A Systematic Review of Observational Studies, Demonstrating Smoking among School Going Adolescents

Shafquat Rozi1*, Gillian Lancaster2, Sadia Mahmud3, Zahid Ahmad Butt4, Talat Roome5, Nida Zahid6

1Department of Community Health Sciences, Aga Khan University, Karachi, Pakistan
2Department of Mathematics and Statistics, Lancaster University, Lancashire, United Kingdom
3Visiting Faculty, Department of Medicine Consultancy, Department of Pediatrics & Child Health, Aga Khan University, Karachi, Pakistan
4School of Population and Public Health, University of British Columbia, Vancouver, Canada
5Department of Pathology, Dow University of Health Sciences, Karachi, Pakistan
6Department of Surgery, Aga Khan University, Karachi, Pakistan

Email: *shafquat.rozi@aku.edu

Abstract

Objective: To characterize the methods of design and analysis currently adopted in survey research of school-based observational studies for smoking, and to identify the common pitfalls made by researchers. Methods: The systematic review was conducted in 2009 and consisted of observational studies in school settings published between January 2005 and January 2009. Smoking status was the main outcome of interest. Following Cochrane style, five steps were followed: setting selection criteria for studies and conducting a literature search; review of abstracts; review of complete articles; data extraction and quality assessment of included studies; and, finally, synthesis of studies. Results: Of the 292 abstracts retrieved, 45 (15.4%) articles were selected for the final review. Inconsistencies were found in the definition of smoking behaviour which impeded generalisability. Individual-level factors had importance, but environmental level factors were also important in studying the aetiology of smoking. Results showed that studies inappropriately reported sample size estimation and important confounding factors. Hierarchical linear modelling, random effects modelling and structural equation modelling were employed in comparatively few studies. Conclusions: There were concerns regarding data analysis of complex surveys. Fifty five percent of reviewed studies ignored environmental effects which may have produced unreliable inferences. Multi-level analysis assisted in understanding school-level effects.
1. Introduction

Tobacco is one of the greatest public health issues of modern times. It consumes a considerable amount of resources of the healthcare system in Pakistan for both treatment and prevention [1]. Currently, about 5.4 million deaths occur every year due to tobacco use [1]. More than 8 million tobacco deaths are expected by 2030 if urgent action is not taken. The dilemma is that 80% of tobacco-related deaths within a few decades will occur in the developing countries [1]. This devastating shift to the developing countries is due to the fact that the global tobacco industry is targeting young adults [2]. Currently, 150 million adolescents are tobacco users worldwide. According to recent studies, school environment has an impact on outcomes of adolescent behaviour that include substance abuse and committing crime [3]. A study from Karachi, Pakistan, reported prevalence of tobacco use among school-going children at 16.1% [4] [5].

Studies from Pakistan have documented the association of betel quid, areca and tobacco use with head and neck cancers [6] [7]. Although tobacco kills far more people than human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS), unfortunately funding for tobacco control is less than that is required for its prevention [8]. About 27 countries have imposed tax rates of more than 75% of the retail price of tobacco to control its use [9]. There is no data available to measure the expenditure for treating diseases caused by tobacco use in Pakistan.

Mostly, survey-based studies have been designed to discuss health and smoking issues which are carried out either in communities, clinics, hospitals or schools. To consider environmental/school-level factors affecting adolescent smoking, surveys often employ multi-stage cluster sampling that would incur a hierarchical structure of the population.

If hierarchical data is analysed using analytical techniques that will not consider clustering of data, this will produce misleading inferences and conclusions regarding the association of smoking with other predictors. A researcher may need to consider intra-class correlation (ICC) in the sample size calculation, if accounting for clustering in the structure of the study. About 55% studies have revealed that a multi-level modelling approach shows that the environmental factors have considerable contribution in variation of smoking prevalence between schools besides individual-level factors [10]. A study suggested that there was more similarity in children within school compared to between schools [10].

