Teenage smoking: why it is important to evaluate reported evidence

Said Shahtahmasebi, PhD.
Good Life Research Centre Trust, Christchurch, New Zealand. radisolevoo@gmail.com

Keywords: adolescent health related behaviour, smoking, evidence-based,
Received: 31/10/2018; Revised: 5/11/2018; Accepted: 10/11/2018

Introduction

In a series of papers published in the Dynamics of Human Health (DHH) (Berridge, 2014a, 2014b, 2014c) (being republished in DHH, issues 2 & 4 of Vol 5 and issue 1 of Vol 6) the importance of adopting an appropriate statistical model was highlighted. In this article I demonstrate that simply applying an appropriate analytical technique to a set of data does not necessarily make evidence. Furthermore, I will discuss the key issues, statistical concepts and substantive theory that is necessary in evaluating and interpreting results. These applications are illustrated using teenage smoking from an adolescent health related behaviour survey study.

Data

The data comes from a secondary source. In 1992, around 60 secondary schools in the former Yorkshire Regional Health Authority geographical boundary (UK), agreed to take part in a health behaviour related survey repeated every two years. Years 9 and 11 (age range 11-16 years) in these schools were surveyed using a health related behaviour questionnaire [Balding, 1992 #3]. The survey questionnaires were anonymous. The questionnaire covered topics related to the attitudes and behaviour of the pupils with regard to health e.g. physical exercise and out of school activities, nutrition, social contacts, dealing with problems, attitudes to and the use of drugs (including smoking and drinking). The outcome variable "smoking" is a dichotomous dummy variable and takes the value "0" if a non-smoker (those who claimed they had never smoked or had given up smoking) and the value "1" if a smoker (those who claimed they smoked either regularly/occasionally). A full analysis of this data set is reported elsewhere (Shahtahmasebi, 2003).

Statistical modelling

In this case the outcome is the smoking status of students with two possible outcomes; smoker (1) or non-smoker (0). Therefore, an appropriate model to fit to these data is the logistic model.

An important issue arises when exploring the effect of a variable on smoking. In order to assess the effect of a variable on smoking, e.g. age, the effect of other variables such as sex, household composition, and peer pressure must be accounted for. A forward iterative process for model selection was adopted. The relationship between smoking and variables were examined one at a time and the variable with the smallest p-value was entered in the model. Then the process was repeated with the remaining variables. Again, the variable with the smallest p-value was selected to enter the model. Those variables which were not significant at 5% significance level were excluded from the subsequent round. The process of elimination continued until there were no variables left significant at the 5% significance level.

Results

The results of model selection are shown in Table 1. Based on these results it is quite easy to conclude that "best friend" has the largest effect on smoking followed by the variables "have partner"; "how feel with opposite sex"; "which parent live with"; and the two "worrying" variables. Results in Table 2 also suggest that compared to a reference group the odds of being a smoker increased for those who claimed their best friend smokes (OR 14.16, 95% CI 12.23-16.36), for those who claimed to have a partner (e.g. up to 6 months OR 3.41, 95% CI 2.36-4.92), those who claimed to be at ease with the opposite sex (OR 1.89, 95% CI 1.47-2.49), those who claimed to worry a lot about money problems (OR 1.71, 95% CI 1.43-2.05) and those who claimed to worry a lot about family problems (OR 1.37, 95% CI 1.15-1.62). On the other hand, the odds of being a smoker decreased for those who claimed to be happy with their body shape (OR 0.80, 95% CI 0.69-0.93) and those who claimed to consider health often when choosing food (OR 0.44, 95% CI 0.30-0.63). Pupils who lived with foster parents appear to have an increased risk of nearly four times that of pupils who lived with both parents.

The result for "which parent live with" can be explained as past behaviour leading to a selection bias; it is plausible that smoking may well have started while in care prior to placement with foster parents (Royal College of Physicians, 1992).

