

## Dose-response relationship of lung cancer to amount smoked, duration and age starting

John S Fry, Peter N Lee, Barbara A Forey, Katharine J Coombs

John S Fry, Peter N Lee, Barbara A Forey, Katharine J Coombs, PN Lee Statistics and Computing Ltd., Sutton, Surrey SM2 5DA, United Kingdom

**Author contributions:** Lee PN, Fry JS and Forey BA planned the study; Literature searches were carried out by Coombs KJ, assisted by Lee PN and Forey BA; Data entry was carried out by Coombs KJ and checked by Forey BA, or carried out by Forey BA and checked by Lee PN; Where appropriate, difficulties in interpreting published data or in the appropriate methods for derivation of RRs were discussed by Forey BA and Lee PN; The statistical analyses were conducted by Fry JS along lines discussed and agreed with Lee PN; Lee PN and Fry JS jointly drafted the paper, which was critically reviewed by Forey BA and Coombs KJ.

Supported by Philip Morris Products SA

Correspondence to: Peter N Lee, MA, Director, PN Lee Statistics and Computing Ltd., 17 Cedar Road, Sutton, Surrey SM2 5DA, United Kingdom. [peterlee@pnlee.co.uk](mailto:peterlee@pnlee.co.uk)

Telephone: +44-20-86428265 Fax: +44-20-86422135

Received: April 2, 2013 Revised: May 9, 2013

Accepted: August 4, 2013

Published online: August 26, 2013

### Abstract

**AIM:** To quantify smoking/lung cancer relationships accurately using parametric modelling.

**METHODS:** Using the International Epidemiological Studies on Smoking and Lung Cancer database of all epidemiological studies of 100+ lung cancer cases published before 2000, we analyzed 97 blocks of data for amount smoked, 35 for duration of smoking, and 27 for age started. Pseudo-numbers of cases and controls (or at risk) estimated from RRs by dose level formed the data modelled. We fitted various models relating  $\log_e$  RR to dose (d), including  $\beta d$ ,  $\beta d^Y$  and  $\beta \log_e (1 + Wd)$ , and investigated goodness-of-fit and heterogeneity between studies.

**RESULTS:** The best-fitting models for  $\log_e$  RR were

$0.833 \log_e [1 + (8.1c/10)]$  for cigarettes/d (c),  $0.792 (y/10)^{0.74}$  for years smoked (y) and  $0.176 [(70 - a)/10]^{1.44}$  for age of start (a). Each model fitted well overall, though some blocks misfitted. RRs rose from 3.86 to 22.31 between  $c = 10$  and 50, from 2.21 to 13.54 between  $y = 10$  and 50, and from 3.66 to 8.94 between  $a = 30$  and 12.5. Heterogeneity ( $P < 0.001$ ) existed by continent for amount, RRs for 50 cigarettes/d being 7.23 (Asia), 26.36 (North America) and 22.16 (Europe). Little heterogeneity was seen for duration of smoking or age started.

**CONCLUSION:** The models describe the dose-relationships well, though may be biased by factors including misclassification of smoking status and dose.

© 2013 Baishideng. All rights reserved.

**Key words:** Smoking; Lung neoplasms; Dose-response; Meta-analysis; Review; Amount smoked; Duration of smoking; Age at starting to smoke

**Core tip:** This paper, for the first time, meta-analyses smoking/lung cancer dose-relationships. Based on data from 71 studies published before 2000, single parameter models were fitted to summarize how the RR increased with increasing amount smoked, longer duration of smoking, and earlier age of starting to smoke. Overall, the models fitted well. Little heterogeneity was seen for duration of smoking or age of start, but the rise in RR with amount smoked was much steeper in North America and Europe than in Asia. The fitted models can be used to more precisely estimate the lung cancer risk from smoking.

Fry JS, Lee PN, Forey BA, Coombs KJ. Dose-response relationship of lung cancer to amount smoked, duration and age starting. *World J Meta-Anal* 2013; 1(2): 57-77 Available from: URL: <http://www.wjgnet.com/2308-3840/full/v1/i2/57.htm> DOI: <http://dx.doi.org/10.13105/wjma.v1.i2.57>

## INTRODUCTION

We recently carried out a systematic review<sup>[1]</sup> of the evidence relating smoking to lung cancer incorporating all 287 studies published before 2000 involving a minimum of 100 lung cancer cases. We refer to this as “our earlier review”. In that review, we assessed evidence concerning amount smoked per day, duration of smoking, and age of starting to smoke. Data are typically available as blocks of RRs for differing levels of the dose-response measure, each compared to never smokers. Comparing meta-analysis estimates for low, medium and high exposure, we clearly demonstrated a dose-response existed. For example, for amount smoked by current smokers, random-effects RR estimates are 4.71 (95%CI: 4.14-5.37,  $n = 86$ ) for about 5 cigs/d, 9.83 (95%CI: 8.60-11.24,  $n = 54$ ) for about 20 cigs/d, and 17.10 (95%CI: 14.62-19.99,  $n = 62$ ) for about 45 cigs/d. Here “about 5 cigs/d” combined results for dose ranges including 5 but not 20 cigs/d, “about 20 cigs/d” considered ranges including 20 but not 5 or 45 cigs/d, and “about 45 cigs/d” ranges including 45 but not 20 cigs/d. This approach has limitations. First, formal statistical comparison of the RRs at the different levels is not possible as the RRs are not independent, having the same denominator. Second, the analyses do not use all the information available. Thus, results for ranges wholly between 5 and 20 cigs/d or wholly between 20 and 45 cigs/d are ignored, as are results for ranges covering two or more of the “key values” of 5, 20 and 45 cigs/d. Also, linearity, or other shapes of the relationship, is not assessed. Dose-response relationships for years quit are considered in a separate paper<sup>[2]</sup>.

Here, we study dose-response in more detail by fitting models to the various dose-response blocks to estimate parameters which can be meta-analyzed and used to assess heterogeneity. We follow the approach previously used<sup>[3]</sup> to quantify the dose-response relationship between environmental tobacco smoke exposure and lung cancer risk, developing a variant of it for age of starting. We restrict attention to the data considered in our earlier review<sup>[1]</sup>. Rather than also considering results for ever smokers, we restrict attention to current smokers, giving a more homogeneous dataset and one showing a stronger dose-relationship. All our analyses are of overall lung cancer risk, no attempt being made in the present paper to fit models for specific histological types.

