10

Epidemiological Analysis of Variation in Phenotypic Definitions: A Proof of Concept Using an Example of a Cessation Phenotype

Kay Wanke and Erik Augustson

Traditional behavioral genetic studies based on phenotypes of observed smoking behavior often lack specificity and are subject to classification bias. This chapter explores the use of an epidemiological approach for modeling smoking phenotypes, based on transitions along the smoking trajectory and prior exposure, which may yield more accurate phenotype definitions for future genetic studies.

Three studies are presented that examine improved phenotypes in relation to numerous variables for smoking behavior and comorbid conditions:

- *An analysis of male Finnish smokers from the* Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC), *examining the behavior of sustained quitters, relapsers, and never quitters relative to several baseline variables involving smoking history, behavioral/psychological symptoms, and alcohol use*
- *Cross-sectional data from the* Tobacco Use Supplement to the U.S. Census Bureau's Current Population Survey (TUS-CPS), *comparing sustained quitters, relapsers, smokers with no quitting history, and sustained quitters on measures of smoking behavior and nicotine dependence*
- *Data from the* Third National Health and Nutrition Examination Survey (NHANES III) *data set, comparing smoking subgroups on the basis of consumption levels and quit attempts with independent variables involving alcohol use, insomnia, and depression*

The results of these analyses demonstrated substantial variations in measured variables between these more tightly controlled phenotypes. Results such as these lend support to the idea of creating more specific, epidemiologically based phenotypes of tobacco use, which, in turn, may correlate more closely with genetic variations.

Introduction

The overall goal of this chapter is to demonstrate the utility of epidemiological approaches for defining phenotypes by narrowing the inclusion criteria, and thereby, it is hoped, reducing misclassification bias. Specifically, a series of analyses explore if changes in phenotype definition affect key indicators of the smoking behavior of interest. The analyses presented here use sustained cessation as an exemplar behavior. This chapter contributes to the discussion of various approaches to defining phenotypes, which is a crucial element of the questions this monograph addresses. The analyses in this chapter will provide an example of how epidemiology can play an active role in understanding smoking behavior and the impact of genetic factors.

A long-standing and solid literature supports the role of genetic factors in smoking behavior,[1–4] and subsequent linkage studies continue to provide highly suggestive findings.[5–8] A review by Munafó and colleagues[9] found that despite this foundation, overall behavioral genetic studies of smoking have produced only a limited body of consistent results. In part, this likely reflects the multiple influences associated with smoking, including the potential additive effects of a large number of genes. However, this lack of solid findings regarding the impact of specific genes is also indicative of a number of methodological limitations within the field of behavioral genetic studies of tobacco use.[9] Among the various identified obstacles to progress in the field is the key observation that smoking phenotypes have typically been poorly defined with respect to important behaviors and exposures, leading to a high likelihood of classification bias.[9] Of particular concern is the inclusion of individuals who have been either only minimally exposed or unexposed to nicotine.[10] The potential importance is highlighted by findings from studies that have attempted to remediate this methodological problem by using more restrictive phenotypes.[11–14] In doing so, it appears that convergent results are beginning to be demonstrated.

Although studies of genetic factors of smoking behavior are a prominent area of research in which phenotypic definitions are crucial, this same fundamental problem exists in purely behavioral and epidemiological studies as well. Some studies have used strategies similar to that presented here to address this issue.[15–17]

Given the historic limitations in the field and the potential advances indicated by attempts to remediate the issues associated with defining phenotypes, a delineated method for doing so may be beneficial to consider. To address the underlying methodological limitations of many approaches to defining phenotypes, a specific strategy is proposed. It is suggested that phenotypes be defined with two features: (1) behavior identified by transitions along the smoking trajectory and (2) adequate exposure to a precursor of the behavior of interest. The rationale for using each of these is discussed below.

Transitions along the Smoking Trajectory

Smoking, like many behaviors, can be conceptualized as occurring along a trajectory, beginning with experimentation, and moving on to initiation, regular use, dependence, and then attempted cessation with success or relapse (figure 10.1). Common points of clinical concern (e.g., initiation, dependence) are labeled on the trajectory. Figure 10.1 shows examples of how various stages along the smoking trajectory represent choice points at which one could define a phenotype based on smoking behavior (e.g., lifetime smoked <100 cigarettes versus ≥100 cigarettes). Hypothesized potential underlying genetic influences are also linked to areas along

Smoking Phenotypes: The Perils of Casting Too Broad a Net

Traditionally, phenotypes for many behavioral genetic studies have been based on broad classifications of behavior such as never versus ever smokers, current versus former smokers, or current smokers versus nonsmokers, which includes both former and never smokers. Although these types of phenotypes have strong face validity, there are a number of serious drawbacks, including the potential for biasing of results due to misclassification. One of the most significant problems with these phenotypes is that they are uninformative regarding exposure to the behavior of interest and, indeed, lack specificity regarding what that behavior may be for the analysis.[a] Research by Saccone and colleagues[b] found changes in their results when unexposed individuals were removed from analyses of a smoking phenotype based on sib-pairs. Although changes in sample size may have affected the findings, the potential impact of including minimally and unexposed individuals in behavioral genetic studies of substance abuse represents a serious methodological issue in need of further investigation.[a] Given this, perhaps it should not come as a surprise that these traditional phenotypes have been problematic in behavioral genetic studies. As noted elsewhere in this chapter, the potential effect of poorly defined phenotypes is demonstrated by studies that have used phenotypes more specifically defined by a particular smoking behavior.[c,d,e,f]

[a]Saccone, N. L., E. L. Goode, and A. W. Bergen. 2003. Genetic analysis workshop 13: Summary of analyses of alcohol and cigarette use phenotypes in the Framingham Heart Study. *Genetic Epidemiology* 25 Suppl. 1: S90–S97.

[b]Saccone, N. L., R. J. Neuman, S. F. Saccone, and J. P. Rice. 2003. Genetic analysis of maximum cigarette-use phenotypes. *BMC Genetics* 4 Suppl. 1: S105.

[c]Bierut, L. J., P. A. Madden, N. Breslau, E. O. Johnson, D. Hatsukami, O. F. Pomerleau, G. E. Swan, et al. 2007. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Human Molecular Genetics* 16 (1): 24–35.

[d]Saccone, S. F., A. L. Hinrichs, N. L. Saccone, G. A. Chase, K. Konvicka, P. A. Madden, N. Breslau, et al. 2007. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Human Molecular Genetics* 16 (1): 36–49.

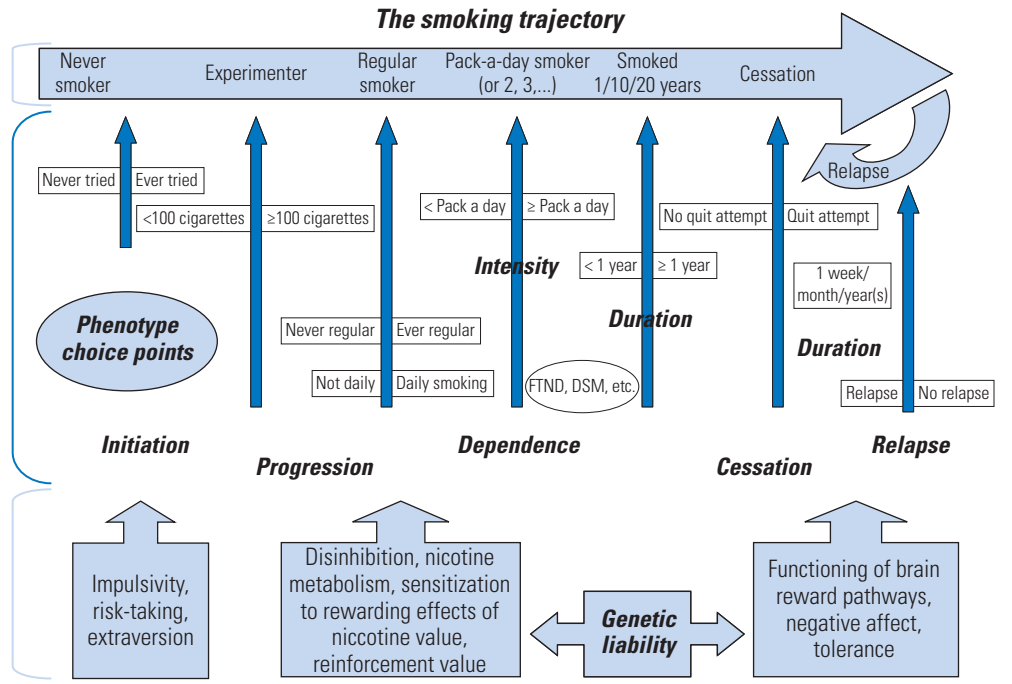
[e]Saccone, S. F., M. L. Pergadia, A. Loukola, U. Broms, G. W. Montgomery, J. C. Wang, A. Agrawal, et al. 2007. Genetic linkage to chromosome 22q12 for a heavy-smoking quantitative trait in two independent samples. *American Journal of Human Genetics* 80 (5): 856–66.

