijuana use be viewed as a possible risk factor for subsequent initiation of daily cigarette smoking in women.

Keywords Marijuana/cannabis · Tobacco · Adolescence · Emerging adulthood · Propensity score

Introduction

Despite significant public health achievements in reducing cigarette smoking, smoking continues to be a leading cause of preventable morbidity and mortality in the USA (US Department of Health and Human Services 2014). Smoking contributes to approximately 480,000 deaths in the USA each year. Because adolescence is a vulnerable period for initiation of cigarette smoking (US Department of Health and Human Services 2012), the public health and tobacco control sectors undertook a comprehensive set of actions in the 2000s to discourage youth smoking, including the following: advertising and marketing restrictions; enforcement of minimum-age purchasing requirements; school- and family-based prevention programs; indoor smoking bans; and tax increases on cigarettes (Farrelly et al. 2013; Kostova et al. 2010; Lantz 2000). These actions contributed to substantial declines in adolescent smoking. Cigarette smoking was at its highest in 1976, when 39% of high school seniors reported past 30-day smoking. The prevalence was only slightly lower by 1996 (34%), but declined to 11% by 2015, a historical low (Johnston et al. 2016).

While cigarette smoking has drastically declined among youth (e.g., Johnson et al. 2018), declines in smoking have been more modest among adults (Jamal et al. 2016). From 2005 to 2015, the prevalence of cigarette smoking decreased

✉ Trang Quynh Nguyen
tnguye28@jhu.edu

[1] Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

[2] Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

[3] Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

[4] Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

by 47% (24–13%) among emerging adults (ages 18–24) and by 26% (24%–18%) among early adults (ages 25–44). Eighty percent of all adult smokers report that they began smoking before age 18. Importantly, this means that 20% began smoking during adulthood, signifying that adulthood may have been overlooked as a critical period for initiation (U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration,, and Center for Behavioral Health Statistics and Quality 2014). To guide smoking prevention strategies, it is necessary to identify factors contributing to smoking initiation both in adolescence and in early adulthood. In the present study, we investigate adolescent marijuana use as a risk factor for cigarette smoking in these periods. There are plausible mechanisms that explain how marijuana use might lead to cigarette smoking, as well as preliminary data to support an association.

Adolescent Marijuana Use and Subsequent Cigarette Smoking

It is well-established that marijuana use and cigarette smoking have substantial overlap in terms of lifetime use, concurrent use, and co-use (Agrawal et al. 2011, 2012; Kristman-Valente et al. 2017; Patton et al. 2005). Historically, most adolescents initiate cigarette smoking prior to using other substances; this pattern is the basis for the "gateway theory." However, changes in the prevalence of use of cigarettes and marijuana have given rise to a "reverse gateway" sequence, meaning that marijuana use may precede cigarette smoking (Agrawal et al. 2011, 2012; Badiani et al. 2015; Brook et al. 2015; Fairman et al. 2018; Humfleet and Haas 2004; Patton et al. 2005; Swift et al. 2012; Timberlake et al. 2007). As the lifetime prevalence of cigarette smoking among US high school students declined from 2001 through 2013 (from 63.9 to 41.1%), marijuana use stayed relatively stable (from 42.4 to 40.7%). In fact, 2013 marks the first year that the lifetime use prevalences were equivalent for cigarette smoking and marijuana use (Johnson et al. 2015). As the sequencing of initiation of drugs changes, marijuana may be increasingly more likely to be initiated before cigarette smoking.

There are several possible explanations for a marijuana-then-cigarette sequence of use, with marijuana use starting in adolescence and cigarette use occurring by early adulthood. First, there are likely shared psychosocial and genetic factors that increase risk for both marijuana use and cigarette smoking (Agrawal et al. 2012), and the timing of onset of the two behaviors could be simply idiosyncratic. Second, those with a history of smoking marijuana may be habituated to the sensation of inhaling combusted smoke, and so, the process of initiating cigarette smoking is familiar and could be considered pleasurable (Agrawal et al. 2012; Agrawal and Lynskey 2009; Hu et al. 2006). Third, there may be important disincentives for using marijuana in adulthood given that it is a controlled substance, such as concerns about being a responsible parent or partner, or about drug testing at work. Therefore, cigarettes may be preferable because of their legal status. Based on these theoretical explanations, we expect that adolescent marijuana use will emerge as an independent risk factor for smoking that may continue into adulthood.

A small body of research examined the association between adolescent marijuana use and subsequent cigarette smoking. Several studies found positive associations: in an Australian cohort ($n = 818$), respondents who reported marijuana use (but not cigarette smoking) in adolescence had an increased risk for cigarette smoking initiation in adulthood (Patton et al. 2005). In a New York City cohort ($n = 1332$), respondents who used marijuana from adolescence up to ages 28–30 were more likely to later report tobacco dependence at ages 31–33 (Brook et al. 2015). In a nationally representative US youth cohort (the Add Health Study), Timberlake et al. (2007) found that adolescent marijuana use was positively associated with both daily smoking and nicotine dependence in adulthood in a subset of older participants. A couple of null findings have also been reported, which seem to suffer from methodological limitations. Timberlake et al. did not find the same associations for a younger group of participants; however, it might be the case that this sub-sample was simply too young (18–22 at follow-up) for the effect to accumulate sufficiently to be seen. Mendel et al. (2012) reported no association between adolescent marijuana use and cigarette smoking in adulthood, but their models for smoking adjusted for variables that may be mediators, so the null estimated association is likely to be biased. Using a cross-lag structural equation model to examine the influence of marijuana use and cigarette smoking on each other over time (from age 13 to age 33) in a Seattle sample ($n = 808$), Kristman-Valente et al. (2017) found some evidence of cigarette smoking influencing subsequent marijuana use (but not vice versa) during adolescence, but found evidence of mutual influence in adulthood, with each behavior at 24 predicting the other behavior at 27. Notably, none of these studies examined sex-specific associations, despite the fact that there are sex differences in the determinants and prevalence of both cigarette smoking and marijuana use across the lifecourse (Azofeifa et al. 2016; Jamal et al. 2016; Johnson et al. 2015). In summary, empirical research suggests that marijuana use in adolescence may be a risk factor for subsequent regular cigarette smoking, but the evidence is limited.