## MATERIALS AND METHODS

### *The International Evidence on Smoking and Lung Cancer database*

All analyses use the International Evidence on Smoking and Lung Cancer database, fully described in our earlier review<sup>[1]</sup>. Papers considered were published before 2000, described studies of 100+ cases, and provided RR estimates for one or more smoking indices. We use the term RR generically to describe alternative RR estimates, *e.g.*,

odds ratio or hazard ratio. Lee *et al*<sup>[4]</sup> gives details of the structure and data entry rules for the database.

### *Data selection and blocks*

The data considered here comprise blocks of RRs, each relative to never smokers, for all lung cancer (or occasionally near equivalent definitions, each including squamous cell carcinoma and adenocarcinoma) for three measures of dose-response among current smokers: amount smoked, duration and age of starting. Where possible, blocks by sex or by sex and race were considered. Except for amount smoked, blocks by age were considered, if available. Covariate-adjusted RRs were preferred to unadjusted RRs. Each block includes an estimate of the RR and 95%CI: for each level of the measure. The data recorded per block included study type, sex, location, publication year, age range (at baseline for prospective studies), product smoked [any product, cigarettes +/- other products (*i.e.*, pipes, cigars), cigarettes only], never smoker definition (never any product, never cigarettes). For each RR, the range of the measure was also recorded.

### *Pseudo-numbers*

We used the method of Hamling *et al*<sup>[4]</sup> on each block to estimate the pseudo-table of numbers of cases, and either controls (for case-control studies) or at risk (for prospective studies) which correspond to the observed RRs and 95%CIs. The method was applied even to unadjusted RRs. This estimation requires, in addition to the given RRs and 95%CIs, estimates of the proportion of never smokers among the controls/at risk and of the ratio of total controls/at risk to total cases, as well as starting values for the numbers of never smoking cases and controls/at risk. These estimates were also recorded on the database. The pseudo-table forms the basic data for fitting the models used, and estimating the overall current smoking RR.

### *Midpoints for levels of exposure*

For amount smoked, midpoint estimates for each exposure level were derived using standard distributions, as described by Fry *et al*<sup>[3]</sup> when relating lung cancer risk to amount smoked by the husband. For US studies, the distribution derived from published data for two large CPS I and CPS II studies<sup>[5]</sup>, while for non-US studies, it was that given in Appendix III of International Smoking Statistics<sup>[6,7]</sup>.

For duration of smoking and age of starting the midpoints were based on US NHANES III<sup>[8]</sup>, selecting data for subjects for the given sex, age and range of values of the relevant dose-response measure.

### *Statistical models*

For each measure, the data analyzed consist of blocks, each containing the pseudo-numbers and the estimated midpoint exposures for each of  $\ell$  exposure levels, and for never smokers. The methodology varies by dose-response measure, as described below.

**Table 1** Models used to relate risk to dose

$\log_e(\text{RR})$	=	$\beta_1 d$	(linear)
$\log_e(\text{RR})$	=	$\beta_1 d + \beta_2 d^2$	(quadratic)
$\log_e(\text{RR})$	=	$\beta_1 d + \beta_2 d^2 + \beta_3 d^3$	(cubic)
$\log_e(\text{RR})$	=	$\beta_1 d^Y$	(power)
$\log_e(\text{RR})$	=	$\beta_1 \log_e(d)$	(log)
$\log_e(\text{RR})$	=	$\beta_1 \exp_e(d)$	(exponential)
$\log_e(\text{RR})$	=	$\beta_1 \log_e(1 + Wd)$	(log-with-baseline)

### Amount smoked

The Greenland and Longnecker method<sup>[9,10]</sup> was used to fit functional forms relating RR to dose (midpoint amount smoked). We fitted the models expressing dose,  $d$ , in units of 10 cigs/ $d$ .

In the simplest application, the RR is predicted by  $\log_e \text{RR}(d) = \beta d$ ,  $\beta$  and SE ( $\beta$ ) are estimated separately per block, and estimates of  $\beta$  and SE ( $\beta$ ) are then combined using inverse-variance weighted random-effects or fixed-effects meta-analysis<sup>[11]</sup>. This model implies that a fixed dose increment increases risk by a fixed factor. The method can be used with  $d$  replaced by a function of  $d$ , such as  $d^{1/2}$ ,  $d^2$ , or  $\log(d + 1)$ .

Greenland *et al.*<sup>[9]</sup> describe a more general, “pool-first”, method in which all the blocks are considered in a single analysis. The method gives the same results for the model  $\log \text{RR} = \beta d$ , but allows direct fitting for other functional forms.

As the best model was initially unclear, we first tried various models (Table 1) using the pool-first method, comparing deviances to assess which models fitted the overall data better. For the “power” and “log-with-baseline” models, the parameters  $Y$  or  $W$  could not be fitted directly, but an iterative method was adopted, comparing deviances for a range of values.

For the models with lowest deviance, the simpler approach was then used to estimate  $\beta_1$  (and  $\beta_2$ ,  $\beta_3$ ) and its standard error (SE) for each block. For a particular model, goodness-of-fit for a block was tested by comparing observed and fitted number of cases (and for case-control studies also the observed number of controls) at each level of amount smoked (including never smokers). The fitted values were estimated as described in Goodness of fit<sup>[11]</sup>. As also described there, the sum of (observed - fitted)<sup>2</sup>/fitted over levels was taken as an approximate chi-squared on  $\ell - 1$  degrees of freedom (df) for prospective studies or on  $2\ell - 1$  df for case-control studies. Information on overall goodness-of-fit was derived by summing observed and fitted values over blocks for never smokers and for specified levels of amount smoked, and similarly deriving an approximate chi-squared statistic. Plots of observed and predicted RRs per block were also examined.

### Duration of smoking

The approach used was as for amount smoked.

### Age of starting to smoke

For smokers of a given age, age of starting ( $a$ ) and dura-

tion ( $y$ ) are directly related. We used the same basic approach, replacing  $y$  by  $70 - a$  to produce a duration-like measure. As this produced a relatively good fit, we did not attempt sensitivity analyses replacing  $y$  by  $60 - a$  or  $80 - a$ .

### Regression analyses

Sources of heterogeneity were studied by inverse-variance weighted regression of  $\beta$ . Between block variation was examined one factor at a time (simple regression), and using forward stepwise methods. The factors used were study type, sex, location, publication year, midpoint age (at baseline for prospective studies), smoking product and study size. The deviance of the fitted models indicated the extent to which heterogeneity was explained.