[f]Uhl, G. R., Q. R. Liu, T. Drgon, C. Johnson, D. Walther, and J. E. Rose. 2007. Molecular genetics of nicotine dependence and abstinence: Whole genome association using 520,000 SNPs. *BMC Genetics* 8:10.

the trajectory. As shown in figure 10.1, a number of factors are likely to be more or less associated with behavior at a specific point on the trajectory and with transition to the next stage. Therefore, it is crucial that researchers identify the specific point they are studying and link this point to a specific behavior of interest. For example, given this trajectory of behavior, one might be interested in factors, including potential underlying genetic factors, specifically associated with the question of why some people progress beyond initiation to dependence and others do not. As such, multiple phenotypes of interest are defined by transitions along the trajectory within the broad category of smoking.

Considering specific points along the smoking trajectory suggests that inconsistent findings within the tobacco behavioral genetic literature could be, in part, accounted for by differences in which behavior was studied. If the impact of a genetic trait on smoking varies depending on the specific behavioral point on the trajectory, then a given polymorphism associated with the genetic influence may

Figure 10.1 Phenotype Choice Points along the Smoking Trajectory



Note. FTND = Fagerström Test for Nicotine Dependence; DSM = *Diagnostic and Statistical Manual of Mental Disorders.*

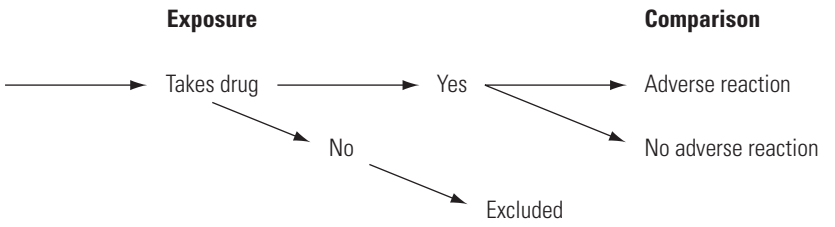
or may not be associated with "smoking." For example, if nicotine metabolism were associated with nicotine dependence, but not with risk of relapse, then polymorphisms associated with nicotine metabolism would be positively associated with smoking only in studies in which nicotine dependence was the basis of the phenotype.

Defining phenotypes based on choice points along the trajectory also suggests that by considering the mechanism by which a gene was believed to influence smoking behavior, one could hypothesize a priori its potential as a candidate gene for that specific behavior along the smoking behavior trajectory. For example, Lerman and colleagues[18] reported the results of a trial investigating the role of the *OPRM1*N40D* variant on cessation. They found that smokers with the variant allele were significantly more likely to be abstinent at the end of the trial and to report fewer symptoms of mood disturbance and less weight gain postcessation. These results suggest that at least some functions of the *OPRM1* gene as it relates to smoking behavior are potentially implicated specifically in the maintenance of smoking abstinence.

Consideration of Adequate Exposure

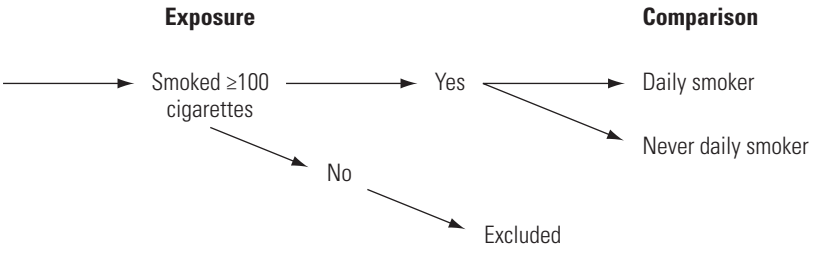
The second feature used by this approach to define phenotypes is to ensure that adequate exposure occurred. Lack of exposure has been identified as a potentially significant methodological challenge that may be particularly important in studies of smoking behavior.[10] For the purposes of phenotype definition, it is suggested that adequate exposure in all individuals used in a study can be accounted for by including only those who have had exposure to a

Figure 10.2 Example of Phenotypic Comparison for Drug Response Using the Proposed Approach



Note. Comparison is only among those who had adequate exposure, defined in this example as having taken the drug.

Figure 10.3 Example of Phenotypic Comparison for Progression to Daily Smoking (Behavior of Interest) Using the Proposed Approach

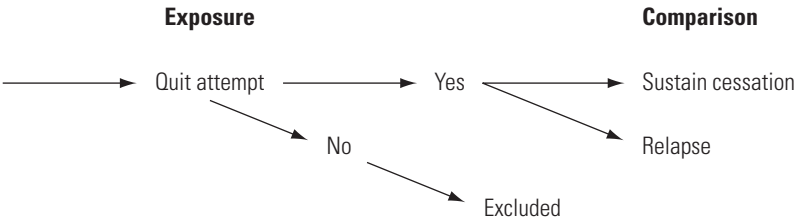


Note. Comparison is only among those who had adequate exposure to precursor of the behavior of interest. Exposure is defined as having smoked ≥100 cigarettes in lifetime.

precursor of the behavior of interest along the smoking trajectory. Conversely, all individuals who have not passed through the previous stage on the trajectory must be excluded. A hypothetical example from pharmacology highlights the importance of appropriate exposure. If the goal of the study were to identify features of individuals at risk of adverse reactions to a medication, one would administer the medication to all of the participants and then compare the study measures among those who did versus those who did not have an adverse reaction. Any individual not taking the drug would be uninformative regarding possible adverse reactions (figure 10.2)

This point is relevant to behavioral genetic studies of smoking behavior. For example, in attempting to understand factors associated with progression from experimentation to daily smoking, individuals who had never experimented with cigarettes would be uninformative. That is, their risk to progress to regular use would be unknown because they had never been exposed.[10] So the appropriate analysis would compare individuals who progressed to regular smoking versus those who did not, but only among those individuals who had experimented with cigarettes based on a specific "exposure"—for example, had smoked ≥100 cigarettes in their lives (figure 10.3). However, a common phenotype previously used in such studies is current smokers versus nonsmokers;[9] the latter includes individuals who have experimented, but did not progress, and individuals who were never smokers. Similarly, in attempting to identify factors associated with relapse versus successful cessation, one would include only smokers who had made a quit

Figure 10.4 Example of Phenotypic Comparison for Sustained Smoking Cessation (Behavior of Interest) Using the Proposed Approach



Note. Comparison is only among those who had adequate exposure to precursor of behavior of interest. Exposure is defined as having made a quit attempt.

attempt as only these individuals would be informative (figure 10.4). However, the classic comparison is often between former smokers and current smokers,[9] which includes smokers who have never tried to quit.

The Proposed Approach

Strategy and Considerations

Given the above, the approach presented here proposes that phenotypes be formulated in the following way. First, a behavior of interest associated with a specific stage of the smoking continuum must be defined. Second, the comparison group must consist of individuals who have had adequate exposure to the behavior of the previous stage along the continuum. The remainder of this chapter focuses on an example using sustained smoking cessation as the phenotype of interest. Smoking cessation is defined by two separate behaviors: initiation of a quit attempt and sustaining that quit attempt.[19] This then represents a choice point that can be used to define inclusion in the phenotype analysis. The behavior of interest in the example used here is sustaining cessation. Adequate exposure to a precursor of the behavior of interest is defined by having made a quit attempt. Thus, this phenotype is based on a comparison of individuals who do and do not maintain cessation only among smokers who have made a quit attempt (figure 10.4).