Propensity Score Methods and Causal Inference

The relationship between adolescent marijuana use (exposure) and subsequent daily smoking (outcome) is likely confounded by many factors. Most of the above-mentioned studies, appropriately, adjusted for large sets of potential confounders, including demographic variables such as age, sex, race, Hispanic ethnicity, and other psychosocial and contextual factors such as depression, anxiety, anti-social behavior, substance use, parental

marital status, and parental smoking. The method of adjustment used in these studies was multiple regression, while a powerful statistical tool, regression analysis is susceptible to a problem called *extrapolation*, whereby individuals in one group are compared to individuals in another group who are substantially dissimilar. The risk of extrapolation is relevant in studying the effect of adolescent marijuana use on cigarette smoking, because adolescents who use marijuana are likely to be substantially different from those who do not. Extrapolation from one group to another can lead to effect estimates that are very sensitive to the specific regression model used and rely on an assumption that the regression model can accurately extrapolate from one group to the other.

Causal inference methods that explicitly seek to improve the similarity (also called *balance*) between exposed and unexposed groups with respect to the distribution of the potential confounders limit extrapolation (Ho et al. 2007) and thus provide more accurate estimates of the causal effect of the exposure on the outcome. A class of such methods are based on the *propensity score* (PS), i.e., the probability of being exposed given observed confounder values, estimated via a model for the exposure with the confounders as predictors (Rosenbaum and Rubin 1983; Stuart 2010). The PS has the property that conditional on it, the distribution of the observed confounders is similar between the exposed and the unexposed. This means we can use the PS to improve balance on the confounders, using one of several methods: weighting, matching, or subclassification based on the PS; the sample that has been weighted/matched/subclassified to improve balance is then used to estimate the causal effect of the exposure on the outcome. These methods are more appropriate than standard regression analysis when the exposed and the unexposed groups (e.g., marijuana users and non-users) are dissimilar in pre-exposure characteristics.

The other advantage of causal inference methods over regression analysis is the explicit consideration of the target of inference, i.e., the group for which the causal effect is estimated. The three common choices are as follows: (i) the full sample (or the population represented by the full sample), (ii) the exposed group (or those in the population that are exposed), and (iii) the unexposed group (or those in the population that are unexposed). The first choice is relevant when studying a natural exposure that could affect anyone in the sample/population, or an intervention that there is policy interest in administering to everyone. This is not a relevant option for our work because we cannot "administer" marijuana use to all adolescents, and we anticipate that there may be individuals or subgroups of the population that have little to no chance of marijuana use. The second choice is arguably more appropriate, which calls for the estimation of the effect of marijuana use on marijuana users, i.e., how the outcome in marijuana users, on average, is different from the outcomes they would experience had they not used marijuana. This is more generally called the *average causal effect on the exposed* (or the *average treatment effect on the treated* if the exposure is an intervention or treatment). The third type of effect, the effect of marijuana use on non-users—had they used marijuana—is not of practical interest.

Current Study

The purpose of this study is to examine whether marijuana use in adolescence increases risk of subsequently initiating daily cigarette smoking (in late adolescence or in early adulthood)—for those who used marijuana in adolescence but had never smoked cigarettes daily. This is an important research question, because the loosening of restrictions on marijuana policy in concert with declining use of cigarettes in adolescence could mean that, over time, adolescent marijuana use may become an increasing relevant risk factor for cigarette smoking in adulthood. We use the Add Health sample with outcome data ascertained in Wave 4, i.e., longer follow-up than was available to Timberlake et al. (2007). Given the potential for strong confounding when examining the links between marijuana use in adolescence and cigarette smoking in adulthood (as marijuana users and nonusers are different in many respects), we use propensity score (PS) weighting to maximize balance between users and non-users on a wide range of observed confounders and conduct a sensitivity analysis for unobserved confounding. These represent an advance over existing work in this area.

Methods

Study Sample

Data are from the National Longitudinal Study of Adolescent to Adult Health (Add Health), a nationally representative cohort study based in the USA (Harris 2013). Respondents were recruited during the 1994–1995 school year (wave 1) when they were in grades 7–12 and last surveyed in 2007–2008 (wave 4). A wide range of information was collected from respondents, e.g., demographic and socioeconomic factors, health-related behaviors including substance use, and physical and mental health.

The Add Health sampling strategy consisted of a systematic random sample of high schools and "feeder" schools (i.e., middle schools whose students matriculate into the selected high school). A total of 134 schools participated (79% of those sampled). An in-school survey was completed by 90,118 students, and 20,745 students participated in an additional, detailed in-home interview (75.6 and 79.5% of eligible students, respectively). During the in-home interview, 85% of students' parents (usually the primary female caregiver) were also interviewed (85%, $n = 17{,}760$) (Harris 2009, 2013; Resnick et al. 1997). Three subsequent follow-up interviews were conducted, including a wave 4 in-

home interview in 2008–2009. For this study, we used data from the wave 1 in-home interview (including parent survey) and the wave 4 in-home interview.

We restricted the analysis sample to those respondents who at wave 1 were aged 16–18 years (i.e., those who were past their 16th but had not reached their 19th birthday) and had no prior history of daily cigarette smoking. The restriction to no prior daily cigarette smoking permitted us to evaluate the effect of marijuana use on the initiation of daily cigarette smoking. The sample therefore included those without and those with a smoking history but who had never smoked daily. We included only those with data on exposure and outcome, as recommended by White et al. (2011), who argues that imputing such variables only add noise. The analysis sample included 2928 women and 2731 men.

