SMOKING

Educational attainment and cigarette smoking: a causal association?†

Stephen E Gilman,1,2* Laurie T Martin,3 David B Abrams,4 Ichiro Kawachi,1 Laura Kubzansky,1 Eric B Loucks,5 Richard Rende,6 Rima Rudd1 and Stephen L Buka7

Accepted 28 November 2007

Background Despite abundant evidence that lower education is associated with a higher risk of smoking, whether the association is causal has not been convincingly established.

Methods We investigated the association between education and lifetime smoking patterns in a birth cohort established in 1959 and followed through adulthood (n = 1311). We controlled for a wide range of potential confounders that were measured prior to school entry, and also estimated sibling fixed effects models to control for unmeasured familial vulnerability to smoking.

Results In the full sample of participants, regression analyses adjusting for multiple childhood factors (including socioeconomic status, IQ, behavioural problems, and medical conditions) indicated that the number of pack-years smoked was higher among individuals with less than high school education [rate ratio (RR) = 1.58, confidence interval (CI) = 1.31, 1.91]. However, in the sibling fixed effects analysis the RR was 1.23 (CI = 0.80, 1.93). Similarly, adjusted models estimated in the full sample showed that individuals with less than high school education had fewer short-term (RR = 0.40; CI = 0.23, 0.69) and long-term (RR = 0.59; CI = 0.42, 0.83) quit attempts, and were less likely to quit smoking (odds ratio = 0.34; CI = 0.19, 0.62). The effects of education on quitting smoking were attenuated in the sibling fixed effects models that controlled for familial vulnerability to smoking.

Conclusions A substantial portion of the education differential in smoking that has been repeatedly observed is attributable to factors shared by siblings that contribute to shortened educational careers and to lifetime smoking trajectories. Reducing disparities in cigarette smoking.

† A previous version of this work was presented at the Society for Epidemiologic Research/North American Congress of Epidemiology, June 2006, Seattle, WA, USA.

* Corresponding author. Department of Society, Human Development, and Health, 677 Huntington Avenue, Boston, MA 02115, USA. E-mail: sgilman@hsph.harvard.edu

1 Department of Society, Human Development, and Health, Harvard School of Public Health, Boston, MA, USA.
2 Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA.
3 Child Trends, Washington, DC, USA.
4 Office of Behavioral and Social Science Research, National Institutes of Health, Bethesda, MD, USA.
5 Department of Psychiatry and Department of Epidemiology, McGill University, Montreal, Canada.
6 Department of Psychiatry and Human Behavior, Brown Medical School and Butler Hospital, Providence, RI, USA.
7 Department of Community Health, Brown University, Providence, RI, USA.
smoking, including educational disparities, may therefore require approaches that focus on factors early in life that influence smoking risk over the adult life span.

**Keywords**  Education, smoking, causality, disparities

---

**Introduction**

Educational differences in smoking were not observed prior to the discovery that smoking causes death. However, following the 1964 Surgeon General’s report and subsequent public health campaigns, researchers established a large educational difference between those who do and do not smoke. In the United States, for example, the annual decline in smoking prevalence between 1974 and 1985 in the National Health Interview Survey was approximately five times higher among the most educated than among the least educated. The gap in smoking rates between individuals with the highest and lowest levels of education is greater than at any time in the past. Educational differences in the risk for daily smoking initiation, development of nicotine dependence, and in smoking cessation contribute to this gap.

A causal effect of education on smoking could provide an additional input into policy decisions regarding tobacco control. However, the possibility that the association between education and smoking is not causal casts doubt on the public health relevance of studies showing a link between education and smoking. This possibility was raised as early as 1982 by Farrell and Fuchs, who concluded that because smoking patterns were established by age 17, they could not be influenced by years of schooling. They argued instead that ‘there are apparently one or more ‘third variables’ that affect both smoking and years of schooling’, (p. 228) and hypothesized that such ‘third variables’ include intelligence and time discount, e.g. the propensity to refrain from smoking in anticipation of future health benefits. Few of the numerous studies reporting on educational differences in smoking attempted to rule out these or other alternative explanations for the association between lower educational attainment and higher rates of smoking.

