Original Investigation

Epidemiological evidence on count processes in the formation of tobacco dependence

David A. Barondess, Ph.D., Emily M. Meyer, Ph.D., Prashanthi M. Boinapally, M.D., Brian Fairman, B.S., & James C. Anthony, Ph.D.

Department of Epidemiology, College of Human Medicine, Michigan State University, East Lansing, MI

Corresponding Author: David A. Barondess, Ph.D., Department of Epidemiology, College of Human Medicine, B601 West Fee Hall, East Lansing, MI 48824, USA. Telephone: 517-353-8623; Fax: 517-432-1130; E-mail: barondes@msu.edu

Received December 10, 2009; accepted April 9, 2010

Abstract

Introduction: Large-sample epidemiological studies of tobacco cigarette smoking routinely assess so-called “lifetime prevalence” of tobacco dependence. This work delves into the earliest stages of smoking involvement, focusing on newly incident tobacco cigarette smokers in the very recent past, and examines hypothesized subgroup variation in count processes that become engaged once smoking starts. Here, the term “count process” has two components: (a) whether smoking will be persistent and (b) the rate of smoking, conditional upon membership in a latent class of smokers who will persist, as estimated under the zero-inflated Poisson (ZIP) model for complex survey data.

Methods: We estimate these ZIP parameters for nationally representative samples of newly incident smokers in the United States (all with smoking initiation within 24 months of assessment). Data are from the 2004–2007 National Surveys on Drug Use and Health.

Results: Once cigarette smoking started, roughly 40%–45% persisted, and the estimated median rate was five smoking days/30 days, conditional on membership in the latent class of persistent smokers. Among non-Hispanic recent-onset cigarette smokers, Whites, Black/African Americans, Asians, and Native American/Alaskan Natives did not differ, but recent-onset smokers of Hispanic origin and those of Pacific Islander background had comparatively less cigarette involvement.

Discussion: Tobacco prevention and control initiatives may require elaboration in the form of brief interventions, including interpersonal and social transactions that might constrain a mounting frequency of days of smoking before daily smoking starts, and until conventional smoking cessation medication aids become indicated. These very-early stage interventions (VESI) might be mounted within family or peer groups or in the primary care or school settings, but randomized trials to evaluate VESI interventions will be required.

doi: 10.1093/ntr/ntq073

Advance Access published on May 27, 2010

© The Author 2010. Published by Oxford University Press on behalf of the Society for Research on Nicotine and Tobacco. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org

Introduction

Despite years of epidemiological research on tobacco cigarette smoking, smoking remains the single most preventable cause of death worldwide, resulting in more than 5 million deaths per annum (World Health Organization, 2008). According to revised estimates from the 1990 Global Burden of Disease (GBD) project, by 2015, more individuals will die from tobacco-related diseases than from HIV/AIDS (Mathers & Loncar, 2006). In the United States, approximately 440,000 deaths per year, including approximately 82% of lung cancer deaths, are attributable to smoking (Center for Disease Control Prevention, 2008). The GBD projections do not count indirect contributions of tobacco to patterns of unhealthy behavior that, in turn, are linked to significant mortality and morbidity (Ezzati et al., 2002). For example, prior use of tobacco has been linked to later more serious drug use involvement. Smokers are more likely to have the opportunity to try cannabis compared with nonusers (Wagner & Anthony, 2002). Tobacco is one of several drug compounds linked to excess risk of cannabis dependence after the first cannabis experience (Chen, O'Brien, & Anthony, 2005).

It is the early count process that is under study in this project, with a focus on the first months and year, after onset of smoking. For the most part, tobacco-attributable diseases and unhealthy behavior can be traced back to the formation of tobacco or nicotine dependence syndromes, which sustain or propel a count process of accumulating “pack-years” of cigarette smoking. Nonetheless, in the earliest stages of smoking, there is no dependence, and the reverse is true. That is, from the first cigarette onward, it is a count process that sustains or drives the formation of the dependence syndrome, as with other psychoactive drugs (e.g., see Anthony, 2010; Koob & Le Moal, 1997). In this report, we estimate components of this count process using a zero-inflated Poisson (ZIP) model, in a novel application of the ZIP model to the early stages of smoking involvement. It is our intent to initiate this line of research with nationally representative survey sample data on newly incident tobacco smokers who have started to smoke within 24 months of

the date of cross-sectional assessment and to focus the inquiry on an observation interval defined by the number of smoking days in the 30 days prior to assessment.

