Nicotine & Tobacco Research, 2019, 173–179
doi:10.1093/ntr/ntx201
Original investigation
Received December 16, 2016; Editorial Decision August 29, 2017; Accepted September 5, 2017

Original investigation

Using Elastic Net Penalized Cox Proportional Hazards Regression to Identify Predictors of Imminent Smoking Lapse

Robert Suchting PhD[1], Emily T. Hébert DrPH[2], Ping Ma PhD[3], Darla E. Kendzor PhD[2,4], Michael S. Businelle PhD[2,4]

[1]Department of Psychiatry and Behavioral Sciences, University of Texas Health Science Center at Houston, Houston, TX; [2]Oklahoma Tobacco Research Center, Stephenson Cancer Center, Oklahoma City, OK; [3]Division of Population Health, Children's Medical Center, Dallas, TX; [4]Department of Family and Preventive Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK

Corresponding Author: Robert Suchting, PhD, Department of Psychiatry and Behavioral Sciences, University of Texas Health Science Center at Houston, 1941 East Road, Houston TX, 77054, USA. Telephone: 713-486-2521; Fax: 713-486-2618; E-mail: Robert.Suchting@uth.tmc.edu

Abstract

Introduction: Machine learning algorithms such as elastic net regression and backward selection provide a unique and powerful approach to model building given a set of psychosocial predictors of smoking lapse measured repeatedly via ecological momentary assessment (EMA). Understanding these predictors may aid in developing interventions for smoking lapse prevention.

Methods: In a randomized-controlled smoking cessation trial, smartphone-based EMAs were collected from 92 participants following a scheduled quit date. This secondary analysis utilized elastic net-penalized cox proportional hazards regression and model approximation via backward elimination to (1) optimize a predictive model of time to first lapse and (2) simplify that model to its core constituent predictors to maximize parsimony and generalizability.

Results: Elastic net proportional hazards regression selected 17 of 26 possible predictors from 2065 EMAs to model time to first lapse. The predictors with the highest magnitude regression coefficients were having consumed alcohol in the past hour, being around and interacting with a smoker, and having cigarettes easily available. This model was reduced using backward elimination, retaining five predictors and approximating to 93.9% of model fit. The retained predictors included those mentioned above as well as feeling irritable and being in areas where smoking is either discouraged or allowed (as opposed to not permitted).

Conclusions: The strongest predictors of smoking lapse were environmental in nature (e.g., being in smoking-permitted areas) as opposed to internal factors such as psychological affect. Interventions may be improved by a renewed focus of interventions on these predictors.

Implications: The present study demonstrated the utility of machine learning algorithms to optimize the prediction of time to smoking lapse using EMA data. The two models generated by the present analysis found that environmental factors were most strongly related to smoking lapse. The results support the use of machine learning algorithms to investigate intensive longitudinal data, and provide a foundation for the development of highly tailored, just-in-time interventions that can target on multiple antecedents of smoking lapse.

Introduction

Cigarette smoking is implicated in almost one-half million premature deaths and $289 billion in total economic costs in the United States each year.[1] Although the life expectancy of smokers is 11–12 years shorter compared with nonsmokers, quitting smoking by age 39 reduces the years of life lost by nearly 90%.[2] In 2010, the majority of smokers in the United States indicated that they would like to quit smoking, however, only 6.2% successfully quit.[3] Smoking cessation rates are even lower among those of low socioeconomic status (SES).[4–6] Research suggests that low SES smokers are just as likely as high SES smokers to try to quit smoking and to use smoking cessation aids,[5] yet they are far less successful in quitting due to a number of factors, including greater neighborhood disadvantage,[7,8] lower self-efficacy for quitting smoking,[7] higher nicotine cravings,[7,9] higher levels of stress and boredom,[10] and greater exposure to prosmoking social environments.[11]

Recent research has explored factors related to smoking lapse in greater detail than was previously possible.[12,13] Methods such as ecological momentary assessment (EMA), in which behaviors and experiences are assessed in natural environments in real-time[14] allow for the careful examination of environmental, psychological, and social cues that surround smoking behaviors.[15] EMA reduces recall bias by evaluating the situational context of smoking in real-time,[16] facilitating the identification and study of cues that may not otherwise be identified as predictors of smoking, as well as the specific temporal patterns preceding a smoking event.[14] EMA research has identified a number of situational antecedents to smoking lapse. For example, studies have demonstrated that negative affect (including stress and anxiety),[17–19] smoking cravings,[20,21] the presence of others smoking,[22] the consumption of food, caffeine,[23] and alcohol,[21] and proximity to tobacco retail outlets,[24,25] are all associated with smoking cessation and lapse. Furthermore, EMA has demonstrated that many triggers, such as alcohol consumption, negative affect, and proximity to others smoking, interact to exert combined effects that contribute to smoking lapse.[26]