Establishing causality in the absence of experimental data is a virtually intractable problem; however, data arising from ‘natural experiments’ can be used to overcome some of the limitations of observational studies when it is not feasible to conduct a randomized trial. One such natural experiment occurs in the context of a family study, by comparing outcomes of ‘discordant siblings’. In this design, siblings with different levels of education (i.e. discordant for education) are compared with respect to their smoking outcomes; this design removes the confounding effects of factors that are shared by both siblings. This includes both environmental and genetic factors that siblings have in common, but not factors such as individual-specific experiences in the social environment or genetic susceptibility that is not shared with siblings. The objectives of the current study are to investigate the association between education and smoking using analyses adjusting for potential confounders measured prior to school entry, and using sibling fixed effects models that adjust for unmeasured familial vulnerability.

**Methods**

Participants were offspring of pregnant women enrolled in the National Collaborative Perinatal Project (NCPP) between 1959 and 1966. The original aims of the NCPP were to identify the developmental consequences of pregnancy and delivery complications. Detailed social and medical histories were obtained from mothers at the time of enrolment. Information on offspring birth outcomes and subsequent growth and development was obtained several times during the first year of life, and again at ages 4 and 7.

The New England Family Study (NEFS) was established to locate and interview the adult NCPP offspring at the Providence, Rhode Island and Boston, Massachusetts sites (n = 17921) between 2001–04. Participants in the current study were selected through a multi-stage sampling procedure as part of the Brown-Harvard Transdisciplinary Tobacco Use Research Center, which involved a core assessment interview and three component studies. Screening questionnaires were mailed to 4579 of the 15721 Boston and Providence NCPP offspring who survived until age 7. Of the 3121 questionnaires returned (68.2%), 2271 were eligible for participation based on the combined inclusion criteria of the three component studies. In total, we enrolled 1674 NCPP offspring. Participants enrolled in the NEFS had a somewhat higher level of education (e.g. 64.1% with at least some college education) than participants who were eligible but not enrolled (e.g. 51.8% with at least some college education). Data from 49 individuals were excluded from the final sample because of participation in a pilot version of the survey (n = 4) or because of problems with the interview administration (n = 45). This yielded 1625 completed adult assessments. The analysis sample for the current
study was restricted to participants who reported having smoked at least once in their lifetime and had complete data on all key study variables.

**Measures**

**Educational attainment**

Education was assessed during the NEFS follow-up interview and was classified according to five categories: (i) less than high school or GED; (ii) high school degree; (iii) high school degree plus additional technical training or certificate; (iv) some college and (v) college degree.

**Smoking**

Smoking histories were obtained by the Life Interview of Smoking Trajectories and Quitting Methods Questionnaire, developed by the Methods and Measurement core of the Brown-Harvard Transdisciplinary Tobacco Use Research Center. These instruments obtain detailed information on participants’ experiences with smoking beginning from experimentation, progression to regular smoking, levels of consumption, nicotine dependence and patterns of quit attempts. Regular smoking was defined as a positive response to the question ‘Did you ever become a weekly smoker (that is, smoke at least once per week for two months or longer)?’ We created a summary measure of cigarette consumption using data on participants’ smoking intensity and duration during their heaviest smoking phase; similar to measures of ‘pack-years’, this was calculated as the number of years of participants’ heaviest smoking phase × number of cigarettes per day/20. Nicotine dependence was defined according to Diagnostic and Statistical Manual, Fourth Edition, criteria (DSM-IV), and was assessed using the Composite International Diagnostic Interview (CIDI). Smoking desistance (among participants who became regular smokers) was defined by the number of 24 h quit attempts, number of 3 month abstinence periods and smoking cessation. Cessation was coded as positive if participants did not smoke during the year preceding the interview. Participants also provided information on their ages of onset of regular smoking and nicotine dependence symptoms, and age of smoking cessation.

**Potential confounders**

Information on early childhood factors was collected during the NCPP upon the mother’s enrollment and again during the 7-year follow-up interview. Measures of parental socioeconomic status from both assessments include occupation, educational attainment and an indicator of household income below the United States poverty threshold. Additional characteristics of early childhood environment between birth and age 7 were household crowding, family disruption by age 7, number of moves and father’s unemployment. Maternal smoking during pregnancy, defined as the maximum number of cigarettes smoked per day, was coded as none, less than one pack (1–19 cigarettes), and a pack or more (20+ cigarettes).

Measures of early childhood physical health included birthweight, number of medical conditions during the child’s first year of life and history of asthma by age 7. Psychological development included full-scale IQ at age 7, psychologist’s abnormal behaviour rating at age 7 and the presence of neurological soft signs at 7 years.

