Nicotine yield versus salivary cotinine concentration Health Survey for England data Report 1 Jarvis et al., 2001 and HSE1998

Authors : Peter Lee and Jan Hamling

Date : 2 June 2004

Executive summary

Jarvis et al., 2001 reported the results of analyses based on data from the 1998 Health Survey for England relating cotinine to the nicotine yield of the cigarette smoked by current smokers. After adjusting for cigarettes/day, age, sex, body mass index, car ownership, housing tenure, unemployment, occupational class and educational qualifications, they reported only a weak (though statistically significant) relationship of cotinine to nicotine yield that explained only 0.79% of the variation in cotinine.

We have attempted to reproduce the results presented by Jarvis *et al*. There were some problems in doing so, because it became clear that Jarvis *et al* had coded nicotine yield from brand differently from the coding in the HSE data file, and that for some brands the coding by Jarvis *et al* had apparently been in error. Nevertheless, we were able to reproduce their findings quite closely. After adjustment for cigarettes/day and the potential confounders they considered, we estimated that the slope of the cotinine/nicotine yield relationship was 59.2 (95% CI 27.4-91.0) whereas they estimated it was 71.0 (41.3-100.6).

We also studied the effect adjustment for individual covariates had on the estimated slope. Adjustment for cigarettes/day had the largest effect but adjustment for social class and education also substantially reduced the slope estimate.

The analyses by Jarvis *et al* were not entirely appropriate, as they were based on untransformed nicotine and cotinine values despite the evidence that cotinine is more log-normally than normally distributed, and despite considerable previous work based on log transformed data.

Results of additional analyses are described. These allowed a number of conclusions:

(i) Cigarettes/day explains a substantial part of the variation in cotinine levels, with the log cotinine/log cigs/day relationship quite close to a linear one.

(ii) Nicotine yield explains much less of the variation, with the log cotinine/log nicotine yield relationship somewhat erratic.

(iii) The "compensation index" for nicotine yield is estimated as 0.85 (S.E. 0.05) after adjustment for cigarettes/day, age, sex, body mass index and alcohol consumption. Alternative estimates, adjusting for other combinations of potential confounding variables confirmed that compensation was substantial but incomplete.

(iv) Time last smoked a cigarette is a strong predictor of the cotinine level, though its importance is less when cigarettes/day is also included in the model.

(v) Other variables are much less strong predictors of cotinine level, with little variance in cotinine explained by the social class variables, once the other direct smoking-related variables are included in the model.

Index

1. Introduction

The Health Survey for England is an annual survey of a representative sample of households in England. The questionnaire includes questions on the brand of cigarette currently smoked by the subject. From this a cigarette nicotine rating can be derived by reference to standard brand yield tables (as produced by the Laboratory of the Government Chemist (LGC)). The 1998 survey included nurse visits during which saliva samples were taken. These samples were analysed for cotinine concentration. It is therefore possible to investigate the association between machine-measured cigarette nicotine yield and subjects' cotinine concentration.

This association is reported in Jarvis et al., 2001, (subsequently referred to as Jarvis or "the paper"). The paper's conclusions are that:

> "Smokers' tendency to regulate nicotine intake vitiates potential health gains from lower tar and nicotine cigarettes. Current approaches to characterizing tar and nicotine yields of cigarettes provide a simplistic guide to smokers' exposure that is misleading to consumers and regulators alike and should be abandoned."

The objectives of this report are:

1) To reproduce the results given in Jarvis;
2) To study the effects of the individual covariates used in that paper; and
3) To carry out alternative analyses using log(cotinine) as the response variable and log(nicotine) as the explanatory variable.

The analyses reported here make use of the data from the 1998 Health Survey for England. We have access to data from the equivalent surveys for other years (1999-2001), each of which includes salivary cotinine data. Further investigations of the association of nicotine yield with cotinine level could therefore be carried out, using the data from these surveys.

2. Initial investigations

2.1 Data exclusions

Jarvis restricts attention to subjects who stated at both the initial interview and at the nurse visit that they smoked cigarettes, for whom the cigarette brand "usually smoked" had a known nicotine yield and who had a "measured cotinine concentration" available.

Note that these restrictions do not exclude smokers of mixed products (manufactured cigarettes plus any combination of cigars, pipes and hand-rolled cigarettes). It would be possible to exclude manufactured cigarette smokers who also smoked cigars or pipes. Unfortunately the structure of the questionnaire makes it impossible to identify smokers of both manufactured and hand-rolled cigarettes.

Based on investigating the numbers of subjects included in analysis in Jarvis it seems that all subjects with a cotinine value are included even though 14 of these subjects apparently have an invalid cotinine value (two variables exist in the HSE1998 data, one for cotinine result, the other for valid cotinine result).

The analyses reported below also use these restrictions, but the 14 subjects with apparently invalid cotinine are excluded.

2.2 Nicotine coding

Our investigations relating to Table 1 and Fig. 1 of Jarvis highlighted differences between the nicotine yields available in the HSE1998 data and those used in the paper's analyses. These differences related to the use of different LGC surveys for coding brand nicotine yield. The HSE1998 nicotine yield data were derived from LGC survey 41 while the yields used in Jarvis were derived from LGC survey 42. This is rational because survey 41 related to brands tested during 1997 (and so was available during the year of the survey) while survey 42 related to the yield of the cigarettes actually smoked during the survey year.

While investigating this issue several problems were identified. Some of the codings in the HSE1998 data do not correspond to survey 41 values. Similarly some of the codings used in the Jarvis paper are not valid compared with survey 42. Finally, the process of comparing brand names given in the HSE1998 data with the brands listed in the LGC surveys involves a good amount of guesswork as to the most appropriate category to assign to each survey brand name.

We could therefore do various analyses:

- using the HSE codings as given or our best guess at the Jarvis et al., 2001 codings
- using corrected versions of either of these two codings
- using the subset of subjects whose nicotine codings are reasonably indisputable (for either survey 41 or survey 42).

The analyses reported below use our best guess at the Jarvis codings except that 19 subjects who smoked Silk Cut Extra Mild 100s are not included.

[This brand is listed in LGC survey 41 with a nicotine yield of 0.333 mg/cigarette. In the HSE1998 data the brand is (presumably accidentally) coded as 0.072 mg/cigarette (the rating given in survey 41 for Silk Cut Super Low King Size). In LGC survey 42 the brand name is not reported but (by comparing Fig. 1 of the paper and our attempt at reproducing it) it seems that Jarvis coded these subjects' nicotine yield as 0.048 mg/cigarette – the yield given in that survey for Silk Cut Super Low King Size. This example gives some indication of the problems involved in deciding the nicotine rating to assign to each individual brand, and the scale of the differences between the various possible codings.]

2.3 Ranges of values and real variables

Several of the variables used in these analyses are held as real values. For example, 'Number of cigarettes smoked a day' is a weighted average of number of cigarettes smoked on weekdays and number smoked at weekends and is given to at least 5 decimal places. Similarly, nicotine yield is given to 3 decimal places.

The ranges given in Tables 1 and 2 of the paper are therefore misleading. In Table 1 the nicotine range 0.4-0.75 is taken to mean 0.400-<0.760. The cigarettes/day ranges in Table 2 of the paper give falsely low upper values for cigarettes/day ranges – for example the range given as 8-12 should read 8-<13.