As an additional consideration, given that epidemiology is fundamentally based on tracking behaviors associated with the distribution of public health outcomes, this approach focuses on observable behavior. (For additional examples and further discussion of behavioral taxonomies, see Gifford and Humphreys,[20] Gifford and colleagues,[21] Silva,[22] and Follette;[23,24] also see other articles in the December 1996 special issue of the *Journal of Consulting and Clinical Psychology*.) In the case of behavioral genetics, phenotypes function as a means to define categories of people who have certain features and behaviors. This is distinct from identifying why people have those features. The distinction is important in that it is inappropriate to include elements in the definition of a phenotype that address underlying causal differences between groups (see further discussion in Silva[22] and Skinner[25]). The underlying causes of why groups differ on the outcome of interest—in this case, genetic influences—are what the comparisons of phenotypes hope to elucidate. So, it becomes a circular argument to include them in the definition of the phenotype itself. Hypothesized underlying

causal mechanisms are appropriate for defining endophenotypes (see other chapters in this volume), which serve to bridge the phenotype-gene causal associations.

Testing the Approach

The fundamental assumption behind the argument for using more tightly defined phenotypes is that the detection of genetic influences on key smoking behaviors will be substantially improved; that is, reducing classification bias improves the ability to detect associations. Although this logic seems sound, it is an empirical question, and analyses are needed to justify that this approach truly does improve the ability to find replicable associations. Ultimately, the test of this question will need to be carried out using genotyping of various phenotypes within a single data set. However, this methodology is resource intensive, and the "proof of concept" analyses performed in this chapter are a necessary starting point.

Simply stated, the basic research question is, does it matter if "improved" phenotypes are used or not? For the purposes of this chapter, evidence of "improvement" is demonstrated when, as the phenotype changes, factors associated with the behavior of interest change in an a priori, potentially clinically meaningful direction. The rationale is that changes in factors associated with the phenotype are indicative of changes in the underlying endophenotype and genes associated with the behavior. In the examples presented in this chapter, "improved" definitions are detected by finding changes in characteristics predictive of smoking cessation that vary as the definition of the comparison groups changes. The characteristics chosen to serve as independent variables for the analyses presented in this chapter have been previously associated or may be associated with ability to sustain smoking cessation: (1) markers of nicotine dependence[26–30] and (2) the presence of comorbid conditions including heavy alcohol use,[31,32] anxiety/depression,[33–36] and insomnia.[37,38]

Epidemiology offers a potentially powerful set of methodological tools for addressing the importance of phenotypic definition. Various phenotypes can be defined on the basis of clearly observable behaviors, and analyses can be performed to assess how key indicators and outcomes change as the phenotypic definitions change. In addition, there are multiple, large-scale data sets based on a variety of samples that have measures of smoking across the behavioral continuum along with a wide array of potential covariates. This allows for considering a number of comparison phenotypes within a data set as well as potentially confirming the findings across data sets.

To test this approach, a series of analyses across multiple data sets are presented comparing different definitions of smoking cessation (table 10.1). Multiple data sets were analyzed for a variety of reasons. First, by applying the strategy to a number of different data sets, a basic replication of the validity of the approach was performed. Second, each of these data sets had varying methodological strengths and weaknesses; considering the impact of phenotypic variation across the data sets provided broader coverage across the methodologies of the surveys. Next, although all of these data sets assessed smoking behavior, no large-scale data sets exist that were designed to focus on smoking behavior. As such, none of the smoking assessments are comprehensive. Analyzing multiple data sets allowed for covering some of the gaps present in any single survey. Each survey also assessed different aspects of sustained cessation and independent variables of potential interest. Again, performing an analysis across the data sets allowed for broader coverage of variables associated with nicotine dependence and comorbid conditions, as well as definitions of cessation. Lastly, because of methodological differences in each data set, it was possible

Table 10.1 A Comparison of the Data Sets and Variables Used in the Analyses Presented in This Chapter

	ATBC	TUS-CPS	NHANES III
Year collected	1985–1993	2003	1988–1994
Sample size per group	~1,380	1,500	Varies
Groups			
Sustained quitters	X	X	X
Current smokers		X	X
Relapsers	X	X	X
	X	X	
Independent variables			
Age at smoking onset	X	X	X
Years smoked	X	X	X
Cigarettes per day	X	X	X
Pack-years	X		
Inhale when smoking	X		
Nicotine dependence		X	
Time to first cigarette		X	
Alcohol use/history	X		X
Anxiety	X		
Depression	X		X
Insomnia	X		X

Note. ATBC = Alpha Tocopherol, Beta-Carotene Cancer Prevention Study; TUS-CPS = Tobacco Use Supplement to the U.S. Census Bureau's Current Population Survey; NHANES III = Third National Health and Nutrition Examination Survey; Pack-years = the number of years of smoking multiplied by the number of packs of cigarettes smoked per day.

to vary the specific analyses that were performed to broaden the test of this general strategy for defining phenotypes.

Method

General Analytic Approach

The primary objective of these analyses was to perform a proof of concept of the approach to defining phenotypes laid out in this chapter. This was done by comparing the results of between-group analyses with smoking phenotype as the dependent variable and measures of nicotine dependence and comorbid conditions as the independent variable. Individuals who had successfully sustained cessation were the primary phenotype. With the use of classic contingency table analysis, comparisons were performed between those who had sustained cessation and either (1) phenotypes defined by more traditional approaches or (2) extreme examples of phenotypes that would not meet the inclusion criteria proposed here (i.e., individuals who had never made a quit attempt). The analyses were then repeated using the proposed strategy of "improved" phenotypic definition to assess how the results change. In general, results presented are based on χ^2 and *t*-test univariate analyses by using data from the ATBC, the TUS-CPS, and NHANES III (table 10.1). Each of these data sets and analyses is discussed separately below.

ATBC Analyses

Data Source

The first set of analyses presented in this chapter was performed by using a sample drawn from the ATBC.[39] Between 1985 and 1993, this longitudinal, population-based chemoprevention trial enrolled 29,133 Finnish male current smokers between the ages of 50 and 69 years into a randomized primary prevention trial to assess whether alpha-tocopherol or beta-carotene would reduce cancer incidence. At baseline, all individuals in the ATBC study smoked at least five cigarettes per day and were generally in good health. Median age at entry into the study was 57 years, the median number of cigarettes smoked per day (CPD) was 20, and the reported median years of smoking was 36. Exclusion criteria for the ATBC included a history of cancer, significant cardiac diagnoses, cirrhosis, chronic alcoholism, and significant psychiatric diagnoses. Diagnosis of lung cancer during the trial was an additional exclusion for the analyses because of its impact on smoking cessation[40,41] and high mortality.

Within the ATBC, extensive medical history data based on participant self-report and medical examination were collected at baseline, and participants were followed for five to eight years with three scheduled follow-up visits per year (i.e., every four months). At each follow-up, participants were queried about health and smoking status since their last visit.

Phenotype Comparisons

All individuals in the ATBC were established heavy smokers, so the focus of analyses within this data set was on how the results might vary between phenotypic definitions that did and did not consider the issue of adequate exposure. Three groups were defined: sustained quitters, relapsers, and never quitters. Sustained quitters and relapsers were exposed to the key behavior of having made at least one quit attempt during the trial.[15] Sustained quitters were defined as men who reported not smoking at all for at least 10 consecutive follow-up visits, equal to 40 months or more of abstinence. Relapsers were men who reported that they made a quit attempt but sustained it for no more than one consecutive interval and had a confirmed relapse. Of the original ATBC sample, approximately 30% of the participants made a quit attempt during the trial. Of these, 1,379 men met the definition of a sustained quitter and 1,388 met the definition of a relapser. They were therefore included in this analysis. An additional group of 1,380 men who did not make a quit attempt during the trial were randomly selected to serve as the never quitter comparison group.

