SUMMARY OF FINDINGS
FROM THE SMOKING SURVEY

The final (and arguably most painful!) stage in any research project involves writing up the findings. We have not done this, for the Smoking Survey was only ever intended as an aid to learning, not as an end in itself, and the various mistakes and compromises that were made along the way certainly limit its value as a genuine research resource.

Nevertheless, even allowing for the errors and the cock-ups, the survey did generate some interesting material, and much of this has been analyzed in this book. Drawing on the 8 hypotheses which survived into Part II, we have used the Smoking Survey data in Part III to illustrate various different procedures and statistical tests, and in doing so, some interesting and even at times surprising results have emerged. Perhaps, therefore, we should briefly summarize what we think we have learned about smoking attitudes and behaviour from this survey.

Before we begin, though, we must again emphasise one of the major weaknesses of the whole study. This survey was based on a quota sample design, and although the final sample did approximately reproduce the quotas for sex and age, we have no way of knowing whether it was representative of the population from which it was drawn on any other parameters. All the tests of statistical significance which have been used in this book assume a probability sample design, which means, strictly speaking, that we cannot make inferences from our sample statistics to the wider population, even when differences appear 'significant' (although in practice we have done, just as other researchers often do when they use quota samples).

This drawback, together with the fact that interviewing was relatively uncontrolled and unmonitored, leads us to urge caution in using the results that we have reported for any serious academic purpose. We actually believe that, taken as a whole, the survey was reasonably reliable and has produced fairly valid findings. Nevertheless, its weaknesses (particularly at the crucial early stages) are such that it is probably

safer to regard it as a very elaborate pilot study than as a finished product in its own right.

The hypotheses

Ten hypotheses were developed in the appendix to Chapter 1, and eight of these were operationalized in Chapter 3 (the other two were dropped in Chapter 2 as inappropriate for a survey design to investigate). Seven of these eight have subsequently been tested in Chapters 8 through Appendix I.

The original hypotheses are listed in full in the appendix to Chapter 1. We summarize them here in Table S.1, together with the tests that were applied to them and the outcome of the analysis.

Table S.1: The Smoking Survey hypotheses, with the statistical tests that were used and whether the hypotheses were supported or rejected

Hypothesis	Statistical tests used	Hypothesis supported or rejected?
1. Parental socialization	Two-way crosstabulation using Chi Square	Rejected*
2. Peer group socialization	Two-way crosstabulation using Chi Square	Supported**
5. Stress	1. T-test 2. Logistic regression (Appendix H)	Supported**
6. Self-interest	Mann-Whitney U test	Supported
7. Age effect	None	Not tested
8. Social class effect	1. Regression 2. Spearman's rho	Class predicts smoking: supported. Class predicts attitude: rejected.
9. Zealous converts	Pearson correlation	Supported
10. Cognitive dissonance	Two-way crosstabulation using Chi Square	Supported**

* No significant association between parents' smoking behaviour and respondents' smoking behaviour but a significant association between mothers' smoking behaviour and respondents' smoking behaviour.

** Line of causation has not been established

Let us run through each of these findings in a little more detail:

Hypothesis 1: Parental socialization

Our first hypothesis emerged as one of the most interesting that we tested and serves nicely as an example of the way in which hypothesis testing can throw up findings that beg further exploratory research.

In Chapter 8 we set about testing this hypothesis by crosstabulating whether the respondent smokes by whether his or her parents smoked when the respondent was a child. The results from the Chi Square test showed no significant association between the two variables ($p > 0.05$) and our hypothesis was therefore rejected. This finding was shown to be consistent with previous research which has also noted a lack of association between parental smoking behaviour during childhood and the respondents' smoking behaviour in adulthood (e.g. A. Marsh and J. Matheson, *Smoking Attitudes and Behaviour*, London, OPCS, 1983).

It was at this point, however, that we decided to *explore* the data and delve a bit deeper by controlling for a third variable – gender – to see if the association between respondents' smoking behaviour and their parents' smoking history was being masked by the gender of the respondent. After all, it seemed reasonable to speculate that girls might only be influenced to smoke if their mothers smoked, and that boys might only be influenced to smoke if their fathers smoked.