The current review study was planned to determine the methodological approaches in recent surveys, and to identify the common pitfalls in the methodology, especially design and analysis, in school-based observational studies for teenage smoking.
2. Material and Methods

The systematic review was conducted in 2009 and comprised of observational studies in school settings published between January 2005 and January 2009 with smoking status as the main outcome of interest. Following the Cochrane methodology [11], five steps were followed: setting selection criteria for studies and conducting a literature search; review of the abstracts; review of the complete articles; data extraction and quality assessment of included studies; and, finally, synthesis of studies. The focus was not on pooling estimates.

Eligibility criteria

In the first step, the eligibility criteria for the inclusion of the studies and the search strategy to be used for selection of the studies for literature review were defined.

Inclusion criteria

The search was limited to peer review published studies in English language. All the studies selected for this systematic review were observational; cohort retrospective, prospective, case-control, longitudinal and cross-sectional with population aged 11 - 18 years in school settings from both the developed and the developing world with smoking status as the main outcome of interest.

Exclusion Criteria

We excluded non-peer reviewed articles, letters to editors, conference proceedings and articles published in a language other than English. All studies done on adolescents smoking in hospitals or clinics or communities were also excluded from the review. If the main outcome of interest of the study was something other than smoking then the study was not included in the systematic review.

Search Strategy for Identification of Studies

A comprehensive literature search was performed in both general databases, such as PubMed free access database, Embase, and subject-specific databases such as: PSYCINFO. Moreover, to retrieve publications reporting observational studies for smoking among teenage children attending school, we performed a combined search strategy that included the following terms as both medical subject heading (MeSH) terms and text words: Adolescent in MeSH; Adolescents; Adolescents, Female; Adolescents, Male; Teenagers; Teens; Youth; 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8; Schools in MeSH; Primary Schools; Schools, Secondary; Secondary Schools; 10 or 11 or 12 or 13; Smoking in MeSH; Cigarette Smoking; Smoking Tobacco; Tobacco Smoking; Tobacco Smoking Pollution; 15 or 16 or 17 or 18 or 19; and 9 and 14 and 20.

Filters were also used for observational studies during the literature search defined by the SIGN group and University of York [12]. This filtered the extracts of the observational studies from the retrieved lists after the POL search which is the population (adolescents), outcome (smoking) and location (secondary schools). Moreover, in the second step articles that were fulfilling our eligibility criteria were extracted (Figure 1).
This resulted in selection of articles for second-level review of the eligibility of the retrieved full papers. One investigator carried out the second level review but uncertainty about inclusion of studies was resolved with the other reviewer. To validate the inclusion of studies, two reviewers independently reviewed the inclusion process on a 10% sample selected randomly by the principal investigator. Finally, 100% agreement was achieved on inclusion status.

In the 3rd step one investigator extracted data for the included studies on a pre-designed data-abstraction form. Most questions on the form were pre-coded. Validation of data-abstraction was performed by selecting a random sample of the included studies for the second reviewer to check (5 papers). In case of a disagreement, the relevant field was checked on all abstraction forms and a mutually agreed description was achieved. Data were primarily extracted to assess design and analysis, including whether there was a reported sample size calculation or if analysis had been conducted appropriately or not. We also assessed whether the investigators had provided details of sample size calculation and analysis considering clustering of data and intra-class correlation or not. Place of publication, date of publication, population characteristics, detailed description of outcomes, study design, sampling strategy and detail of individual and school-level factors relating to adolescents smoking were also extracted. The research team discussed any discrepancies.

Both internal and external validity were assessed during methodological quality assessment. The interpretation of the findings of a study depended on design, conduct and assessment (internal validity), as well as on populations and outcome measures (external validity). In the final step, the findings of the literature review were synthesized and the key points were summarised in tabular form.

**Assessment of quality of study**

The quality of each study was assessed independently. Any disagreements on
study quality were resolved through discussion with the other reviewer to confirm whether or not the study had achieved the quality dimension [13] [14]. The quality of reviewed studies based on internal and external validity was assessed which aimed at ascertaining whether the methods used in the design and conduct of the studies were likely to prevent systematic error. For this purpose, study design, method of analysis, sample size calculation, whether the researchers had discussed the limitations of their study or not, selection bias, consideration of potential confounders of the study (age, gender, socio-economic status), the quality of reporting, the generalizability of the results, conclusion, and recommendations based on study results were the factors assessed.