2.4 Confounders

Table 2 of Jarvis gives results adjusted for "potential confounders". These are listed as cigs/day (both grouped and as a continuous variable), the physical characteristics of the subject (age, sex and body mass index (BMI)) and several social class indicators (car ownership, housing tenure, unemployment, occupational class and educational qualifications). Jarvis gives no other details of the variables used.

The derivation of a grouped variable for number of cigs/day (similar to that used in Jarvis) has already been discussed in section 2.3. When attempting to reproduce the results cited by Jarvis we used the five consumption groups defined in the paper (0-7, 8-12, 13-17, 18-22 and 23+ cigs/day). For our additional analyses we derived a 20-level grouped variable, using a standard RoeLee function (%RANK).

For the physical characteristics of the subject there were obvious choices of variable, described in the HSE1998 database as "Age last birthday", "Sex" and "Valid BMI – inc. estimated >130kg". Of these only sex is categorical.

The derivation of the social class indicators was more open to interpretation. The variables used by us were respectively:

- Car or van available to you or members of household (Yes/No);
- Household tenure (Own it outright, buying it with the help of a mortgage, pay part rent and part mortgage, rent it, live here rent free);
- Unemployed (Yes/No);
- Social class (I, II, IIIN, IIIM, IV, V, Armed Forces, Not fully described); and
- Highest educational qualification (7 levels).

All of these are variables available in the HSE1998 database except for the variable Unemployed. This was derived from the HSE1998 variable "Activity status for last week". From this the categories On a government scheme for employment, Waiting to take up paid work, Looking for paid work, Intending to look for work and Permanently unable to work were taken to represent unemployment. Those in full time education, those doing unpaid work, the retired and those looking after a home or family (or doing "something else") were not classified as unemployed.

3. Objective 1 - Reproducing the results given in Jarvis

Table 1 below shows how various smoker characteristics vary by the grouped nicotine yield of the cigarette. It shows the results given in Table 1 of the paper and the results of our attempts to reproduce those results.

Our results are quite close to those given in the paper. The main differences are in the lowest nicotine category. This is where our exclusion of Silk Cut Extra Mild 100s smokers would have an effect. The differences in numbers of subjects are accounted for by this exclusion, the exclusion of subjects with invalid cotinine values and (presumably) our not coding a nicotine value for a few additional subjects (around 10 people) who were so coded in the paper.

Table 2 below shows the regression analysis results given in Jarvis and the results of our attempts at equivalent analyses.

In attempting to compare the results reported by Jarvis and those we have calculated ourselves, a number of points should be made. Firstly, where the results are unadjusted for potential confounders and are calculated within a given grouping of cigarette consumption, there is no ambiguity about the analysis to conduct - it is a simple linear regression of saliva cotinine (y) on cigarette nicotine yield (x) with no constraint for the line to pass through the origin. Here, with the exception particularly of the estimates for the 0-7 cigs/day group where the intercept and slope values are rather different (Jarvis - intercept 32.6, slope 88.6; our analysis - intercept 90.6, slope 69.2), the estimates of intercept, slope and variance explained seem reasonably similar.

Second, it was not quite clear what analysis was conducted by Jarvis for the results they reported as "univariate" analysis and cigarette consumption "all." Were they adjusted for the five groupings of cigs/day? Inasmuch as the description of this analysis in their Table 2 mentions no way in which it is different from the analysis shown in Figure 1 of their paper, and the results quoted are different, we assumed that this Table 2 analysis (but not the Figure 1 analysis) was adjusted for grouped cigs/day as this would summarize the preceding results. However this may be the wrong assumption.

Third, it is not clear what is meant by "intercept" in Jarvis's Table 2 for analyses that are adjusted for covariates. As we understand it from the description in the statistical methods section, salivary cotinine is estimated by a linear model involving the sum of:

(a) a fitted "mean" value,

(b) for each categorical variable, a fitted offset value corresponding to each level of that variable except the first, and

(c) for each continuous variable, the product of a fitted slope times the value of that variable.

Thus, for example, in a model involving nicotine yield, sex and social class only, one might have a model in which saliva cotinine is fitted by the model:

y = 100
+40 * nicotine yield
-20 if female (and 0 if male, the first level)
+10 if middle class
+30 if lower class (and 0 if upper class, the first level).

One cannot fit a term for each level of each categorical variable, if over-definition (or "aliasing") is to be avoided (since, for example, exactly the same predictions could be reached by adding 50 to the mean and subtracting 50 from both the gender coefficients). However there is a problem in deciding what the "intercept" actually is for such a model. In our work we have taken it as the "mean," so it corresponds to the fitted line for people in the first category of all the grouped variables (male, upper class in the example). But Jarvis may have used some other definition. Inasmuch as our interest is much more in the slope and the percent variance explained, which are unaffected by the definition of the intercept, this is, however, not particularly important.

Overall, the results seem generally quite close to those quoted in the paper.

For the 'All, unadjusted' analysis a plot of cotinine against nicotine yield was generated (see Figure 1 below). This is equivalent to Fig. 1 of the paper. The pattern of points is similar to that given in the paper's Fig. 1 except that:

- All the points in the first column (around 0.05 mg/cig) except the one with the highest cotinine value are missing from our plot. These represent the 19 Silk Cut Extra Mild 100s smokers whom we excluded.
- Our plot shows approximately 3 points between 900 and 1000 ng/ml cotinine which are missing from Fig. 1. The paper makes no mention of excluding such points.

Also shown below (see Figure 2) is a plot of residuals against nicotine yield, again for the 'All, unadjusted' analysis. This plot shows variance

increasing with increasing nicotine yield which means that the model used is not strictly valid for the data (the model's assumption of constant variance is not satisfied). Appropriately transformed nicotine and cotinine values (e.g. log(nicotine) and log(cotinine)) would give a more valid model.

4. Objective 2 – The effects of adjustment for individual covariates

Table 3 below shows the results of regressions adjusted for each of the confounders in turn, both with no other adjustment and additionally adjusted for cigs/day. The confounders investigated are those used in the paper plus alcohol and time last smoked. Previous work suggested that alcohol was an important confounder of the cotinine-nicotine association. Time last smoked has values Within the last 30 minutes; Within the last 31-60 minutes; Over an hour ago, but within two hours; Over two hours ago, but within 24 hours; and More than 24 hours ago. We expected that the time over which the subject's body had been processing the most recent dose of nicotine would have an important effect on the measured cotinine level.

4.1 Effects on the deviance

The results for adjustment by a single variable are given first in Table 3. Of the physical characteristics variables, age gave the greatest improvement in fit of the model (implied by the greatest significance of change in deviance). Sex and BMI also made highly significant improvements to the fit of the model. Of the social class indicators, education and social class were highly significant. The other social class indicators made no significant difference to the model. Adjusting for alcohol made only a small and insignificant difference. Time last smoked made much more difference to the fit of the model than any of the physical characteristics or social class covariates (as assessed by the drop in deviance per degree of freedom used) and, in turn, the effect of time last smoked was much less than that of cigs/day.

When cigarettes/day was also adjusted for, age, BMI, education and time last smoked still made considerable additional difference to the fit of the model. Adjusting for sex or social class was now of borderline significance.

Unlike when adjusting for alcohol alone, adjusting for alcohol when cigarettes/day was already adjusted for was highly significant, so alcohol should be adjusted for in subsequent analyses.