With EMA monitoring of mood and environment in real-time, it has become possible to develop just-in-time, mobile, adaptive interventions to prevent smoking lapse.[27] However, the development of effective just-in-time interventions relies on accurate identification of moments of high-lapse risk. Businelle et al.[18] demonstrated that lapse risk estimation was feasible in socioeconomically disadvantaged smokers, using a weighted algorithm that identified 80% of all first lapses based on the presence of six risk factors in the 4-hour period leading up to the lapse. Such algorithms enable the delivery of highly specific, tailored interventions that may help reduce risk for lapse.[28] While existing research has identified real-time antecedents of lapse (e.g., smoking urge,[18,19,26,29] stress,[18,19,30–32] negative affect[32]), small numbers of variables (e.g., 2–5 variables) in these studies have typically been selected on the basis of theory,[20] or because of their high prevalence in assessments where smoking was indicated.[33] This practice may ignore important patterns in the data that may be useful for new hypothesis generation and may overlook variables that are related to lapse.

Machine learning algorithms can be used to reduce a large set of covariates to a smaller set of the strongest predictors of an outcome such as smoking lapse.[34] Such an approach is especially valuable for the analysis of EMA data, which involves repeated assessment of multiple variables surrounding a specific event or behavior.[35] For example, Dumortier et al. utilized machine learning methods to classify smoking urge into low and high urge states based on situational cues reported via EMA.[36] With intensive, longitudinal data, an algorithmically optimized model may be beneficial because it can maximize the predictive accuracy of the outcome through reduction of statistical concerns such as multicollinearity. Narrowing to a smaller set of predictors may improve our understanding of the most important smoking lapse risk factors, leading to interventions that target relevant risk factors in real-time, or help individuals to avoid situations and/or cope with thoughts/emotions that are most likely to result in lapse.

The present study utilizes machine learning to optimize prediction of time to smoking lapse using algorithms that have not previously been applied to smoking lapse EMA data. The primary purpose of this exploration was to derive a parsimonious statistical model for predicting the moment of first lapse among socioeconomically disadvantaged smokers attending a smoking cessation treatment clinic.

Methods

Participants and Procedure

The present secondary analysis includes data from 92 socioeconomically disadvantaged smokers collected via smartphone-based EMA. The parent study[18] and EMA procedures[37] have been detailed at length elsewhere. However, in brief, participants were recruited from a tobacco cessation clinic at a safety net hospital in Dallas, TX, between August 2011 and April 2013. Participants were randomized to receive either usual care, which included group counseling and pharmacotherapy, or usual care plus small financial incentives (i.e., contingency management) for biochemically verified smoking abstinence. Participants were followed weekly from 1 week pre-quit to 4 weeks after their quit date. At baseline, participants received smartphones and were asked to complete EMAs five times per day (four random assessments and one daily diary) for two consecutive weeks (1 week pre-quit, 1 week post-quit). Daily diaries were prompted by the study phone once a day, 30 minutes after the participant's pre-set waking time. Random assessments were prompted four times a day during the participant's pre-set waking hours. EMAs that were prompted by the smartphone could be delayed for 5 minutes (up to three times per EMA) and no prompted EMAs were pushed within 1 hour of another prompted EMA. Participants were also asked to self-initiate EMAs when they had an urge to smoke and when they were about to smoke or already smoked during the post-quit period. Participant-initiated EMAs were not included in the present analysis. This study was approved by the Institutional Review Boards at the University of Texas Southwestern Medical Center and the University of Texas School of Public Health.

Measures

Demographics

At the baseline visit, participants completed assessments of descriptive characteristics including age, sex, race, ethnicity, and smoking history. Intervention group membership (i.e. contingency management vs. usual care) was indicated in a dichotomous variable.

EMA Measures

Smartphone-based EMA questions were administered to participants for each of the first 7 days following the scheduled quit date.