3. Results

Table 1 presents the methodological characteristics of the 45 studies, including: Setting, school type and grades, sample size, study design, and the sampling strategy. Of the 292 abstracts retrieved, 45 (15.4%) articles were selected for final review. About 32.4% of the published studies were conducted in the United States, followed by Canada (10.8%), China (8.6%), the United Kingdom (8.6%), and India (8.6%), while the remaining 32.8% were in other countries including Malaysia, Norway, Scotland, Lebanon, Taiwan, Australia, Sudan, Iceland, Brazil, New Zealand, Belgium, Greece, Netherland and Germany. The countries involved in joint projects were the USA, the UK, Greece, Romania, Denmark, Israel and other European countries.

The studies included were conducted in public and private schools, vocational schools, and technical education schools. However, most of the reviewed studies did not clearly state the proportion of public and private schools that they had included in their study. The studies under review included 77.7% cross-sectional and 22.3% prospective cohort or longitudinal follow-up studies. However, in some longitudinal studies (8.8%), the length of follow-up was not reported.

Twenty percent of studies comprised two-stage or multi-stage cluster sampling, 11.1% studies employed random sampling with stratification, 13.3% studies reported simple random sampling, 15.5% reported convenience sampling and 15.5% derived data from other projects for secondary analysis and did not clearly state their sampling strategy. However, 11.3% studies reported other sampling techniques such as systematic sampling, and single-stage cluster sampling, while the rest of the studies (13.3%) did not discuss sampling strategy.

The reviewed studies determined samples from schools in specific regional areas, whereas few studies obtained samples at national level. Some schools were selected using systematic random sampling. Mostly school-based studies conducted surveys on a particular school day. Those adolescents who were absent on the day of the interview or had dropped out of school were excluded, which may have biased the results and generalisability of the findings.