4.2 Effects on the slope

When no adjustment was made for cigarettes/day or potential confounders, the slope of cotinine on nicotine was estimated as 134.89. When adjustment was made for any one of the following factors – age, BMI, car available, tenure, unemployed and alcohol – the slope was little affected, with estimates in the range 138.03 to 132.45. However, adjustment for sex or for social class reduced the slope estimate by about 10. Adjustment for education reduced it even more substantially, to 108.58, while adjustment for all the factors Jarvis included (other than cigarettes/day) reduced the slope estimate to 100.00. Adjustment for time last smoked gave an estimate of 98.25. Adjustment for cigarettes/day gave the greatest reduction in the slope – from 134.89 to 86.93.

With cigarettes/day adjusted for, additional adjustment for age, BMI, car available, tenure, unemployment and alcohol had little effect on the slope estimate. Sex also had little effect, indicating that the effect on the slope described in the previous paragraph was explained by the different cigarette consumption of men and women. Adjusting for time last smoked reduced the slope estimate much less when cigs/day was already in the model than when it was not. This is explained by the strong correlation between the two variables. The factors that reduced the slope estimate most, when cigs/day was already adjusted for, were social class (reducing the estimate to 75.66) and education (reducing it to 67.13). When adjustment was made for cigs/day and all the variables considered by Jarvis, the slope estimate reduced to 59.20.

5. Objective 3 - Additional regression analyses using log cotinine and log nicotine or grouped nicotine

Multiple linear regression analysis was used to fit models of the form:

$$Y_i = a_0 + \sum_{j=1}^{m} a_j X_{ij} + e_i$$

where Y_i is the log cotinine value for subject i (i = 1, …N), X_{ij} is the value of the j^{th} predictor variable (j = 1, …M) for subject i and a_j (j = 0, …M) are coefficients to be estimated. e_i is the error term, assumed to be normally distributed with variance σ^2.

Various possible predictor variables (or groups of variables) are considered in the modelling:

Nicotine yield This was included either as log nicotine yield or as a grouped variable with 20 levels. Note that where log nicotine yield was included as a predictor variable, the estimated coefficient *a* can be taken as equal to (1-I) where I is the compensation index (I = 1 implies full compensation, with cotinine independent of yield, while I = 0 implies no compensation, with cotinine directly proportional to yield).

Cigarette consumption This was included either as log cigs/day or as a grouped variable with 20 levels.

Main covariates Age (as a continuous variable), sex, body mass index (continuous), as described in section 2.4 above, together with the additional variable Units of alcohol/week (also continuous).

Social class The five variables described in section 2.4.

Time last smoked As described in section 4 above.

Analyses were restricted to subjects with complete data on all the relevant variables; who were current cigarette smokers, both at initial

interview and at the time of the nurse visit; who were regular smokers (smoking at least one cigarette per day); and for whom valid cotinine and nicotine yield values were available.

Table 4 below shows the various models fitted and gives the corresponding deviances and degrees of freedom (DF). From an examination of this table, various conclusions are evident at first glance:

(a) Including cigarettes/day in the model causes a large drop in deviance, the great part of this drop being caused by the log cigs/day variable on its own. For example, in the analyses adjusted for the main covariates (B), the drop in deviance is 520.14 when models B1 and B2 are compared (introducing log cigs/day) and is only a further 15.46 when models B2 and B3 are compared (introducing the 20 level cigs/day variable).

(b) Including time last smoked in the model also causes a large drop in deviance, particularly when cigs/day is not included in the model. Thus, comparing models B1 and D1 (without cigs/day included), the drop is 592.23 but comparing models B2 and D2 (with it included as a single variable) it is only 226.93.

On the other side of the coin, the effect of including cigs/day in the model is much less when time last smoked is also included. Clearly, they are highly correlated, as the heavier the smoker the more likely they would have smoked recently.

(c) The drop in deviance associated with the main covariates is clearly more than expected by chance, but is much less than that associated with cigs/day or time since last smoked. Thus, comparing models A1 and B1 the drop is 73.50. The drop in deviance is less when cigs/day is included in the model, perhaps because of the confounding of alcohol and cigarette consumption.

(d) The drop in deviance when social class variables are included is not that great, especially when the other direct smoking-related variables are included. Thus, when comparing models B1 and C1, where none of the other direct smoking variables is in the model (only age, sex, BMI and alcohol being included), the drop in deviance is by 98.17 on 17 d.f. or 5.77 per DF. However, comparing models D9 and E9, with all the other smoking variables in, the drop is much smaller, at only 16.75 on 19 DF, or 0.88 per DF. This is only slightly greater than the residual deviance per DF of 0.64 from model E9. This suggests that the social class variables have little independent effect, and that most, if not all, of the association of saliva cotinine with social class is due to confounding by smoking.

(e) The drop in deviance associated with nicotine yield is much less than that associated with cigarette consumption (e.g. compare the deviances for models B2 and B4 or for models B3 and B7). The drop in deviance associated with log nicotine yield only appears to be a relatively small part of that associated with yield as a grouped variable.

These conclusions are expanded upon in further Tables and Figures described below.

Table 5 gives a summary of results relating log cotinine to log nicotine yield. Regardless of which set of variables are included in the model, there is evidence of an association, which is more significant if cigarettes/day is not included in the model. With adjustment for the main covariates only, the association is highly significant ($p<0.001$), with the compensation index estimated as 0.72 (S.E. 0.06). With adjustment also for cigs/day, the significance reduces to $p<0.01$ and the compensation index increases to 0.85 (S.E. 0.05). All the analyses are suggestive of substantial, but incomplete compensation.

In the analyses in Table 2, the estimates of the percent variance in saliva cotinine explained by brand nicotine yield are quite low. However in view of the large variability between saliva cotinine levels in subjects smoking the same brand at the same level of cigarette consumption, such percent variance estimates could not possibly be that high, even if in fact there were an exact linear relationship between saliva cotinine and brand nicotine yield. It seems unreasonable to express percent variance explained, in the context in which Jarvis *et al* seem to be interpreting it, using a metric where 100% is clearly unachievable.

Table 6 pursues an alternative, and perhaps more meaningful, way of expressing the relationship, by comparing the effects of including log nicotine yield and grouped nicotine yield as predictor variables of log cotinine. If there were a perfect linear relationship between log cotinine and log nicotine yield, then the drop in deviance explained by log nicotine (L on 1 DF) would be 100% of the drop in deviance explained by grouped nicotine yield (G on 19 DF). It can be seen that the percentages expressed in this way are substantially larger than those shown in Table 2, but remain not that high. L as a percentage of G is always larger than the 5.26% (1/19) expected by chance but no more than 24.2% in any of the analyses shown and as low as 8.0% in some of the analyses.

The relationship of cotinine to nicotine yield is illustrated further in Figure 3 (adjusted for the main variables) and Figure 4 (adjusted for the main variables and grouped cigarettes/day) which show the values of the fitted coefficients by grouped nicotine. There is some evidence of an increase in log cotinine with increasing log nicotine yield, but the pattern is clearly erratic.

Table 7 summarises the results relating log cotinine to number of cigarettes/day. Clearly, all the analyses show a very highly significant relationship, with the fitted effect of log cigs/day at least 15 times its standard error. The association is little affected by adjustment for the main covariates, social class or nicotine yield, but is clearly weakened by adjustment for time

last smoked. The drop in deviance due to log cigs/day (L) is a very major part of the drop due to cigarettes/day as a grouped variable (G), with L between 94.6% and 97.1% of G depending on the model. This suggests that the relationship between log cotinine and log cigs/day is very nearly linear.

This is illustrated further in Figure 5 where adjustment is made for the main variables only. The relationship can be seen as steep and quite linear.