Readers not familiar with the ZIP model may require clarification of what is meant by the term “count process.” Here, the first component involves whether smoking will be persistent once it starts. For example, some who smoke a cigarette never try another cigarette, ever. For other newly incident smokers, the first smoking experience serves a reinforcing function such that the probability of smoking the next cigarette is very high and the elapsed time from first to second cigarette is quite short. During any specified weekly or monthly interval of time after onset of smoking, nonpersistent smoking can be measured by the number of days of smoking during that interval, but the newly incident smokers with zero days of smoking during that interval can be sorted as members of two latent classes: (a) a latent class of newly incident smokers who never will smoke again and (b) a latent class of newly incident smokers who will persist in smoking, despite having not smoked in the interval under observation. In epidemiological research on subgroups of the population, we can use the ZIP model to estimate the mean probability or odds of being a member of these two classes, and subgroup variation in the odds is characterized by the “persistence parameter (PP)” of the ZIP model. The existence of the second latent class is posited on the basis of the excess number of zeroes in the count distribution for days of smoking in the interval under observation.

The second component of the count process involves the rate of smoking during the interval of observation, conditional on membership in the latent class of newly incident smokers who will persist in smoking, despite not having smoked in the interval under observation. Here, we can use the ZIP model to estimate subgroup variation in the rate of smoking during the interval of observation and form rate ratios (RR).

Taken together, these two components characterize the count process of very-early stage smoking involvement, which can be studied in complement with the probability of becoming tobacco or nicotine dependent, as explained recently by Anthony (2010). Indeed, in the earliest stages of smoking involvement, the emergence of a dependence syndrome will vary as a function of the count process components; its probability will be driven upward by these components of the count process. If a newly incident smoker tries one cigarette and never smokes again, dependence will not emerge. It is in this sense that the probability of dependence is influenced by the count process. Nonetheless, as dependence syndromes begin to form, there is a feedback loop and the development of the dependence syndrome begins to drive the count process (Rose, Dierker, & Donny, 2010). The feedback loop represents a violation of standard multiple regression models and the study of dose-response relationships in toxicology and pharmacological research. For this reason, it is a mistake to plot the probability, prevalence, or odds of being tobacco dependent as a function of the elements of the count process (i.e., a mistake in the form of a model misspecification). Also, it would be a mistake to plot the count process estimates as a function of whether a smoker had developed tobacco or nicotine dependence. Instead, if we are to examine these outcomes of newly incident smoking, it will be necessary to have longitudinal data, modeled in a multivariate fashion that includes the feedback loops.

With respect to tobacco smoking in general and nicotine dependence in particular, previously conducted large-sample epidemiological studies have routinely examined the sociocultural milieu of past drug encounters—the so-called “quality and quantity” of tobacco-related experiences (Brook, Saar, Zhang, & Brook, 2009). Here, we hypothesize that once smoking starts, persistence in tobacco involvement will vary across ethnic groups, as might the rates of smoking. As such, this study extends our group’s previous work on the earliest stages of tobacco dependence (Storr, Reboussin, & Anthony, 2004).

In these initial models, we posit a small set of subgroup variations, with subgroups defined in terms of time-fixed, or relatively time-invariant variables (e.g., sex, race/ethnicity), that will need to be taken into account as potential confounders when the epidemiological research shifts toward time-varying characteristics (e.g., years of schooling, socioeconomic position). When subgroup variation is observed for either count process parameter (in relation to the potentially confounding variables), we also explore whether this variation might be attributed to imbalances in the calendar year of first cigarette smoking and the age/age-of-onset nexus. We note that adolescent-onset smoking has been associated with higher lifetime prevalence of sustained smoking and nicotine dependence, but as noted by Breslau and Peterson (1996), this relationship appears to be a function of elapsed time from smoking onset to the date of assessment. Here, we constrain elapsed time to be a value less than 24 months such that the expectation is for no excess risk for adolescent-onset smokers.

A few words about this age/age-of-onset nexus may be in order. Because we have restricted the sample to recent-onset smokers, the age of the smoker on the date of survey assessment is tightly correlated with the age-of-onset for smoking (i.e., correlation >.9). As such, multicollinearity thwarts concurrent regression modeling of both variables. This is what we mean by the age/age-of-onset nexus in this context.

Methods

This project’s survey data on newly incident tobacco smokers come from the 2004–2007 National Surveys on Drug Use and Health (NSDUH), which are cross-sectional sample surveys conducted each year in the United States, in a fashion that allows borrowing of survey data across multiple years (e.g., see O’Brien & Anthony, 2005). The NSDUH collects information on use of alcohol, tobacco, and other drugs and related issues, including drug dependence in the recent past, mental health, and general health concerns. Detailed NSDUH methods descriptions can be found in readily accessible online documentation (e.g., <http://www.oas.samhsa.gov/NSDUH/Methods.cfm>, last accessed 18 November, 2009).