### Statistical analysis

No multiple testing adjustments were made, significant being defined as  $P < 0.05$ . However, results showing stronger evidence of a relationship ( $P < 0.01$ , or  $P < 0.001$ ), and sometimes weaker evidence ( $P < 0.1$ ) are also distinguished, where appropriate. All data entry and most statistical analyses were carried out using ROELEE version 3.1 (available from PN Lee Statistics and Computing Ltd., 17 Cedar Road, Sutton, Surrey SM2 5DA, United Kingdom). Some analyses used Excel 2003.

## RESULTS

### Studies considered

For each of the 71 studies providing the data used, Studies<sup>[11]</sup>, gives the six character reference code (REF); a brief description incorporating the location, characteristics of the population studied, study design, and study duration; the total number of lung cancers studied; and the measures for which data are analyzed.

### Amount smoked

Details of blocks used for each measure are given in Blocks<sup>[11]</sup>. These include study REF, sex (and where applicable race), study type, location, product smoked, definition of unexposed group, adjustment factors used, current smoker RR, and total numbers of cases in smokers.

The 97 blocks derive from 69 studies, 45 providing results for a single block, 22 results by sex, and 2 (DORGAN and HUMBLE) results by sex and race. 55 blocks (56.7%) are for males, 34 (35.1%) females, and 8 (8.2%) both sexes. 48 (49.5%) are from prospective studies. 43 (44.3%) are from North American studies, 32 (33.0%) from Europe and 17 (17.5%) from Asia, the remaining 5 (5.2%) from South America, Africa or Australasia. Five different combinations of product *vs* unexposed occur: cigarettes  $\pm$  other products *vs* never any product (32 blocks 33.0%), cigarettes  $\pm$  other products *vs* never cigarettes (29, 29.9%), any product *vs* never any product (20, 20.6%), cigarettes only *vs* never any product (13, 13.4%) and cigarettes only *vs* never cigarettes for (3, 3.1%). Of the 8 blocks for sexes combined, 4 (50.0%) concern RRs adjusted for sex, while 64 of the full 97 blocks (66.0%)

**Table 2** Amount smoked by current smokers (cigarettes per day) - dose-response data

Block: Study	Amount smoked groupings <sup>1</sup>	Mean values	RRs <sup>2</sup>
1: AKIBA	1-14, 15-24, 25+	8.11, 19.19, 34.63	3.50, 6.10, 19.10 M
2: AKIBA	1-14, 15+	8.11, 24.09	3.60, 5.80 M
3: ARCHER	1-19, 20, 21+	10, 20, 31.83	3.53, 6.09, 8.52 M
4: AXELSS	20	20	43.30
5: BENSHL	1-9, 10-19, 20+	4.85, 12.73, 26.03	4.00, 9.05, 10.95 M
6: BEST	1-9, 10-20, 21+	4.85, 15.92, 31.83	10.00, 16.41, 17.31 M
7: BOUCOT	1-20, 21-40, 41+	13.38, 29.02, 53.33	54.09, 78.56, 161.70 M
8: BRETT	1-14, 15-24, 25+	8.11, 19.19, 34.63	2.55, 4.25, 8.00 M
9: BROSS	1-20, 21+	13.38, 31.83	4.91, 7.20 M
10: BUFFLE	1-19, 20, 21+	10.20, 31.83	5.60, 11.84, 22.10 M
11: CEDERL	1-7, 8-15, 16+	4.29, 11.61, 25.41	3.40, 7.50, 11.90 M
12: CEDERL	1-7, 8-15, 16+	4.29, 11.61, 25.41	2.83, 7.74, 7.56
13: CHANG	1-10, 11-20, 21+	7.08, 17.67, 31.83	5.02, 10.60, 8.26
14: CHANG	1-10, 11-20, 21+	7.08, 17.67, 31.83	3.03, 4.87, 8.21 M
15: CHOW	1-19, 20-29, 30+	10, 21.35, 38.39	13.88, 21.87, 44.48 M
16: COMSTO	1-19, 20-39, 40+	10, 22.90, 45.71	12.42, 18.16, 24.92 M
17: COMSTO	1-19, 20-39, 40+	10, 22.90, 45.71	7.45, 17.35, 13.27
18: CORREA	1-20, 21+	13.38, 31.83	9.30, 25.30 M
19: CPSI	1-9, 10-19, 20-39, 40+	4.85, 12.73, 22.90, 45.71	4.51, 8.41, 14.30, 17.49 M
20: CPSII	1-9, 10-19, 20, 21-39, 40, 41+	4.85, 12.73, 20, 26.71, 40, 53.33	12.22, 14.52, 21.59, 22.72, 24.14, 45.52 M
21: CPSII	1-9, 10-19, 20, 21-39, 40, 41+	4.85, 12.73, 20, 26.71, 40, 53.33	3.89, 8.33, 14.21, 21.40, 19.31, 18.22
22: DARBY	1-14, 15-24, 25+	8.11, 19.19, 34.63	73.47, 95.43, 142.69 M
23: DARBY	1-14, 15-24, 25+	8.11, 19.19, 34.63	15.70, 21.50, 41.62 M
24: DEAN3	1-12, 13-22, 23+	7.65, 18.42, 33.04	5.46, 7.42, 21.66 M
25: DEAN3	1-12, 13-22, 23+	7.65, 18.42, 33.04	3.16, 8.42, 24.24 M
26: DEKLER	1-14, 15-24, 25+	8.11, 19.19, 34.63	19.40, 23.00, 32.50 M
27: DOLL2	1-14, 15-24, 25+	8.11, 19.19, 34.63	5.20, 10.60, 22.40 M
28: DOLL2	1-14, 15-24, 25+	8.11, 19.19, 34.63	1.29, 6.43, 29.71 M
29: DORANT	1-9, 10-19, 20+	4.85, 12.73, 26.03	8.52, 27.22, 36.24 M
30: DORGAN	1-19, 20+	10, 26.03	9.13, 20.65 M
31: DORGAN	1-19, 20+	10, 26.03	26.67, 72.46 M
32: DORGAN	1-19, 20+	10, 26.03	6.55, 24.13 M
33: DORGAN	1-19, 20+	10, 26.03	7.43, 41.43 M
34: DORN	1-9, 10-20, 21-39, 40+	4.85, 15.92, 26.71, 45.71	4.02, 9.92, 17.19, 22.75 M
35: ENGELA	1-4, 5-9, 10-14, 15-19, 20+	2.5, 6.5, 10.88, 15.83, 26.03	1.40, 4.10, 7.00, 11.00, 15.00 M
36: ENGELA	1-4, 5-9, 10-14, 15+	2.5, 6.5, 10.88, 24.09	12.00, 12.00, 24.00, 26.00
37: ENSTRO	1-9, 10-19, 20, 21-39, 40+	4.85, 12.73, 20, 26.71, 45.71	4.74, 7.68, 13.65, 16.08, 19.41 M
38: ENSTRO	1-9, 10-19, 20, 21+	4.85, 12.73, 20, 31.83	2.15, 4.31, 9.48, 16.47 M
39: GAO2	1-19, 20-29, 30+	10, 21.35, 38.39	3.36, 7.54, 10.63 M
40: GILLIS	1-14, 15-24, 25-34, 35-49, 50+	8.11, 19.19, 28.13, 39, 53.33	4.50, 7.60, 8.60, 9.70, 7.80
41: HAENSZ	1-20, 21+	13.38, 31.83	1.77, 5.15 M
42: HAMMO2	1-19, 20+	10.00, 26.03	9.15, 10.39 M
43: HAMMON	1-9, 10-20, 21-39, 40+	4.85, 15.92, 26.71, 45.71	7.44, 8.42, 17.91, 20.64 M
44: HIRAYA	1-9, 10-19, 20+	4.85, 12.73, 26.03	2.06, 4.00, 6.24 M
45: HIRAYA	1-9, 10-19, 20+	4.85, 12.73, 26.03	2.25, 2.56, 4.47 M
46: HITOSU	1-14, 15-24, 25+	8.11, 19.19, 34.63	2.08, 2.82, 4.68 M
47: HITOSU	1-14, 15+	8.11, 24.09	3.11, 3.17 M
48: HOLE	1-14, 15-24, 25-34, 35+	8.11, 19.19, 28.13, 44.38	5.47, 8.90, 10.75, 7.49
49: HUMBLE	1-19, 20+	10, 26.03	9.20, 24.70 M
50: HUMBLE	1-19, 20+	10, 26.03	11.60, 26.10 M
51: HUMBLE	1-19, 20+	10, 26.03	19.20, 16.00
52: HUMBLE	1-19, 20+	10, 26.03	18.50, 36.90 M
53: KAISE2	1-19, 20+	10, 26.03	4.47, 10.34 M
54: KAISE2	1-19, 20+	10, 26.03	7.61, 22.12 M
55: KAISER	1-19, 20-40, 41+	10, 24.32, 53.33	6.58, 17.24, 20.91 M
56: KAISER	1-19, 20-40, 41+	10, 24.32, 53.33	3.42, 7.98, 12.63 M
57: KANELL	1-10, 11-20, 21-35, 36+	7.08, 17.67, 26.71, 45.71	1.71, 7.06, 20.39, 34.22 M
58: KATSOU	1-20, 21+	13.38, 31.83	2.26, 7.46 M
59: KAUFMA	1-14, 15-24, 25-34, 35-44, 45+	8.11, 19.19, 28.13, 39, 53.33	8.00, 15.00, 28.00, 43.00, 60.00 M
60: KINLEN	1-14, 15-24, 25+	8.11, 19.19, 34.63	10.61, 14.14, 21.74 M
61: KNEKT	1-14, 15+	8.11, 24.09	5.00, 12.70 M
62: KOO	1-10, 11-20, 21-30	7.08, 17.67, 25.88	1.36, 7.29, 1.52
63: LIAW	1-10, 11-20, 21+	7.08, 17.67, 31.83	3.10, 3.60, 8.30 M
64: LIDDEL	1-19, 20+	10, 26.03	3.33, 5.02 M
65: MACLEN	1-9, 10-19, 20-29, 30+	4.85, 12.73, 21.35, 38.39	1.36, 3.41, 4.16, 5.00 M
66: MACLEN	1-9, 10-19, 20+	4.85, 12.73, 26.03	0.76, 3.44, 3.84
67: MATOS	1-14, 15-24, 25+	8.11, 19.19, 34.63	1.60, 8.00, 15.00 M