In Table 8, the complete fitted model is shown for model B2 (as defined in our Table 4). It can be seen that cotinine levels rise significantly with age and are higher in men than in women, but are not so strongly related to BMI or alcohol consumption. Similar conclusions are reached with alternative models including other aspects of smoking.

6. Further possible analyses

There are a number of extensions to results presented here which might give interesting results.

For example, analyses could be performed excluding smokers of cigarettes and pipes or cigars, or excluding those who smoked less than 5 cigarettes/day or last smoked a cigarette more than 24 hours before the saliva sample was taken.

It would also be possible to investigate the effects of "families" of cigarette brands, rather than nicotine yield alone which combines a number of brands from different manufacturers into one category.

As mentioned previously, the analyses could be extended to combine data from several years' data from the HSE series.

7. Summary and conclusions

Jarvis et al., 2001 reported the results of analyses based on data from the 1998 Health Survey for England relating cotinine to the nicotine yield of the cigarette smoked by current smokers. After adjusting for cigarettes/day, age, sex, body mass index, car ownership, housing tenure, unemployment, occupational class and educational qualifications, they reported only a weak (though statistically significant) relationship of cotinine to nicotine yield that explained only 0.79% of the variation in cotinine.

We have attempted to reproduce the results presented by Jarvis *et al*. There were some problems in doing so, because it became clear that Jarvis *et al* had coded nicotine yield from brand differently from the coding in the HSE data file, and that for some brands the coding by Jarvis *et al* had apparently been in error. Nevertheless, we were able to reproduce their findings quite closely. After adjustment for cigarettes/day and the potential confounders they considered, we estimated that the slope of the cotinine/nicotine yield relationship was 59.2 (95% CI 27.4-91.0) whereas they estimated it was 71.0 (41.3-100.6).

We also studied the effect adjustment for individual covariates had on the estimated slope. Adjustment for cigarettes/day had the largest effect but adjustment for social class and education also substantially reduced the slope estimate.

The analyses by Jarvis *et al* were not entirely appropriate, as they were based on untransformed nicotine and cotinine values despite the evidence that cotinine is more log-normally than normally distributed, and despite considerable previous work based on log transformed data.

Results of additional analyses are described. These allowed a number of conclusions:

(i) Cigarettes/day explains a substantial part of the variation in cotinine levels, with the log cotinine/log cigs/day relationship quite close to a linear one.

(ii) Nicotine yield explains much less of the variation, with the log cotinine/log nicotine yield relationship somewhat erratic.

(iii) The "compensation index" for nicotine yield is estimated as 0.85 (S.E. 0.05) after adjustment for cigarettes/day, age, sex, body mass index and alcohol consumption. Alternative estimates, adjusting for other combinations of potential confounding variables confirmed that compensation was substantial but incomplete.

(iv) Time last smoked a cigarette is a strong predictor of the cotinine level, though its importance is less when cigarettes/day is also included in the model.

(v) Other variables are much less strong predictors of cotinine level, with little variance in cotinine explained by the social class variables, once the other direct smoking-related variables are included in the model.