Comparison of the never quitters to sustained quitters served as the phenotype approach that did not address the issue of exposure to precursors of the behavior of interest, allowing for exploration of the potential impact of misclassification of previous quit history. A more tightly defined phenotypic comparison was made of relapsers versus sustained quitters. In addition, relapsers were compared to never quitters to assess if differences among current smokers based on quit history were present.

Independent Variables

The analysis presented here is based solely on the relationship of baseline variables to sustained cessation. Variables included in this analysis were those associated with smoking history, behavioral/psychological symptoms, and alcohol use. Smoking history included age of smoking onset, years smoked, cigarettes per day, pack-years,* and frequency of inhaling while smoking. Behavioral and psychological

symptoms reported as having occurred during the last four months included anxiety, depression, problems concentrating, and insomnia. Alcohol consumption was converted to mean grams per day on the basis of reported average number of drinks consumed daily during the last year.

Analysis

For each comparison, unadjusted analyses were performed on the basis of a classical contingency table analysis conducted using SAS 8.2.[42] Reported *p*-values are based on either χ^2 for categorical variables or *t*-test for continuous variables.

ATBC Results

The results of the analyses of the ATBC comparisons are presented in table 10.2. Compared with sustained quitters, never quitters demonstrated heavier smoking histories, more current cigarettes per day, and were more likely to always inhale when smoking. Never quitters also reported more symptoms of coexisting comorbid conditions in that they drank more alcohol per day and were more likely to endorse experiencing depressed mood, problems concentrating, and insomnia during the last four months.

A similar general pattern was noted for sustained quitters compared to relapsers. Relapsers reported higher CPD, more pack-years, and being more likely to always inhale. Relapsers also reported drinking more alcohol per day and being more likely to endorse symptoms of anxiety, depression, and insomnia. However, closer examination of the data revealed that, although the general pattern of responding was similar, the differences between sustained quitters and the other two groups were more extreme for never quitters than for relapsers.

Further evidence of this pattern was demonstrated in the analysis of differences between relapsers and never quitters. Significant differences between relapsers and never quitters were found for all markers of nicotine dependence, as well as mean alcohol use per day, with never quitters always being more extreme. No significant difference between relapsers and never quitters was noted for any of the mood symptoms.

ATBC Discussion

These results suggest that sustained quitters, relapsers, and never quitters are distinct

Table 10.2 Results from the Analysis of the ATBC Study Data

	Never quitters *n* = 1,380	Relapsers *n* = 1,388	Sustained quitters *n* = 1,379
Age at smoking onset (mean)	19.1†***	20.0	20.3
Years smoking (mean)	36.7†***	33.9	33.4
Cigarettes per day (mean)	21.7†***	19.1***	17.2
Pack-years (mean)	40.1†***	33.0***	29.2
Always inhale (% yes)	55.3%†***	50.2%**	44.2%
Alcohol (mean grams/day)	19.6†*	17.0***	12.5
Feelings of anxiety (% yes)	22.4%	23.7%*	19.9%
Feelings of depression (% yes)	15.1%*	15.0%*	11.2%
Problems concentrating (% yes)	12.5%*	11.4%	9.3%
Insomnia (% yes)	20.1%***	19.7%***	13.6%

Note. Comparison group vs. sustained quitters: $^{*}p \le 0.01$; $^{**}p \le 0.001$; $^{***}p \le 0.0001$. Relapsers vs. never quitters: $^{\dagger}p \le 0.01$.

phenotypic categories in that markers of nicotine and alcohol dependence varied across the groups. At first, it was suspected that men from this sample with no quit history—that is, never quitters—had an "unknown" liability to relapse and, therefore, hypothetically would fall in the middle of the continuum of "relapse liability," with relapsers and sustained quitters anchoring the ends of this continuum. This assumed that never quitters were ultimately made up of individuals who would become sustained quitters or relapsers should they make a quit attempt. That never quitters were more extreme than relapsers on markers of dependence suggests that they may in fact have greater liability to relapse and that not making a quit attempt during the eight years of the ATBC trial may be related to their dependence. Knowledge of quit history before entering the trial might help to further understand this possibility, but the ATBC did not collect this information, so prior experience with quitting is unknown.

Both relapsers and never quitters differed from sustained quitters on symptoms of mood disruption, but not from each other on this cluster of symptoms. Although there are a variety of potential interpretations, it is interesting to note that these symptoms are similar to those often experienced during withdrawal from nicotine.[43] It may be that the higher frequency of the symptoms at baseline has important implications for whether an individual will attempt to quit smoking and the subsequent risk for relapse.

The ATBC data set has a number of features that contribute to its usefulness in this series of analysis. The trial has a large sample such that appropriate power was available for the analyses. The sample was followed for an extended period with excellent follow-up rates. This allowed for considering sustained cessation across a time frame much longer than is typically assessed. In addition, the ATBC was not a smoking cessation trial, so it affords an opportunity to observe self-initiated cessation attempts and sustaining those attempts. The study also assessed a variety of variables of potential interest, allowing for analyses of comorbid conditions.

The ATBC data set has limitations in regard to the question of appropriate strategies to define phenotypes. Common across all data sets used in this study and cessation research in general, having made a quit attempt during the trial was based solely on self-report. Research suggests that individuals typically are truthful about smoking behavior in contexts in which no negative consequences are associated with smoking status. As the above-mentioned lack of data on prior cessation experience highlights, the ATBC was fundamentally a trial testing a nutritional cancer prevention intervention. As such, although all participants were established heavy smokers, available smoking variables are limited. The variable used to define cessation attempts is somewhat imprecise, especially in terms of accurate length of cessation in fewer than four-month intervals. No data are available regarding what resources may have been used during a reported quit attempt. This could clearly affect cessation success, although nicotine replacement therapy and other pharmacological interventions known to affect cessation success were not widely available or used during the time frame of the ATBC study. In addition, because of the lack of data regarding quit attempts before entering the trial, misclassification of quit history (the fundamental prior exposure variable) is still possible despite the long follow-up period during the trial. Given that all participants were heavy smokers, a limited range of smoking behaviors can be compared for the purposes of considering various phenotypes and the impact of varying definitions. Given some of these limitations, an effort was made to confirm the usefulness of this approach in two additional data sets: TUS-CPS and NHANES III.

TUS-CPS Analyses

Data Source

In a fashion similar to that performed in the ATBC, cross-sectional data from the 2003 TUS-CPS[44] were used in a second analysis investigating how varying phenotypic definitions may affect the results. The CPS, administered by the U.S. Census Bureau, uses a multistage probability sample design to produce reliable national and state estimates on labor force characteristics among the U.S. civilian noninstitutionalized population aged 15 years and older. Every three years since the 1980s, the National Cancer Institute and the Centers for Disease Control and Prevention sponsor the TUS to be conducted in conjunction with the CPS. The TUS collects data on tobacco use and related attitudes and practices among those who completed the CPS. In 2003, the TUS collected detailed data on smoking, former smoking, and quitting behaviors from approximately 250,000 respondents in the months of February, June, and November. Details of the sampling methods for the TUS-CPS are reported elsewhere.[45]

Phenotype Comparisons

To mimic the above analyses using ATBC data and typical behavioral genetic studies focusing on specific homogeneous populations, a similar set of samples was selected, limited to white males. The sample size also was limited to 1,500 per group to approximate that used in the ATBC analysis and to reflect the sample size of the parameters that might be seen in a behavioral genetics study. In this case, four groups were defined to be used in the comparison: all current smokers (current), established daily smokers with and without a history of making a quit attempt (relapsers and no quit, respectively), and sustained quitters. Relapsers, no quit, and sustained quitters all had histories of being daily smokers for at least five years, and sustained quitters reported having sustained cessation for one to five years. The current smoking group included respondents who indicated that they smoked "every day" or "some days."

The analyses of current versus sustained quitters and no quit versus sustained quitters represented comparisons in which exposure to the precursor of the behavior of interest (having made a quit attempt) was not part of the phenotype definition. As in the analysis of the ATBC data, analysis of relapsers versus sustained quitters was the more tightly defined comparison. For each group, 1,500 white males were randomly selected, again to create a similar set of subsamples to that used for the ATBC analyses and behavioral genetic studies.