68: MIGRAN	1-9, 10-19, 20, 21+	4.85, 12.73, 20, 31.83	4.01, 4.24, 5.14, 5.93 M
69: MIGRAN	1-9, 10-19, 20	4.85, 12.73, 20	4.88, 6.53, 7.48 M
70: MRFITR	1-19, 20-39, 40+	10, 22.90, 45.71	10.86, 50.12, 56.43 M
71: NAM	1-24, 25+	14.06, 34.63	6.70, 10.27 M
72: NAM	1-24, 25+	14.06, 34.63	9.06, 16.65 M
73: PARKIN	1-14, 15+	8.11, 24.09	3.90, 5.20 M
74: PERSH2	1-9, 10+	4.85, 20.90	5.76, 11.34 M
75: PETO	1-14, 15+	8.11, 24.09	5.50, 9.49 M
76: PEZZO2	1-20, 21-40, 41+	13.38, 29.02, 53.33	8.00, 44.39, 112.13 M
77: PEZZOT	1-20, 21-40, 41+	13.38, 29.02, 53.33	7.40, 70.00, 246.50 M
78: PRESCO	1-14, 15+	8.11, 24.09	10.20, 19.96 M
79: PRESCO	1-14, 15+	8.11, 24.09	6.36, 10.08 M
80: SEGI2	1-9, 10-19, 20-29, 30-39, 40+	4.85, 12.73, 21.33, 31.07, 45.71	2.10, 3.10, 3.40, 6.90, 7.90 M
81: SEGI2	1-9, 10-19, 20+	4.85, 12.73, 26.03	2.90, 1.44, 1.03
82: SHAW	1-19, 20+	10, 26.03	6.31, 30.48 M
83: SOBUE	1-19, 20-29, 30+	10, 21.35, 38.39	3.52, 4.00, 4.55 M
84: SPEIZE	1-4, 5-14, 15-24, 25-34, 35+	2.5, 9.42, 19.19, 28.13, 44.38	2.70, 5.20, 12.60, 15.70, 22.00 M
85: STOCKW	1-19, 20-40, 41+	10, 24.32, 53.33	6.67, 14.51, 28.84 M
86: SVENSS	1-10, 11-20, 21+	7.08, 17.67, 31.83	4.60, 12.60, 59.00 M
87: TENKAN	1-14, 15-24, 25+	8.11, 19.19, 34.63	15.86, 20.25, 24.97 M
88: TSUGAN	1-15, 16-35, 36+	9.33, 22.45, 45.71	0.90, 1.22, 1.66
89: TULINI	1-14, 15-24, 25+	8.11, 19.19, 34.63	6.02, 12.00, 27.30 M
90: TULINI	1-14, 15-24, 25+	8.11, 19.19, 34.63	8.17, 26.30, 38.70 M
91: TVERDA	1-9, 10-19, 20+	4.85, 12.73, 26.03	2.14, 3.32, 6.56 M
92: TVERDA	1-9, 20+	4.85, 26.03	4.53, 18.00 M
93: WAKAI	1-19, 20-20, 30+	10.00, 21.35, 38.39	1.80, 4.01, 9.19 M
94: WU	1-20, 21+	13.38, 31.83	3.25, 8.48 M
95: WYNDE6	1-10, 11-20, 21-30, 31+	7.08, 17.67, 25.88, 43.06	6.80, 11.16, 17.32, 28.22 M
96: WYNDE6	1-10, 11-20, 21-30, 31+	7.08, 17.67, 25.88, 38.39	3.75, 11.97, 21.64, 39.14 M
97: YAMAGU	1-20, 21+	13.38, 31.83	3.75, 12.14 M