The results were still not significant, but an interesting pattern to the data did emerge that warranted yet further investigation. It appeared that males (but not females) were more likely to smoke if their mothers (but not their fathers) had smoked. Our results seemed to show a significant relationship ($p < 0.05$) between men's smoking behaviour and their mothers' smoking history.

This finding needs to be regarded as very tenuous. For a start, the cell variations were not huge (54 per cent of males smoked if their mothers had smoked, but 46 per cent smoked if their mother had not smoked – a difference of just 8 per cent) and the association only just reached our significance level of $p = 0.05$ (see Table 8.7). Furthermore, the conditions of the Chi Square test had been violated in that we were

no longer conducting a single test on a predicted outcome, but had applied a series of tests in an exploration of possible associations in the data. To have any faith in the validity of this finding, it really would be necessary to reproduce it on a second, independent sample.

We returned to this relationship between respondents' smoking behaviour and their mothers' smoking history in Appendix I when we looked at loglinear modelling. But when we ran a loglinear analysis on the three variables (smoking, mothers' smoking behaviour and gender), a rather different picture emerged. The results indicated that a simpler set of relationships could accurately represent the relationship between the three variables, for the best fitting model only included two pairs of associations (between the respondents' smoking behaviour and the respondents' gender, and between respondents' smoking behaviour and mothers' smoking history). The conclusion from the loglinear analysis was, therefore, that there is an association between respondents' smoking behaviour and their mothers' smoking history *regardless* of their gender. Inclusion of the three variable interaction term in the model only had a relatively small effect on the overall fit of the model.

We also considered the impact of mothers' smoking history on people's smoking when we analyzed the influences on the probability of being a smoker by means of a logistic regression model in Appendix H. In this case, we were interested in comparing the effect of maternal smoking with that of other predictor variables including self-reported stress and the proportion of friends who smoked. The results showed that its relative effect was weak and, controlling for the other variables, its association with people's probability of smoking statistically non-significant.

It seems that we must conclude that the hypothesis has been refuted, not only in its original form (that parental smoking influences everybody, females as well as males), but also in its amended form (that maternal smoking influences males only). This is, however, an issue where it would still be interesting to gather further data, and we shall be happy to add your findings to our web site if you decide to investigate this topic on a fresh sample.

Hypothesis 2: Peer group socialization

In Chapter 8, we sought to test this hypothesis by means of a crosstabulation between smoking behaviour and the proportion of family and friends who smoke. The results showed a clear pattern of association. The relationship was statistically significant and the strength of the relationship was moderately strong. The hypothesis was therefore supported.

The contingency table showed that 70% of smokers reported that most of their friends also smoked, but this was true of only 28% of non-smokers. Not surprisingly, the Chi Square test indicated a highly significant ($p < 0.001$) result. The strength of this association was then confirmed by a phi test which reported a moderate association of 0.49 (on a scale of –1 to +1). Put in terms of odds, respondents who reported that many of their friends and family smoke were nearly eight and a half times more likely to smoke than not.

In Appendix H, we extended our analysis by running a logistic regression to assess the relative impact of parents' smoking behaviour in childhood (see hypothesis 1), respondents' reported level of stress (hypothesis 5) and the proportion of the respondents' peer group that smoke (hypothesis 2). The analysis demonstrated that, regardless of whether the respondents' parents smoked or the stress level which the reported, the proportion of their friends who smoke had a significant and relatively strong impact on the probability that they themselves smoked. It can be calculated from the coefficients in Table H.5 that, if most or all friends smoke, then there is only a 9.5 per cent probability of being a non-smoker, even if the mother did not smoke in childhood and the reported stress level is zero. If, however, virtually no friends smoke, this probability soars to 70%.

An important caveat to all of these findings, however, is that the association between smoking behaviour and peer group smoking does not enable us to draw any conclusions about the line of causation. In the logistic regression, we treated respondents' smoking behaviour as the dependent variable and peer group smoking behaviour as the independent variable, but our hypothesis suggested that the line of causation could plausibly run either way. People might be encouraged to take up and

keep up smoking if their friends smoke, but it could equally be the case that people who smoke seek out friends who share their habit.