Figure 1 Plot of cotinine against nicotine and unadjusted regression

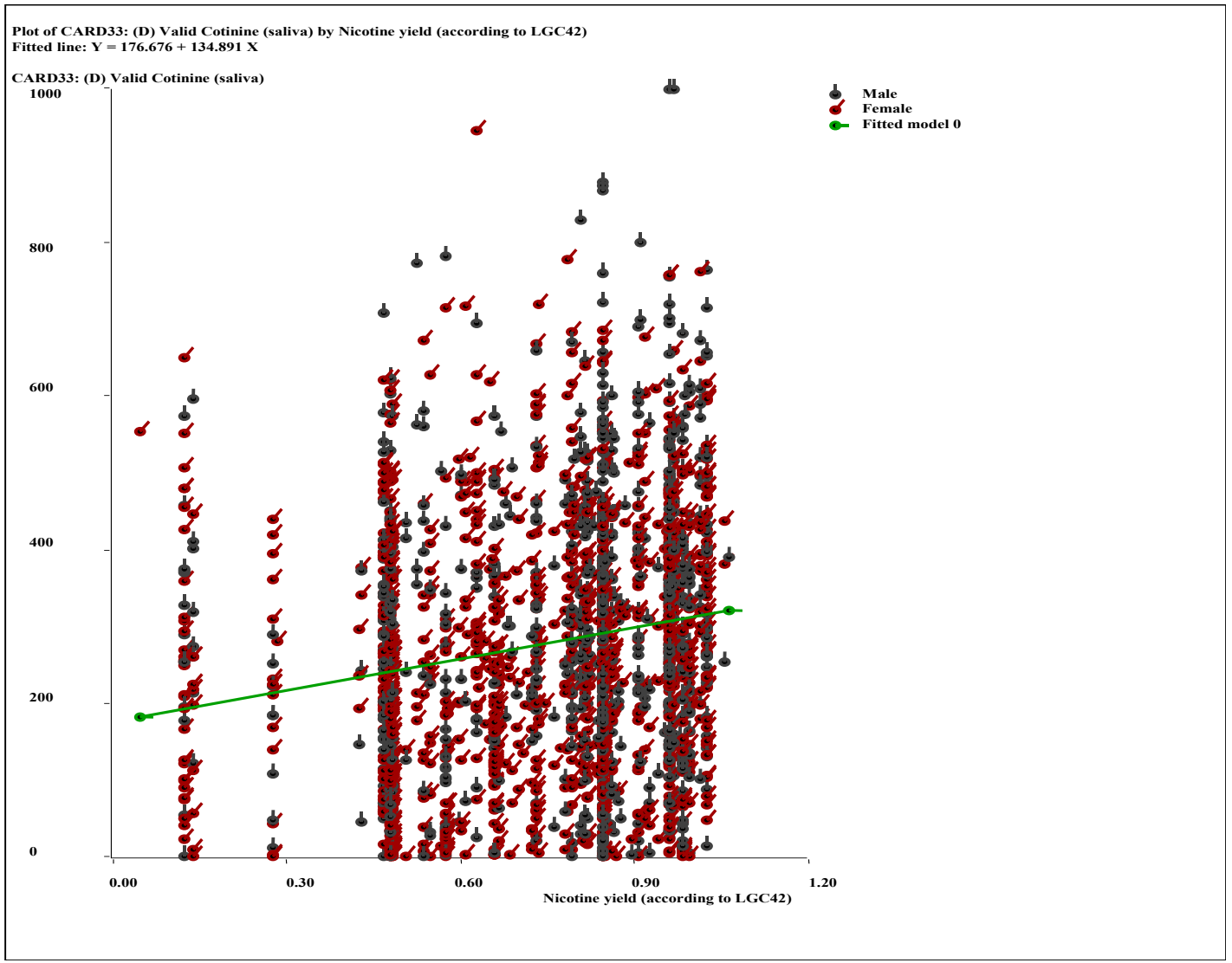

Figure 2 Plot of residuals of the unadjusted regression

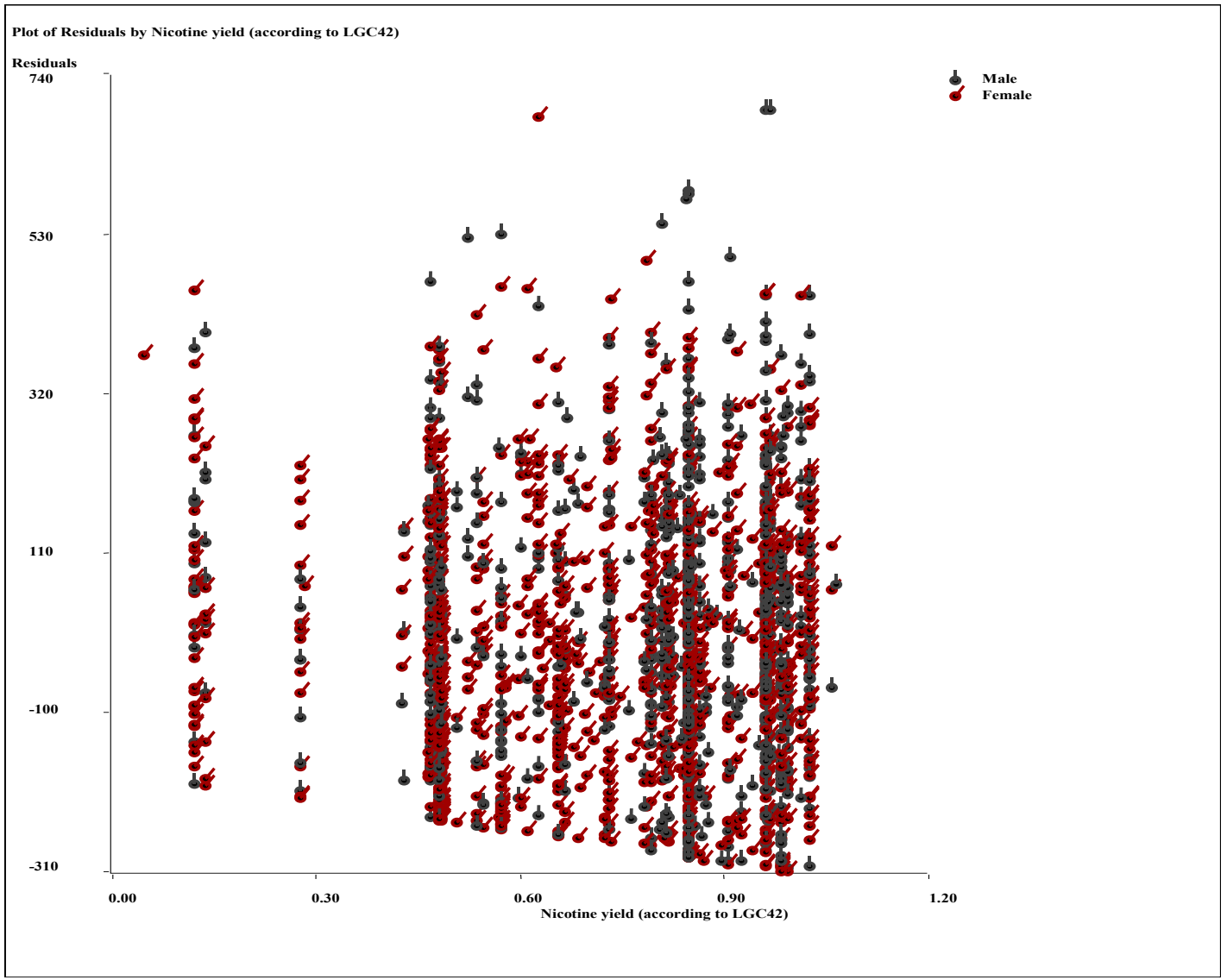

Figure 3 Fitted coefficients for grouped nicotine, adjusted for the main variables

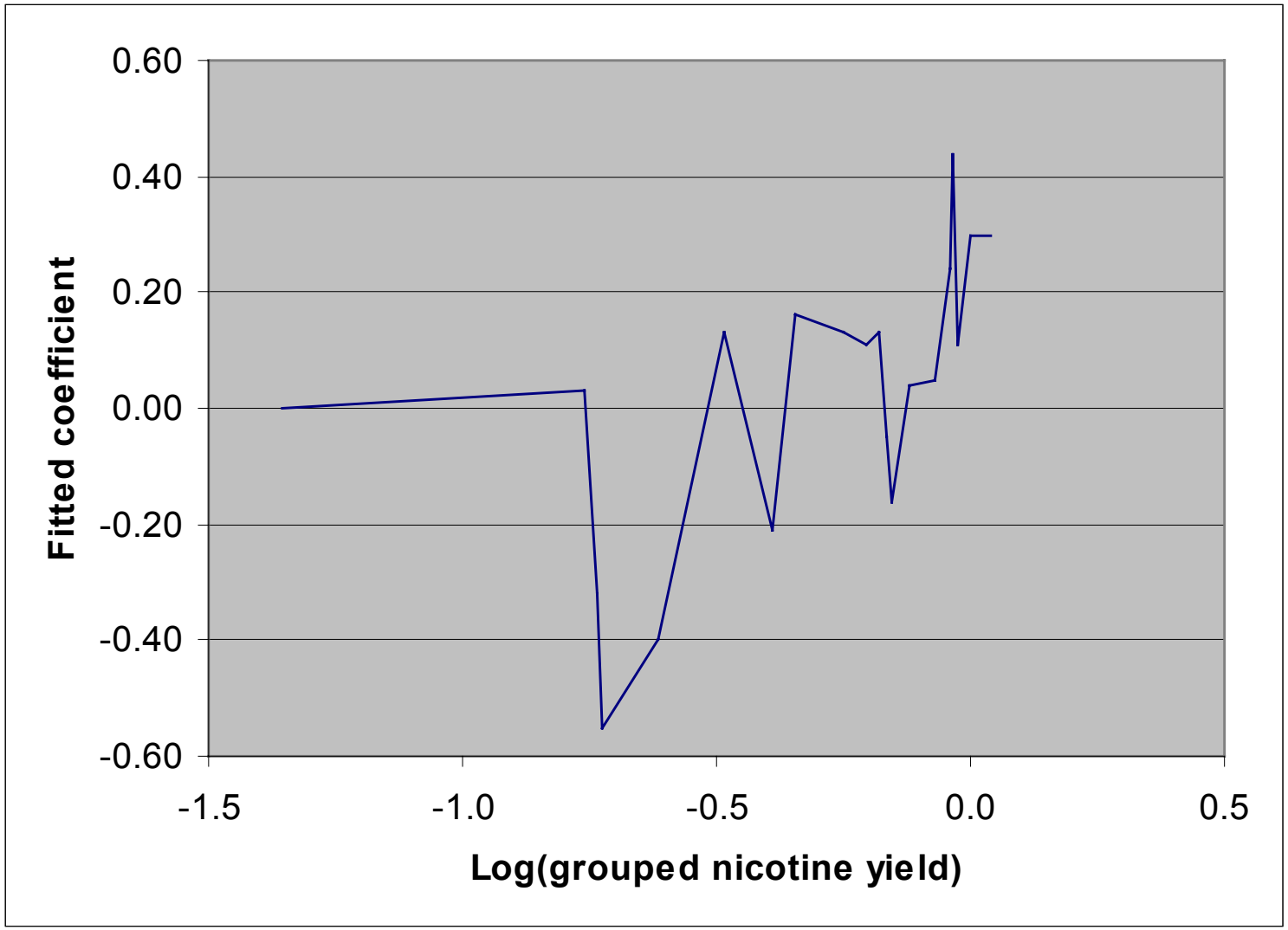

Figure 4 Fitted coefficients for grouped nicotine, adjusted for the main variables and grouped cigarettes/day