Independent Variables

The TUS-CPS includes questions associated with a variety of issues associated with smoking behavior and tobacco control. For this analysis, measures were used of number of CPD, age of onset of regular smoking, years smoked, time to first cigarette in the morning, and items from the Shiffman Nicotine Dependence Syndrome Scale (NDSS). In addition to CPD and age of onset for regular smoking, time to first cigarette was used as an indicator of nicotine dependence. Evidence from factor analysis strongly supports the use of this item as an overall marker for nicotine dependence,[46–50] and it has previously been used in population-based surveys as an indicator of nicotine dependence.[51] An abbreviated version of the NDSS[52,53] is also available on the 2003 TUS-CPS and is asked of all current smokers and former smokers who have quit within five years. The TUS includes NDSS questions about difficulty delaying smoking, willingness to go out in a storm to get cigarettes, experiencing craving, and willingness to go outside to smoke even if it interferes with an activity. TUS scores these items on a two-point scale (1 = yes, 2 = no),

and the items were summed to provide a total with lower scores indicating less nicotine dependence.

Statistics

The analysis was conducted using SAS 8.2[42] and SUDAAN 8.0.1.[54] SUDAAN was used to account for the complex sample design and the weights of the respondents. Standard contingency table analysis and univariate techniques were used in this analysis. All variables presented from this analysis are continuous, and reported *p*-values are based on *t*-tests by using corrected standard errors derived from the SUDAAN PROC DESCRIPT procedure.

TUS-CPS Results

Results from this analysis are presented in table 10.3. Results from the less restrictive phenotypes are presented first. Compared to current smokers, sustained quitters demonstrated earlier age of onset, higher levels of CPD, longer smoking histories, and higher levels of nicotine dependence as assessed by the NDSS items. Time to first cigarette was not significantly different between current and sustained quitters. In contrast, sustained quitters had longer time to first cigarette, lower level of CPD, and a later age of onset compared to no quit smokers. Sustained quitters continued to have a lower score on the NDSS items, indicating more frequent symptoms associated with nicotine dependence. Total years smoked was not significantly different between the two groups.

Analysis of relapsers versus sustained quitters, the more restrictive phenotype comparison, showed no significant difference between the groups on age of onset, CPD, and years smoked. Time to first cigarette was significantly different between relapsers and sustained quitters, with relapsers reporting a shorter mean interval to first cigarette. As with the other comparisons, sustained quitters had significantly lower scores on the NDSS items.

TUS-CPS Discussion

Findings from this analysis also point to the importance of tightly defined phenotypes in that the pattern of results varies depending on the comparison group. Use of the traditional comparison of all current smokers versus those who had sustained cessation for at least a year suggested possible higher levels of nicotine dependence as assessed by age of onset, CPD, years smoked, and total score on NDSS items. However, analysis of smokers who had progressed to established daily smoking but had not had exposure to the

Table 10.3 Results from the Analysis of TUS-CPS Data

	Current	No quit	Relapser	Sustained quitters
Age of onset	17.4*	16.4**	16.9	17.1
Cigarettes per day	19.9***	23.4**	20.9	21.2
Years smoked	23.2*	26.0	25.7	25.2
Minutes to first cigarette[a]	48.6	37.8**	43.5*	53.6
Nicotine dependence[b]	5.7***	5.6***	5.5***	5.1

Note. A randomly selected sample of 1,500 for each group was used in these analyses. All reported data are for mean response to variable. Comparison group versus sustained quitters: $^{*}p \le 0.01$; $^{**}p \le 0.001$; $^{***}p \le 0.0001$.

[a]Time to first cigarette in the morning in minutes.

[b]Total score of abbreviated Nicotine Dependence Syndrome Scale items. Lower scale is indicative of higher level of nicotine dependence.

pre-behavior of interest (a quit attempt) suggested a reverse pattern, with later age of onset, fewer CPD, and longer interval to first cigarette, for the sustained quitters. Although there appeared to be a trend of an increase of markers of nicotine dependence among smokers from sustained quitters versus relapsers versus no quit smokers, in this sample no substantial evidence was found of differences between relapsers and sustained quitters except in time to first cigarette and score on the abbreviated NDSS.

The abbreviated NDSS is unique to the TUS among the data sets discussed in this chapter. Inconsistent with the general pattern of responses from the other variables used as markers for nicotine dependence, sustained quitters demonstrated more dependence than any of the other three groups. The reasons for this are unclear. One possibility is that the various groups of current smokers were asked about their current experience, while former smokers were asked to recall their previous experience of nicotine-dependence symptoms. Given that the sample of former smokers used in the analysis consisted of individuals who had quit at least one year ago, and up to five years ago, it may be that recall bias played a role in these discrepant results. It may also be that this finding is unique to the samples selected and not reflective of the overall national sample at large.

The TUS-CPS is useful for the proof of concept of the analytic approach being explored in this chapter in that it has a large sample size, includes a wide range of smokers in the sample, and contains a number of items specifically related to smoking behavior, including lifetime history of having made a quit attempt. However, given the cross-sectional nature of the data and the structure of the questions associated with smoking cessation, the survey is of limited use for tracking and understanding cessation in detail. For example, no data are available regarding the details of quit history, such as the time frame of having made an attempt, unless it occurred within the last year, nor is information available on what cessation methods have been used across the person's quit history. Also, items from the NDSS are only asked of former smokers if they attempted to quit within the last five years. As with all the data sets used in this chapter, smoking status and quit history were based solely on self-report. In addition, the absence of questions related to potentially important comorbid conditions such as alcohol use and psychiatric symptoms represents an important gap in the types of questions that can be addressed by this survey.

A final analysis of this approach for defining phenotypes was performed by using the NHANES III data set. This nationally representative data set includes questions relevant to comorbid conditions and so can serve as a potential replication of some of the key findings associated with these behaviors found in the ATBC and TUS-CPS analyses. As an additional "proof of concept" in this analysis, the impact of changing comparisons was explored within a more traditional psychological/epidemiological study in which the selected sample was not based on gender or race/ethnicity, but rather, on groups selected from the population-based, nationally representative sample as a whole.

NHANES III Analyses

Data Source

This sample was selected from NHANES III, conducted from 1988 to 1994 on a cross-sectional representative sample of the U.S. civilian noninstitutionalized population aged two months and older living in households. Only individuals over the age of 18 years were included in the present analysis. Detailed descriptions of the

sample design and operation of the survey have been published elsewhere.[55] For this study, smoking information was obtained from the Household Adult Questionnaire administered in participants' homes.

Phenotype Comparisons

Three phenotypes were defined for the purposes of these analyses on the basis of increasingly restrictive inclusion criteria to reflect progression from regular use to cessation and relapse: all current daily smokers (daily, n = 4,990), current daily smokers who reported having made a quit attempt that lasted for at least one year (relapsers, n = 1,157), and former smokers who had been abstinent for longer than one year (sustained quitters, n = 4,463). The definitions were nonmutually exclusive, and daily smokers included all individuals who also qualified as relapsers. Of importance, the NHANES III only asked about quit attempts that lasted at least one year. Therefore, former smokers had to have sustained cessation for one year or longer to be included in the analyses, and the sample of relapsers included only individuals who had sustained cessation for at least one year before returning to daily smoking. For this analysis, daily smokers versus sustained quitters served as the comparisons of less restrictive phenotypes, while relapsers versus sustained quitters represented the more tightly defined phenotype.

Independent Variables

Similar to the previous analyses, available variables were selected associated with nicotine dependence and comorbid conditions believed to affect sustained cessation. Smoking variables included in the analysis were age of smoking onset, years smoked, and cigarettes per day. NHANES III also provides information about a wide range of health and behavioral symptoms. For this analysis, current mean drinks per day, history of heavy drinking (5+ drinks per day), symptoms of insomnia, and lifetime diagnosis of major depression were included. Diagnosis of major depression was obtained from the Diagnostic Interview Schedule (DIS) administered as part of the NHANES III examination. Methodological constraints of the NHANES were such that only adults under the age of 40 years received the DIS, so reported results for lifetime history of major depression are limited to only the age range of 18–39 years.