Sample sizes ranged from 384 to 91,778 school-children. About 20% studies employed single-stage or multi-stage cluster sampling. However, clustering was
Table 1. Summary of methodological aspects and analytical issues of studies reviewed.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Country &amp; Setting</th>
<th>Study design</th>
<th>Sampling technique &amp; sample size</th>
<th>Method of analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bidstrup PE et al. 2008 [15]</td>
<td>Denmark, 118 Public schools 230 classes, 2 classes per school with 13 pupils per class</td>
<td>Cross-sectional 2004</td>
<td>Random sampling of schools (Data derived from Danish national institute of Public health in 2008) 1426 boys 1487 girls</td>
<td>Multilevel logistic regression</td>
</tr>
<tr>
<td>Bird Y et al. 2006 [9]</td>
<td>Juancz, Mexico, 6 Public &amp; private schools. Data Collection year was not mentioned.</td>
<td>Cross-sectional study</td>
<td>Random sampling stratified by SES to select schools. All 6th grade students were selected 242 boys 264 girls</td>
<td>Multiple logistic regression by SES, school setting &amp; gender</td>
</tr>
<tr>
<td>Bird Y et al. 2007 [16]</td>
<td>Ciudad Juarez, Mexico,6 Public &amp; private schools in 2000</td>
<td>Cross-sectional study</td>
<td>Random sampling stratified by SES to select schools. All 6th grade students were selected 242 boys 264 girls</td>
<td>Chi-squared test, ANOVA and logistic regression</td>
</tr>
<tr>
<td>Bond L et al. 2007 [17]</td>
<td>Victoria, Australia, 26 Secondary schools. Year of data collection not mentioned.</td>
<td>Longitudinal study</td>
<td>Data derived from Gatehouse project 1124 boys 1276 girls</td>
<td>Prevalence estimates, &amp; Logistic regression using robust” “information sandwich” estimates of standard errors to account for clustering within schools.</td>
</tr>
<tr>
<td>Carson NJ et al. 2005 [18]</td>
<td>Northern Virginia Philadelphia, USA, 5 Public high schools. Year of data collection not mentioned.</td>
<td>Cross-sectional</td>
<td>Data derived from another Longitudinal study of behavioural predictors of smoking, 967 adolescents</td>
<td>SEM, EFA, CFA, model fit assessed</td>
</tr>
<tr>
<td>Chang F et al. 2006 [19]</td>
<td>Taipei, Taiwan, 16 Vocational high schools 39 classes from 2000-2002</td>
<td>Longitudinal study 3 years</td>
<td>Probability proportionate to size systematic random sampling to draw schools and classes. All 10th grade students selected 1695 boys and girls</td>
<td>GEE</td>
</tr>
<tr>
<td>Conti DV et al. 2006 [21]</td>
<td>China, 62 Middle schools 83 Upper schools from Oct 2002-Dec 03</td>
<td>Longitudinal study</td>
<td>Multistage cluster sampling stratified by school type 5042 boys 5606 girls</td>
<td>Multilevel random coefficient modeling</td>
</tr>
<tr>
<td>Damianaki A et al. 2008 [22]</td>
<td>Semi urban area in Crete, Greece, 15 Public schools from 2004-05</td>
<td>Cross-sectional</td>
<td>Random sampling stratified by rural/urban status to select schools and students 386 boys 507 girls</td>
<td>Stepwise logistic regression, Mann-Whitney, t-test</td>
</tr>
<tr>
<td>Difranza JR et al. 2008 [23]</td>
<td>Urban and suburban region in central Massachusetts, USA, 2 Schools in 2007</td>
<td>Cross-sectional</td>
<td>Convenience sampling 1055 boys and girls</td>
<td>2-tailed Kendall’s tau b test for correlations</td>
</tr>
<tr>
<td>Study</td>
<td>Country/Region</td>
<td>Type of Study</td>
<td>Data Source and Time Period</td>
<td>Methodology and Analysis</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------------</td>
<td>------------------------</td>
<td>-----------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>Droomers M et al. 2005 [25]</td>
<td>New Zealand,</td>
<td>Longitudinal study</td>
<td>Data derived from Dunedin multidisciplinary health behavior project</td>
<td>Longitudinal Logistic GEE</td>
</tr>
<tr>
<td></td>
<td>Number and type of schools not stated, April 1972, and March 1973</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>El-Roueiheb Z et al. 2008 [27]</td>
<td>Beirut, Lebanon, 3 Public, 10 private schools from 2003-04</td>
<td>Cross-sectional</td>