As we explained in Chapter 9, the issue of causality cannot finally be answered with the kind of data that we have collected, which is cross-sectional (i.e. the data relate to one slice in time). Unlike longitudinal data, cross-sectional data provide little evidence on the timing of events. If we wanted to go beyond our finding that there is an association between the two variables to make a confident statement about the direction of causation, we would need to conduct further research that allows the temporal ordering of these variables to be specified.

Hypothesis 5: Stress

This hypothesis states that the more stressed people claim to be, the more they are likely to smoke. As with hypothesis 2, so too here, we allowed for the possibility that the line of causation could run in either direction (it is possible that high stress leads people to smoke, or that people who smoke might claim to be more stressed). Therefore, the hypothesis merely stated that there would be an association between the two without specifying its direction, and we ended up testing this both ways round.

In our analysis using the t-test (in Chapter 9) we assumed that the dependent variable was level of stress and we took smoking behaviour as the independent variable. We therefore tested whether the mean level of stress varied significantly between smokers and non-smokers, and we found that it did. The mean stress level for non-smokers was about one point lower (on the ten point scale) than for smokers, and this difference was found to be highly significantly ($p = < 0.001$). Therefore, the hypothesis has been supported.

We also tested this hypothesis with the line of causation running the other way (i.e. with stress predicting whether the respondent smokes) when we ran a logistic regression. We found that an individual with the lowest possible recorded level of stress (a score of zero on the scale) had a 33 per cent probability of being a smoker, but there was a 68 per cent probability of being a smoker if the respondent reported

the highest possible level of stress (a score of ten on the scale). The result of this analysis therefore appears to support the second version of the hypothesis where smoking behaviour is held to reflect stress levels.

Later in Appendix H we ran another logistic regression where we included peer group smoking behaviour and mothers' smoking history along with stress as predictors of the probability of the respondent smoking. The standardized coefficients showed that level of stress had less than a third of the impact on the probability of smoking compared with the variable measuring proportion of friends and family who smoke, but nearly four times the impact of mothers' smoking history. When the other two independent variables are controlled for, the unstandardized coefficient for stress in Table H.5 is rather smaller than it had been when stress had been the only variable in the model (a coefficient of 0.1638 compared with a coefficient of 0.1247). This suggests that some of the effect of stress on the probability of smoking found in the first model is shared in the later model with the other variables (with which it co-varies to some degree).

Hypothesis 6: Self-interest

The hypothesis that smokers will express a greater tolerance towards smoking than will non-smokers was tested in Chapter 9 using a Mann-Whitney U test. The analysis showed that the mean rank of smokers on the summary variable measuring attitude towards smoking was higher than for non-smokers (i.e. on average, smokers expressed a higher level of tolerance). The difference between the two groups was found to be highly significant ($p = < 0.001$), and the hypothesis was therefore upheld.

Having found that smokers tend to be more tolerant towards smoking than non-smokers the next stage in the analysis might be to try to establish why the difference in attitude exists. Simple self-interest seems a plausible inference to make, but is it actually true? To answer this, we would need data on why people hold the attitudes they do – something we do not have in the Smoking Survey.

As so often happens in research, in testing one hypothesis we not only refine our understanding of the phenomenon under investigation, but we also open up new avenues of inquiry and new hypotheses to test.

Hypothesis 7: Age effect

This hypothesis predicts that younger people will be more tolerant in their attitudes towards smoking than older people. This is a hypothesis that we did not test directly, although we did look at it indirectly in Appendix G when we constructed our path model (Figure G.2). There we saw that there is no direct effect of age on attitudes, although there is some evidence of a cumulative indirect effect running mainly through the proportion of friends who smoke. In other words, younger people do seem to be rather more tolerant of smoking than older people are, but this mainly reflects the fact that they are more likely to have friends who smoke.

If you wish to test hypothesis 7 directly, you will need to select two variables – age group (var00034) and combined attitude score (*attit8*) – which are both measured at the ordinal level (although, arguably, *attit8* could also be treated as an interval variable). This means that an appropriate test would be Spearman's rho.