Measures

The exposure variable was *marijuana use* at age 16–18 years and was assessed at the wave 1 interview. We specified marijuana use as having smoked marijuana at least once in the past 30 days AND having used marijuana three or more times in one's lifetime. This definition ensures that those in the marijuana user group were not adolescents who had used marijuana only a couple of times. We will refer to participants with the exposure as "marijuana users" and those without as "non-users".

The outcome variable was *initiation of daily cigarette smoking* between wave 1 and wave 4 (when participants were aged 29–31). We defined daily cigarette smoking as smoking at least one cigarette every day for a 30-day period. This variable was ascertained based on wave 4 reports of having ever been a daily smoker. As the sample includes only those with no daily smoking history at wave 1, we interpret such reports to mean daily smoking was initiated after wave 1. Note that initiation of daily smoking between wave 1 and wave 4 does not necessarily indicate current smoking at wave 4.

Our analyses incorporated a range of covariates, i.e., baseline (wave 1) variables that may confound the exposure-outcome relationship. These include (1) participant characteristics, specifically race, Hispanic ethnicity, age, cigarette smoking history, current tobacco chewing, depressive symptoms, delinquency, history of school suspension, measures of academic achievement; (3) parent/household characteristics, including parent educational attainment, and whether there was a cigarette smoker in the adolescent's home; and (3) daily smoking by best friends. These variables are described in detail in Table 1.

Data Analysis

Data analysis was conducted separately for males and females. The method of analysis was PS weighting before fitting a model to estimate the effect of the exposure on the outcome. This was done after we addressed missing data by multiple imputation. After the main analysis, we conducted a sensitivity analysis to assess how sensitive the estimated effect is to unobserved confounding.

Missing Data Procedures There were missing data for both males and females: math and English grades and vocabulary test score (missing 11.3, 4.0, and 4.7%, respectively), parent education level, and the presence of a cigarette smoker in the home (missing 16.7 and 16.5%); other covariates were missing 0 to 1%. Given the richness of the covariates, we assumed the data to be missing at random and implemented multiple imputation by chained equations using the R package "mice" (van Buuren and Groothuis-Oudshoorn 2011), using linear imputation models for continuous variables and logistic imputation models for categorical variables. Exposure and outcome variables were included in the imputation procedure, as recommended by White et al. (2011). Ten imputations were created each for the male and female samples. The analysis (i.e., PS weighting and weighted outcome modeling—see below) was conducted on each imputed dataset, and estimates were pooled using Rubin's (1987) combining rules.

PS Weighting As the aim of the analysis is to estimate the effect of marijuana use on the group of marijuana users (i.e., comparing the outcomes observed in marijuana users to the outcomes we would expect to observe in that group had they not used marijuana), we want to reweight the marijuana non-users to look similar to marijuana users with respect to the observed covariates. The idea is that after weighting, the outcomes in the weighted non-users are similar to the outcomes we would expect to see in the users had they not used marijuana. The weights used for this purpose are based on an estimated PS model, here the model of marijuana use given the covariates. To compute these weights, we used the R package "twang" (Ridgeway et al. 2015), which employs generalized boosted regression (a nonparametric modeling method) to estimate the PS model. The twang package incorporates the study's sampling weights by multiplying them with PS weights; we refer to the resulting weights as *combined weights*. We checked whether the weighting (using the combined weights) was successful in improving covariate balance between marijuana users and non-users using each covariate's absolute standardized mean difference, defined as the difference in the variable's means between the two groups divided by its standard deviation among the marijuana users. Absolute standardized mean differences larger than 0.2 are considered large and undesirable due to the resulting extrapolation (Stuart 2010).

Outcome Model Estimation After PS Weighting We then fit a model for the outcome in the reweighted sample. This is a

Table 1 Description of the baseline covariates (wave 1) incorporated in the analysis

Covariate	Description
Race	Race data were from five self-identified race items on the survey: White, Black/African American, Asian/Pacific Islander, American Indian/Native American and Other race. For the sample summary (see Results and Table 2) and in the outcome model (see Data Analysis), we used a six-category race variable with the same five categories plus a "multi-racial" category for participants who reported more than one racial identity. In the PS weighting to balance covariates (see Data Analysis), we used three binary variables to represent race: two of the original binary variables (White and Black/African American) plus a third variable combining Asian/Pacific Islander, American Indian/Native American and Other. This collapsing of these identities was less than ideal but needed to avoid sparse data (due to small numbers of marijuana users with these identities) which make weighting challenging.
Hispanic ethnicity	Binary variable.
Age	Continuous age in years, i.e., (interview date – birth date)/365.25
Cigarette smoking history	Ordinal variable with four levels: never tried smoking cigarettes; never smoked a whole cigarette; smoked a whole cigarette for the first time as a teenager (at age 13+); smoked a whole cigarette for the first time before teenage.
Tobacco chewing	Binary variable indicating any tobacco chewing in the past 30 days.
Depressive symptoms score	Sum score on a modified version of the Center for Epidemiology Studies Depression scale (Radloff 1977). This version has 19 instead of 20 items.
Delinquency score	Mean of 15 items reporting involvement in anti-social behavior, such as lying, stealing or fighting, each item on a response scale from 0 = never to 3 = five-or-more-times (Cronbach's alpha = 0.84) (Goodman and Whitaker 2002).
History of school suspension	Binary variable.
Average math grade	Letters grades earned in subject during the past school year, categorical variable: A, B, C, D/F.
Average English grade	Letters grades earned in subject during past school year, categorical variable: A, B, C, D/F.
Vocabulary	Summary score from an abridged version of the Peabody Picture Vocabulary Test—Revised (Dunn and Dunn 1981).
Parental educational attainment	Parent/caregiver education level (from parent survey). The question included 10 levels. The variable was coded to five categories: less than highschool, high school diploma/GED, attended vocational school, attended college and completed college.
Smoker in home	Binary variable indicating whether there is a smoker in the home (parent survey).
Best friend(s) daily smoking	Having at least one of three best friends who smoked daily.

logistic model that includes the exposure as well as the covariates as predictors; this helps adjust for any residual imbalance in these covariates. The regression coefficient of the exposure variable from this model, exponentiated, is the odds ratio (OR) of the outcome associated with marijuana use conditional on the covariates. This model was fit using the R package "survey" (Lumley 2004) to incorporate the study's complex design and the combined weights.