Table 1. Candidate Predictors for Penalized Regression

#	Name	Text	Measurement scale
1	Urge1	I have an urge to smoke.	Strongly disagree (1) to strongly agree (5)
2	Urge2	I really want to smoke.	
3	Urge3	I need a cigarette.	
4	Affect1	I feel irritable.	
5	Affect2	I feel happy.	
6	Affect4	I feel frustrated/angry.	
7	Affect5	I feel sad.	
8	Affect6	I feel worried.	
9	Affect7	I feel miserable.	
10	Affect8	I feel restless.	
11	Affect9	I feel stressed.	
12	Affect10	I feel hostile.	
13	Affect11	I feel calm.	
14	Cig Avail	Cigarettes are available to me.	Not at all (1) to easily available (5)
15	Social1	Other people are around.	Yes/no
16	Social2	I am around other people and at least one of them is smoking.	Yes/no
17	Social3	Are you interacting with people?	Yes/no
18	Social4	I am interacting with people and at least one of them is smoking.	Yes/no
19	Restriction1	Is smoking allowed where you are?	Forbidden (1) discouraged (2) allowed (3)
20	Consumption1	I ate something within the last 15 minutes.	Yes/no
21	Consumption2	I drank alcohol within the last hour.	Yes/no
22	Expectancies1	I am confident that I could do something OTHER THAN SMOKE to improve my mood.	Strongly disagree (1) to strongly agree (5)
23	Expectancies2	I am confident that SMOKING would improve my mood	
24	Motivation1	I am motivated to AVOID smoking.	
25	Motivation2	I am committed to being smoke free.	
26	Abstinence Self Efficacy1	I am confident in my ability to AVOID smoking.	

The present analysis focused on 26 core items that were prompted at each EMA (i.e., one daily diary, four random assessments). Table 1 provides the full list of EMA items and their respective measurement scales.

Smoking Status. Smoking was measured at every EMA and at in-person visits (on the quit day and 1 week post-quit). In-person visits included breath carbon monoxide (CO) measurement (Vitalograph carbon monoxide monitor). Those who self-reported abstinence since 10:00 PM the night before the quit date and had a negative breath CO test (<10 ppm) were considered abstinent.[19,30,37] Similar criteria (breath CO <8 ppm and self-reported abstinence) established abstinence at the 1-week follow-up visit. Participants with ambiguous or inconsistent assessments (e.g., reported lapse via EMA, but denied lapse during in-person visit) were excluded from the present analysis to maximize the accuracy of the available data.

Data Analytic Strategy

Penalized Cox Proportional Hazards Regression

Penalized cox proportional hazards regression was used to predict first cigarette use following the established quit day by engaging in variable selection from a set of 26 possible time-varying predictors (Table 1). For regression models with many predictors, regression coefficients may have high variance, particularly when the predictors are correlated to some extent. Penalization alleviates this issue by restricting the size of regression coefficients through the use of a complexity parameter that controls shrinkage. Various types of penalization exist; for a summary see Hastie, Tibshirani, and Friedman.[38] The present analysis utilized the "elastic net," a penalty that compromises between the ridge regression and lasso penalties: as in ridge regression, the elastic net may shrink model coefficients without indiscriminant elimination of correlated predictors, and similar to the lasso, model coefficients may shrink all the way to zero and result in a de facto variable selection process. Elastic net-penalized cox proportional hazards regression was performed using the "penalized" package version 0.9-47[39] in the R statistical computing environment.[40]

Model Approximation

The model fit by elastic net proportional hazards regression provided regularized (shrunken) parameter estimates and eliminated some variables from the set of predictors. This model may be simplified further using a process called model approximation.[41,42] This process sacrifices some degree of model fit in exchange for increased parsimony and an improved parameter-to-sample size ratio. A simplified model is developed using backward elimination from the penalized model. Backward elimination is a machine learning algorithm that examines the fit of the model (as measured by Akaike information criteria [AIC]) that results from removing variables. The AIC provides a measure of goodness-of-fit with a penalty for model complexity.[43] Each variable is removed from the model in turn, and the model fit is assessed. The variable (if any) that most improves the model in its deletion (determined by lowest AIC value) is then removed. More variables are then removed in the same fashion until variable deletion no longer improves the model. A reduced model that provides around 95% of the fit (via R^2 or pseudo-R^2) may be considered a successful approximation. The present analysis used the StepAIC() function in the R package *MASS* version 7.3-45[44] to engage in backward elimination from the penalized model. Reduced Cox proportional hazards regression models were fit and violations of the proportional hazards assumption were tested using the coxph() and cox.zph() functions respectively of the R package "survival" version 2.39–5.[45]

Results

Sample Characteristics

Analyses were restricted to the 92 participants that had a directly identifiable first lapse (n = 52) or were verified as nonlapse (n = 40).[18] The first moment of smoking lapse was nonidentifiable in 54 out of 146 participants included in the parent study (not included in the present analysis). Participants were primarily female (56.5%), black (62.0%), were 51.9 years old (SD 7.4), had a total household income of less than $20 000 per year (82.7%), and smoked an average of 18 cigarettes per day (SD 8.5). Participants completed 2876 EMAs that were prompted by the study phone (86.8% of all prompted EMAs, M = 38.3 EMAs per participant) over the course of the 1-week post-quit period. For the 92 participants included in the study, lapse (vs. nonlapse) group did not differ as a function of total number of EMAs provided (p = .85). Of the 2876 total post-quit EMAs, 2065 were retained after removing EMAs that were completed after the first identifiable lapse and EMAs that were missing data (n = 106 EMAs) on any of the 26 candidate predictors used in the complete-case survival analysis. Following first lapse, 11.54% (n = 6) of participants were subsequently abstinent for the remainder of the trial following their first lapse, while 88.46% (n = 46) experienced further lapse(s).