In brief, the NSDUH study population is derived from non-institutionalized community-dwelling U.S. residents aged 12 years and older. As such, the sampling approach involves multistage area probability sampling of noninstitutional dwelling units (DUs), including homeless shelters and other group quarters. After sampling and recruitment of eligible designated respondents (DRs) within each sampled DU, data primarily are collected via an audio computer-assisted self-interview method, in or near the DR’s home, with the DR keying answers directly

into a laptop computer to maintain confidentiality of responses. The NSDUH survey participation level is measured in relation to a weighted response rate that is based on the DR who completed the assessment within 90 days after release of the assessment for fieldwork. Between 2004 and 2007, the weighted screening rate ranged from 89% to 91%, while the weighted interview response rate ranged from 74% to 77%. The study protocols for data gathering and analyses were reviewed and approved by the cognizant institutional review boards for protection of human subjects in research.

The focus of this study is on those DRs whose confidential survey responses indicated onset of tobacco cigarette use within 24 months prior to the date of assessment. Accordingly, we excluded the following DRs from analysis: (a) 97,015 DRs who had never smoked cigarettes and (b) 116,390 DRs with onset of tobacco smoking more than 24 months prior to assessment. In the aggregate sample, there were 8,816 recent-onset cigarette users (ROCU), with more than 99% answering NSDUH questions about persistence of tobacco use ($n = 8,698$); 118 who did not answer were excluded from the analyses. The remaining respondents are those who had begun to use tobacco within 24 months of assessment and about whom there was a measurable count process for past 30-day use, with all others treated as a right-or-left censored subpopulation for variance estimation purposes (see [StataCorp, 2007](#), with reference to the “svy subpop” command structure). Figure 1 provides a flowchart depicting the complete identification process for ROCUs in the four NSDUH samples, years 2004–2007.

The key response variable in this research is the number of cigarette smoking days in the 30-day interval just prior to the date of assessment, which is one manifestation of the count process of accumulating tobacco exposure among the ROCUs. As such, this count process is expected to depend upon sex, race/ethnicity, and age group on date of assessment, which closely reflects age-of-onset of tobacco use in a sample of newly incident tobacco users. Calendar year of first use, with an observed range of 2002–2007, also has been examined as a control variable, although no year-by-year variation was anticipated.

Figure 1. Process for identifying recent-onset cigarette users (ROCU). NSDUH = National Surveys on Drug Use and Health.

The plan for data analysis was organized in relation to standard “explore, analyze and estimate, explore” cycles, in which the first exploratory steps involve Tukey-style box-and-whisker plots and other exploratory analyses to shed light on the underlying distributions of each response variable and covariate of interest, and which motivated us to investigate methods for zero-inflated counts. In the initial analysis/estimation step, the task was to estimate regression coefficients for each suspected determinant of the count process, for which the statistical approach was a form of the generalized linear model known as ZIP count regression. This ZIP regression model yields estimates for two parameters that reflect (a) whether tobacco use had persisted until the month of assessment and (b) the rate of tobacco use per 30 days (in that month), conditional on the persistent use outcome ([Hilbe, 1999](#)). In the PP, we have a comparison of subgroups with respect to the log odds of having become a nonsmoker of cigarettes by the date of survey assessment, analogous to a contrast between two latent classes of ROCUs: (a) ROCUs with actualized or potential persistence of tobacco use into the 30-day observation interval and (b) ROCUs who had entered a class of “Subsequently Always Nonsmokers” by that time (i.e., they had stopped smoking by the time of the 30-day interval). This PP is gauged in relation to a regression slope that is null at 0.0; a statistically robust negative sign on the PP for a subgroup indicates an inverse association and an overrepresentation of that subgroup in the “Subsequently Always Nonsmokers” latent class, with an underrepresentation in the other latent class, which encompasses those with nonzero smoking days as well as those who remain “Subsequently Potential Smokers.”

In the other parameter, we have a comparison of subgroups with respect to the rate of tobacco use in the 30 days prior to assessment, given the persistence outcome. Exponentiation of this ZIP slope parameter estimate yields an RR, conditional on the persistence outcome, such that a statistically robust RR is inverse when its value is less than 1.0 (e.g., similar in metric to the odds ratio or other RR estimates). In these analyses, due attention has been given to (a) weighting and sample selection probabilities, (b) the multistage nested structure of the sample, and (c) variance estimation via Stata “svyset” and “subpopulation” approaches with Taylor series linearization ([StataCorp, 2007](#)). The statistical approach extends the count regression model to include covariates.