<sup>1</sup>In some studies, amount smoked is based on cigarette equivalents for cigars and pipes; <sup>2</sup>M indicates a strictly monotonic rise in RR with increasing amount smoked.

concern age-adjusted RRs. Race and/or other factors were adjusted for in 28 (28.9%) blocks.

Table 2 gives for each block the levels used to categorize amount smoked and the corresponding estimated mean values and RRs for each level. The RRs reveal an obvious trend for risk to rise with amount smoked. Of the 96 blocks with more than one level, 84 (87.5%) show a strictly monotonic increase in RR. However, considerable variation is evident in the RR for the highest exposure.

Table 3 gives the pool-first results investigating model suitability. The exponential model is particularly poor, explaining only 21.75% of the overall deviance in the estimates of log RR. The log model is also relatively poor. The linear, quadratic and cubic models are better. However, despite involving more parameters, the cubic model explains less of the overall deviance than do the best-fitting power or log-with-baseline models. The residual deviance is lowest for the log-with-baseline model, the best-fitting W value explaining 94.12% of the overall deviance, though the best-fitting power model explains almost as much (93.95%).

Fit Amount Smoked<sup>[11]</sup>, gives full details for the further analyses carried out using the linear, and best-fitting power and log-with-baseline models. These include 95% CIs for the RRs in Table 2, and observed and fitted numbers by level for each block.

For each of these models, Table 4 compares the observed and fitted numbers of cases summed over blocks

for never smokers and for current smokers by amount smoked. The linear model fits poorly, overestimating cases for never smokers and 30+ cigs/d smokers and underestimating for 1-30 cigs/d smokers, the model implying a far steeper increase with amount smoked than observed. This is consistent with the block-specific goodness-of-fit tests, 63 showing misfits significant at  $P < 0.05$ . This model is clearly inadequate for amount smoked.

Although Table 4 shows highly significant ( $P < 0.001$ ) misfit to both the power and log-with-baseline models, the misfit is not substantial, with observed and expected numbers generally agreeing to a few percent.

For each block, and both models, Table 5 gives fitted values of  $\beta_1$  and SE and goodness-of-fit  $P$  values. A number of blocks show significant ( $P < 0.05$ ) misfit, these tending to be the same blocks for both models. We comment on those 15 blocks where the  $P$  value for the log-with-baseline model is  $< 0.01$  (Fit Amount Smoked<sup>[11]</sup> and Table 2 for further details). These divide into various categories. Three blocks (19: CPS I, 34: DORN, 37: ENSTRO males) involve very large numbers of cases (Table 3) where the model appears to fit quite well, though in block 19: CPS I the observed flattening of response for 40+ cigs/d is not well fitted. Seven blocks (6: BEST, 20: CPS II males, 22: DARBY males, 43: HAMMON, 60: KINLEN, 74: PERSH2, 87: TENKAN) show a marked risk increase for the lowest level of amount smoked, but the slope subsequently flattens. In contrast the reverse is true for five blocks (24: DEAN3 males, 38: ENSTRO fe-

**Table 3** Comparing the suitability of different models relating log RR to amount smoked by current smokers, expressed as d = cigarettes per day/10

Model	Parameter value <sup>1</sup>	Fitted coefficient(s) (SE)	Deviance	DF	Deviance explained (%)
Null	-	-	24894.53	97	
Linear: log RR = $\beta_1 d$	-	$\beta_1 = 0.6107$ (0.0046)	7265.32	96	70.82
Quadratic: log RR = $\beta_1 d + \beta_2 d^2$	-	$\beta_1 = 1.4121$ (0.0130), $\beta_2 = -0.1792$ (0.0027)	2907.39	95	88.32
Cubic: log RR = $\beta_1 d + \beta_2 d^2 + \beta_3 d^3$	-	$\beta_1 = 2.1915$ (0.0266), $\beta_2 = -0.6346$ (0.0138), $\beta_3 = 0.0633$ (0.0019)	1779.05	94	92.86
Power: log RR = $\beta_1 d^Y$	Y = 0.32	$\beta_1 = 1.8922$ (0.0124)	1512.49	96	
	Y = 0.33	$\beta_1 = 1.8691$ (0.0122)	1506.19	96	
	Y = 0.34	$\beta_1 = 1.8457$ (0.0121)	1506.07	96	93.95
	Y = 0.35	$\beta_1 = 1.8222$ (0.0119)	1511.92	96	
	Y = 0.36	$\beta_1 = 1.7986$ (0.0118)	1523.54	96	
	Y = 0.50	$\beta_1 = 1.4673$ (0.0097)	2179.71	96	
	Y = 1.00	$\beta_1 = 0.6107$ (0.0046)	7265.32	96	
	Y = 2.00	$\beta_1 = 0.0969$ (0.0010)	14739.19	96	
Log: log RR = $\beta_1 \log d$	-	$\beta_1 = 1.2265$ (0.0107)	11674.70	96	53.10
Exponential: log RR = $\beta_1 \exp d$	-	$\beta_1 = 0.0120$ (0.0002)	19480.21	96	21.75
Log-with-baseline: log RR = $\beta_1 \log(1 + Wd)$	W = 7.5	$\beta_1 = 0.8520$ (0.0056)	1466.21	96	
	W = 7.7	$\beta_1 = 0.8456$ (0.0055)	1465.37	96	
	W = 7.9	$\beta_1 = 0.8394$ (0.0054)	1464.88	96	
	W = 8.0	$\beta_1 = 0.8364$ (0.0055)	1464.76	96	
	W = 8.1	$\beta_1 = 0.8334$ (0.0054)	1464.71	96	94.12
	W = 8.2	$\beta_1 = 0.8305$ (0.0054)	1464.73	96	
	W = 8.3	$\beta_1 = 0.8277$ (0.0054)	1464.82	96	

<sup>1</sup>Note that we only sought the best-fitting value of Y to two decimal places and of W to one decimal place.