**Analysis procedures**

Analyses were conducted among participants who had smoked at least once in their lifetime, and from whom complete data were available on the early childhood measures obtained during the NCPP. The study was limited to lifetime smokers because experimentation with cigarettes was pervasive in the sample (91.1%), and because our primary concern was with smoking intensity and frequency rather than lifetime abstinence.

Discrete-time survival analysis was used to investigate the development of nicotine dependence (among regular smokers) and smoking cessation (also among regular smokers). Educational attainment was coded as a time-varying covariate, with person-years prior to school completion coded as ‘in-school’; the coefficients for education therefore reflect the effects of participants’ final educational attainment on nicotine dependence and smoking cessation. Analyses of cigarette consumption (pack-years) and quit attempts were conducted in the sample of regular smokers using negative binomial regression. Regression coefficients from this model, when exponentiated, indicate the ratio of the number of quit attempts (or rate of consumption) associated with a unit change in each covariate.

For each outcome, we present the results of two sets of analyses: (i) analyses based on the full analysis sample and (ii) analyses of siblings discordant for education, i.e. individuals with a sibling in the NEFS follow-up study who had a different level of education. For the survival and negative binomial regressions in the full sample, a random intercept for each family was included to account for the interdependence of data from siblings using the GLIMMIX procedure in SAS. In the discordant sibling analyses, we used conditional logistic regression to estimate discrete-time survival models, and estimated fixed effects negative binomial models of count data. These approaches adjust the effect estimates for all between-family variation in the smoking outcomes.

**Results**

Of the 1625 participants who were interviewed, 1445 (88.9%) provided data on all childhood covariates.
A total of 92.1% of these participants \((n = 1311)\) reported lifetime smoking, and therefore comprised the analysis sample for the current study. A comparison of demographic characteristics between the full interviewed sample of 1625 and the analysis sample of 1331 lifetime smokers is shown in Table 1. The samples are similar with respect age, sex, race/ethnicity and the number of siblings per family. The mean (SD) age of the analysis sample is 39.1 years (1.8); the sample is 59.5% females \((n = 780)\), and 84.0% Whites \((n = 1101)\). 10.8% of the sample has less than a high school education \((n = 142)\), while one-third has a college degree \((n = 365)\). The number of siblings in the full and analysis samples is also shown in Table 1. The analysis sample represents 1036 families; 793 participants did not have a sibling in the study, whereas the remaining 518 participants represent 243 families. The age range of siblings is an approximate indicator of the extent of shared environmental experiences in early childhood, as siblings closer in age are likely to have more shared experiences than siblings further apart. The age difference between siblings in the analysis sample was quite narrow [mean (SD): 2.2 (1.3) years].

Patterns of smoking uptake and quitting are shown in Table 2. A total of 63.3% of the sample progressed to regular smoking \((n = 826)\). Among regular smokers, 66.1% \((n = 506)\) subsequently met diagnostic criteria for nicotine dependence and 43.1% \((n = 355)\) quit smoking. The majority of participants who became regular smokers did so in the teen years [mean age (SD): 15.9 (4.1)], with the mean (SD) onset of nicotine dependence occurring 9.4 (7.0) years afterwards. The mean age (SD) of smoking cessation among regular smokers who had quit by the adult follow-up interview was 27.7 (6.7). On average, regular smokers reported smoking 8.4 (9.6) pack-years, made 9.6 (19.6) quit attempts that lasted at least 24 h, and achieved 1.6 (2.1) periods of abstinence that lasted at least 3 months. The remaining columns in Table 2 show the proportion of participants with missing data for each outcome, the number of participants included in each analysis, and the distribution of the smoking outcomes among siblings discordant for educational attainment, New England Family Study.