In this work, we stress precision of the study estimates with a focus on 95% CIs. The p values are also presented as an aid to interpretation. We note that the generally large subgroup sample sizes are conducive to small p values, although some subgroups under study are small, perhaps too small to be considered except on a very exploratory basis (e.g., Native Hawaiians/Other Pacific Islanders with $n = 32$ ROCUs in this sample). That is, for most subgroup contrasts, the statistical power to detect small variations is quite robust, given the subgroup cell sizes.

Results

Table 1 describes the study sample, which reveals that the ROCUs in the combined 2004–2007 NSDUH samples were predominantly male (52%), adolescent-onset smokers aged 12–17 years old (65%), with nearly two-thirds non-Hispanic White (nHW, 65%). Before presentation of the ZIP regression

Table 1. Selected characteristics of recent-onset cigarette users ($n = 8,816$). Data from the National Surveys on Drug Use and Health, United States, 2004–2007

Sample characteristics	Unweighted number of persons	Unweighted %
Sex		
Female	4,229	48
Male	4,587	52
Age group (age of onset) ^a , years		
12–17	5,700	65
18–25	3,035	34
26–34	56	<1
35+	25	<1
Race/ethnicity		
Non-Hispanic (nH) White	5,713	65
nH Black	987	11
nH Native American	148	2
nH Native Hawaiian/Other Pacific Islanders	32	<1
nH Asian Americans	234	3
nH More than one race/ethnicity	285	3
Hispanic	1,417	16
Year of first use		
2002–2003	1,842	21
2004	2,225	25
2005	2,270	26
2006–2007	2,479	28

^aAge of smoking onset (ASO) was the same value as age of assessment or was 1 year different (i.e., ASO = Age or ASO = Age – 1).

parameter estimates, it is worth mentioning that an estimated 42% of the ROCUs from the NSDUH 2007 sample had smoked on at least one of the days within the 30-day interval prior to the date of assessment (95% *CI* = 37%–48%). Among the ROCUs whose cigarette smoking occurred during that 30-day interval, the estimated median number of days of smoking during those 30 days was just 5 days. Corresponding estimates for the 2006 NSDUH were 41% (95% *CI* = 38%–44%) and median of 5 days; for the 2005 NSDUH, 42% (95% *CI* = 36%–48%) and median of 5 days; and for the 2004 NSDUH, 42% (95% *CI* = 36%–49%) and median of 5 days (data for these individual years are not presented in table format). These estimated proportions are from appropriately weighted survey data; the 95% *CI*s are based on the Taylor series linearization variance estimation method for complex survey data. It follows that in the United States, an estimated two of five ROCUs demonstrably persisted in smoking beyond the occasion of first smoking.

As shown in Table 2, before covariate adjustment, the ZIP regression estimates are consistent with no male–female differences in the persistence parameter (uPP) or in the rate of cigarette use (uRR), conditional on the persistence outcome ($p = .3$, $p = .1$, respectively). With respect to age and age-of-onset (with 12- to 17-year olds as the reference group), also before covariate adjustment, the PP is statistically robust and is inverse for the 18- to 25-year-old subgroup (uPP = -0.4 , 95% *CI* = -0.5 to -0.2 , $p < .001$). With respect to the ZIP results involving race/ethnicity, prior to covariate adjustment and with the nHW ROCUs as the reference group, Hispanics had a positive (unadjusted) persistence parameter (uPP = 0.2 , 95% *CI* = 0.03 – 0.4 , $p = .033$). Recalling that the ZIP uPP indicates the likelihood of belonging to the latent class of Subsequently Always Non-

smokers, we can say that those in the 18- to 25-year-old age group of ROCUs were less likely than the adolescent-onset smokers to have become members of the Subsequently Always Nonsmokers latent class (uPP = -0.4 , 95% *CI* = -0.5 to -0.2 , $p < .001$). Thus, among the newly incident recent-onset smokers, the 18- to 25-year olds actually were more likely to persist in tobacco smoking compared with adolescent-onset counterparts. Turning to the race/ethnicity issue, Hispanic ROCUs were more likely to become members of the Subsequently Always Nonsmokers latent class, as shown by a positive sign on the uPP estimate (uPP = 0.2 , 95% *CI* = 0.03 – 0.4 , $p = .033$); they were thus less likely to persist in smoking. With covariate adjustment for sex, age, and year of first cigarette use, there were no changes in these parameter estimates (i.e., uPP = aPP = -0.4 for 18- to 25-year olds and uPP = aPP = 0.2 for Hispanics; see Tables 2 and 3 for corresponding *CI*s and p values).