**Table 4** Amount smoked by current smokers - observed and fitted lung cancers for the linear, best power and best log-with-baseline model, with  $\beta_1$  fitted separately for each block

Midpoint amount smoked (cigs/d)	Observed <sup>1</sup>	Fitted <sup>2</sup>		
		Linear model	Best power model	Best log-with-baseline model
< 5	1249.17	1023.00	1297.42	1173.88
5 to < 10	2579.62	2156.31	2595.37	2539.78
10 to < 15	6125.69	4749.92	6276.74	6299.68
15 to < 20	7940.18	6678.14	8009.06	8156.50
20 to < 30	18138.36	15724.45	17468.99	17678.51
30 to < 40	3858.94	4106.12	3743.23	3701.31
40+	7703.88	8860.42	8115.77	7949.95
Never smoked	6649.61	10947.08	6738.86	6745.84
Total	54245.45	54245.45	54245.45	54245.45
Fit statistic <sup>3</sup>		3792.53	65.29	56.78

<sup>1</sup>Observed pseudo-number of lung cancer cases, summed over blocks; <sup>2</sup>Fitted pseudo-number of lung cancer cases, summed over blocks. For each study, the fitted number of cases for each block is calculated from the numbers at risk and the fitted RRs by amount smoked, derived from the fitted value of  $\beta_1$ ;

<sup>3</sup>Based on summation of (observed-fitted)<sup>2</sup>/fitted, the summation also including terms for the observed and fitted total numbers of controls (not shown). The statistic can be considered to be approximately chisquared on 12 DF and is significant at  $P < 0.001$  for all three models.

males, 57: KANELL, 76: PEZZO2, 77: PEZZOT) with the RR for the highest exposure greater than predicted from the response at lower levels. For some of the 15 blocks, the number of cases in never smokers is relatively low (less than 10 in 6 of them) and the best-fitting model gives rather different fitted numbers, so the fitted block of RRs appears substantially different from that observed. For example, in block 6: BEST where the observed pseudo-number of cases in never smokers is 6.88, and the observed RRs are 10.00, 16.41 and 17.31 for 1-9, 10-20 and 21+ cigs/d, the fitted number of cases in never smokers is 23.61 and the fitted RRs are 2.47, 4.46 and 6.47.

Table 6 presents results of weighted simple regression

analyses of  $\beta_1$  for the log-with-baseline model. There is highly significant ( $P < 0.001$ ) variation by continent, with  $\beta_1$  much lower for Asian studies, and by study size, larger studies giving higher  $\beta_1$  values. Some variation is also seen for sex ( $P < 0.05$ ), study type, publication year and midpoint age ( $P < 0.1$ ), but not with product definition or unexposed group. Table 6 also presents predicted RRs at 20 cigs/d. The variation by continent is clear.

In a forward stepwise analysis (not shown), continent remained highly significant ( $P < 0.001$ ), but no other factor remained significant at  $P < 0.05$ . The association with study size seems due to a strong correlation with continent.

**Table 5** Amount smoked by current smokers - fitted values of  $\beta_1$  and SE, and  $P$  values for goodness-of-fit tests for the best-fitting power model and log-with-baseline model