Table 1. Standard logistic regression: odds ratios for the model of smoking prevalence with their appropriate 95% Confidence Limits after controlling for other factors

Explanatory variables	Lower	Odds Ratio	Upper
Age			
12-13	1.00	1.00	1.00
14-15	1.59	1.84	2.13
Sex			
male	1.00	1.00	1.00
female	1.68	1.96	2.30
Which parent live with			
both parents	1.00	1.00	1.00
mother only	0.92	1.13	1.40
father only	0.92	1.45	2.31
mother and step-father	1.18	1.48	1.84
father and step-mother	0.78	1.31	2.19
foster parents	1.78	3.74	7.87
other	0.79	1.31	2.14
Whether drinks			
no	1.00	1.00	1.00
yes	2.19	2.56	3.00
Whether at least one family smokes			
no	1.00	1.00	1.00
yes	1.40	1.62	1.89
Best friend smokes			
no	1.00	1.00	1.00
yes	12.23	14.16	16.36
Have partner			
never had one	1.00	1.00	1.00
not at the moment	1.44	2.02	2.85
yes, few weeks	2.25	3.25	4.71
yes, up to 6 months	3.21	4.77	7.09
yes, up to a year	1.72	2.80	4.53
yes, > 1 year	2.10	3.17	4.78
How feel with opposite sex			
very uneasy	1.00	1.00	1.00
a little uneasy	1.01	1.32	1.74

Explanatory variables	Lower	Odds Ratio	Upper
at ease	1.43	1.89	2.49
Happy with body shape			
no	1.00	1.00	1.00
yes	0.69	0.80	0.93
Considers health when choosing food			
never	1.00	1.00	1.00
sometimes	0.50	0.62	0.77
quite often	0.38	0.48	0.62
very often	0.23	0.32	0.43
always	0.30	0.44	0.63
Worry about money problems			
never/hardly ever	1.00	1.00	1.00
a little	1.12	1.33	1.58
quite a lot/a lot	1.43	1.71	2.05
Worry about family problems			
never/hardly ever	1.00	1.00	1.00
a little	0.97	1.17	1.42
quite a lot/a lot	1.15	1.37	1.62

Evaluation of results

A major problem with surveys is that some variables go unobserved. Variables are omitted because they are either unmeasurable or difficult to measure. For example, in cross-sectional studies it is not possible to include temporal dependencies or past behaviour. Omitted variables are responsible for spurious relationships between the observed characteristics and the outcome variable; often leading to the overestimation of the relationships between the response (in this case, smoking) and the explanatory variables. Thus, if not accounted for, omitted variables lead to erroneous results and mis-conclusions (Shahtahmasebi, 2003; Shahtahmasebi & Berridge, 2005).

To demonstrate the impact on results from cross-sectional data the selected variables in the model reported in Table 1 were divided into subjective and objective measures. Subjectively measured variables often rely on the respondent's own assessment such as self-reporting of health, fitness and emotional variables. Clearly, measurements reported by the respondents will be influenced by the dynamics of their own characteristics, life experiences, state of mind, wellbeing and other personal characteristics at the time of reporting, providing a proxy for omitted variables. In simple terms such variables carry measurement errors, i.e. the actual value of the parameter being measured (e.g. fitness) will be compounded with values from a mixture of the individual's emotional, psychological, and overall state of health.
A problem arises over the inclusion of subjectively measured social and emotional factors (arbitrarily headed socio-environmental and socio-psychological respectively) in the analysis. Social circumstances will have an impact on the prevalence of teenage smoking, in part, by affecting emotional variables e.g. the desire to belong to a peer group (Sussman & Dent, 1994).

It is highly plausible that the inclusion of such variables in the model will lead to complex correlations and interactions between the explanatory variables on one hand and between explanatory variables and the error component on the other. Such relationships in the model will lead to a well-known specification error (Shahtahmasebi, 2003; Shahtahmasebi & Berridge, 2005) which means the results shown in Table 1 cannot be taken at face value!
In the absence of longitudinal data, to get an idea of the existence of complex multicollinearity in our cross-sectional data, we can assess the role of subjective variables by forcing the direction of causality from objective variables to the outcome (smoking)

(Shahtahmasebi, 1995; Shahtahmasebi *et al.*, 1992; Wenger *et al.*, 1995). Therefore, models were fitted with and without socio-environment and socio-psychological variables. Firstly a model of objective variables was constructed based on demographic variables. Secondly, social variables were introduced to this model, and, thirdly socio-psychological variables were then added to the second model. The results of this modelling process are shown in Table 2.