Turning to the ZIP RR estimates, for which a statistically null value is 1.0 and inverse associations have values under 1.0, we found that conditional on the persistence outcome, among non-Hispanics, the Native Hawaiian/Other Pacific Islanders (NHOP) ROCUs experienced a lower rate of cigarette tobacco use per 30 days prior to assessment compared with the White ROCUs reference subgroup (uRR = 0.5 , 95% *CI* = 0.2 – 0.8 , $p < .001$). After adjustment for sex, age, and year of first smoke covariates, the adjusted rate ratio (aRR) and p values remained nearly identical to those in the unadjusted model and the aRR estimate remained inverse relative to the null value of 1.0 (aRR = 0.4 , 95% *CI* = 0.2 – 0.8 ; $p < .001$). The p value is small, but there is likely some degree of statistical fragility here, due to the small number of ROCUs in this NHOP race/ethnicity subgroup.

Table 2. Estimated bivariate associations linking subgroup membership with zero-inflated Poisson regression parameters

Past 30-day use	uPP ^a	95% CI	<i>p</i>	uRR ^b	95% CI	<i>p</i>
Sex						
Male	−0.1	−0.2 to 0.1	.320	1.1	1.0–1.2	.138
Female (ref)	0	—	—	1.0	—	—
Age group (years)						
12–17 (ref)	0	—	—	1.0	—	—
18–25	−0.4	−0.5 to −0.2	<.001	1.3	1.2–1.5	<.001
26–34	−0.1	−0.8 to 0.6	.780	1.7	1.1–2.6	.018
35+	−0.8	−1.9 to 0.3	.159	2.6	1.8–3.7	<.001
Race/ethnicity						
Non-Hispanic (nH) White (ref)	0	—	—	1.0	—	—
nH Black	−0.1	−0.3 to 0.1	.293	1.1	0.9–1.2	.241
nH Native American	0.5	−0.1 to 1.0	.091	0.9	0.6–1.3	.368
nH Native Hawaiian/Other Pacific Islanders	−0.4	−1.7 to 0.8	.538	0.5	0.2–0.8	<.001
nH Asian Americans (nHAA)	−0.02	−0.4 to 0.4	.958	1.1	0.7–2.6	.552
nH More than one race/ethnicity	0.03	−0.4 to 0.4	.892	0.9	0.7–1.2	.640
Hispanic	0.2	0.03–0.4	.033	0.9	0.7–1.0	.023
Year of first cigarette use						
2002 (ref)	0	—	—	1.0	—	—
2003	−0.1	−0.4 to 0.3	.677	0.9	0.7–1.2	.375
2004	−0.2	−0.5 to 0.2	.303	0.7	0.5–1.0	.019
2005	−0.1	−0.5 to 0.2	.476	0.8	0.6–1.0	.068
2006	−0.2	−0.5 to 0.2	.291	0.7	0.6–1.0	.025
2007	−0.3	−0.7 to 0.1	.161	0.6	0.4–0.8	.002

Note. Data for recent-onset cigarette users from the National Surveys on Drug Use and Health, United States, 2004–2007.

^aUnadjusted persistence parameter from the inflate part of the model.

^bUnadjusted rate ratio from the count part of the model.

Discussion

Studying persistence and rate of smoking in this nationally representative sample of ROCUs, we estimate that for every five individuals who start smoking, two persist in smoking with at least 1 day of tobacco smoking in the month prior to assessment; their estimated median number of smoking days in that month is 5 days. Compared with the adolescent-onset newly incident smokers, 18- to 25-year-old ROCUs had an increased rate of tobacco use during the 30-day interval prior to assessment, conditional on the PP (i.e., more smoking days). With respect to race/ethnicity, compared with nHWs, and conditional on the PP, Hispanic ROCUs and those of NHOPI heritage had a lower rate of cigarette use during the 30 days prior to assessment. Again, conditional on the PP, the rates of smoking in 2004–2007 may have declined for the ROCUs who started in those years, relative to those starting in 2002.