<td>Random sampling stratified by district within Beirut to select schools. All students were interviewed from schools</td>
<td>Multivariable regression analysis with missing data</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1089 boys, 1328 girls</td>
<td></td>
</tr>
<tr>
<td>Ellickson PL et al. 2008 [28]</td>
<td>California and Oregon, USA, 30 Secondary schools from 2000-2005</td>
<td>Longitudinal study over years,</td>
<td>Data derived from ALERT project</td>
<td>Multivariate regression</td>
</tr>
<tr>
<td></td>
<td>North Carolina USA, Wave 1: 8 middle schools, 2 comprehensive k-8 schools, 3 alternative schools Wave 2: 6 high schools, Middle &amp; high grade from 2002-2004</td>
<td></td>
<td>51960 adolescents for smoking analysis. 2000 for problem behaviour analysis</td>
<td></td>
</tr>
<tr>
<td>Ennett ST et al. 2008 [29]</td>
<td>Seychelles, Indian ocean, 10 public schools 2 private schools, from global youth tobacco survey</td>
<td>Cross-sectional</td>
<td>Two stage cluster sampling design to obtain students Schools: probability proportional to school size Student: equal probability sampling</td>
<td>Percentage or proportion and CI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>620 boys, 654 girls</td>
<td></td>
</tr>
<tr>
<td>Faeh D et al. 2006 [30]</td>
<td>Cincinnati, Ohio, USA, 1 public high &amp; 1 junior high public schools, date of study not mentioned. Hawaii, 5 states Oahu, the big island of Hawaii, Maui, Molokai and Kawi, 22 Public &amp; private schools from 2000-2002</td>
<td>Longitudinal study</td>
<td>Convenience sampling Wave 1: 5220 boys and girls Wave 5: 5017</td>
<td>Three level hierarchical growth models, Bayesian information criterion (BIC), Random intercept and random effect model</td>
</tr>
<tr>
<td>Finkelstein DM et al. 2006 [31]</td>
<td>China 7 cities, Hanbin, Shenyang, Wuhan, Chengdu, Kunming, &amp; Hangzhou, Qingdao, 147 Middle schools &amp; high schools, date of study not mentioned.</td>
<td>Cross-sectional</td>
<td>Convenience sampling 844 boys 851 girls</td>
<td>Chi-squared test, ANOVA with Tukey’s test, logistic regression analysis’ and GEE</td>
</tr>
<tr>
<td>Glanz K et al. 2007 [32]</td>
<td>China 7 cities, Hanbin, Shenyang, Wuhan, Chengdu, Kunming, &amp; Hangzhou, Qingdao, 147 Middle schools &amp; high schools, date of study not mentioned.</td>
<td>Longitudinal study</td>
<td>Convenience sampling 844 boys 851 girls</td>
<td>Chi.-square &amp; t-test, GEE, PROC logistic regression</td>
</tr>
<tr>
<td>Grenard JL et al. 2006 [33]</td>
<td>6 countries in Norway including Oslo, 2 Southern inland 3 Northern countries, Public and private schools from 2000-04</td>
<td>Longitudinal study 1 year</td>
<td>Multistage cluster sampling 6991 boys 7387 girls</td>
<td>Chi-squared test t-test, paired t-test, Multilevel analysis &amp; ICC, generalized linear mixed models</td>
</tr>
<tr>
<td>Grotvd L et al. 2008 [34]</td>
<td>Convenience sampling 7762 boys 7768 girls</td>
<td>Cross-sectional</td>
<td>Convenience sampling 7762 boys 7768 girls</td>
<td>Generalized linear model with binominal distribution</td>
</tr>
<tr>
<td>Study ID</td>
<td>Country/Region</td>
<td>Study Design</td>
<td>Sample Characteristics</td>
<td>Methodology</td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>--------------</td>
<td>------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Gutschoven K et al. 2005 [35]</td>
<td>Flanders, Belgium, 15 Schools. Data collected from Leuven Study on Media and Adolescent Health (SOMAH)</td>
<td>Cross-sectional</td>
<td>All students were selected from 1st &amp; 4th year (Data derived from the media and adolescent health project)</td>
<td>ANOVA with post hoc Tukey's test</td>
</tr>
<tr>
<td>Henrikson L et al. 2008 [38]</td>
<td>California USA, 135 High schools in 2005-06</td>
<td>Cross-sectional</td>
<td>Random sampling of schools</td>
<td>Multiple regression for complex sampling design</td>
</tr>
<tr>
<td>Hublet A et al. 2006 [39]</td>
<td>5 European countries (Bulgaria, Croatia Greece, Romania, Slovenia) and UK. High schools in 1999</td>
<td>Cross-sectional</td>
<td>Cluster sampling of schools &amp; classes</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Kokkevi A et al. 2006 [40]</td>