Kaplan–Meier Plot

A Kaplan–Meier plot for the survival function is provided in Figure 1. The plot details the time to smoking lapse in the present study; the x-axis is the time elapsed (in hours) from the start of the quit attempt (from 0 to 168 over the 7-day period) and the y-axis is the probability of abstinence. As the curve shows (with 95% confidence bands in dashed lines), 40% of participants lapsed by the end of the first day (around hour 24) and another 17% of participants lapsed from that point to the end of the 7-day EMA data collection period (at hour 168).

Penalized and Approximated Models

Elastic net-penalized cox proportional hazards regression was used to model time to first cigarette following the established quit date.

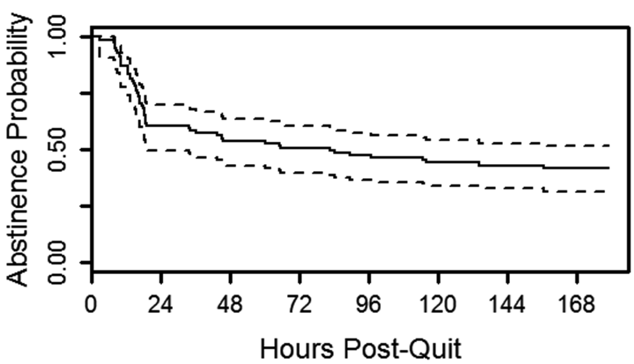

Figure 1. Kaplan–Meier plot of survival probability by hours post-quit.

The elastic net was fit using a combination of optimized lambda values for lasso and ridge regression. Lambda values were optimized by averaging across 10 repetitions of 10-fold cross-validation to minimize variance in estimation.

The elastic net-penalized model retained 17 of 26 candidate predictors of lapse. Shrunken coefficients for all retained predictors are provided in Table 2. Model coefficients may be interpreted in the same manner as OLS regression coefficients, whereby higher values indicate larger magnitude of effect and the accompanying sign indicates the direction of that effect. In the elastic net-penalized model, the highest magnitude effects were from responding "yes" to having consumed alcohol in the past hour and responding "yes" to being around a smoker.

The predictors selected by the elastic net-penalized model were then included in an unpenalized cox proportional hazards regression model to establish a baseline for comparison during model approximation. The baseline model was then reduced by backward elimination (see Supplementary Material for a full account). The approximated model retained six predictors: feeling irritable, having cigarettes available, being around a smoker, being in an area where smoking is discouraged, being in an area where smoking is permitted, and having consumed alcohol in the past hour. The approximated model fit provided 93.9% of the fit of the baseline model (determined via pseudo-R^2 index provided by the coxph() function in R). Model coefficients and hazard ratios are provided in Table 3. As the present model utilized longitudinal format data, deviations from proportional hazards would indicate changes in the influence of each predictor over time.[46] However, graphical analysis of Schoenfeld residual plots as well as statistical analysis using the cox.zph() function in the R package "survival" did not indicate any violations of proportional hazards in the approximated model. As such, model coefficients may be interpreted as the average magnitude of each predictor's influence on the survival rate over time.

The effect of intervention group (usual care vs. contingency management) was assessed for potential influence on the relationship between the set of candidate predictors and time to lapse in all analyses. The intervention group variable was selected by both the penalized and the approximated statistical models and was a statistically reliable predictor of time to lapse in the approximated model (p = .016). The relationship between the other predictors and time to lapse remained very similar in terms of coefficient magnitude and direction compared with the analyses that did not include the incentive variable. Further exploration of the incentive group variable is beyond the scope of the present research; the primary findings and discussion thus focus on the relationship between the predictors and time to lapse across the groups.