Before a more detailed discussion of these findings, several of the more important limitations merit attention. To begin, the NSDUH subgroup includes a number of infrequently occurring race/ethnicity subgroups whose sample sizes sometimes are quite small, despite borrowing of information across multiple years. The estimated regression coefficients associated with these subgroups are statistically fragile and require cautious interpretation. They are presented here primarily as signposts for future research with larger samples. Second, in past years, it was possible to make use of encrypted neighborhood and local area

indicators within the NSDUH public use dataset in order to match on, hold constant, or statistically model neighborhood characteristics. The result was a more fine-grained understanding of individual-level vulnerabilities that might relate to race/ethnicity variation in initiation or persistence in daily use (e.g., [Lillie-Blanton, Anthony, & Schuster, 1993](#); [Petronis & Anthony, 2003](#)). These important encrypted local area indicators are no longer distributed with the NSDUH public use datasets, which hinders efforts to complete analyses that previously were possible. Until those variables are once again made available for research, our capacity to shed light on observed race/ethnicity associations is constrained. Having better control over race/ethnic differences is important, as has been observed by a recent study ([Luo et al., 2008](#)), which found that the number of cigarettes smoked per day and nicotine dependence (as reflected in relapse) vary with race—Blacks becoming cigarette dependent with fewer cigarettes per day compared with Whites. From our study, it now is possible to point toward race/ethnicity subgroups where tobacco count processes vary.

Notwithstanding limitations of this type, in this research, we have identified some potentially interesting sociodemographic variations in a nationally representative sample of ROCUs. The absence of a male–female difference in the count process of smoking involvement is an interesting addition to the growing literature on sex and gender differences in relation to smoking and drug taking generally (e.g., see [Wagner & Anthony, 2007](#)). As noted by others (e.g., [Reitzel et al., 2009](#)), light and

Table 3. Estimated covariate adjusted associations linking subgroup membership with zero-inflated Poisson regression parameters

Past 30-day use	aPP ^a	95% CI	<i>p</i>	aRR ^b	95% CI	<i>p</i>
Sex						
Male	−0.1	−0.2 to 0.1	.344	1.1	1.0–1.2	.079
Female (ref)	0	—	—	1.0	—	—
Age group (years)						
12–17 (ref)	0	—	—	1.0	—	—
18–25	−0.4	−0.5 to −0.2	<.001	1.3	1.2–1.5	<.001
26–34	−0.1	−0.8 to 0.6	.759	1.8	1.3–2.7	.002
35+	−0.9	−2.0–0.2	.134	1.0	1.8–4.1	<.001
Race/ethnicity						
Non-Hispanic (nH) White (ref)	0	—	—	1.0	—	—
nH Black	−0.1	−0.2 to 0.1	.521	1.0	0.9–1.1	.920
nH Native American	0.5	−0.1 to 1.1	.095	0.8	0.6–1.2	.182
nH Native Hawaiian/Other Pacific Islanders	−0.4	−1.8 to 0.9	.541	0.4	0.2–0.8	<.001
nH Asian Americans (nHAA)	0.1	−0.3 to 0.5	.654	0.9	0.7–1.2	.454
nH More than one race/ethnicity	0.03	−0.4 to 0.4	.884	1.0	0.7–1.2	.782
Hispanic	0.2	0.04–0.4	.024	0.8	0.7–1.0	.004
Year of first cigarette use						
2002 (ref)	0	—	—	1.0	—	—
2003	−0.1	−0.4 to 0.3	.625	0.9	0.7–1.2	.340
2004	−0.2	−0.5 to 0.1	.239	0.7	0.5–0.9	.004
2005	−0.1	−0.4 to 0.2	.480	0.7	0.5–1.0	.022
2006	−0.2	−0.5 to 0.2	.285	0.7	0.5–0.9	.011
2007	−0.3	−0.7 to 0.1	.178	0.5	0.4–0.7	.001

Note. Data for recent-onset cigarette users from the National Surveys on Drug Use and Health, United States, 2004–2007. aRR = adjusted rate ratio.

^aAdjusted persistence parameter from the inflate part of the model.

^baRR from the count part of the model.

nondaily smoking has been observed among smokers of Hispanic heritage, and this is a phenomenon that deserves additional study. To the best of our knowledge, there is no prior epidemiological evidence on persistence and rates of smoking among recent-onset smokers of Native Hawaiian or Pacific Islander background, but as noted by Kaholokula (2008), there might be unique sociocultural facets of the Pacific Islander experience that promotes less persistence or lower rates. As observed by Breslau and Peterson (1996), there is no apparent adolescent-onset excess risk of becoming a persistent smoker once elapsed time since first smoke is held constant; it is the 18- to 25-year olds who have a greater rate of smoking, conditional on the PP, perhaps due to greater disposable income, unless the ban on underage smoking has become more successful in recent years.