Sensitivity Analysis with a More Detailed Exposure Definition

Note that our definition of marijuana use is broad, inclusive of users with varying frequency of use. For a sensitivity analysis, we split the marijuana use group into frequent users (those who used at least 4 days in the past 30 days) and infrequent users (those who used less frequently). Using nonuse as the comparison condition, we estimated the effect of frequent marijuana use and infrequent marijuana use on subsequent initiation of daily cigarette smoking for the frequent and infrequent users separately and compared results to the main analysis.

Sensitivity Analysis for Unobserved Confounding

Interpretation of the conditional OR as the causal effect of marijuana use relies on the assumption that conditional on the covariates we adjust for, there are no unobserved confounders (Rosenbaum and Rubin 1983). Although we have included many relevant covariates and thus are confident about the estimates, the assumption of no unobserved confounding cannot be tested. We evaluated how sensitive the estimated effect was to unobserved confounding, using the method presented in VanderWeele and Arah (2011). Unobserved confounding was represented conceptually as due to a binary variable U that is independent of the observed covariates and that influences both the exposure and the outcome. We specified U's associations with the exposure and the outcome to be equal to the maximum of the observed covariates' conditional associations with the exposure and the outcome, respectively. We varied the prevalence of U in the unexposed group to assess how the estimated effects of marijuana use changed.

Results

At wave 1, 12.0% of males and 7.1% of females reported marijuana use. The prevalence of subsequent initiation of daily cigarette smoking was higher among those who used marijuana (47.1% of males, 46.9% of females) than among those who did not (32.2% of males, 21.9% of females). Table 2 shows summary statistics for the baseline covariates, by gender and exposure condition (these statistics were all survey-weighted, so they represent the composition of the population that the sample was drawn from). t tests (for continuous

Table 2 Means (SDs) and proportions of covariates, by exposure condition, stratified on sex

	Females			Males		
	Used marijuana		*p* value	Used marijuana		*p* value
	Yes (*n* = 213) Percent[a]	No (*n* = 2715) Percent[a]		Yes (*n* = 322) Percent[a]	No (*n* = 2409) Percent[a]	
Race			.98			.001
White	61.8	63.8		61.9	65.1	
Black	21.0	20.4		19.2	19.0	
Asian	3.6	4.0		1.6	4.4	
Native American	1.4	0.9		2.8	0.3	
Multi-race	4.0	3.7		4.3	3.3	
Other	8.1	7.0		7.5	10.7	
Hispanic ethnicity	13.6	14.5	.90	13.4	13.0	.88
Age in years, mean (SD)	17.4 (.8)	17.4 (.8)	.56	17.5 (.8)	17.4 (.8)	.03
Cigarette smoking history			< .001			< .001
Never tried a cigarette	17.2	52.7		23.1	50.7	
Never smoked a whole cig.	14.6	17.3		14.4	17.3	
1st whole cig. in teenage	59.3	24.8		45.3	22.3	
1st whole cig. before teenage	8.9	5.0		16.7	8.0	
Chewing tobacco	4.0	2.2	.34	19.1	10.5	.001
Depression score, mean (SD)	17.1 (6.6)	16.0 (5.4)	.08	15.6 (5.2)	14.6 (4.6)	.03
Delinquency index, mean (SD)	5.99 (5.39)	2.40 (2.96)	< .001	8.67 (7.25)	3.46 (4.05)	< .001
History of school suspension	26.4	15.2	.01	55.4	29.0	< .001
Average math grade			< .001			< .001
A	14.5	25.2		9.2	24.4	
B	21.4	29.6		26.9	28.6	
C	22.8	22.8		24.7	21.9	
D/F	23.1	9.7		19.2	14.2	
Average English grade			.02			< .001
A	26.2	37.0		13.2	24.5	
B	35.0	36.4		27.9	38.0	
C	23.3	15.4		33.8	23.2	
D/F	9.3	4.4		16.7	9.4	
Vocabulary test, mean (SD)	101 (12)	101 (16)	.89	102 (14)	103 (15)	.51
Parent educational attainment			.45			.57
Less than high school	8.3	14.8		10.6	13.6	
High school diploma/GED	28.8	28.3		27.9	23.9	
Attended vocational school	11.1	7.9		7.8	9.6	
Attended college	14.2	15.1		14.5	15.4	
Completed college	20.8	18.5		20.6	22.6	
Smoker in home	42.3	32.5	.06	41.0	33.2	.01
Best friend(s) smoking daily	64.3	36.3	< .001	70.7	38.9	< .001

SD = standard deviation

[a] Proportions, means, and SDs were all conditional on the column (e.g., among female marijuana users) and estimated using survey weights

p values are from *t* tests for continuous variables and chi-square tests for categorical variables

variables) and chi-square tests (for categorical variables) show that marijuana users and non-users differed significantly on many of the covariates, including smoking history, current tobacco chewing, delinquency score, school suspension history, academic achievement, presence of a smoker in the home, and friends' smoking behavior.