Discussion

The present study applied two machine learning/data mining algorithms (elastic net-penalized cox proportional hazards regression and backward elimination) to an intensive longitudinal dataset to predict time to smoking lapse during a quit attempt. The algorithms generated two statistical models, one to optimize predictive power and another to maximize parsimony. The elastic net Cox proportional hazards regression (or simply, "penalized") model retained 17 of 26 candidate predictors of time to smoking lapse. This study is among the first to identify many of these 17 EMA variables as predictors of time to first smoking lapse. The covariates with the highest magnitude coefficients (the strongest predictors) included having consumed alcohol in the previous hour, being near a person that is

Table 2. Selected Predictors of Time to Smoking Lapse and Penalized Coefficients from Elastic Net Model

Variable	Coefficient
Consumption2—I drank alcohol within the last hour.	0.4054
Social2—I am around other people and at least one of them is smoking.	0.3628
Social4—I am interacting with people and at least one of them is smoking.	0.2995
Cig Avail—Cigarettes are available to me.	0.2426
Urge2—I really want to smoke.	0.2148
Restriction1–3—Is smoking allowed where you are?	0.1784
Restriction1–2—Is smoking discouraged where you are?	0.1429
Affect1—I feel irritable.	0.1269
Abstinence Self Efficacy1—I am confident in my ability to avoid smoking.	−0.0921
Affect5—I feel sad.	0.0897
Consumption1—I ate something within the last 15 minutes.	0.0543
Motivation1—I am motivated to AVOID smoking.	−0.0386
Urge1—I have an urge to smoke.	0.0374
Affect11—I feel calm.	−0.0373
Expectancies2—I am confident that smoking would improve my mood.	−0.0214
Affect9—I feel stressed.	0.0133
Motivation2—I am committed to being smoke free.	−0.0078

Coefficients sorted by magnitude from highest to lowest.

Table 3. Approximated Model Coefficients and Hazard Ratios for Time to Smoking Lapse

	Coef.	SE	z	p	Hazard ratio	Hazard ratio, 95% CI	
Affect1—I feel irritable.	0.3727	0.1210	3.0800	.0021	1.4516	1.1450	1.8400
Cig Avail—Cigarettes are available to me.	0.3976	0.1519	2.6180	.0089	1.4883	1.1050	2.0040
Social2—I am around other people and at least one of them is smoking.	1.0477	0.3004	3.4880	.0005	2.8512	1.5820	5.1370
Restriction1–2—Is smoking discouraged where you are?	2.1482	0.6435	3.3380	.0008	8.5694	2.4280	30.2460
Restriction1–3—Is smoking allowed where you are?	1.6591	0.6271	2.6460	.0082	5.2545	1.5370	17.9600
Consumption2—I drank alcohol within the last hour.	1.5779	0.3208	4.9190	.0000	4.8448	2.5840	9.0850

smoking, and interacting with a person that is smoking. Backward elimination ("approximated" model) further reduced the penalized model to retain six predictors: having consumed alcohol in the previous hour, being around a person that is smoking, being in an area where smoking is discouraged (as opposed to forbidden), being in an area where smoking is permitted (as opposed to forbidden), having cigarettes easily available, and feeling irritable.

The predictors selected by the reduced model were mostly environmental in nature (as opposed to factors internal to the participant such as their current urges, affect, motivation, or self-efficacy). As participants encountered high-risk environments, they were more likely to lapse. These results are consistent with Shiffman et al.,[21] who found that lapses were more likely to occur when cigarettes were easily available, when smoking was permitted, and when participants were in the presence of other smokers. Likewise, Deiches et al.[31] reported that the easy availability of cigarettes was related to almost 75% of lapses. Alcohol use has also been consistently reported as a predictor of smoking lapse[47] and has strong associations with both smoking and urges to smoke.[48] Alcohol consumption may also affect the reward processes involved with smoking, as Piasecki et al. reported that participants experienced increased pleasure and decreased punishment from smoking a cigarette while drinking.[49] The reasons that individuals place themselves in high-lapse risk situations are unclear. Marlatt and Gordon[50] have hypothesized that individuals often make "apparently irrelevant decisions" that lead to high-lapse risk situations or places, such as spending time with other smokers, or going to a bar where smoking is allowed. Such decisions may allow a person to disavow personal responsibility for the lapse.[50] Future smoking cessation interventions should emphasize the importance of avoiding high-risk situations to those making quit attempts. These results may also provide support for the need of smoke-free policies for workplaces, restaurants, and bars that would enable smokers trying to quit to reduce exposure to high-risk cues.