Some readers might wish for an immediate application of these findings in clinical translational research, with practical implications for either professional practice in the clinic or the public health context (see, e.g., Caraballo, Yee, Gfroerer, & Mirza, 2008; Kellam & Anthony, 1998; O'Loughlin, Karp, Koulis, Paradis, & DiFranza, 2009; Rose et al., 2010; Zang & Wynder, 1998). An alternative perspective is one in which this type of research can be viewed as a form of “basic science” research without immediate translations into the practice settings. In this context, these research findings are based upon a novel methodological look at the process of starting and stopping cigarette smoking (i.e., the PP and non-PP estimated in this study). The immediate application must be in relation to

possible replications in new samples here in the United States and elsewhere. The most salient long-term applications involve genetics and gene–environment interactions research (e.g., Kendler, Schmitt, Aggen, & Prescott, 2008) guiding us toward predictive and explanatory models that help account for why some who start smoking cigarettes become persistent smokers, while others never persist and develop tobacco dependence.

In this initial and novel application of ZIP models in research on count processes in the early stages of tobacco cigarette smoking, we have focused our attention on newly incident tobacco smokers (i.e., those with a relatively short elapsed time since first cigarette smoked, less than 24 months) in order to build up the sample size for optimal statistical precision of the epidemiological estimates and for optimal statistical power of our tests for subgroup variation. Notwithstanding the large sample sizes and favorable statistical power, we found very little evidentiary support for subgroup variation (e.g., in relation to male–female differences and most race/ethnicity subgroup contrasts). The most noteworthy subgroup variation involved statistically significant differences for small subgroups that require replication before anyone should be confident about their interpretation. On this basis, we now will proceed toward analyses of newly incident smokers with even smaller units of elapsed time since first cigarette (e.g., less than 3 months), in confidence that the resulting constraints on sample size should not be hiding large effects due to reduced statistical power. As such, the work we have completed to date

sets the stage for a future line of national scale epidemiological research that involves an application of the ZIP model with the time intervals for initiation of smoking brought toward zero values so as to more closely approximate what might be found in an intensive prospective follow-up study of a nationally representative sample of newly incident smokers (e.g., with follow-up assessments every 1–3 months). This type of short-interval follow-up study with a nationally representative sample of newly incident smokers never has been accomplished before. If it is to be accomplished, then the field will need the estimates from the cross-sectional approximation in order to design the prospective research (e.g., to estimate statistical precision and power for that new research). If the prospective short-interval follow-up research on a national scale is never completed, then the resulting cross-sectional estimates may be the best we can afford.

In this research's clarification of modest-sized subgroup variations in the persistence and RR parameters of the ZIP model, we also set the stage for more probing research on the genetic and gene–environment interactions that might have a bearing on the persistence (or nonpersistence) of smoking once it starts, as well as these influences on the rates of smoking, conditional on membership in the persistence class. With or without addition of these genetic elements in a prediction equation for persistence and the rate of smoking, conditional on persistence, there is anticipated utility in the prediction equations in public health planning for future tobacco prevention and control initiatives.

Funding

National Institute on Drug Abuse awards K05DA015799, T32DA021129, R01DA016558, and MSU research funds from the Provost's Office.

Declaration of Interest

None declared.

Acknowledgments

The authors wish to thank the following individuals, whose various efforts contributed to the success of this research: Drs. Mirjana Radovanovic and Hui Cheng, O. A. Adelaja (all of whom provided general conceptual assistance), and Jessica VanDenBerge (copy editing), all from Michigan State University. We also are grateful to the United States Substance Abuse and Mental Health Services Administration Office of Applied Studies, which administers the NSDUH and arranges for timely release of the NSDUH public use datasets. The authors wish to acknowledge the project's funding sources.