Block: Study	Log-with-baseline model <sup>1</sup> log RR = $\beta_1 [1 + (8.10c/10)]$			Power model <sup>1</sup> log RR = $\beta_1 [(c/10)^{0.34}]$		
	$\beta_1$	SE $\beta_1$	$P$ (fit) <sup>2</sup>	$\beta_1$	SE $\beta_1$	$P$ (fit) <sup>2</sup>
1: AKIBA	0.6599	0.0712	NS	1.4882	0.1609	NS
2: AKIBA	0.6099	0.0660	NS	1.3445	0.1453	NS
3: ARCHER	0.6713	0.1274	NS	1.4932	0.2835	NS
4: AXELSS	1.3245	0.2214	NS	2.9770	0.4976	NS
5: BENSHL	0.7133	0.1169	NS	1.6337	0.2688	NS
6: BEST	0.5678	0.0824	0.0000	1.3483	0.1948	0.0001
7: BOUCOT	0.8332	0.2140	NS	1.7439	0.4482	NS
8: BRETT	0.6637	0.1146	NS	1.4946	0.2575	NS
9: BROSS	0.6090	0.0735	NS	1.3514	0.1634	NS
10: BUFFLE	0.9349	0.1040	NS	2.0810	0.2313	NS
11: CEDERL	0.8191	0.0726	NS	1.8388	0.1634	NS
12: CEDERL	0.7697	0.0693	NS	1.6833	0.1520	NS
13: CHANG	0.6772	0.1522	NS	1.5044	0.3402	NS
14: CHANG	0.6255	0.1165	NS	1.3937	0.2595	NS
15: CHOW	1.0002	0.1050	NS	2.2134	0.2311	NS
16: COMSTO	0.8497	0.1552	NS	1.8491	0.3399	NS
17: COMSTO	0.8857	0.1172	NS	1.9560	0.2601	NS
18: CORREA	0.9945	0.0483	NS	2.2052	0.1069	NS
19: CPS I	0.8314	0.0348	0.0002	1.8278	0.0773	0.0000
20: CPS II	0.8262	0.0331	0.0000	1.8467	0.0739	0.0000
21: CPS II	0.8972	0.0304	0.0153	1.9874	0.0677	0.0006
22: DARBY	0.8198	0.1213	0.0000	1.9085	0.2770	0.0000
23: DARBY	1.0879	0.0841	NS	2.4434	0.1879	NS
24: DEAN3	0.8271	0.0724	0.0009	1.8968	0.1639	0.0042
25: DEAN3	0.8617	0.0801	(0.0569)	1.9152	0.1789	0.0368
26: DEKLER	0.5704	0.1839	NS	1.3114	0.4155	NS
27: DOLL2	1.0464	0.0644	(0.0824)	2.3531	0.1443	NS
28: DOLL2	1.1001	0.1698	NS	2.4424	0.3745	0.0091
29: DORANT	1.1082	0.0867	0.0322	2.5183	0.1972	0.0270
30: DORGAN	0.9728	0.0937	NS	2.1814	0.2101	NS
31: DORGAN	1.4168	0.1835	NS	3.1557	0.4086	NS
32: DORGAN	0.9636	0.0595	NS	2.1418	0.1325	NS
33: DORGAN	1.0359	0.1844	NS	2.2976	0.4094	NS
34: DORN	0.9024	0.0171	0.0001	2.0099	0.0381	0.0000
35: ENGELA	1.0083	0.1270	NS	2.3043	0.2995	NS
36: ENGELA	1.0897	0.1561	0.0206	2.5654	0.3553	NS
37: ENSTRO	0.8261	0.0312	0.0000	1.7910	0.0684	0.0000
38: ENSTRO	0.8229	0.0252	0.0000	1.8368	0.0564	0.0000
39: GAO2	0.7037	0.1042	NS	1.5526	0.2304	NS
40: GILLIS	0.5811	0.0729	NS	1.2704	0.1614	NS
41: HAENSZ	0.3382	0.0805	NS	0.7563	0.1796	NS
42: HAMMO2	0.5198	0.1167	0.0112	1.1870	0.2641	0.0138
43: HAMMON	0.8032	0.0748	0.0038	1.8185	0.1677	0.0116
44: HIRAYA	0.5974	0.0337	NS	1.3374	0.0756	NS
45: HIRAYA	0.4424	0.0501	NS	0.9729	0.1100	NS
46: HITOSU	0.4573	0.1182	NS	1.0324	0.2654	NS
47: HITOSU	0.4988	0.1223	NS	1.0992	0.2686	NS
48: HOLE	0.6142	0.1082	NS	1.3410	0.2413	NS
49: HUMBLE	1.0435	0.1431	NS	2.3386	0.3207	NS
50: HUMBLE	1.0378	0.2633	NS	2.3306	0.5907	NS
51: HUMBLE	0.8922	0.1393	NS	1.9997	0.3114	NS
52: HUMBLE	1.2268	0.2487	NS	2.7318	0.5532	NS
53: KAISE2	0.7599	0.1018	NS	1.6991	0.2278	NS
54: KAISE2	1.0098	0.1107	NS	2.2592	0.2478	NS
55: KAISER	0.7685	0.0477	0.0247	1.6130	0.1010	0.0033
56: KAISER	0.6882	0.0570	NS	1.4932	0.1238	NS
57: KANELL	0.8982	0.0647	0.0000	1.9758	0.1442	0
58: KATSOU	0.4517	0.1260	NS	1.0089	0.2809	NS
59: KAUFMA	1.0712	0.0583	NS	2.3560	0.1278	NS
60: KINLEN	0.6021	0.0616	0.0017	1.3838	0.1399	0.0055
61: KNEKT	0.8644	0.1264	NS	1.9618	0.2876	NS
62: KOO	0.4309	0.1424	NS	0.9299	0.3133	NS
63: LIAW	0.5538	0.0918	NS	1.2384	0.2045	NS
64: LIDDEL	0.5136	0.0742	NS	1.1535	0.1666	NS

65: MACLEN	0.4973	0.1511	NS	1.0875	0.3355	NS
66: MACLEN	0.4301	0.1161	NS	0.9350	0.2577	NS
67: MATOS	0.8315	0.1122	NS	1.8390	0.2485	NS
68: MIGRAN	0.4590	0.1448	NS	1.0568	0.3286	NS
69: MIGRAN	0.7246	0.2207	NS	1.6547	0.4987	NS
70: MRFTIR	0.6382	0.2225	0.0338	1.1984	0.4461	0.0152
71: NAM	0.6883	0.0711	NS	1.5157	0.1569	NS
72: NAM	0.8481	0.0690	NS	1.8809	0.1532	NS
73: PARKIN	0.6129	0.0554	NS	1.3556	0.1222	NS
74: PERSH2	0.8420	0.0381	0.0004	1.9065	0.0854	0.0446
75: PETO	0.6262	0.1606	NS	1.4651	0.3737	NS
76: PEZZO2	1.3641	0.1251	0.0051	2.9784	0.2710	0.0118
77: PEZZOT	1.5483	0.1569	0.0005	3.4045	0.3415	0.0014
78: PRESCO	0.8289	0.1002	(0.0759)	1.9305	0.2318	NS
79: PRESCO	0.7383	0.0942	NS	1.6725	0.2122	NS
80: SEGI2	0.5679	0.0991	NS	1.2849	0.2208	NS
81: SEGI2	0.1503	0.1297	NS	0.3513	0.2865	NS
82: SHAW	1.1326	0.1121	NS	2.5298	0.2509	NS
83: SOBUE	0.4048	0.0594	NS	0.8893	0.1309	NS
84: SPEIZE	0.8772	0.0390	NS	1.9509	0.0870	NS
85: STOCKW	0.8822	0.0096	NS	1.9383	0.0210	NS
86: SVENSS	0.9255	0.1158	NS	2.0479	0.2573	NS
87: TENKAN	0.6211	0.1092	0.0001	1.4407	0.2484	0.0003
88: TSUGAN	0.1104	0.1169	NS	0.2461	0.2574	NS
89: TULINI	1.0019	0.0901	NS	2.2394	0.2008	NS
90: TULINI	1.2489	0.0862	(0.0728)	2.8447	0.1983	0.0098
91: TVERDA	0.6063	0.0773	NS	1.3657	0.1746	NS
92: TVERDA	0.9309	0.1838	NS	2.1244	0.4198	NS
93: WAKAI	0.6776	0.1099	(0.0831)	1.5073	0.2430	NS
94: WU	0.5916	0.1017	NS	1.3183	0.2263	NS
95: WYNDE6	0.9181	0.0373	NS	2.0237	0.0820	NS
96: WYNDE6	0.9796	0.0371	0.0102	2.1767	0.0825	0.0060
97: YAMAGU	0.6608	0.1223	NS	1.4756	0.2723	NS

<sup>1</sup>c = cigarettes/d; <sup>2</sup>Not significant (NS) indicates  $P \geq 0.1$ .  $P$  values in the range  $0.05 \leq P < 0.1$  are shown in brackets.