The two statistical models generated by the machine learning algorithms in the present study satisfied different goals. The penalized model provided shrunken regression coefficients (some all the way to zero), reducing issues related to multicollinearity and eliminating the weakest predictors. The subsequent approximated model further reduced the number of predictors to a much smaller set of covariates, maximizing the interpretability of the model (as there are fewer variable relationships to understand) at the cost of 6.1% of variance explained. Neither model should be considered to be the correct model; rather, each algorithm provides a unique model that may be useful in different contexts. To the extent that future samples feature participants with similar demographics to the present sample, the penalized model may be more appropriate. In particular, the penalized model identified several predictors with high-magnitude coefficients that were not selected by the approximated model, including interacting with someone who is smoking, having a strong desire to smoke, and being confident to avoid smoking. These predictors merit greater consideration among similar populations, and highlight the usefulness of the penalized model over and above the approximated model in providing a more complete picture for predicting first lapse. Alternatively, the higher parsimony approximated model sacrifices complexity in favor of generalizability, and therefore may be of greater utility in dissimilar samples.

Smoking interventions may be improved through consideration of the present findings. First, the systematic machine learning method can be used to refine future EMA interventions by limiting lengthy assessments to only the most pertinent variables related to the outcome. Studies that use EMA typically impose substantial burden on participants due to the intensive, repeated nature of sampling,[16] and shorter assessments would likely increase participation and compliance. Second, future interventions may be improved by focusing on the strongest predictors of lapse, such as encouraging skills to cope with risky situations, environments, and feelings of irritability. Interventions may range from simple reminders (i.e., promote awareness of the risks associated with particular situations) to contingency management reinforcement targeted at smoking and/or the situational risk factors associated with smoking such as alcohol use. Further, smartphones used to collect EMA data may be used to push targeted interventions in real-time. For example, Businelle et al. developed a mobile app that delivered automated smartphone messages tailored to the participant's current level of lapse risk and lapse triggers.[51] Similarly, Naughton et al.[52] developed a smartphone app that delivered smoking cessation messages based on geolocation of a participant's own high-risk smoking areas.

The current study is an important first step toward the development of highly sophisticated and innovative interventions that can simultaneously identify and intervene on multiple antecedents of smoking lapse. For example, these analyses allow for the efficient examination of within-subject differences in lapse risk and the inclusion of more potential antecedents than is possible with traditional methods. As a result, future interventions could be highly personalized based on specific cognitive, affective, or behavioral patterns or GPS coordinates. These types of interventions may be especially valuable for smokers of low SES, who despite having lower quit rates,[4,5] have high rates of smartphone ownership.[53]

The primary limitation to the present study included limited external validity due to the specificity and size of the sample. The majority of the sample (82.7%) had a total annual household income of less than $20 000; thus, the results may not be generalizable to smokers of higher SES. For example, low SES smokers may experience different social and environmental contexts that have a stronger influence on smoking behavior. Research has demonstrated that smokers of low SES do not experience the same social pressures not to smoke as high SES smokers,[11] and that their workplace environments, including higher stress, or a lack of smoking policies, may also contribute to the risk of smoking lapse.[10] Further, the EMA data collection was limited to 7 days post-quit, and the presence of incentives to participate in the study may have influenced results. Future research may address these limitations by including more direct measures of smoking during the quit attempt, lengthening the data collection period, and assessing participant responses in the absence of incentives. Researchers may also pursue interventions that target the strongest predictors of smoking lapse for individuals attempting to quit in real-time. Additionally, the present analysis could not establish the temporal precedence necessary to assess causality. While the longitudinal structure of the data resulted in the use of all data prior to a given lapse, the type of survival modeling performed here necessitated usage of data that was also concurrent with the lapse itself. Future research may attempt to uncover patterns in the data that can address causality. Finally, the present research found that financial incentives such as contingency management (as opposed to usual care) may affect time to lapse in the presence of the other predictors selected by the penalized and approximated models. Future research should further explore these relationships for the purpose of better predicting imminent lapse and possibly intervening before lapse occurs.

Supplementary Material

Supplementary data is available at *Nicotine & Tobacco Research* online.

Funding

This work was supported by The University of Texas Health Science Center School of Public Health and American Cancer Society grants MRSGT-12-114-01-CPPB (to MSB) and MRSGT-10-104-01-CPHPS (to DEK).

Declarations of Interests

None declared.