### Table 1

Demographic characteristics of participants in the New England Family Study (NEFS) and participants in the analysis sample

<table>
<thead>
<tr>
<th></th>
<th>All NEFS participants (n = 1625)</th>
<th>Analysis sample (n = 1311)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent ((n))</td>
<td>Percent ((n))</td>
</tr>
<tr>
<td><strong>Age, Mean (SD)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>39.1 (1.9)</td>
<td>39.1 (1.8)</td>
</tr>
<tr>
<td><strong>Female sex</strong></td>
<td>59.2 (962)</td>
<td>59.5 (780)</td>
</tr>
<tr>
<td><strong>Race/ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>83.5 (1354)</td>
<td>84.0 (1101)</td>
</tr>
<tr>
<td>Black</td>
<td>9.6 (155)</td>
<td>9.1 (119)</td>
</tr>
<tr>
<td>Other</td>
<td>7.0 (113)</td>
<td>6.9 (91)</td>
</tr>
<tr>
<td><strong>Educational attainment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than high school or GED</td>
<td>11.4 (185)</td>
<td>10.8 (142)</td>
</tr>
<tr>
<td>High school degree</td>
<td>13.7 (222)</td>
<td>13.8 (181)</td>
</tr>
<tr>
<td>High school plus additional training</td>
<td>16.1 (262)</td>
<td>16.7 (219)</td>
</tr>
<tr>
<td>Some college</td>
<td>30.0 (488)</td>
<td>30.8 (404)</td>
</tr>
<tr>
<td>College graduate</td>
<td>28.8 (468)</td>
<td>27.8 (365)</td>
</tr>
<tr>
<td><strong>Number of siblings per family</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>56.8 (923)</td>
<td>60.5 (793)</td>
</tr>
<tr>
<td>2</td>
<td>34.5 (560)</td>
<td>32.8 (430)</td>
</tr>
<tr>
<td>3</td>
<td>7.2 (117)</td>
<td>5.7 (75)</td>
</tr>
<tr>
<td>4</td>
<td>1.2 (20)</td>
<td>0.6 (8)</td>
</tr>
<tr>
<td>5</td>
<td>0.3 (5)</td>
<td>0.4 (5)</td>
</tr>
</tbody>
</table>

### Table 2

Distribution of cigarette smoking outcomes and patterns of missing data in the full sample, and the sample of siblings discordant for educational attainment, New England Family Study

<table>
<thead>
<tr>
<th>Smoking uptake</th>
<th>Full analysis sample</th>
<th>Discordant siblings only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD) or per cent ((n))</td>
<td>Sample size</td>
</tr>
<tr>
<td>Onset of regular smoking(^a)</td>
<td>63.3 (826)</td>
<td>1305</td>
</tr>
<tr>
<td>Onset of nicotine dependence(^b)</td>
<td>66.1 (506)</td>
<td>765</td>
</tr>
<tr>
<td>Mean cigarette consumption (\text{pack-years})(^b)</td>
<td>8.4 (9.6)</td>
<td>787</td>
</tr>
<tr>
<td>Smoking desistance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean 24 h quit attempts(^b)</td>
<td>9.6 (19.6)</td>
<td>785</td>
</tr>
<tr>
<td>Mean 3 month abstinence periods(^b)</td>
<td>1.6 (2.1)</td>
<td>807</td>
</tr>
<tr>
<td>Smoking cessation(^b)</td>
<td>43.1 (355)</td>
<td>824</td>
</tr>
</tbody>
</table>

\(^a\)The prevalence of regular smoking was estimated in the sample of lifetime smokers \((n = 1311)\).

\(^b\)Estimated in the sample of regular smokers.
discordant siblings. While the distributions of the smoking outcomes in the sample of discordant siblings are similar to those in the full analysis sample, the sample size is smaller.

**Smoking trajectories in relation to educational attainment**
The majority of participants who became regular smokers (84.7%, \( n = 700 \)) did so prior to completing school. Therefore, we do not present analyses of the relation between education and smoking initiation. Among participants who became regular smokers, there was no association between educational attainment and the development of nicotine dependence according to DSM-IV criteria. In contrast, among participants who became regular smokers, there was a strong education gradient in lifetime smoking patterns (Table 3). Participants without high school degrees smoked \( \sim 50\% \) more pack-years than participants with college degrees; rate ratios (RRs) in the full sample were 1.63 (1.38–1.92) in Model I (unadjusted), and 1.58 (1.31–1.91) in Model II (controlling for participant demographic factors and childhood covariates assessed during the NCPP). However, the RR for the effect of low education on pack-years was attenuated in the sibling fixed effects model [1.23, confidence interval (CI): 0.80–1.93].