References

- Anthony, J. C. (2010). Novel phenotype issues raised in cross-national research on drug dependence. *Annals of the New York Academy of Sciences*, 1187, 354–370.
- Breslau, N., & Peterson, E. L. (1996). Smoking cessation in young adults: Age at initiation of cigarette smoking and other suspected influences. *American Journal of Public Health*, 86, 214–220.
- Brook, J. S., Saar, N. S., Zhang, C., & Brook, D. W. (2009). Familial and non-familial smoking: Effects on smoking and nicotine dependence. *Drug and Alcohol Dependence*, 101, 62–68.
- Caraballo, R. S., Yee, S. L., Gfroerer, J., & Mirza, S. A. (2008). Adult tobacco use among racial and ethnic groups living in the United States, 2002–2005. *Preventing Chronic Disease: Public Health Research, Practice, and Policy*, 5, 1–9.
- Center for Disease Control and Prevention. (2008). Smoking-attributable mortality, years of potential life lost, and productivity losses—United States, 2000–2004. *Morbidity and Mortality Weekly Report*, 57, 1226–1228.
- Chen, C-Y, O'Brien, M., & Anthony, J. C. (2005). Who becomes cannabis dependent soon after onset of use? Epidemiological evidence from the United States: 2000–2001. *Drug and Alcohol Dependence*, 79, 11–22.
- Ezzati, M., Lopez, A. D., Rodgers, A., Vander Hoorn, S., Murray, C. J., & Comparative Risk Assessment Collaborating Group. (2002). Selected major risk factors and global and regional burden of disease. *Lancet*, 360, 1347–1360.
- Hilbe, J. (1999). Zero-inflated Poisson and negative binomial regression. *Stata Technology Bulletin*, 47, 37–47.
- Kaholokula, J. (2008). Culturally informed smoking cessation strategies for Native Hawaiians. *Nicotine & Tobacco Research*, 10, 671–681.
- Kellam, S. G., & Anthony, J. C. (1998). Targeting early antecedents to prevent tobacco smoking: Findings from an epidemiologically based randomized field trial. *American Journal of Public Health*, 88, 1490–1495.
- Kendler, K. S., Schmitt, E., Aggen, S. H., & Prescott, C. A. (2008). Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. *Archives of General Psychiatry*, 65, 674–682.
- Koob, G. F., & Le Moal, M. (1997). Drug abuse: Hedonic homeostasis dysregulation. *Science*, 278, 52–58.
- Lillie-Blanton, M., Anthony, J. C., & Schuster, C. R. (1993). Probing the meaning of race/ethnic group comparisons in crack cocaine smoking. *Journal of the American Medical Association*, 269, 993–997.
- Luo, Z., Alvarado, H., Hatsukami, D. K., Johnson, E. O., Bierut, L. J., & Breslau, N. (2008). Race differences in nicotine dependence in the Collaborative Genetic Study of Nicotine dependence (COGEND). *Nicotine & Tobacco Research*, 10, 1223–1230.
- Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine*, 3, e442.
- O'Brien, M. S., & Anthony, J. C. (2005). Risk of becoming cocaine dependent: Epidemiological estimates for the United State, 2000–2001. *Neuropsychopharmacology*, 30, 1006–1018.

O'Loughlin, J. O., Karp, I., Koulis, T., Paradis, G., & DiFranza, J. (2009). Determinants of first puff and daily cigarette smoking in adolescents. *American Journal of Epidemiology*, 170, 585–597.

Petronis, K. R., & Anthony, J. C. (2003). A different kind of contextual effect: Geographical clustering of cocaine incidence in the USA. *Journal of Epidemiology and Community Health*, 57, 893–900.

Reitzel, L. R., Costello, T. J., Mazas, C. A., Vidrine, J. I., Businelle, M. S., Kendzor, D. E., et al. (2009). Low-level smoking among Spanish-speaking Latino smokers: Relationships with demographics, tobacco dependence, withdrawal, and cessation. *Nicotine & Tobacco Research*, 11, 178–182.

Rose, J. S., Dierker, L. C., & Donny, E. (2010). Nicotine dependence symptoms among recent onset adolescent smokers. *Drug and Alcohol Dependence*, 106, 126–132.

StataCorp. (2007). *Stata statistical software: Release 10.0*. College Station, TX: StataCorp LP.

Storr, C. L., Reboussin, B. A., & Anthony, J. C. (2004). Early childhood misbehavior and the estimated risk of becoming tobacco-dependent. *American Journal of Epidemiology*, 160, 126–130.

Wagner, F. A., & Anthony, J. C. (2002). Into the world of illegal drug use: Exposure opportunity and other mechanisms linking the use of alcohol, tobacco, marijuana, and cocaine. *American Journal of Epidemiology*, 155, 918–925.

Wagner, F. A., & Anthony, J. C. (2007). Male-female differences in the risk of progression from first use to dependence upon cannabis, cocaine, and alcohol. *Drug and Alcohol Dependence*, 86, 191–198.

World Health Organization. (2008). *World Health Statistics Report—2008*. Retrieved November 18, 2009, from <http://www.who.int/whois/whostat/2008/en/index.html>

Zang, E. A., & Wynder, E. L. (1998). Smoking trends in the United States between 1969–1995 based on patients hospitalized with non-smoking-related diseases. *Preventative Medicine*, 27, 854–861.