Figure 1 shows how PS weighting improved covariate balance between marijuana users and non-users across the imputed datasets. For both males and females, there was substantial covariate imbalance before PS weighting, especially with respect to smoking history, tobacco chewing, delinquency, school suspension, grades, parent educational attainment,

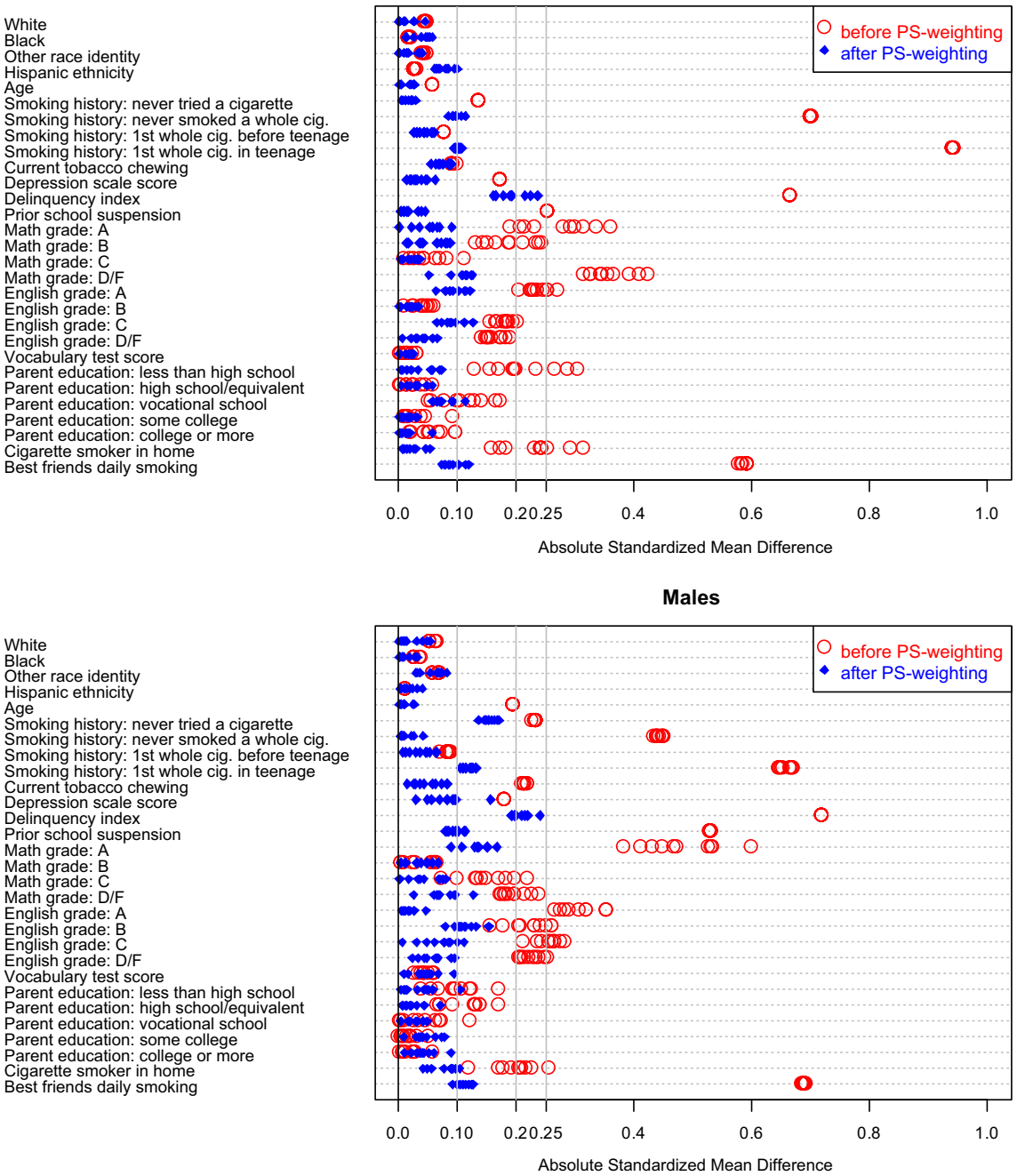

Fig. 1 Covariate balance before and after propensity score (PS) weighting across the 10 imputed datasets for each sample

presence of a smoker in the household, and daily smoking by best friends. After PS weighting, balance was improved on all covariates, as shown in smaller absolute standardized mean differences. In three female and four male imputed datasets, the post-PS-weighting absolute standardized mean difference for delinquency was slightly larger than 0.2; yet, this was a substantial improvement from the pre-PS-weighting value of about 0.7.

After PS weighting, the prevalence of daily cigarette smoking by wave 4 among marijuana non-users was 41.9% for males and 34.3% for females. These are substantially closer to the prevalence observed among marijuana users. The outcome model, fit to the reweighted data, estimated that for female adolescent marijuana users, marijuana use increased the odds of subsequent initiation of daily cigarette smoking by 1.71 times [95% CI = (1.13, 2.59)]. For male adolescent marijuana users, the estimated effect was modest and not statistically significant [OR = 1.13, 95% CI = (0.73, 1.76)].

Results from the sensitivity analysis with the split exposure groups are consistent with results from the main analysis.

Among females, OR = 2.29, 95% CI = (1.06, 4.93), for frequent marijuana users, and OR = 1.69, 95% CI = (1.03, 2.77), for infrequent users. Among males, OR = 1.12, 95% CI = (0.69, 1.83), for frequent users, and OR = 1.20, 95% CI = (0.68, 2.12), for infrequent users. The confidence intervals, as expected, are larger due to the smaller sizes of the exposure groups (89 female and 177 male frequent users and 124 female and 145 male infrequent users).