References

1. U.S. Department of Health and Human Services. *The Health Consequences of Smoking—50 Years of Progress. A Report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2014.
2. Jha P, Ramasundarahettige C, Landsman V, et al. 21st-Century hazards of smoking and benefits of cessation in the United States. *N Engl J Med*. 2013;368(4):341–350.
3. Centers for Disease Control and Prevention. Quitting smoking among adults--United States, 2001–2010. *MMWR Morb Mortal Wkly Rep*. 2011;60(44):1513–1519.
4. Gilman SE, Abrams DB, Buka SL. Socioeconomic status over the life course and stages of cigarette use: initiation, regular use, and cessation. *J Epidemiol Community Health*. 2003;57(10):802–808.
5. Kotz D, West R. Explaining the social gradient in smoking cessation: it's not in the trying, but in the succeeding. *Tob Control*. 2009;18(1):43–46.
6. Reid JL, Hammond D, Boudreau C, Fong GT, Siahpush M; ITC Collaboration. Socioeconomic disparities in quit intentions, quit attempts, and smoking abstinence among smokers in four western countries: findings from the International Tobacco Control Four Country Survey. *Nicotine Tob Res*. 2010;12 (Suppl):S20–S33.
7. Businelle MS, Kendzor DE, Reitzel LR, et al. Mechanisms linking socioeconomic status to smoking cessation: a structural equation modeling approach. *Health Psychol*. 2010;29(3):262–273.
8. Ma P, Businelle MS, Balis DS, Kendzor DE. The influence of perceived neighborhood disorder on smoking cessation among urban safety net hospital patients. *Drug Alcohol Depend*. 2015;156:157–161.
9. Businelle MS, Kendzor DE, Reitzel LR, et al. Pathways linking socioeconomic status and postpartum smoking relapse. *Ann Behav Med*. 2013;45(2):180–191.
10. Albertsen K, Hannerz H, Borg V, Burr H. The effect of work environment and heavy smoking on the social inequalities in smoking cessation. *Public Health*. 2003;117(6):383–388.
11. Paul CL, Ross S, Bryant J, Hill W, Bonevski B, Keevy N. The social context of smoking: a qualitative study comparing smokers of high versus low socioeconomic position. *BMC Public Health*. 2010 April 27;10:211.