Short-term (24 h) quit attempts and long-term (3 month) periods of abstinence were less frequent among participants with a high school degree or less, with approximately half as many quit attempts as those participants who graduated from college. Adjusted rate ratios (and 95% CI) for short- and long-term quit attempts were 0.40 (0.23–0.69) and 0.59 (0.42–0.83), respectively. The effects of less than high school educational attainment on short-term (RR: 1.04; CI: 0.39–2.82) and long-term (RR: 0.61; CI: 0.26–1.46) quit attempts were attenuated in the sibling models. We observed a similar pattern in the analyses of smoking cessation, wherein low education was associated with a lower odds of cessation in Models I (OR: 0.22; CI: 0.13–0.37) and II (OR: 0.34; CI: 0.19–0.62), but not among discordant siblings in Model III.

**Discussion**
The objective of the current study was to evaluate evidence for a potential causal effect of educational attainment on lifetime patterns of cigarette smoking. This association has emerged repeatedly from epidemiologic studies in the United States, Europe, and several emerging market economies. The relevance of this evidence for public health and social policy will remain uncertain until researchers can establish whether educational attainment is itself a causal factor in the aetiology of cigarette smoking rather than alternative factors that confer vulnerability to attain less education and to smoke.

We evaluated whether there is evidence consistent with a causal association between education and smoking patterns by controlling for a wide range of potential confounders that were measured prior to school entry and by analysing differences in the smoking outcomes of siblings with different levels of education. After controlling for measured confounders, lower education was associated with more pack-years of smoking, fewer quit attempts and a lower likelihood of cessation. The effect of education was characterized by a higher risk of smoking frequency/intensity among participants without high school degrees compared to those who had graduated college. However, after controlling for between-family variation in smoking outcomes in the sibling fixed effects analyses, evidence for an effect of education on smoking was substantially weaker. CIs were also wider among discordant siblings due to the smaller number of participants in these analyses.

Among smokers, there was no association between education and nicotine dependence. The absence of an association between education and nicotine dependence among regular smokers in the NEFS sample is consistent with findings of Breslau et al. from the National Comorbidity Survey, but is different from the study of Hu et al. of a much younger sample of smokers from the National Longitudinal Survey of Adolescent Health. Inconsistencies may be due to age differences between the samples and to differences in the conceptualization of nicotine dependence.

Results from this study underscore the importance of differentiating the processes underlying the progression to regular smoking and the pathways to desistance. This necessitates a developmentally sensitive analysis that takes into account the age distributions of each smoking transition. While aspects of the school environment may contribute to the development of regular smoking, it is unlikely that level of completed education was a major contributor to smoking initiation given the early onset of smoking. However, once regular smoking behaviours were established, our initial results indicated that educational attainment had a significant impact on adult smoking trajectories as indexed by pack-years, quit attempts and cessation. Adjusting for a wide range of factors measured prior to school entry (e.g. parental socioeconomic status, maternal smoking, IQ) had little impact on the effect estimates for educational attainment. Rather, the attenuation of the effects of education in the discordant sibling analyses suggests that other, unmeasured factors operating at the family level (i.e. shared by siblings), contribute to the relation between education and smoking behaviours.
<table>
<thead>
<tr>
<th>Educational attainment</th>
<th>Cigarette consumption&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Number of 24-hr quit attempts&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Number of 3-month abstinence periods&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Smoking cessation&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model I&lt;sup&gt;d&lt;/sup&gt; (Full sample) RR (95% CI)</td>
<td>Model II&lt;sup&gt;e&lt;/sup&gt; (Discordant siblings) RR (95% CI)</td>
<td>Model III&lt;sup&gt;f&lt;/sup&gt; (Discordant siblings) RR (95% CI)</td>
<td>Model I&lt;sup&gt;d&lt;/sup&gt; (Full sample) RR (95% CI)</td>
</tr>
<tr>
<td>Less than high school or GED</td>
<td>1.63 (1.38, 1.92)</td>
<td>1.58 (1.31, 1.91)</td>
<td>1.23 (0.80, 1.93)</td>
<td>0.38 (0.23, 0.62)</td>
</tr>
<tr>
<td>High school degree</td>
<td>1.46 (1.24, 1.72)</td>
<td>1.41 (1.19, 1.67)</td>
<td>1.28 (0.91, 1.81)</td>
<td>0.70 (0.43, 1.14)</td>
</tr>
<tr>
<td>High school plus additional training</td>
<td>1.43 (1.23, 1.67)</td>
<td>1.34 (1.15, 1.57)</td>
<td>1.10 (0.78, 1.56)</td>
<td>0.69 (0.44, 1.09)</td>
</tr>
<tr>
<td>Some college</td>
<td>1.36 (1.19, 1.55)</td>
<td>1.33 (1.16, 1.53)</td>
<td>1.22 (0.93, 1.60)</td>
<td>0.58 (0.39, 0.87)</td>
</tr>
<tr>
<td>College graduate</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<sup>a</sup>Cigarette consumption defined as the number of pack-years smoked during heaviest smoking phase. RRs obtained from negative binomial regression models, offset by the logarithm of the number of years’ duration of heaviest smoking phase.