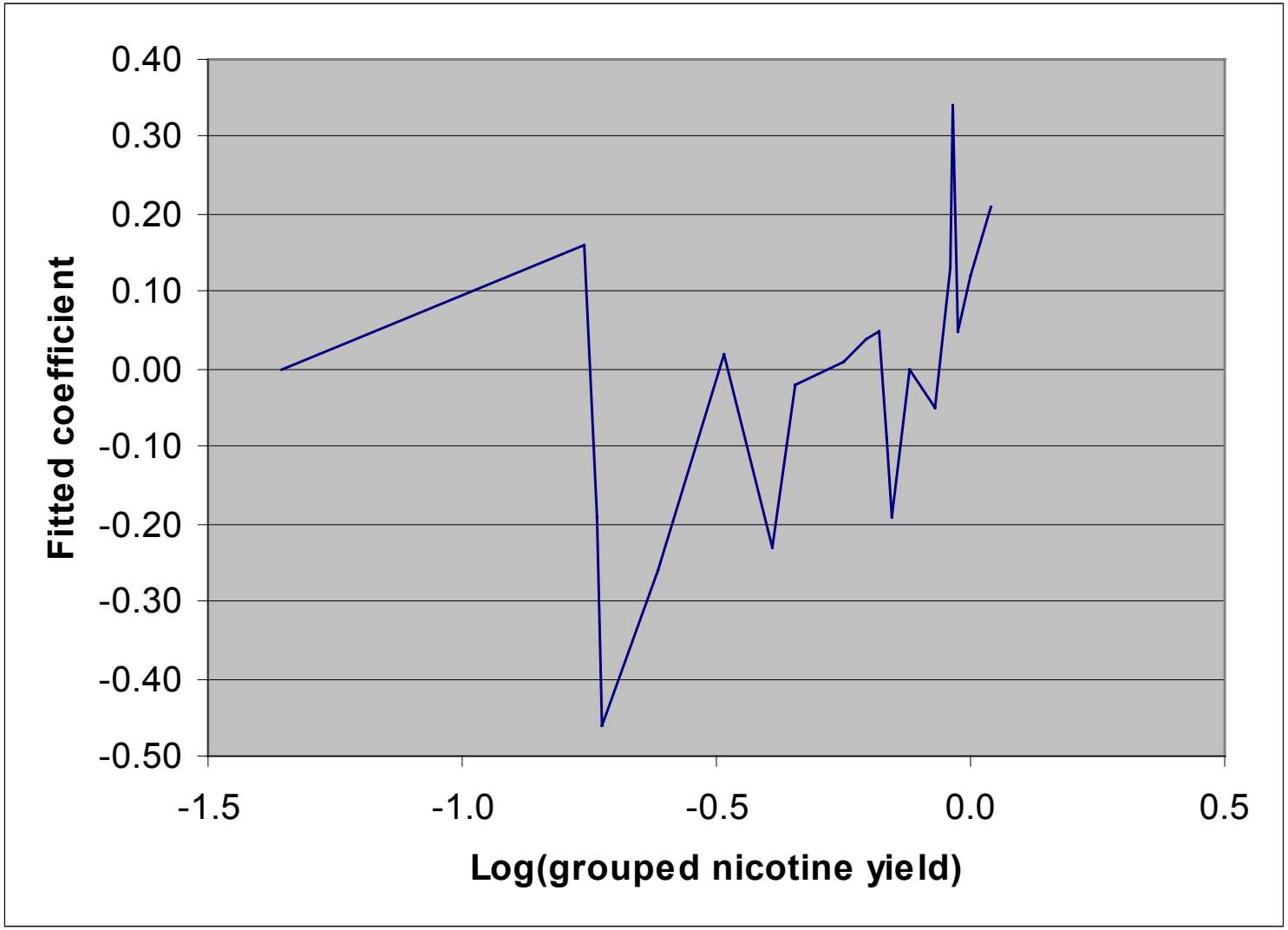

Figure 5 Fitted coefficients for grouped cigarettes/week, adjusted for the main variables

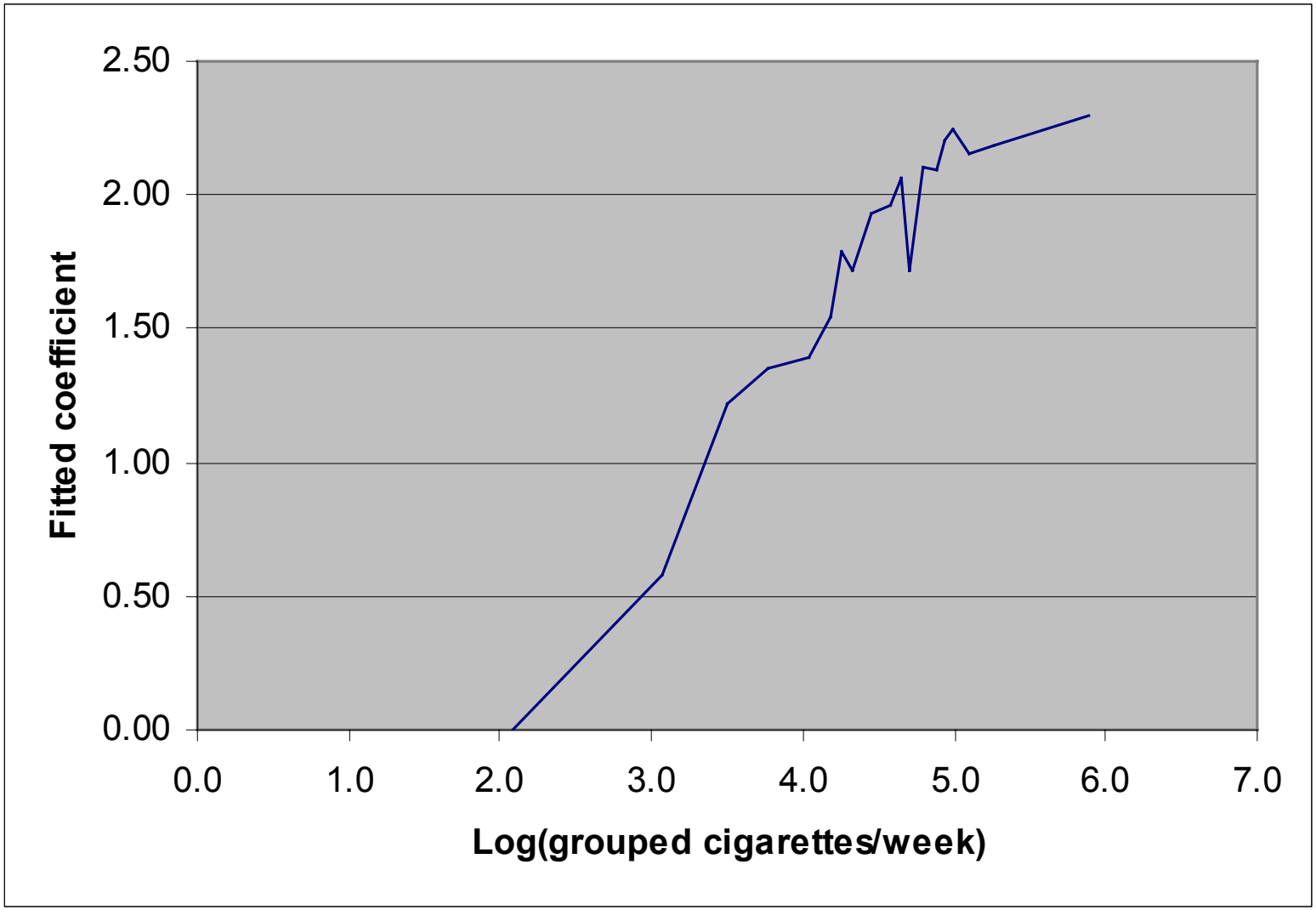

Table 1 Characteristic of smokers by nicotine yield

Results from Table 1 of Jarvis ("The paper") and our attempts to reproduce those results ("Our work").