To assess how sensitive the estimated effect for female marijuana users is to potential unobserved confounding, we first examined the conditional associations of the baseline covariates with exposure and outcome. We found that never-having-tried-a-whole-cigarette (prevalence 31.8% among female marijuana users) was most strongly associated with both the exposure (OR = 0.344) and the outcome (OR = 0.388); this was not surprising as smoking history was expected to be an important confounder. If we assume that an unobserved confounder U had associations with the exposure and outcome of equal strengths to these, with prevalence between 10 and 50% among non-users, then the effect of marijuana use on subsequent daily cigarette smoking varies accordingly as depicted in Fig. 2, which shows that the estimated effect is relatively robust.

Discussion

The prevalence of adolescent cigarette smoking has declined substantially over the past 20 years, while adult smoking has declined at a much slower pace. It is therefore important to look at additional factors—beyond adolescent smoking—that contribute to smoking in adulthood. Adolescent marijuana use is one such factor. We examined the effect of adolescent marijuana use on the risk of subsequently becoming a daily cigarette smoker (over late adolescence and early adulthood) among a nationally representative sample. We used PS-weighting analysis, a method that improves the balance between the exposed and unexposed groups with respect to observed confounders (an improvement over standard regression analysis), which is a more rigorous approach than previous studies.

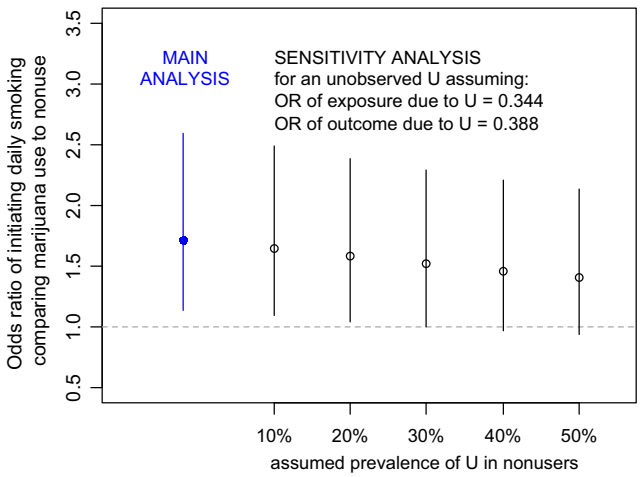

Fig. 2 Sensitivity of the estimated effect—of wave 1 marijuana use on initiating daily cigarette smoking by wave 4—for female marijuana users, to an unobserved confounder U

Our findings show that females who used marijuana in adolescence (but had no prior history of daily cigarette smoking) had an increased risk of initiation of daily smoking by age 29–31 [OR = 1.71, 95% CI = (1.13, 2.59)], a result that is robust to unobserved confounding. Among males, the estimated effect was smaller and not statistically significant [OR = 1.13, 95% CI = (0.73, 1.76)]. While we did not statistically test a hypothesis regarding differences by sex (which would require combined modeling rather than our separate analyses), these different findings are interesting and highlight the importance of stratifying on sex when examining the link between marijuana use and cigarette smoking. It is unclear why we observed an association between adolescent marijuana use and later daily smoking among women, but not men. One possibility is that the marijuana use and cigarette smoking may serve different functions for women compared to men which differentially affects the transition from the regular use of one substance to the other. Tobacco research has found that women are more likely than men to smoke in response to negative feelings or stress (US Department of Health and Human Services 2002). It is possible that some female adolescents were using marijuana to handle stress, and as they became young adults, they were more likely than their male peers (who also used marijuana in adolescence) to transition to using cigarettes to deal with stress. Another possibility is a conjecture based on the theory that as adolescent marijuana users become adults, they may use cigarettes as a substitution for marijuana because cigarettes become more accessible—not only physically but also socially and legally—while marijuana becomes less so; and that this may be reinforced by adult role-related concerns, e.g., about being a responsible parent or partner or about drug testing at work. As women on average marry earlier than men, their earlier involvement in partner, spouse, and parent roles may be a factor encouraging a transition from an illicit to a licit substance.

Finally, there is a third reason that might explain this finding. In this cohort, adolescent marijuana use was assessed in 1994–1995, when rates of both marijuana and cigarette use were substantially lower among women and girls than among men and boys. Marijuana use was uncommon among girls; thus, the girls who used marijuana in this sample may have been more inclined to go against the norm and may be more likely to be smokers in adulthood. This is a theory about unobserved confounding, i.e., even though we tried to create a weighted comparison group of non-users who were similar to the female marijuana users with respect to the observed covariates (demographic variables, psychosocial variables such

as delinquency, depressive symptoms, school suspension, academic achievements, as well as smoking history and parent education and smoking status), the two groups may still have been different in some important confounders leading to differential rates of both exposure and outcome. While the observed delinquency measure is a measure of behavior against the norm, there could potentially be differential measurement error, where those with higher true values (more common among marijuana users than non-users) might have been likely to under-report them. Also, the covariates do not capture non-delinquent types of departure from the norms (e.g., being creative and artistic), which might be associated with substance use. While we cannot rule out unobserved confounding, the sensitivity analysis we conducted suggests that the effect for female adolescent marijuana users is robust to this problem.

Study Limitations We highlight two particular limitations. The first one is the lack of a measure of smoking frequency among study participants who smoked. While the study sample was restricted to those without a daily smoking history, we expect variation with respect to smoking history. Our smoking history variable, however, only provides differentiation at the low end of the spectrum (never having tried a cigarette, having tried but not smoked a whole cigarette, and having smoked a whole cigarette) but not at the high end of the spectrum (e.g., smoking only occasionally vs. near daily smoking). This means smoking frequency among those who smoked (albeit less than daily) was an unmeasured confounder. While we could not adjust for this variable directly, the inclusion of a broad range of covariates in the analysis (tobacco chewing, smoking by friends and family, and various psychosocial, behavioral and academic measures) is likely to help partially adjust for it, because these variables are likely correlated with smoking frequency. With the female sample where a positive effect of adolescent marijuana use on subsequent daily smoking initiation was found, we conducted a sensitivity for unobserved confounding. This sensitivity analysis found that the estimated effect is quite robust to unobserved confounding, which lessened our concern about this unmeasured confounder.