Statistics

The analysis was conducted using SAS 8.2[42] and SUDAAN 8.0.1.[54] SUDAAN was used to account for the complex sample design and the weights of the respondents. Standard contingency table analysis and univariate techniques were used in this analysis based on SUDAAN procedures. Preliminary analyses revealed large differences in current age among the three groups (data not shown). To address this, age adjustment was performed for all comparisons.

NHANES III Results

The results for the comparisons from NHANES III are presented in table 10.4. The general pattern was similar between daily smokers and relapsers when compared to sustained quitters with almost all comparisons being statistically significant. In considering the pattern of responses, some differences were noted. Relapsers reported higher cigarette consumption and were more likely to have histories of major depression and sleep disturbance. Relapsers had a slightly earlier age of regular smoking onset. Daily smokers reported higher mean daily alcohol consumptions and were slightly more likely to report a history of drinking five or more drinks per day.

NHANES III Discussion

Although the results of this analysis were not as clear, the potential usefulness of the

Table 10.4 Results from the Analysis of NHANES III Data

	Current smokers	Relapsers[a]	Sustained quitters[b]
Age onset (mean)	17.0***	16.9*	17.2
Years smoked (mean)	21.0***	20.0***	20.7
CPD (mean)	24.1***	27.7**	21.8
ETOH drinks per day (mean)[c]	4.4***	3.7**	2.7
HX 5+ ETOH drinks per day (% yes)	22.0***	20.6**	16.2
HX depression (% yes)[d]	11.6*	14.0*	7.6
HX insomnia (% yes)[d]	30.1	34.4*	25.2

Note. Data from all analyses are adjusted for age of participant. CPD = cigarettes smoked per day; ETOH = alcohol; HX = history. Comparison group versus sustained quitters: $^{*}p \leq 0.05$; $^{**}p \leq 0.005$; $^{***}p \leq 0.0005$.

[a]Relapsers were defined as current daily smokers who reported having a history of being abstinent for at least one year.

[b]Sustained quitters were former smokers who had been abstinent for at least one year.

[c]Only asked of individuals who reported alcohol use.

[d]Age-restricted question.

more restrictive phenotype comparison appears generally supported. In general, similar patterns of statistical significance were seen among the two comparisons. However, the levels of significance are affected by the sizable sample size variations, and so it is difficult to interpret the meaningfulness of the statistical significance. Consideration of variation in patterns of responses suggests that there were differences between the two phenotypic approaches.

Although fewer markers than desirable of nicotine dependence were available in this data set, NHANES does provide indications of potential alcohol abuse/dependence and psychiatric comorbidity. Relapsers began smoking at an earlier age and smoked more cigarettes per day. They were more likely to be diagnosed with a history of major depression on the basis of report of symptoms and to report sleep disturbance. However, symptoms associated with alcohol abuse were more likely to be reported among the all daily smoker group. Thus, there appears to be an increase in markers of dependence and comorbidity as the comparison moves along the smoking trajectory and becomes increasingly restrictive, moving from all daily smokers to smokers who relapsed despite having maintained abstinence for a year or longer. This suggests that (1) the role of nicotine and alcohol dependence potentially increases as one moves along the smoking trajectory and (2) analyses based on phenotypic definitions that do not consider progression along the smoking continuum are more likely to experience misclassification and thereby less likely to detect effects of potentially important genetic influences.

The distinctions between the phenotypes in the NHANES analysis may have been affected by a number of methodological issues. First, the analyses on this data set were not limited to only white men. It was not possible to explore any impact of this difference because of sample size limitations in the data set. A larger issue that likely had a significant impact is the restricted nature of the available cessation question. In NHANES III, the cessation question only asks about quit attempts that lasted at least one year, so no data are available on the much more typical quit attempt followed by relapse within a few weeks.[56] This suggests that a large number of individuals included in the daily smoker phenotype would have qualified as relapsers had more refined data been available. In addition, as individuals

who have maintained cessation for a year or more may be more similar to sustained quitters than individuals who relapse more quickly, differences between individuals who relapse versus those who sustain cessation may have been masked in the analysis of this data set.

Strengths of this data set include that NHANES is a large-scale, nationally representative sample, and the smokers included in the sample are more diverse than those seen in the ATBC sample of older, heavy smoking, Finnish men. NHANES also collects data on a number of potentially important comorbid conditions and provides a clinical diagnosis of major depression, rather then mere endorsement of symptoms, giving NHANES some advantages over both the TUS-CPS and the ATBC.

As noted above, a significant limitation of the NHANES data set is that cessation is only assessed if it lasts a year or longer. In addition, individuals may be in the process of sustaining cessation but have not reached the one-year mark, and so are dropped from the analysis. Next, the constraints of the NHANES methodology are such that not all individuals are asked all questions; in particular, this affects the depression data included in this sample. This may represent a significant issue in that the clinical structured interview for diagnosing major depression was not administered to individuals older than 39 years of age. Given that many individuals do not successfully quit smoking until they are in their 40s, the relation between depression and the behavior of interest (sustaining cessation) may be misrepresented. As with the ATBC, NHANES is not a smoking-specific survey, and the smoking questions are limited in scope. Lastly, the information provided in NHANES is self-reported and cross-sectional. However, despite these limitations, a number of findings consistent with those seen from the ATBC and TUS-CPS data sets emerged.

Summary

This chapter has presented an approach for refining phenotype definitions in the hopes of reducing "noise" associated with misclassification that subsequently reduces researchers' ability to detect small but important genetic influences on smoking behavior. The strategy employed by this approach is based on two features. First, an observable behavior of interest that identifies a specific point along the smoking continuum must be chosen. Then, the comparison group must be defined such that it also has exposure to the precursor of the behavior of interest, excluding all individuals who have not progressed to that point on the smoking continuum. In the case of the example presented here, sustained smoking cessation was the behavior of interest, and the comparison group comprised smokers who had made a quit attempt but subsequently relapsed. In multiple data sets, the results of analyses of the improved phenotype were contrasted to classic, more broadly defined phenotypes and to phenotypes that lacked the key behavioral exposure.

This series of analyses has only considered a few of the possible comparisons that could have been used to validate the proof of concept for this approach. The goal of the current analyses was to demonstrate that even a small difference in definition could affect the subsequent results and the likelihood that associations between behavior and genes would be detected. As such, it was not the intent to fully explore this issue in each data set, although a number of interesting issues regarding the association of various markers of dependence and comorbidity were suggested.

Although it is argued that the phenotypes represent an improvement over those typically used in a number of studies, it may be that even more refined phenotypes are

needed to truly advance the field of genetics of smoking behavior. For example, some work has indicated that consideration of average CPD is insufficient to differentiate smokers.[13,57–59] The results of a study by Saccone and colleagues[13] on the genetic linkage of heavy smoking and chromosome 22q12 suggest that even more precise phenotypes may be needed. They found that use of maximum cigarettes smoked in 24 hours was a more useful indicator of nicotine dependence and a possible important marker of genetic variation among smokers.

In considering the data from across the three data sets, there were consistent limitations that may influence the findings. First, typical of epidemiological surveys, all data used were self-report. Although self-report of smoking status and quit history is consistently used in the literature and considered to be largely reliable in the context of most surveys, the findings would be strengthened if confirmation of smoking status was available.

Next, only limited information was available regarding the details of current and previous smoking behavior. This is a challenge across the available data sets that address smoking. Specific to this project on sustained cessation and relapse, no information was available on methods used to quit smoking. The techniques used to quit smoking could clearly have an impact on cessation success. If methods used varied significantly between sustained quitters and relapsers, the potential influence of the factors on which these analyses focused—that is, markers of nicotine dependence and comorbidity—would be misstated in the findings. This would affect efforts to identify the role of specific underlying genetics polymorphisms. However, the impact of this lack of available data is likely mitigated by the observation that two of the data sets (ATBC, NHANES) were collected during time frames when pharmacological cessation interventions were not widely in use.

Lastly, although the ATBC data set was available for longitudinal analysis, the ATBC is not primarily a project of tobacco use and has a restricted range of smoking behaviors and ages of participants. Longitudinal data from a study that focuses on smoking behavior would be extremely valuable, providing detailed information on smoking as it varies across time.