Although, the final model in both Tables 1 & 2 are the same, in this round of analysis we were initially only interested in the model fitting process. It is during this process that we will gain more insight into the interrelationship between the various types of variables. The role different variables play in smoking can be examined by comparing results from the three models in Table 2.

The inclusion of socio-environmental factors (model 2) has a major impact on model 1 (Table 2): there are significant changes in parameter estimates of the variables "gender", "age", "which parent live with" - and variables reflecting social status "where live" and "social class" are no longer significant and drop out of the model. While some increases in parameter estimates are to be expected when adding new significant variables to the logistic regression model, the large decrease in parameter estimates confirms that the effect of the variables "age" and "which parent live with" have substantially reduced.

This is consistent with socio-environmental variables having an intervening effect between age, parent(s) and smoking. Similarly, when socio-psychological variables are added to model 2 (see model 2 to model 3, Table 2), a modest decrease in parameter estimates of demographic and socio-environmental variables can be noted. This decrease is consistent with socio-psychological variables having an intervening effect between demographic and socio-environmental variables and smoking.

Table 2. Model fitting results for the model of smoking - N=9230

	Model 1		*Model 2*		*Model 3*	
Explanatory variables	Parameter estimate	Standard error	Parameter estimate	Standard error	Parameter estimate	Standard error
Demographic factors						
Age						
12-13	0.00		0.00		0.00	
14-15	1.06	0.06	0.68	0.07	0.61	0.08
Gender						
male	0.00		0.00		0.00	
female	0.52	0.06	0.68	0.07	0.67	0.08
Which parent live with						
both parents	0.00		0.00		0.00	
mother only	0.47	0.08	0.23	0.10	0.12	0.11
father only	0.69	0.18	0.49	0.23	0.37	0.24
mother and step-father	0.80	0.09	0.53	0.11	0.39	0.11
father and step-mother	0.77	0.20	0.44	0.25	0.27	0.26
foster parents	1.63	0.29	1.56	0.37	1.32	0.38
other	0.73	0.19	0.38	0.24	0.28	0.25
Where live						
town/city centre	0.00					
town/city suburb	-0.30	0.13				
small town/city centre	-0.08	0.11				
small town/city sub.	-0.26	0.11				
in village	0.10	0.10				
outside town/village	0.20	0.13				

Social class						
high	0.00					
medium	0.15	0.10				
low	0.25	0.09				
Socio-environmental factors						
Whether drinks						
no			0.00		0.00	
yes			1.11	0.08	0.94	0.08
Whether at least one family smokes						
no			0.00		0.00	
yes			0.60	0.08	0.47	0.08
Best friend smokes						
no			0.00		0.00	
yes			2.81	0.07	2.65	0.07
Socio-psychological factors						
Have partner						
never had one					0.00	
not at the moment					0.70	0.17
yes, few weeks					1.18	0.19
yes, up to 6 months					1.56	0.20
yes, up to a year					1.03	0.25
yes, > 1 year					1.15	0.21
How feel with opposite sex						
very uneasy					0.00	
a little uneasy					0.28	0.14
at ease					0.64	0.14
Happy with body shape						
no					0.00	
yes					-0.22	0.08
Worry about money problems						
never/hardly ever					0.00	
a little					0.29	0.09
quite a lot/a lot					0.54	0.09
Worry about family problems						
never/hardly ever					0.00	
a little					0.16	0.10
quite a lot/a lot					0.31	0.09

For example, it can be seen that the change in the parameter estimates of the variable “best friend smokes” from model 2 to model 3 (Table 2) is over two times its standard error. This change is consistent with variable “best friend” having an intervening effect between smoking and demographic variables. If there is a ‘true’ best friend effect it is too complex to distinguish with cross-sectional data. It is plausible that prior to taking up smoking such pupils may have had lower self-esteem, a wish to gain confidence, a desire to belong to a peer group and possibly lacked social and parental guidance.