The second limitation is the dated nature of the data. This study's respondents, who were 16–18 in 1994–1995 and 29–31 in 2008, were likely very different from today's youth. We opted to use this sample because Add Health is one of the only large nationally representative samples of adolescents with comprehensive information about behavioral health, and therefore, one of the few studies available to examine our research questions. Also, studying long-term outcomes, such as initiation of daily smoking through late adolescence into adulthood, also requires a long follow-up period, meaning that the measurement of marijuana exposure in adolescence has to go back years in the past. This is a general challenge of the study of any long-term behavior process. Yet extrapolation from the study's finding to the present—how adolescent marijuana smoking today may affect cigarette smoking in the future—requires careful consideration of how the context has changed and is still changing. Two theories mentioned earlier about the transition from marijuana use to tobacco use were as follows: (1) through marijuana smoking, people get used to inhaling smoke, which makes cigarette smoking easier, which leads some people to smoking cigarettes and becoming dependent; and (2) when adolescent marijuana users become adults, they may smoke cigarettes as a substitution for marijuana because cigarettes become more accessible while marijuana is illegal and less accessible. The first theory is a biological mechanism that is likely to continue to operate the same way now as in the past 20 years. Given the current historical low prevalence of adolescent cigarette smoking (Johnston et al. 2016), however, its effect may be amplified, and if this is the main mechanism, marijuana use may be an increasingly important pathway to tobacco addiction. The second theory, on the other hand, may become less and less relevant over time, as with increasing access to and decriminalization of marijuana, the push to substitute marijuana (with tobacco cigarettes) should decrease. Also, now that cigarette smoking is stigmatized and de-normalized, the pull towards it is likely weaker than before. In addition, both theoretical pathways hold only if smoking continues to be the dominant mode of consumption for marijuana. Although youth are increasingly using a diverse set of modes of consumption, the best evidence indicates that 95% of marijuana users smoke marijuana (Johnson et al. 2016). Finally, it is perhaps most important to note that given the limitations of the data, this study does not claim to provide the final answer on whether adolescent marijuana use changes the risk of subsequent initiation of daily cigarette smoking. We recommend that this analysis be conducted on multiple longitudinal datasets that follow participants from adolescence through early adulthood, and be repeated as more recent data become available. We also recommend that as more evidence accumulates over time, the results be interpreted in their entirety, taking into account social and legal changes that are relevant to the use of marijuana and tobacco.

Conclusion This study highlights the need for research, both quantitative and qualitative, to understand marijuana-to-tobacco transition processes and how they change over time. This is crucial for the development of relevant and timely policies and interventions to protect and further the gains in tobacco control. Until a better understanding of these processes is achieved, given the possibility that marijuana may be an increasingly important pathway to tobacco use, we recommend that (1) tobacco control surveillance systems should routinely include marijuana use; (2) the tobacco control community would benefit from deeper integration and

collaboration/co-operation with public health colleagues working to address marijuana use; and (3) smoking prevention interventions could increase scope to include marijuana use.

Funding The current study is supported by the National Institute on Drug Abuse (K01DA031738, Johnson).

Compliance with Ethical Standards

Conflicts of Interest The authors declare that they have no conflict of interest.

Research Involving Human Participants and/or Animals The current project did not involve data collection. We used de-identified data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

Informed Consent Non-applicable, because the current project did not involve data collection.

References

Agrawal, A., Budney, A. J., & Lynskey, M. T. (2012). The co-occurring use and misuse of cannabis and tobacco: A review. *Addiction, 107*(7), 1221–1233.

Agrawal, A., & Lynskey, M. T. (2009). Tobacco and cannabis co-occurrence: Does route of administration matter? *Drug and Alcohol Dependence, 99*(1–3), 240–247.

Agrawal, A., Scherrer, J. F., Lynskey, M. T., Sartor, C. E., Grant, J. D., Haber, J. R., et al. (2011). Patterns of use, sequence of onsets and correlates of tobacco and cannabis. *Addictive Behaviors, 36*(12), 1141–1147.

Azofeifa, A., Mattson, M. E., Schauer, G., McAfee, T., Grant, A., & Lyerla, R. (2016). National estimates of marijuana use and related indicators — National Survey on drug use and health, United States, 2002-2014. *Morbidity and Mortality Weekly Report, 65*(11).

Badiani, A., Boden, J. M., De Pirro, S., Fergusson, D. M., Horwood, L. J., & Harold, G. T. (2015). Tobacco smoking and cannabis use in a longitudinal birth cohort: Evidence of reciprocal causal relationships. *Drug and Alcohol Dependence, 150*, 69–76.

Brook, J. S., Lee, J. Y., & Brook, D. W. (2015). Trajectories of marijuana use beginning in adolescence predict tobacco dependence in adulthood. *Substance Abuse, 36*(4), 470–477.

Dunn, L. M., & Dunn, L. M. (1981). *Manual for the peabody picture vocabulary test-revised*. Circle Pines, MN: American Guidance Service.

Fairman, B. J., Johnson, R. M., & Furr-Holden, C. D. M. (2018). When cannabis is used before tobacco or alcohol: Demographic predictors and associations with heavy use, cannabis use disorder, and other drug outcomes. Prevention Science. Manuscript under review.

Farrelly, M. C., Loomis, B. R., Han, B., Gfroerer, J., Kuiper, N., Couzens, G. L., … Caraballo, R. S.. (2013). A comprehensive examination of the influence of state tobacco control programs and policies on youth smoking. American Journal of Public Health, 103(3), 549–555.

Goodman, E., & Whitaker, R. C. (2002). A prospective study of the role of depression in the development and persistence of adolescent obesity. *Pediatrics, 109*(3), 497–504.