<sup>b</sup>Rate ratios obtained from negative binomial regression models.

<sup>c</sup>Odds ratios obtained from discrete-time survival models including person-years at risk from age at regular smoking through age at cessation or censoring/age at interview.

<sup>d</sup>Model I includes a random intercept for each sibling set.

<sup>e</sup>Model II includes a random intercept for each sibling set, and adjusts for age, sex, race/ethnicity, age at regular smoking, parental occupation, parental education, household poverty, household crowding, parental separation, residential instability, paternal unemployment, maternal smoking during pregnancy, IQ, abnormal behaviour ratings, neurological soft signs, birthweight, medical conditions during first year of life and childhood asthma.

<sup>f</sup>Model III is a fixed-effects/conditional analysis among siblings discordant for education, with adjustment for age, sex, age at regular smoking, maternal smoking during pregnancy, IQ, abnormal behaviour ratings, neurological soft signs, birthweight, medical conditions during first year of life and childhood asthma.

<sup>g</sup>Less than high school or GED and high school degree categories combined because of sparse data.
Limitations

Our analyses rely on retrospective reports of lifetime smoking histories. Self-reported smoking status has a high level of agreement with serum cotinine-defined smoking status.\textsuperscript{45,46} The correspondence between self-reported smoking and serum cotinine concentrations has not been found to differ according to education or other indicators of socioeconomic status.\textsuperscript{45,47} With respect to cigarette consumption, studies comparing consumption levels reported retrospectively to those obtained prospectively show a moderate to high level of agreement.\textsuperscript{18,48,49} Fewer studies examined retrospectively reported ages of smoking initiation, although there is some evidence in support of \textsuperscript{50,51} and against\textsuperscript{52} the reliability of such reports.

The NCPP cohort was not designed to be a representative sample of all births in Rhode Island and Massachusetts, and the adult offspring of NCPP participants who were included in the current study on the basis of several layers of inclusion criteria cannot be considered a representative sample of adults from this geographic area. The discordant sibling analyses are limited by small sample size, thereby reducing the precision of effect estimates. As these analyses rely solely on within-family variation, ignoring all variation in smoking outcomes between-families, the consequences of lower sample size on statistical power are potentially substantial.

While information was available on parental socioeconomic status and early childhood physical and cognitive development, information was not available on all potential confounders. To the extent that these potential confounders represent vulnerability to smoking that is shared by siblings, their effects were accounted for in the fixed effects analyses. The discordant sibling design does not control for exposures that are not shared by the siblings. A potentially relevant class of non-shared factors includes those that cause siblings to attain different levels of education. However, controlling for covariates such as IQ, medical conditions and parental employment in these analyses strengthens the discordant sibling design, as these would be a likely source of sibling differences in education.

Implications of our findings for understanding the relation between education on smoking

Our analyses of education differentials in smoking in the full sample are consistent with prior studies showing strong protective effects of schooling on smoking behaviours; these effects remained even after adjusting for a wide range of potential confounding factors measured prior to school entry. The fixed effects models that we estimated among siblings in the NEFS suggest that some portion of the effect of education on smoking is due to unmeasured familial vulnerability. Several prior studies\textsuperscript{53–55} have also attempted to control for unobserved confounders. Using the method of instrumental variables, they reported statistically significant effects of education on smoking, although the magnitude of the effects was substantially smaller than in standard regression models. The validity of the instrumental variable approach rests on identifying appropriate instruments—in this case, factors that influence smoking exclusively through education. Sander,\textsuperscript{53} however, used family background characteristics as instruments, which we have shown to predict smoking independent of education.\textsuperscript{11} It is unclear whether using stronger instruments would lead to a further attenuation of the effect of education on smoking. In contrast, the fixed effects method used in our study reduces confounding due to shared sibling vulnerability, but suffers from a loss of efficiency because estimates are based solely on within-family variation in smoking behaviours. Taken as a whole, we are hesitant to conclude that there is no causal effect of education on smoking solely on the basis of the fixed effects analyses in our study, given the strong effects that emerged from the full sample analyses and the limitations of the sibling models that we describe earlier. It is therefore important to identify, and estimate the magnitude of effects of, the specific factors that represent a common source of vulnerability to attain lower levels of education and to persist in smoking through adulthood.