One thing you will need to think about in interpreting your results is whether any association that you might find between age group and attitudes indicates an effect of age as such, or can better be explained as a generation (cohort) effect. Is it that young people are always more tolerant in every generation (in which case, these young respondents can be expected to become less tolerant as they get older), or is it that this generation specifically is more tolerant than others (in which case their attitudes should remain fairly constant as they age)? We suspect it is the former, for more than any other before it, this generation has been brought up in a culture which is hostile to smoking. If, despite this, they are still more tolerant of smoking than their elders, this must almost certainly be a function of youth.

Hypothesis 8: Social class effect

This hypothesis predicts the higher somebody's social class and level of education, the less likely they are to smoke, and the less tolerant they will be of smoking. We need to break down this hypothesis because it consists of four specific predictions:

1. Social class predicts smoking consumption

2. Social class predicts attitude towards smoking

3. Level of education predicts smoking consumption

4. Level of education predicts attitude towards smoking

In this book we have left the last two of these for you to test. The hypothesis that level of education predicts smoking consumption can be tested using a regression, for the dependent variable (smoking consumption) is an interval level, as is our measure for level of education (the respondent's age when he or she completed full-time education). The hypothesis that level of education predicts attitude towards smoking can be tested using Spearman's rho, for although level of education is an interval variable, the attitude variable is (strictly speaking) measured at the ordinal level.

As for class effects, we tested the hypothesis that class predicts level of smoking consumption in Chapter 10 where we ran a regression with cigarette consumption as the dependent variable and occupational class (measured on the Cambridge scale) as the independent variable. Our results showed a moderate and significant correlation between the two variables ($R = 0.376$ and $p = < 0.01$), and we showed that predicted levels of consumption fall as position on the Cambridge scale rises.

This evidence provides support for our hypothesis, although the relationship does not look strong. The adjusted R^2 for the regression was just 0.126 (indicating that the model explained only 13 per cent of the variance). There was also a large standard error, so the best estimate we could make of somebody's smoking consumption based on their occupation required a huge margin of $\pm$ 75 cigarettes a week in order to achieve a 95 per cent confidence level.

In Chapter 10 we also tested whether social class predicts people's attitude towards smoking. This time we decided to take the Registrar General's schema as our class variable, and the measure of attitudes was the combined attitude score, *attit8*. Both of these variables are ordinal measures, so the appropriate test was Spearman's rho. The results of the analysis showed a very weak correlation between the two variables that was not significant. This part of the hypothesis was therefore rejected.

Later in Chapter 10, we began to move beyond our original hypothesis by exploring models that might better predict smoking consumption. We reasoned that, together with social class, four additional variables might predict how many cigarettes smokers consume: gender, parental smoking, peer group smoking and a belief in the benefits of smoking.

The multiple regression procedure fitted a model to the data using the method of forward selection of the independent variables. This produced a best fitting model with just social class and belief about the beneficial effects of smoking predicting the level of smoking consumption. The other three variables (gender, parental smoking and peer group smoking) were not entered into the model.

This model succeeded in explaining 19 per cent of the variance in the amount that smokers smoke. According to their Beta values, both variables have roughly an equal effect on the dependent variable. This is a somewhat better predictive model of smoking consumption than the one based on social class alone (where we were only able to explain 13 per cent of the variance in smoking consumption), although the R^2 is still not very impressive.

Hypothesis 9: Zealous converts

The zealous convert hypothesis holds that ex-smokers will become increasingly intolerant of smoking, the longer the time since they gave up. We tested this hypothesis in Chapter 8 when we correlated the number of months since ex-smokers had stopped smoking with the combined attitude score (here we treated the attitude score as an interval level variable). Our results provided support for the hypothesis, for we found a significant (though not strong) correlation between the two variables

(Pearson's $r = -0.257$). Given that our hypothesis implies a direction of causation (time elapsing since stopping smoking affects level of tolerance towards smoking), we could also have run a regression model, and this is something you might like to try in order to make predictions of respondents' attitudes based on the amount of time that has passed since they gave up smoking.

In Chapter 9 we did return to this hypothesis when we explored whether ex-smokers ever become as intolerant as those who never smoked at all. We tested this using an ANOVA model comparing the mean attitude scores of smokers, ex-smokers and non-smokers. The results showed that all three groups differ significantly in their toleration of smoking with smokers being the most tolerant, followed by ex-smokers and, finally, non-smokers. We therefore conclude that, while giving up smoking seems to raise intolerance towards smoking, ex-smokers do not tend to become as intolerant as those who never smoked.