Despite the limitations of each data set and the varying methodology across the data sets, it was consistently found that changing the comparison groups (e.g., phenotypic definition) affected the nature of the results when considering potential markers of nicotine dependence and factors known to affect cessation. This indicates both that use of classic phenotypes can lead to misclassification errors and that the approach proposed based on trajectory and exposure has merit. Ultimately, the approach presented in this chapter needs to be tested within an analysis that includes genotyping to fully assess its value. Studies using more tightly defined phenotypes are demonstrating more convergent findings in support of the role of specific genes and smoking behavior.[11–14] A more complete test of the approach presented here would examine if varying the phenotypes affected the genetic findings. However, the findings presented here support this approach and lend credence to its adoption.

Conclusions

1. More tightly defined phenotypes of smoking behavior that are based on transitions along the smoking trajectory and adequate prior exposure have the potential to reduce the classification bias and lack of specificity inherent in broader existing phenotypes such as current smoking status. These improved phenotypes, in turn, may lead to closer correlations between smoking behavior and genetic variables in future studies.

2. Studies involving both longitudinal and cross-sectional population data show measurable differences among improved phenotypes, including sustained quitters, relapsers, and never quitters, in key markers such as smoking history, other indices of nicotine dependence, and comorbid conditions such as psychological symptoms and alcohol use.
3. Refined nicotine-dependence phenotypes based on longitudinal characterizations of smoking patterns show promise for further testing in genetic studies in support of potential phenotype-gene causal associations for nicotine dependence. Research indicates the potential need for further refinement of such phenotypes.

References

1. Lessov, C. N., N. G. Martin, D. J. Statham, A. A. Todorov, W. S. Slutske, K. K. Bucholz, A. C. Heath, and P. A. Madden. 2004. Defining nicotine dependence for genetic research: Evidence from Australian twins. *Psychological Medicine* 34 (5): 865–79.
2. Li, M. D., R. Cheng, J. Z. Ma, and G. E. Swan. 2003. A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addiction* 98 (1): 23–31.
3. Madden, P. A., A. C. Heath, N. L. Pedersen, J. Kaprio, M. J. Koskenvuo, and N. G. Martin. 1999. The genetics of smoking persistence in men and women: A multicultural study. *Behavior Genetics* 29 (6): 423–31.
4. True, W. R., A. C. Heath, J. F. Scherrer, B. Waterman, J. Goldberg, N. Lin, S. A. Eisen, M. J. Lyons, and M. T. Tsuang. 1997. Genetic and environmental contributions to smoking. *Addiction* 92 (10): 1277–87.
5. Bergen, A. W., J. F. Korczak, K. A. Weissbecker, and A. M. Goldstein. 1999. A genome-wide search for loci contributing to smoking and alcoholism. *Genetic Epidemiology* 17 Suppl. 1: S55–S60.
6. Gelernter, J., C. Panhuysen, R. Weiss, K. Brady, J. Poling, M. Krauthammer, L. Farrer, and H. R. Kranzler. 2007. Genomewide linkage scan for nicotine dependence: Identification of a chromosome 5 risk locus. *Biological Psychiatry* 61 (1): 119–26.
7. Straub, R. E., P. F. Sullivan, Y. Ma, M. V. Myakishev, C. Harris-Kerr, B. Wormley, B. Kadambi, et al. 1999. Susceptibility genes for nicotine dependence: A genome scan and followup in an independent sample suggest that regions on chromosomes 2, 4, 10, 16, 17 and 18 merit further study. *Molecular Psychiatry* 4 (2): 129–44.
8. Swan, G. E., H. Hops, K. C. Wilhelmsen, C. N. Lessov-Schlaggar , L. S. Cheng, K. S. Hudmon, C. I. Amos, et al. 2006. A genome-wide screen for nicotine dependence susceptibility loci. *American Journal of Medical Genetics Part B, Neuropsychiatric Genetics* 141 (4): 354–60.
9. Munafó, M. R., T. G. Clark, E. C. Johnstone, M. F. G. Murphy, and R. T. Walton. 2004. The genetic basis for smoking behavior: A systematic review and meta-analysis. *Nicotine & Tobacco Research* 6 (4): 583–98.
10. Saccone, N. L., E. L. Goode, and A. W. Bergen. 2003. Genetic analysis workshop 13: Summary of analyses of alcohol and cigarette use phenotypes in the Framingham Heart Study. *Genetic Epidemiology* 25 Suppl. 1: S90–S97.
11. Bierut, L. J., P. A. Madden, N. Breslau, E. O. Johnson, D. Hatsukami, O. F. Pomerleau, G. E. Swan, et al. 2007. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Human Molecular Genetics* 16 (1): 24–35.
12. Saccone, S. F., A. L. Hinrichs, N. L. Saccone, G. A. Chase, K. Konvicka, P. A. Madden, N. Breslau, et al. 2007. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Human Molecular Genetics* 16 (1): 36–49.
13. Saccone, S. F., M. L. Pergadia, A. Loukola, U. Broms, G. W. Montgomery, J. C. Wang, A. Agrawal, et al. 2007. Genetic linkage to chromosome 22q12 for a heavy-smoking quantitative trait in two independent samples. *American Journal of Human Genetics* 80 (5): 856–66.
14. Uhl, G. R., Q. R. Liu, T. Drgon, C. Johnson, D. Walther, and J. E. Rose. 2007. Molecular genetics of nicotine dependence and abstinence: Whole genome association using 520,000 SNPs. *BMC Genetics* 8: 10.
15. Augustson, E. M., K. L. Wanke, S. Rogers, A. W. Bergen, N. Chatterjee, K. Synder, D. Albanes, P. R. Taylor, and N. E. Caporaso. 2008. Predictors of sustained smoking cessation: A prospective analysis of chronic smokers from the Alpha-Tocopherol Beta-Carotene Cancer Prevention Study. *American Journal of Public Health* 98 (3): 549–55.
16. Augustson, E. M., and S. E. Marcus. 2004. Use of the Current Population Survey to characterize sub-populations of continued smokers: A national perspective on the "hardcore" smoker phenomenon. *Nicotine & Tobacco Research* 6 (4): 621–29.
17. Emery, S., E. A. Gilpin, C. Ake, A. J. Farkas, and J. P. Pierce. 2000. Characterizing and identifying "hard-core" smokers: Implications for further reducing smoking prevalence. *American Journal of Public Health* 90 (3): 387–94.
18. Lerman, C., E. P. Wileyto, F. Patterson, M. Rukstalis, J. Audrain-McGovern, S. Restine, P. G. Shields, et al. 2004. The functional mu opioid receptor (OPRM1)

Asn40Asp variant predicts short-term response to nicotine replacement therapy in a clinical trial. *Pharmacogenomics Journal* 4 (3): 184–92.