<td>9 European countries, 2 UK counties and Canada, Number and types of schools not specified. Study from 1990-2002</td>
<td>Cross-sectional</td>
<td>Two stage sampling stratified by school type 16445 boys and girls</td>
<td>Survey logistic regression</td>
</tr>
<tr>
<td>Kristjansson AL et al. 2009 [41]</td>
<td>Iceland, Secondary schools, in March 2006.</td>
<td>Cross-sectional</td>
<td>9th-10th grade students of all Icelandic secondary schools</td>
<td>Pearson’s correlation &amp; logistic regression</td>
</tr>
<tr>
<td>Lee LK et al. 2005 [44]</td>
<td>Malaysia 7 districts, Sermban, Port Dickson, Kualapilah, Jempaltampin&amp; Remban, 14 Schools, from June 2001-August 2002</td>
<td>Cross-sectional</td>
<td>Multistage stratified random sampling, 6 schools from urban, 8 schools from rural 2 classes per school</td>
<td>Chi-squared test &amp; CI</td>
</tr>
<tr>
<td>Mathur C et al. 2008 [46]</td>
<td>Delhi (n = 16) and Chennai (n = 16), India, 32 Public &amp; private school in 2004</td>
<td>Cross-sectional</td>
<td>Convenience sampling 11,642 boys &amp; girls</td>
<td>Mixed effects regression model</td>
</tr>
<tr>
<td>Study Authors and Year</td>
<td>Location</td>
<td>Sample Description</td>
<td>Data Collection Design</td>
<td>Methods</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------</td>
<td>--------------------</td>
<td>------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Muilenburg JL et al. 2008 [47]</td>
<td>Southeastern USA</td>
<td>6 Public schools in 2006</td>
<td>Cross-sectional</td>
<td>Ordered logit regression</td>
</tr>
<tr>
<td>Poulin CC et al. 2007 [49]</td>
<td>Atlantic province (Nova Scotia, New Brunswick, Prince Edward Island, Newfoundland &amp; Labrador, Secondary schools in 2002</td>
<td>Single stage cluster sample stratified by grade and either health regional or school district</td>
<td>Cross-sectional</td>
<td>Two level random effect model, Sensitivity analysis was performed using multinomial logistic regression</td>
</tr>
<tr>
<td>Rodrigue D et al. 2007 [51]</td>
<td>USA, 5 Public schools. Data collected in 2003.</td>
<td>Longitudinal study over 4 years</td>
<td>Not stated</td>
<td>SEM, CFA</td>
</tr>
<tr>
<td>Schnoer CW et al. 2009 [52]</td>
<td>Denmark, Sweden, Norway, Finland, and UK secondary schools 2001-2002</td>
<td>Data derived from a project (The health behaviour in school children) 20,399 boys and girls</td>
<td>Cross-sectional</td>
<td>Binary logistic regression analysis stratified by country</td>
</tr>
<tr>
<td>Sinha DN et al. 2007 [53]</td>
<td>India (North, Eastern, Western, North East), 179 schools in 2006</td>
<td>Two stage cluster sampling Schools and classes selected randomly. All students selected within classes 12086 boys and girls</td>
<td>Cross-sectional</td>
<td>Logistic regression for clustered data</td>
</tr>
<tr>
<td>Sinha DN et al. 2008 [54]</td>
<td>India, 6 regions: North, South, East, West, Central &amp; North East, 2003: 818 schools 2006:179 schools,</td>
<td>Two stage cluster sampling Schools selected proportional to number of students, and classes selected randomly. All students interviewed from selected classes 2003: 68,077 boys and girls 2006: 12,086 boys and girls</td>
<td>Cross-sectional</td>
<td>Logistic regression for correlated data</td>
</tr>
<tr>
<td>Thrasher JF et al. 2008 [55]</td>
<td>Mexico,18 Public &amp; private schools in 2006</td>
<td>Proportional sampling strategy to randomly select schools &amp; classes 3874 boys and girls</td>
<td>Cross-sectional</td>
<td>Multilevel adjusted logistic regression model, SEM</td>
</tr>
<tr>
<td>Wen X et al. 2007 [56]</td>
<td>Huang pu and Guangzhou, China, 6 Secondary schools in 2004 Guangzhou, China, 3 Public schools, 1 Factory school 2 General paid private schools in 2004</td>
<td>Cluster sampling of schools 2021 boys 1850 girls</td>
<td>Cross-sectional</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Wen X et al. 2008 [57]</td>
<td></td>
<td>Multistage sampling with stratification by school type 3957 boys and girls and 2870 parents of these students</td>
<td>Cross-sectional</td>
<td>Chi-squared &amp; Kruskal Wallis</td>
</tr>
</tbody>
</table>