Explanations for education differentials in smoking behaviours include differential valuations of the health consequences of smoking,\textsuperscript{54,56} and differences in access to and effectiveness of cessation treatments.\textsuperscript{57} Lower education may also confer risk for persistent smoking due to lower occupational status and financial strains,\textsuperscript{36} and be a more frequent mechanism for coping with stress among individuals with lower levels of education.\textsuperscript{58}

Evidence that aspects of the school\textsuperscript{59} and neighbourhood environments\textsuperscript{60} increase risk for smoking among adolescents suggests that school experiences are relevant in addition to the educational degree attained. Relevant aspects of the school environment that could contribute to the development of smoking include school smoking policies, affiliation with deviant peers, and economic status.\textsuperscript{61,62} Evidence that smokers have shortened educational careers raises the possibility of a reciprocal effect of educational achievements and smoking behaviours over time.\textsuperscript{63,64}

Inequalities in smoking by educational attainment are a major contributor to educational inequalities in mortality.\textsuperscript{33,65,66} Therefore, tobacco control efforts could have a significant impact on reducing health disparities. The identification of lower education as a marker of ‘hard core’ smoking, both in community and clinical samples,\textsuperscript{57} suggests that multiple types of tobacco control efforts will be needed, including community-wide programs,\textsuperscript{68} workplace\textsuperscript{69} and school interventions.\textsuperscript{70}
Acknowledgements

The authors acknowledge collaborators on the Brown-Harvard TUTURC for the development of the Life Interview of Smoking Trajectories, Quitting Methods Questionnaire and Socioeconomic Status Interview (David Abrams, Laurie Chassin, Melissa Clark, Suzanne Colby, Raymond Niaura, Edmond Shenassa and Saul Shiffman). We sincerely appreciate the efforts and contributions of Michelle Rogers and Kathleen McGaffigan for data management and programming. This research was supported in part by a Transdisciplinary Tobacco Use Research Center (TUTURC) Award (P50 CA084719) and grant AG023397 (Buka) from NIH, and by the Robert Wood Johnson Foundation.

Conflict of interest: None declared.

---

KEY MESSAGES

- There is abundant evidence that cigarette smoking is more prevalent among individuals with lower education.
- It remains unresolved whether low education is a causal risk factor for smoking, or whether educational disparities in smoking are attributable to factors that confer risk for low education and for smoking.
- We investigated the relation between education and smoking in a birth cohort followed through middle adulthood, controlling for a wide range of potential confounders measured prior to school entry. We also estimated fixed effects models to control for unmeasured confounders shared by siblings.
- In fully adjusted models, we observed education differentials in cigarette consumption, frequency of quit attempts, and likelihood of quitting that were similar to prior reports. These differentials were reduced in sibling models controlling for unmeasured confounders.
- A portion of the education differential in smoking is attributable to factors shared by siblings that contribute to shortened educational careers and to lifetime smoking trajectories.

References


45 Vartiainen E, Seppala T, Lillsunde P, Puska P. Validation of self reported smoking by serum cotinine measurement in a community-based study. *J Epidemiol Community Health* 2002;56:167–70.


Commentary: Why are we biased against bias?

Jay S Kaufman

Accepted 31 January 2008

Greater attention to causal inference has been one of the most important trends in social epidemiology over the last decade. The groundwork was laid 35 years ago by Mervyn Susser’s book 'Causal Thinking in the Health Sciences', but growing interest more recently in causal techniques such as potential outcomes models and directed graphs has given the field new capacities for strengthening inference and honing arguments. Many techniques that have been standard in econometrics and the social sciences for years have made their way into social epidemiology in the last decade, including multilevel modeling, propensity score matching and instrumental variables. One such technique, exploited cleverly in the article by Gilman and colleagues, is the fixed effects regression model.

Epidemiologists have long been enthusiastic users of the same conditional estimator used for fixed effects analyses, but only in the context of the matched case-control study. Rather than have a case and a control matched by study design, however, the tradition in social sciences has been to consider exposed and unexposed observations matched by nature.