We further elaborated on this finding by taking account of the possible confounding influence of the smoking behaviour of the peer group. We have already found that attitudes to smoking are related to the smoking behaviour of friends (people with lots of friends who smoke are more tolerant); that attitudes are also related to people's own smoking behaviour (smokers are more tolerant than ex-smokers who are more tolerant than non-smokers); and that smokers tend to associate more with friends who smoke. To disentangle this web of covariation we used a two-way ANCOVA design (because we had one interval level dependent variable as well as two nominal level independent variables).

The results of the analysis showed that whether a respondent was a smoker, ex-smoker or non-smoker is significantly association with their attitude towards smoking regardless of the smoking behaviour of their peers. The smoking behaviour of the peer group was not, however, significantly associated with attitudes towards smoking once the smoking behaviour of the respondent had been controlled for; nor was there significant evidence that the attitudes of smokers, ex-smokers and non-smokers reflect differences in the smoking patterns prevalent in their respective peer groups. The conclusion that we can draw from this is simply that, while having family and friends

who smoke might have some influence on people's attitudes about smoking, by far the more important influence is their own smoking behaviour, now and in the past.

Hypothesis 10: Cognitive dissonance

The last of our hypotheses was that smokers will tend to deny that smoking causes harm to themselves or to others and will instead tend to assert its beneficial effects. The test for this hypothesis involves a simple Chi Square test, since both variables (var00028 and var00003) are measured at nominal level.

Since the analysis we did not conduct this analysis, let us examine the crosstabulation for the two variables (Table S.2). This shows that whereas 79 per cent of non-smokers thought that there were no benefits to be had from smoking, only 45 per cent of smokers thought likewise. If we look at the Chi Square statistic (with continuity correction applied) the association is highly significant ($p = < 0.001$) and phi is 0.354, which indicates a moderate strength of association between the two variables. We can therefore conclude from this analysis that our final hypothesis – hypothesis 10 – has been supported.

Table S.2: A crosstabulation of smoking behaviour by whether the respondent believes smoking has any benefits

Smoking have benefits? * Smoked in last week? Crosstabulation

			Smoked in last week?		
			no	yes	Total
Smoking have benefits?	no	Count	127	66	193
		% within Smoked in last week?	78.9%	44.6%	62.5%
		% of Total	41.1%	21.4%	62.5%
	yes	Count	34	82	116
		% within Smoked in last week?	21.1%	55.4%	37.5%
		% of Total	11.0%	26.5%	37.5%
Total		Count	161	148	309
		% within Smoked in last week?	100.0%	100.0%	100.0%
		% of Total	52.1%	47.9%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	38.663[b]	1	.000		
Continuity Correction[a]	37.215	1	.000		
Likelihood Ratio	39.539	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	38.538	1	.000		
N of Valid Cases	309				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 55.56.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.354	.000
	Cramer's V	.354	.000
N of Valid Cases		309	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Once again, however, we need to remember that evidence for an association does not tell us about the direction of causation. The 'cognitive dissonance' hypothesis suggests that smokers will seek to define smoking as beneficial, but it also suggests that this may in turn reinforce their smoking. In other words, our theory suggests that causation may run in both directions.

A call for further research!

Research papers and monographs that report new findings often end up suggesting that further work needs to be done. This is not simply an example of academics ensuring a continuing demand for their own services – it also reflects a genuine humility on the part of researchers who understand that quantitative data analysis can never be definitive and that all findings are subject to further test and refinement.

In our case, such tentativeness is particularly appropriate, given the small size of our sample, the inexperience of our research team and the non-probability design of the sampling strategy. But we also have another, even better, reason for concluding by calling for further research, and this is that we hope to encourage groups of college students who use this book to replicate the smoking survey on a new sample to let us know the results. The more the hypotheses are tested, and the more the results of our exploratory analyses can be subjected to analysis on fresh samples, the less hesitant we may become about some of the results we have reported in this Appendix.

On this website you will find details of how to send us any results that you may wish to share with us, and with other readers of this book.