not taken into account during sample size estimation by all the 45 studies. Such studies that did not take clustering into account in sample size calculation suffered from a considerable loss of power since the design effect was multiplied by the
sample size calculated under simple random sampling to account for clustering.

However, we made no judgment regarding sample size calculations for studies which did not mention clustering pattern.

The definitions of adolescents’ tobacco use varied from study to study and, hence, it was not possible to aggregate them together. Smoking use among adolescents was categorised into five levels or stages 1) experimentation, 2) regular use, 3) occasional use, 4) frequent use, and 5) heavy use across studies. About 75% of reviewed studies did not discuss level or stage of tobacco use but they explained whether or not adolescents had ever smoked and/or were currently smoking.

While most of the studies applied simple definitions of smoking behaviour, it is imperative to recognise that some research studies also reported progression across stages of tobacco use.

Few studies (4.4%) were purely descriptive (i.e. prevalence, rates or means). Most of the studies reviewed employed traditional methods of analysis e.g., 8.8% performed a t-test, 28.8% multivariable logistic regression and linear regression, 11.1% analysis of variance and 17.7% a chi-squared test. Also, 55% studies correctly accounted for clustering and performed hierarchical linear modelling, random effects modelling, and generalized estimating equation (GEE). Although in the 55% of studies appropriate statistical techniques were used to answer the research questions, some studies employed longitudinal samples, but this was not reflected in their analysis, reflecting lack of competence in advanced statistical methods.

Different statistical software packages were used to analyse the data. Most studies used Statistical Package for the Social Science (SPSS) (28.8%) and Statistical Analysis System (SAS) (24.4%). Some studies reported using STATA (17.7%), Multilevel software (MLWIN) (8.8%), or Survey data analysis (SUDAAN) (4.4%) and 4.4% of studies reported the use of MPLUS for statistical analysis. However, 20% studies did not mention the type of software used by them for their analysis.

Approximately 73.3% of reviewed studies reported about missing information in their datasets. About 17.7% of studies imputed missing values or used a regression technique with imputation and about 28.8% of them excluded the missing information from their analysis. However, 33.2% studies did not discuss anything about the missing data in their studies. About 11.1% of studies had no missing information.

The median prevalence of smoking across studies was 13.1% interquartile range [IQR] = 11.4). The demographic variables included in the review studies were adolescent age and gender, family, socio-economic status and education, and family smoking. The risk factor most commonly associated with teenage smoking leading them to become habitual or regular smokers were senior students smoking in school, peer smoking, public schools, school suspension, retailer near the school, higher level of exposure to movies, pupil’s relation and attitude to school, school size, social environment of the school, and quality of
school teachers. In schools where anti-smoking policies were well implemented, the ratio of adolescents smoking to those who did not was lower compared to those schools where no tobacco monitoring policies were present. Good grades, school connectedness, and knowledge about public policies were prominent school-level protective factors that restricted adolescents from smoking.

The assessment of potential confounders and risk factors is vital. In most of the studies, smoking status of family members was considered. Some studies reported the effect of individual (mother’s and father’s) smoking habits on adolescents’ smoking, while a few observed the combined effect of parents’ smoking on adolescent smoking. Furthermore, the effect of sibling smoking was also assessed in some studies. A few studies combined parent, siblings and other relatives smoking as family member smoking.

The association between parents or family members smoking and adolescent smoking suggested that parental smoking influenced the initiation of adolescent smoking. When family members of adolescent were using tobacco, they had easy access to use it and it made them believe that it was socially acceptable.

Adolescents whose friends were smokers were more likely to smoke compared to those with no peer smoking. The average age of smoking in reviewed studies was approximately 13.1 years. At this age usually peer pressure or peer relationships become stronger than family relationships, so adolescents were more likely to be influenced by the behaviour of friends.

It was found that only a few studies fulfilled the criteria of good quality study. We scored studies according to their assessed validity, thus, the more valid a study was the higher the score it received. The quality of the reviewed studies was below average in approximately one-third (37%) of studies, 48% were marked as average, and 15% as good according to the criteria of strobe [14] and consort guidelines for randomised controlled trials (RCTs) [13]. This reflects an inadequate understanding of the substantive issues underpinning hierarchical data in a survey research.

4. Discussion

Discrepancies were observed in the definition of smoking in the reviewed studies and the tools that were used to measure adolescent smoking were not standardised, which may hinder generalisability. A number of studies were designed to assess the social influences of adolescent smoking along with individual-level covariates. Although we found different study designs, analytical approach and covariates, our results indicate that environmental level factors are as crucial as individual factors for studying the aetiology of smoking.