19. Zhu, S.-H. 2006. Increasing cessation in the population: Quit attempts vs. successful quit attempts. Paper presented at the 13th World Conference on Tobacco OR Health, Washington, DC. http://2006confex.com/uicc/wctoh/techprogram/P227.HTM.
20. Gifford, E., and K. Humphreys. 2007. The psychological science of addiction. *Addiction* 102 (3): 352–61.
21. Gifford, E. V., J. B. Ritsher, J. D. McKellar, and R. H. Moos. 2006. Acceptance and relationship context: a model of substance use disorder treatment outcome. *Addiction* 101 (8): 1167–77.
22. Silva, F. 1993. *Psychometric foundations and behavioral assessment.* Newbury Park, CA: Sage.
23. Follette, W. C. 1997. A behavior analytic conceptualization of personality disorders: a response to Clark, Livesley, and Morey. *Journal of Personality Disorders* 11 (3): 232–41.
24. Follette, W. C. 1996. Introduction to the special section on the development of theoretically coherent alternatives to the DSM system. *Journal of Consulting and Clinical Psychology* 64 (6): 1117–1119.
25. Skinner, B. F. 1938. *The behavior of organisms: An experimental analysis.* Englewood Cliffs, NJ: Prentice Hall.
26. Hymowitz, N., K. M. Cummings, A. Hyland, W. R. Lynn, T. F. Pechacek, and T. D. Hartwell. 1997. Predictors of smoking cessation in a cohort of adult smokers followed for five years. *Tobacco Control* 6 Suppl. 2: S57–S62.
27. Coambs, R. B., S. Li, and L. T. Kozlowski. 1992. Age interacts with heaviness of smoking in predicting success in cessation of smoking. *American Journal of Epidemiology* 135 (3): 240–46.
28. Shiffman, S., M. Hickcox, J. A. Paty, M. Gnys, J. D. Kassel, and T. J. Richards. 1996. Progression from a smoking lapse to relapse: Prediction from abstinence violation effects, nicotine dependence, and lapse characteristics. *Journal of Consulting and Clinical Psychology* 64 (5): 993–1002.
29. Hyland, A., Q. Li, J. E. Bauer, G. A. Giovino, C. Steger, and K. M. Cummings. 2004. Predictors of cessation in a cohort of current and former smokers followed over 13 years. *Nicotine & Tobacco Research* 6 Suppl. 3: S363–S369.
30. Nordstrom, B. L., T. Kinnunen, C. H. Utman, E. A. Krall, P. S. Vokonas, and A. J. Garvey. 2000. Predictors of continued smoking over 25 years of follow-up in the Normative Aging Study. *American Journal of Public Health* 90 (3): 404–6.
31. Hays, J. T., D. R. Schroeder, K. P. Offord, I. T. Croghan, C. A. Patten, R. D. Hurt, D. E. Jorenby, and M. C. Fiore. 1999. Response to nicotine dependence treatment in smokers with current and past alcohol problems. *Annals of Behavioral Medicine* 21 (3): 244–50.
32. Kalman, D., D. Tirch, W. Penk, and H. Denison. 2002. An investigation of predictors of nicotine abstinence in a smoking cessation treatment study of smokers with a past history of alcohol dependence. *Psychology of Addictive Behaviors* 16 (4): 346–49.
33. Hitsman, B., B. Borrelli, D. E. McChargue, B. Spring, and R. Niaura. 2003. History of depression and smoking cessation outcome: A meta-analysis. *Journal of Consulting and Clinical Psychology* 71 (4): 657–63.
34. Black, D. W., M. Zimmerman, and W. H. Coryell. 1999. Cigarette smoking and psychiatric disorder in a community sample. *Annals of Clinical Psychiatry* 11 (3): 129–36.
35. Lasser, K., J. W. Boyd, S. Woolhandler, D. U. Himmelstein, D. McCormick, and D. H. Bor. 2000. Smoking and mental illness: A population-based prevalence study. *JAMA: The Journal of the American Medical Association* 284 (20): 2606–10.
36. Breslau, N., S. P. Novak, and R. C. Kessler. 2004. Psychiatric disorders and stages of smoking. *Biological Psychiatry* 55 (1): 69–76.
37. Colrain, I. M., J. Trinder, and G. E. Swan. 2004. The impact of smoking cessation on objective and subjective markers of sleep: Review, synthesis, and recommendations. *Nicotine & Tobacco Research* 6 (6): 913–25.
38. Smith, M. T., M. I. Huang, and R. Manber. 2005. Cognitive behavior therapy for chronic insomnia occurring within the context of medical and psychiatric disorders. *Clinical Psychology Review* 25 (5): 559–92.
39. *Annals of Epidemiology.* 1994. The Alpha-Tocopherol, Beta-Carotene Lung Cancer Prevention Study: Design, methods, participant characteristics, and compliance. *Annals of Epidemiology* 4 (1): 1–10.

40. Gritz, E. R., R. Nisenbaum, R. E. Elashoff, and E. C. Holmes. 1991. Smoking behavior following diagnosis in patients with stage I non-small cell lung cancer. *Cancer Causes & Control* 2 (2): 105–12.
41. Cox, L. S., C. A. Patten, J. O. Ebbert, A. A. Drews, G. A. Croghan, M. M. Clark, T. D. Wolter, P. A. Decker, and R. D. Hurt. 2002. Tobacco use outcomes among patients with lung cancer treated for nicotine dependence. *Journal of Clinical Oncology* 20 (16): 3461–69.
42. SAS. 2007. SAS products & solutions: Statistical analysis. http://www.sas.com/technologies/analytics/statistics/stat/index.html (accessed December 19, 2008).
43. American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders: DSM–IV–TR.* 4th ed., text rev. Arlington, VA: American Psychiatric Publishing.
44. National Cancer Institute. 2007. What is the TUS-CPS? http://riskfactor.cancer.gov/studies/tus-cps (accessed December 19, 2008).
45. National Cancer Institute. 2007. Where can I get the TUS data, documentation, & questionnaires? http://www.riskfactor.cancer.gov/studies/tus-cps/info.html (accessed December 19, 2008).
46. Nonnemaker, J. M., and G. Homsi. 2007. Measurement properties of the Fagerström Test for Nicotine Dependence adapted for use in an adolescent sample. *Addictive Behaviors* 32 (1): 181–86.
47. Richardson, C. G., and P. A. Ratner. 2005. A confirmatory factor analysis of the Fagerström Test for Nicotine Dependence. *Addictive Behaviors* 30 (4): 697–709.
48. Haddock, C. K., H. Lando, R. C. Klesges, G. W. Talcott, and E. A. Renaud. 1999. A study of the psychometric and predictive properties of the Fagerström Test for Nicotine Dependence in a population of young smokers. *Nicotine & Tobacco Research* 1 (1): 59–66.
49. Radzius, A., J. J. Gallo, D. H. Epstein, D. A. Gorelick, J. L. Cadet, G. E. Uhl, and E. T. Moolchan. 2003. A factor analysis of the Fagerström Test for Nicotine Dependence (FTND). *Nicotine & Tobacco Research* 5 (2): 255–40.
50. Etter, J. F., T. V. Duc, and T. V. Perneger. 1999. Validity of the Fagerström Test for Nicotine Dependence and of the Heaviness of Smoking Index among relatively light smokers. *Addiction* 94 (2): 269–81.
51. Hyland, A., S. Garten, G. A. Giovino, and K. M. Cummings. 2002. Mentholated cigarettes and smoking cessation: Findings from COMMIT. Community Intervention Trial for Smoking Cessation. *Tobacco Control* 11 (2): 135–39.
52. Shiffman, S., and M. A. Sayette. 2005. Validation of the Nicotine Dependence Syndrome Scale (NDSS): A criterion-group design contrasting chippers and regular smokers. *Drug and Alcohol Dependence* 79 (1): 45–52.
53. Shiffman, S., A. Waters, and M. Hickcox. 2004. The Nicotine Dependence Syndrome Scale: A multidimensional measure of nicotine dependence. *Nicotine & Tobacco Research* 6 (2): 327–48.
54. RTI International. 2007. About SUDAAN. http://www.rti.org/sudaan/page.cfm?nav=901.
55. National Center for Health Statistics. 1994. Plan and operation of the Third National Health and Nutrition Examination Survey, 1988–94. Series 1: Programs and collection procedures. *Vital and Health Statistics* 32:1–407.
56. Hughes, J. R., J. Keely, and S. Naud. 2004. Shape of the relapse curve and long-term abstinence among untreated smokers. *Addiction* 99 (1): 29–38.
57. Bergen, A. W., X. R. Yang, Y. Bai, M. B. Beerman, A. M. Goldstein, and L. R. Goldin. 2003. Genomic regions linked to alcohol consumption in the Framingham Heart Study. *BMC Genetics* 4 Suppl. 1: S101.
58. Goode, E. L., M. D. Badzioch, H. Kim, F. Gagnon, L. S. Rozek, K. L. Edwards, and G. P. Jarvik. 2003. Multiple genome-wide analyses of smoking behavior in the Framingham Heart Study. *BMC Genetics* 4 Suppl. 1: S102.
59. Saccone, N. L., R. J. Neuman, S. F. Saccone, and J. P. Rice. 2003. Genetic analysis of maximum cigarette-use phenotypes. *BMC Genetics* 4 Suppl. 1: S105.