Characteristic		Nicotine yield, mg/cigarette		
		0-<0.4	0.4-<0.76	0.76+
No of subjects	The paper	101	715	1215
	Our work	82	699	1207
Mean nicotine yield, mg	The paper	0.14	0.57	0.91
	Our work	0.17	0.56	0.91
Mean tar yield, mg	The paper	1.38	6.68	11.5
	Our work	1.79	7.14	11.8
Mean carbon monoxide yield, mg	The paper	1.60	7.48	13.14
	Our work	1.78	7.74	13.67
Mean number of cigarettes/day	The paper	13.5	13.3	15.5
	Our work	14.0	13.3	15.5
% smoking <5 cigarettes/day	The paper	17.8	16.5	9.1
	Our work	17.1	16.3	9.0
Sex: % male	The paper	31	34	49
	Our work	34	33	49
Mean age, years	The paper	44.8	40.7	39.9
	Our work	45.0	40.8	39.9
% with degree level education	The paper	16	13	5
	Our work	16	13	5
% with no educational qualifications	The paper	26	28	35
	Our work	24	28	35
% unemployed	The paper	2	5	7
	Our work	2	5	7
% manual occupation	The paper	41	52	65
	Our work	26	45	60
% rented accommodation	The paper	22	36	43
	Our work	22	35	42
% with no car ownership	The paper	17	20	27
	Our work	18	20	27

Table 2 Initial regression analyses

Results from Fig. 1 and Table 2 of Jarvis ("The paper") and our attempts to reproduce those results ("Our work").

Analysis	Source of results	No. of subjects	Intercept	Slope (95% CI)	% variance explained
All, unadjusted	The paper	2031	173.5	138.7 (106.8-170.6)	3
	Our work	1988	176.7	134.9 (101.3-168.5)	3.02
0-7 cigs/day, unadjusted	The paper	459	32.6	88.6 (17.7-120.9)	1.4
	Our work	444	90.6	69.2 (15.2-123.2)	1.41
0-7 cigs/day, adjusted	The paper	428	32.6	32.0 (-18.3-82.4)	0.01
	Our work	404	104.7	24.8 (-32.3-81.8)	0.19
8-12 cigs/day, unadjusted	The paper	493	196.4	73.3 (14.9-131.7)	1.1
	Our work	487	185.7	86.1 (26.8-145.5)	1.65
8-12 cigs/day, adjusted	The paper	470	155.0	44.3 (-12.3-100.9)	0.03
	Our work	455	224.2	35.0 (-24.9-94.9)	0.30
13-17 cigs/day, unadjusted	The paper	370	255.7	76.1 (14.8-137.4)	1.14
	Our work	360	258.4	73.0 (10.6-135.3)	1.46
13-17 cigs/day, adjusted	The paper	345	209.6	69.8 (4.7-134.8)	0.10
	Our work	336	221.7	62.8 (-8.7-134.3)	0.95
18-22 cigs/day, unadjusted	The paper	471	304.8	68.3 (-2.0-138.5)	0.6
	Our work	462	297.6	75.3 (5.7-144.8)	0.97
18-22 cigs/day, adjusted	The paper	436	389.8	67.7 (-2.1-137.5)	0.06
	Our work	423	353.0	40.3 (-34.77-114.6)	0.28
23+ cigs/day, unadjusted	The paper	238	300.4	118.2 (20.9-215.5)	2.2
	Our work	235	300.4	116.3 (17.3-215.2)	1.71
23+ cigs/day, adjusted	The paper	212	432.3	122.8 (21.7-223.9)	2.1
	Our work	216	502.4	116.9 (6.76-227.1)	2.24
All, adjusted for grouped cigs/day only	The paper	2031	177.7	132.4 (99.3-165.5)	3.1
	Our work	1988	82.4	80.8 (51.5-110.2)	1.45
All, adjusted	The paper	1891	164.1	71.0 (41.3-100.6)	0.79
	Our work	1834	163.5	59.2 (27.4-91.0)	0.73

Table 3 Regression analyses of cotinine on nicotine showing the effects of adjustment for the confounders individually

Adjustment	Deviance	DF	P value*	Slope	S.E.
None	55570606	1986		134.89	17.14
Age	53910845	1985	<0.00001	138.03	16.90
Sex	55117427	1985	0.00006	126.94	17.19
BMI	52793226	1894	0.00033	134.77	17.53
Car available	55564984	1985	0.65	135.46	17.19
Tenure	55485562	1980	0.57	133.13	17.30
Unemployed	55480184	1985	0.07	132.45	17.19
Social class	52851477	1915	0.00026	122.75	17.73
Education	53505876	1979	<0.00001	108.58	17.24
All potential confounders considered by Jarvis	46410189	1810		100.00	18.06
Alcohol	54421336	1948	0.16	136.95	17.26
Time last smoked	45124749	1981	<0.00001	98.25	15.59
Cigs/day	43129363	1985	<0.00001	86.93	15.24

Adjustment	Deviance	DF	P value**	Slope	S.E.
Cigs/day , Age	42331711	1984	<0.00001	90.51	15.11
Cigs/day , Sex	43046600	1984	0.05	83.95	15.31
Cigs/day , BMI	40307850	1893	<0.00001	83.57	15.47
Cigs/day , Car available	43119091	1984	0.49	86.10	15.29
Cigs/day , Tenure	43051293	1979	0.49	85.51	15.37
Cigs/day , Unemployed	43115942	1984	0.43	86.08	15.28
Cigs/day , Social class	41370734	1914	0.016	75.66	15.83
Cigs/day , Education	41950749	1978	<0.00001	67.13	15.37
Cigs/day, all potential confounders considered by Jarvis	36788125	1809		59.20	16.19
Cigs/day , Alcohol	41615634	1947	<0.00001	87.78	15.23
Cigs/day , Time last smoked	39820994	1980	<0.00001	76.93	14.71

* : Compared with the “None” model ** : Compared with the model adjusted for cigs/day

Table 4 Effect of including various predictor variables for log cotinine on the deviance of the model