Harris, K. M. (2009). The National Longitudinal Study of adolescent to adult health (add health), waves I & II, 1994-1996; wave III, 2001-2002; wave IV, 2007-2009 [machine-readable data file and documentation]. *Chapel Hill, NC: Carolina Population Center, Univeristy of Carolina at Chapel Hill.* https://doi.org/10.3886/ICPSR27021.v9

Harris, K. M. (2013). The Add Health Study: Design and Accomplishments. Retrieved from www.cpc.unc.edu/projects/addhealth/documentation/guides/DesignPaperWIIV.pdf

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15*(3), 199–236.

Hu, M. C., Davies, M., & Kandel, D. B. (2006). Epidemiology and correlates of daily smoking and nicotine dependence among young adults in the United States. *American Journal of Public Health, 96*(2), 299–308.

Humfleet, G. L., & Haas, A. L. (2004). Is marijuana use becoming a "gateway" to nicotine dependence? *Addiction, 99*, 5–6.

Jamal, A., King, B. A., Neff, L. J., Whitmill, J., Babb, S. D., & Graffunder, C. M. (2016). Current cigarette smoking among adults – United States, 2005-2015. *MMWR. Morbidity and Mortality Weekly Report, 65*(44), 1205–1211.

Johnson, R. M., Brooks-Russell, A., Ma, M., Fairman, B. J., Tolliver, R. L., & Levinson, A. H. (2016). Usual modes of marijuana consumption among high school students in Colorado. *Journal of Studies on Alcohol and Drugs, 77*(4), 580–588.

Johnson, R. M., Fairman, B., Gilreath, T., Xuan, Z., Rothman, E. F., Parnham, T., & Furr-Holden, C. D. M. (2015). Past 15-year trends in adolescent marijuana use: Differences by race/ethnicity and sex. *Drug and Alcohol Dependence, 155*, 8–15.

Johnson, R. M., Fleming, C. B., Cambron, C., Brighthaupt, S.-C., Dean, L. T., & Guttmannova, K. (2018). Race/ethnicity differences in trends of alcohol, cigarette, and marijuana use among adolescents in Washington state, 2004–2014. Manuscript submitted for publication.

Johnston, L. D., Malley, P. M. O., Miech, R. A., Bachman, J. G., & Schulenberg, J. E. (2016). Monitoring the future national survey results on drug use, 1975-2015: Overview, key findings on adolescent drug use. Ann Arbor: Institute for Social Research, The University of Michigan. https://doi.org/10.1017/CBO9781107415324.004.

Kostova, D., Ross, H., Blecher, E., & Markowitz, S. (2010). *Prices and cigarette demand: Evidence from youth tobacco use in developing countries* (NBER working paper series no. 15781). Retrieved from http://www.nber.org/papers/w15781.pdf

Kristman-Valente, A., Hill, K. G., Epstein, M., Kosterman, R., Bailey, J. A., Steeger, C. M., … Hawkins, J. D. (2017). The relationship between marijuana and conventional cigarette smoking behavior from early adolescence to adulthood. Prevention Science. https://doi.org/10.1007/s11121-017-0774-4.

Lantz, P. M. (2000). Investing in youth tobacco control: A review of smoking prevention and control strategies. *Tobacco Control, 9*(1), 47–63.

Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software, 9*(8), 1–19.

Mendel, J. R., Berg, C. J., Windle, R. C., & Windle, M. (2012). Predicting young adulthood smoking among adolescent smokers and nonsmokers. *American Journal of Health Behavior, 36*(4), 542–554.

Patton, G. C., Coffey, C., Carlin, J. B., Sawyer, S. M., & Lynskey, M. (2005). Reverse gateways? Frequent cannabis use as a predictor of tobacco initiation and nicotine dependence. *Addiction, 100*(10), 1518–1525.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*(3), 385–401.

Resnick, M. D., Bearman, P. S., Wm Blum, R., Bauman, K. E., Harris, K. M., Jones, J., et al. (1997). Protecting adolescents from harm

findings from the National Longitudinal Study on adolescent health. *JAMA, 278*(10), 823–832.

Ridgeway, G., McCaffrey, D. F., Morral, A., Griffin, B. A., & Burgette, L. (2015). Twang: Toolkit for weighting and analysis of nonequivalent groups. Retrieved from https://cran.r-project.org/web/packages/twang/vignettes/twang.pdf

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons Inc..

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*(1), 1–21.

Swift, W., Coffey, C., Degenhardt, L., Carlin, J. B., Romaniuk, H., & Patton, G. C. (2012). Cannabis and progression to other substance use in young adults: Findings from a 13-year prospective population-based study. *Journal of Epidemiology and Community Health, 66*(7), e26.

Timberlake, D. S., Haberstick, B. C., Hopfer, C. J., Bricker, J., Sakai, J. T., Lessem, J. M., & Hewitt, J. K. (2007). Progression from marijuana use to daily smoking and nicotine dependence in a national sample of U.S. adolescents. *Drug and Alcohol Dependence, 88*(2–3), 272–281.

U.S. Department of Health and Human Services. (2012). *Preventing tobacco use among youth and young adults a report of the surgeon general executive summary*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

U.S. Department of Health and Human Services. (2014). *The health consequences of smoking—50 years of progress: A report of the surgeon general*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health http://doi.org/NBK179276

U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, & Center for Behavioral Health Statistics and Quality. (2014). National Survey on drug use and health. Retrieved from https://doi.org/10.3886/ICPSR36361.v1.

US Department of Health and Human Services. (2002). *Women and smoking: A report of the Surgeon General.*

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3). https://doi.org/10.18637/jss.v045.i03

VanderWeele, T. J., & Arah, O. a. (2011). Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology, 22*(1), 42–52.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine, 30*(4), 377–399.