12. Piasecki TM, Fiore MC, McCarthy DE, Baker TB. Have we lost our way? The need for dynamic formulations of smoking relapse proneness. *Addiction*. 2002;97(9):1093–1108.
13. Piasecki TM, Baker TB. Any further progress in smoking cessation treatment? *Nicotine Tob Res*. 2001;3(4):311–323.
14. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol*. 2008;4:1–32.
15. Schüz N, Cianchi J, Shiffman S, Ferguson SG. Novel technologies to study smoking behavior: Current developments in ecological momentary assessment. *Curr Addict Rep*. 2015;2(1):8–14.
16. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. *Psychol Assess*. 2009;21(4):486–497.
17. Shiffman S, Balabanis MH, Gwaltney CJ, et al. Prediction of lapse from associations between smoking and situational antecedents assessed by ecological momentary assessment. *Drug Alcohol Depend*. 2007;91(2–3):159–168.
18. Businelle MS, Ma P, Kendzor DE, Frank S, Wetter DW, Vidrine DJ. Using intensive longitudinal data collected via mobile phone to detect imminent lapse in smokers undergoing a scheduled quit attempt. *J Med Internet Res*. 2016;18(10):e275.
19. Businelle MS, Ma P, Kendzor DE, et al. Predicting quit attempts among homeless smokers seeking cessation treatment: an ecological momentary assessment study. *Nicotine Tob Res*. 2014;16(10):1371–1378.
20. Shiffman S, Gwaltney CJ, Balabanis MH, et al. Immediate antecedents of cigarette smoking: an analysis from ecological momentary assessment. *J Abnorm Psychol*. 2002;111(4):531–545.
21. Shiffman S, Paty JA, Gnys M, Kassel JA, Hickcox M. First lapses to smoking: within-subjects analysis of real-time reports. *J Consult Clin Psychol*. 1996;64(2):366–379.
22. Shapiro D, Jamner LD, Davydov DM, James P. Situations and moods associated with smoking in everyday life. *Psychol Addict Behav*. 2002;16(4):342–345.
23. Beckham JC, Wiley MT, Miller SC, et al. Ad lib smoking in post-traumatic stress disorder: an electronic diary study. *Nicotine Tob Res*. 2008;10(7):1149–1157.
24. Watkins KL, Regan SD, Nguyen N, et al. Advancing cessation research by integrating EMA and geospatial methodologies: associations between tobacco retail outlets and real-time smoking urges during a quit attempt. *Nicotine Tobacco Res*. 2014; 16(Suppl 2):S93–101.
25. Reitzel LR, Cromley EK, Li Y, et al. The effect of tobacco outlet density and proximity on smoking cessation. *Am J Public Health*. 2011;101(2):315–320.
26. Lam CY, Businelle MS, Aigner CJ, et al. Individual and combined effects of multiple high-risk triggers on postcessation smoking urge and lapse. *Nicotine Tob Res*. 2014;16(5):569–575.
27. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. *Ann Behav Med*. 2016:1–17.
28. McClernon FJ, Roy Choudhury R. I am your smartphone, and I know you are about to smoke: the application of mobile sensing and computing approaches to smoking research and treatment. *Nicotine Tob Res*. 2013;15(10):1651–1654.
29. Holt LJ, Litt MD, Cooney NL. Prospective analysis of early lapse to drinking and smoking among individuals in concurrent alcohol and tobacco treatment. *Psychol Addict Behav*. 2012;26(3):561–572.
30. Bandiera FC, Atem F, Ma P, Businelle MS, Kendzor DE. Post-quit stress mediates the relation between social support and smoking cessation among socioeconomically disadvantaged adults. *Drug Alcohol Depend*. 2016 June 1;163:71–76.
31. Deiches JF, Baker TB, Lanza S, Piper ME. Early lapses in a cessation attempt: lapse contexts, cessation success, and predictors of early lapse. *Nicotine Tob Res*. 2013;15(11):1883–1891.
32. Shiffman S, Waters AJ. Negative affect and smoking lapses: a prospective analysis. *J Consult Clin Psychol*. 2004;72(2):192–201.
33. Piasecki TM, Trela CJ, Hedeker D, Mermelstein RJ. Smoking antecedents: separating between- and within-person effects of tobacco dependence in a multiwave ecological momentary assessment investigation of adolescent smoking. *Nicotine Tob Res*. 2014;16(Suppl 2):S119–S126.
34. Kuhn M, Johnson K. *Applied predictive modeling*. New York: Springer; 2013.
35. Shiffman S. Conceptualizing analyses of ecological momentary assessment data. *Nicotine Tob Res*. 2014;16(Suppl 2):S76–S87.
36. Dumortier A, Beckjord E, Shiffman S, Sejdić E. Classifying smoking urges via machine learning. *Comput Methods Programs Biomed*. 2016 December;137:203–213.
37. Kendzor DE, Businelle MS, Poonawalla IB, et al. Financial incentives for abstinence among socioeconomically disadvantaged individuals in smoking cessation treatment. *Am J Public Health*. 2015;105(6):1198–1205.
38. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. New York: Springer; 2009.
39. Goeman J. Penalized R package, version 0.9–50. 2017.
40. R Core Team. R: A language and environment for statistical computing. 2016. http://www.R-project.org/. Accessed May 24, 2017.
41. Ambler G, Brady AR, Royston P. Simplifying a prognostic model: a simulation study based on clinical data. *Stat Med*. 2002;21(24):3803–3822.
42. Harrell F. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. New York, NY: Springer; 2015.
43. Akaike H. Information theory and an extension of the maximum likelihood principle. Second International Symposium on Information Theory. 1973:267–281.
44. Venables W, Ripley B. *Modern Applied Statistics Using S*. New York, NY: Springer; 2002.
45. Therneau T. A Package for Survival Analysis in S. version 2.38. 2015. http://CRAN.R-project.org/package=survival. Accessed May 24, 2017.
46. Borucka J. *Extensions of cox model for non-proportional hazards purpose*. PhUSE, Paper SP07. Warsaw, Poland: PAREXEL. 2013.
47. Cooney NL, Litt MD, Cooney JL, Pilkey DT, Steinberg HR, Oncken CA. Alcohol and tobacco cessation in alcohol-dependent smokers: analysis of real-time reports. *Psychol Addict Behav*. 2007;21(3):277–286.
48. Piasecki TM, McCarthy DE, Fiore MC, Baker TB. Alcohol consumption, smoking urge, and the reinforcing effects of cigarettes: an ecological study. *Psychol Addict Behav*. 2008;22(2):230–239.
49. Piasecki TM, Jahng S, Wood PK, et al. The subjective effects of alcohol-tobacco co-use: an ecological momentary assessment investigation. *J Abnorm Psychol*. 2011;120(3):557–571.
50. Marlatt GA, Gordon JR. *Relapse Prevention: Maintenance Strategies in the Treatment of Addictive Behaviors*. New York, NY: Guilford Press; 1985.
51. Businelle MS, Ma P, Kendzor DE, Frank SG, Vidrine DJ, Wetter DW. An ecological momentary intervention for smoking cessation: evaluation of feasibility and effectiveness. *J Med Internet Res*. 2016;18(12):e321.
52. Naughton F, Hopewell S, Lathia N, et al. A context-sensing mobile phone app (Q Sense) for smoking cessation: a mixed-methods study. *JMIR Mhealth Uhealth*. 2016;4(3):e106.
53. Anderson M. Technology Device Ownership, 2015. 2016. http://www.pewinternet.org/2015/10/29/technology-device-ownership-2015/. Accessed May 24, 2017.