Model	Predictor variables included: Nicotine Yield	Cigs/day	Age, sex, BMI, alcohol	Social class	Time last smoked	Deviance	DF
A1	-	-	-	-	-	1975.00	1762
A2	-	Log	-	-	-	1429.28	1761
A3	-	Grouped	-	-	-	1412.72	1744
A4	Log	-	-	-	-	1950.79	1761
A5	Log	Log	-	-	-	1422.10	1760
A6	Log	Grouped	-	-	-	1405.69	1743
A7	Grouped	-	-	-	-	1865.61	1743
A8	Grouped	Log	-	-	-	1379.66	1742
A9	Grouped	Grouped	-	-	-	1360.86	1725
B1	-	-	Yes	-	-	1901.50	1758
B2	-	Log	Yes	-	-	1381.36	1757
B3	-	Grouped	Yes	-	-	1365.90	1740
B4	Log	-	Yes	-	-	1879.77	1757
B5	Log	Log	Yes	-	-	1374.64	1756
B6	Log	Grouped	Yes	-	-	1359.64	1739
B7	Grouped	-	Yes	-	-	1811.85	1739
B8	Grouped	Log	Yes	-	-	1341.15	1738
B9	Grouped	Grouped	Yes	-	-	1323.15	1721
C1	-	-	Yes	Yes	-	1803.33	1739
C2	-	Log	Yes	Yes	-	1343.92	1738
C3	-	Grouped	Yes	Yes	-	1328.97	1721
C4	Log	-	Yes	Yes	-	1794.80	1738
C5	Log	Log	Yes	Yes	-	1341.32	1737
C6	Log	Grouped	Yes	Yes	-	1326.33	1720
C7	Grouped	-	Yes	Yes	-	1741.81	1720
C8	Grouped	Log	Yes	Yes	-	1313.74	1719
C9	Grouped	Grouped	Yes	Yes	-	1295.93	1702
D1	-	-	Yes	-	Yes	1309.27	1754
D2	-	Log	Yes	-	Yes	1154.43	1753
D3	-	Grouped	Yes	-	Yes	1146.62	1736
D4	Log	-	Yes	-	Yes	1298.18	1753
D5	Log	Log	Yes	-	Yes	1148.06	1752
D6	Log	Grouped	Yes	-	Yes	1141.00	1735
D7	Grouped	-	Yes	-	Yes	1257.55	1735
D8	Grouped	Log	Yes	-	Yes	1119.91	1734
D9	Grouped	Grouped	Yes	-	Yes	1111.51	1717
E1	-	-	Yes	Yes	Yes	1272.22	1735
E2	-	Log	Yes	Yes	Yes	1131.94	1734
E3	-	Grouped	Yes	Yes	Yes	1124.00	1717
E4	Log	-	Yes	Yes	Yes	1265.89	1734
E5	Log	Log	Yes	Yes	Yes	1128.38	1733
E6	Log	Grouped	Yes	Yes	Yes	1120.83	1716
E7	Grouped	-	Yes	Yes	Yes	1230.42	1716
E8	Grouped	Log	Yes	Yes	Yes	1103.36	1715
E9	Grouped	Grouped	Yes	Yes	Yes	1094.76	1698

Table 5 Effect of including log nicotine yield as a predictor variable of log cotinine

Variables included as adjustment factors				Compensation Index		Drop in	Residual deviance		
Age, sex, BMI, alcohol	Social class	Time last smoked	Cigs/day	Mean	S.E.	deviance	per D.F.	F	p
-	-	-	-	0.71	0.06	24.21	1.11	21.90	<0.001
-	-	-	Log	0.84	0.05	7.18	0.81	8.89	<0.01
-	-	-	Grouped	0.84	0.05	7.03	0.81	8.72	<0.01
Yes	-	-	-	0.72	0.06	21.73	1.07	20.31	<0.001
Yes	-	-	Log	0.85	0.05	6.72	0.78	8.58	<0.01
Yes	-	-	Grouped	0.85	0.05	6.26	0.78	8.01	<0.01
Yes	Yes	-	-	0.82	0.06	8.53	1.03	8.26	<0.01
Yes	Yes	-	Log	0.90	0.05	2.60	0.77	3.37	0.07
Yes	Yes	-	Grouped	0.90	0.05	2.64	0.77	3.42	0.06
Yes	-	Yes	-	0.80	0.05	11.09	0.74	14.98	<0.001
Yes	-	Yes	Log	0.85	0.05	6.37	0.66	9.72	<0.01
Yes	-	Yes	Grouped	0.86	0.05	5.61	0.66	8.53	<0.01
Yes	Yes	Yes	-	0.85	0.05	6.33	0.73	8.67	<0.01
Yes	Yes	Yes	Log	0.88	0.05	3.57	0.65	5.48	0.02
Yes	Yes	Yes	Grouped	0.89	0.05	3.18	0.65	4.87	0.03

Table 6 Comparison of effects of including log nicotine yield and grouped nicotine yield as predictor variables of log cotinine

Variables included as adjustment factors				G. Drop in deviance from grouped nicotine (on 19DF)	L. Drop in deviance from log nicotine (on 1DF)	L as % of G*
Age, sex BMI, alcohol	Social class	Time last smoked	Cigs/ day			
-	-	-	-	109.39	24.21	22.1
-	-	-	Log	49.62	7.18	14.5
-	-	-	Grouped	51.85	7.03	13.6
Yes	-	-	-	89.65	21.73	24.2
Yes	-	-	Log	40.21	6.72	16.7
Yes	-	-	Grouped	42.75	6.26	14.6
Yes	Yes	-	-	61.52	8.53	13.9
Yes	Yes	-	Log	30.18	2.60	8.6
Yes	Yes	-	Grouped	33.05	2.64	8.0
Yes	-	Yes	-	51.72	11.09	21.4
Yes	-	Yes	Log	34.52	6.37	18.5
Yes	-	Yes	Grouped	35.11	5.61	16.0
Yes	Yes	Yes	-	41.81	6.33	15.1
Yes	Yes	Yes	Log	28.58	3.57	12.5
Yes	Yes	Yes	Grouped	29.24	3.18	10.9

*5.26% would be expected if log nicotine was unassociated with log cotinine

Table 7 Relationship between log cotinine and number of cigarettes/day

Variables included as adjustment factors				Fitted effect of log cigs/day		G. Drop in deviance from grouped cigs/day (on 18 d.f.)	L. Drop in deviance from log cigs/day (on 1 d.f.)	L as % of G
Age, sex BMI., alcohol	Social class	Time last smoked	Nicotine yield	Mean	S.E.			
-	-	-	-	0.76	0.03	562.28	545.73	97.1
-	-	-	Log	0.76	0.03			
-	-	-	Grouped	0.76	0.03			
Yes	-	-	-	0.76	0.03	535.60	520.14	97.1
Yes	-	-	Log	0.75	0.03			
Yes	-	-	Grouped	0.73	0.03			
Yes	Yes	-	-	0.73	0.03	474.35	459.41	96.9
Yes	Yes	-	Log	0.73	0.03			
Yes	Yes	-	Grouped	0.71	0.03			
Yes	-	Yes	-	0.49	0.03	162.66	154.85	95.2
Yes	-	Yes	Log	0.48	0.03			
Yes	-	Yes	Grouped	0.47	0.03			
Yes	Yes	Yes	-	0.47	0.03	148.22	140.28	94.6
Yes	Yes	Yes	Log	0.47	0.03			
Yes	Yes	Yes	Grouped	0.45	0.03			

Table 8 Fitted model involving the main covariates and log cigs/day

Variable		Estimate	S.E.	p	Drop in deviance*
Constant		3.78	0.15	<0.001	
Age		0.0059	0.0014	<0.001	43.01
Sex -	male	Reference			23.07
	female	-0.20	0.045	<0.001	
BMI		-0.015	0.0047	0.0012	2.35
Units of alcohol/week		-0.0044	0.0009	<0.001	5.07
Log (cigs/day)		0.76	0.030	<0.001	520.14

* Residual deviance = 1381.36 on 1757 d.f.

Reference

Jarvis, M. J., Boreham, R., Primatesta, P., Feyerabend, C., and Bryant, A. 2001. Nicotine yield from machine-smoked cigarettes and nicotine intakes in smokers: evidence from a representative population survey. *J. Natl. Cancer Inst.* 93:134-138.