In all the reviewed studies, the questionnaire was self-administered by students who not only reported their own smoking habits but also the smoking habits of their parents, family and friends who currently smoke cigarettes. However, it was difficult to report accurately the frequency and number of cigarettes used by their family and friends. Studies reported that parents, siblings and peers
were powerful influencers for adolescent smoking.

As in most of the studies, the interviews were conducted on one particular day, hence it was expected that a few students might be absent on that particular day thus not including them in the study may have produced biased results and may also have an effect on generalisability.

Causal inferences cannot be determined from cross-sectional studies. The predictors of smoking reported by cross-sectional studies were not markedly different from that reported by longitudinal studies.

Confounders and biases are major concerns in observational studies, which need to be addressed appropriately. However, most of the studies failed to measure important confounding factors. In most of the studies, clustering was not considered during sample size calculation which may have caused a reduction in the power. In conclusion, the results of this review highlight concerns about the analysis of complex survey data. Hierarchical linear modelling, random effects modelling and structural equation modelling were used in comparatively fewer studies. About 45% studies ignored the contextual/environmental (e.g. retailer near the school, social environment of school and implementation of smoking policies at school) factors in their studies which may have produced misleading inferences.

The results of reviewed studies guided us in comprehending the effect of school-level factors and variances between schools. Moreover, it also provided us with more intuitive information on school-level covariates that can have an influence on the adolescents smoking behaviour.

One of the limitations of our study is that we only reviewed studies in English. This could have resulted in some studies being excluded from our review.

It is recommended that future studies should consider environmental/social settings from where the individuals are drawn at the design and analysis stage as well as taking clustering into account to come up with a leading inference to deal with public health issues.

5. Conclusion

There were different study designs, analytical approaches and covariates in the studies reviewed, but results indicate that environmental level factors are as crucial as individual factors for studying the aetiology of smoking.

Disclaimer

None.

Source of Funding

None.

Conflicts of Interest

None.
References


[12] Centre for Review and Dissemination University of York. https://www.crd.york.ac.uk/CRDWeb


among Adolescents. Cancer Epidemiology Biomarkers & Prevention, 17, 1862-1871. https://doi.org/10.1158/1055-9965.EPI-07-2773


https://doi.org/10.1007/s11121-008-0087-8

https://doi.org/10.1186/1471-2458-6-169

https://doi.org/10.1016/j.jadohealth.2006.04.011

https://doi.org/10.1080/13557850701234948

https://doi.org/10.1080/14622200600576610

https://doi.org/10.1186/1471-2458-8-322

https://doi.org/10.1080/146220050125260

https://doi.org/10.1186/1471-2458-8-218

https://doi.org/10.1016/j.amepre.2007.02.025

https://doi.org/10.1016/j.ypmed.2008.04.008

https://doi.org/10.1186/1471-2458-6-280

https://doi.org/10.1016/j.drugalcdep.2006.05.018

https://doi.org/10.1111/j.1651-2227.2008.01133.x

in the School Environment and Smoking Onset in Elementary School Students. 
*Cancer Epidemiology Biomarkers & Prevention*, **14**, 1762-1765. 
https://doi.org/10.1158/1055-9965.EPI-05-0065

Schools and Cigarette Access Behaviors among Underage Smoking Students. 
*Annals of Behavioral Medicine*, **33**, 105-111. 
https://doi.org/10.1207/s15324796abm3301_12

School Students in Negeri Sembilan, Malaysia. *Asia-Pacific Journal of Public 

of School Smoking Policies and Student Perceptions of Enforcement on School 
Smoking Prevalence and Location of Smoking. *Health Education Research*, **22**, 
782-793. https://doi.org/10.1080/1462200701767779

https://doi.org/10.1080/14622200701767779

https://doi.org/10.1016/j.ypmed.2007.01.003

Alcohol Review*, **26**, 615-624. https://doi.org/10.1080/09595230701613619

https://doi.org/10.1093/her/cym085

https://doi.org/10.1097/PSY.0b013e31802e0fe


https://doi.org/10.1016/j.amepre.2008.03.036