There is some evidence to suggest that parental influence indirectly predicts lower levels of smoking (Charlton, 1984; Eiser *et al.*, 1989; U.S. Department of Health and Human Services, 1994). The “best friend” effect may not be straightforward to interpret with these data because we have no knowledge of the pupil’s previous smoking habits; they may have been a smoker prior to the friendship.

As most pupils were made aware of the dangers of smoking, these variables may be a proxy for the underlying effect of attitudes to health and smoking. The choice of food represents the health consciousness of pupils suggesting that those who attach importance to health related behaviour have a reduced risk of being a smoker. While the worrying variables serve to

demonstrate the subjective effect of smoking, where smoking leads to the maintenance of smoking (McNeill, 1991; McNeill *et al*., 1987). Again without prior information about the pupils smoking behaviour these results do not constitute evidence for worrying leading to smoking. This association between smoking and worrying/health may help explain the prevalence but not incidence of smoking.

Concluding comments

For research results to inform the policy and decision making process then over and above appropriate study design and analytical methodology, interpretation of results must take account of the dynamics of human behaviour. As demonstrated in this paper, even when the methodology is correct it is unwise to take the results at face value as they are highly likely to lead to erroneous policy formation.

References

Berridge, D. (2014a). Paper 2: Issues arising in the statistical modelling of multivariate health data. *Dynamics of Human Health (DHH), 1*(2), https://journalofhealth.co.nz/wp-content/uploads/2014/2007/DHH_Damon2012.pdf.

Berridge, D. (2014b). Paper 3: The exploratory analysis of data on survival in old age. *Dynamics of Human Health (DHH), 1*(3), https://journalofhealth.co.nz/wp-content/uploads/2014/2010/DHH_Berridge_Paper-2013.pdf.

Berridge, D. (2014c). Paper I: The use of the statistical model as an effective tool in the analysis of multivariate data in health research. *Dynamics of Human Health (DHH), 1*(1), https://journalofhealth.co.nz/wp-content/uploads/2014/2007/DHH_Article-2011_BeridgeD.pdf.

Charlton, A. (1984). The brigantia smoking survey: A general review. Public education about cancer. *UICC Technical Report Series, 77*, 92-102.

Eiser, J., Morgan, M., Gammage, P., & Gray, E. (1989). Adolescent smoking: Attitudes, norms and parental influences. *British Journal of Soc. Physio., 28*, 193-202.

McNeill, A. (1991). The development of dependence on smoking in children. *Psychopharmacology-(Berl), 92*(1), 115-117.

McNeill, A., Jarvis, M., & West, R. (1987). Subjective effects of cigarette smoking in adolescents. *Psychopharmacology-(Berl), 92*(1), 115-117.

Royal College of Physicians. (1992). *Smoking and the young*. London.

Shahtahmasebi, S. (1995). *Statistical modelling of dependency in old age.* Unpublished Ph.D, University of Wales, Bangor, North Wales.

Shahtahmasebi, S. (2003). Teenage smoking: Some problems in interpreting the evidence. *Int J Adolesc Med Health, 15*(4), 307-320.

Shahtahmasebi, S., & Berridge, D. (2005). Teenage smoking: A longitudinal analysis. *Int J Adolesc Med Health, 17*(2), 137-155.

Shahtahmasebi, S., Davies, R., & Wenger, C. (1992). A longitudinal analysis of factors related to survival in old age. *The Gerontologist, 333*, 404-413.

Sussman, S. C., & Dent, W. (1994). Group self-identification and adolescent cigarette smoking: A 1-year prospective study. *Journal of Abnormal Psychology, 103*(3), 576-580.

U.S. Department of Health and Human Services. (1994). *Preventing tobacco use among young people. A report of the surgeon general*: Public Health Service, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

Wenger, G. C., Davies, R., & Shahtahmasebi, S. (1995). Morale in old age: Refining the model. *International Journal of Geriatric Psychiatry, 10*, 933-943.