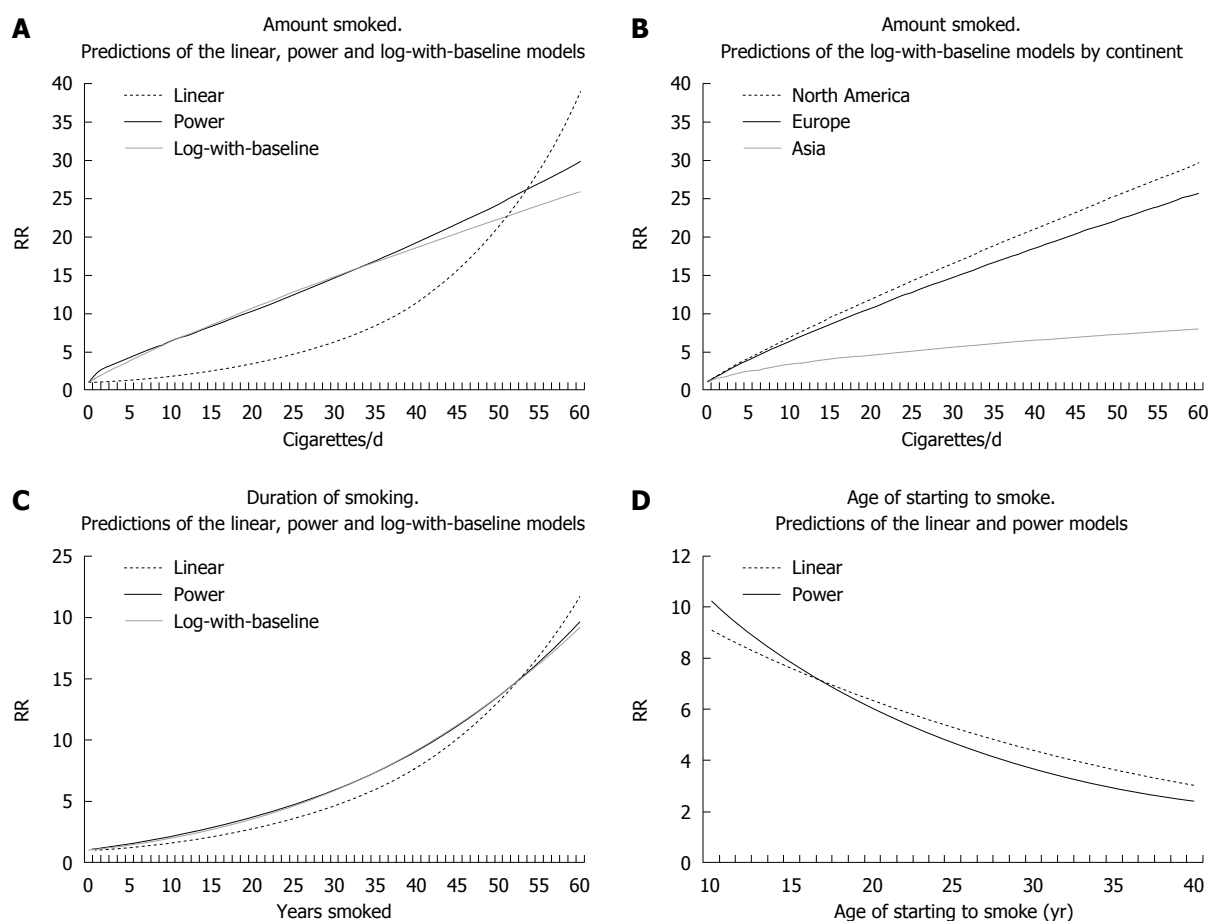
**Table 6 Amount smoked by current smokers - inverse-variance weighted simple regression analyses of  $\beta_1$  based on best-fitting log + baseline model**

Factor	Level	<i>n</i>	$\beta_1$ (95% CI)	<i>P</i> <sup>1</sup>	RR for 20 cigs/d
All		97	0.83 (0.80-0.86)		10.71
Sex	Male	55	0.79 (0.75-0.84)	< 0.05	9.50
	Female	34	0.82 (0.75-0.88)		10.21
	Combined	8	0.89 (0.84-0.94)		12.50
Study type	Case-control	49	0.86 (0.82-0.90)	< 0.1	11.48
	Prospective	48	0.80 (0.76-0.85)		9.85
Continent	North America	43	0.87 (0.84-0.89)	< 0.001	11.48
	Europe	32	0.83 (0.76-0.90)		10.65
	Asia	17	0.53 (0.45-0.61)		4.53
	Other	5	0.80 (0.61-0.99)		9.80
Publication year <sup>2</sup>	< 1990	40	0.83 (0.79-0.88)	< 0.1	10.71
	1990-1994	29	0.77 (0.70-0.84)		8.90
	1995-1999	28	0.87 (0.82-0.92)		11.94
Product <sup>3</sup>	Any product	20	0.75 (0.64-0.85)	NS	8.36
	Cigarettes +/-	61	0.85 (0.81-0.88)		11.09
	Cigarettes only	16	0.82 (0.74-0.90)		10.19
Unexposed	Never cigarettes	32	0.83 (0.76-0.90)	NS	10.49
	Never any product	65	0.84 (0.80-0.87)		10.76
Grouped midpoint age (yr)	< 50	14	0.79 (0.67-0.92)	< 0.1	9.57
	50-59	62	0.85 (0.82-0.89)		11.35
	60+	21	0.79 (0.71-0.84)		9.13
Cases in smokers	< 100	29	0.67 (0.56-0.78)	< 0.001	6.72
	100 to < 200	28	0.73 (0.64-0.83)		8.08
	200 to < 500	16	0.76 (0.65-0.87)		8.72
	500 to < 1000	16	0.80 (0.75-0.86)		9.87
	1000+	8	0.89 (0.85-0.92)		12.42

<sup>1</sup>Probability values for factor considered, presented as < 0.001, < 0.01, < 0.05, < 0.1 or not significant (NS) ( $P \geq 0.1$ ); <sup>2</sup>Of principal publication for the study;

<sup>3</sup>Any product = Smokes cigarettes and/or pipes and/or cigars; Cigarettes +/- = Smokes cigarettes with or without other products (pipes, cigars).





**Figure 1 Model predictions.** A: Amount smoked. Predictions of the linear, power and log-with-baseline models; B: Amount smoked. Predictions of the log-with-baseline models by continent; C: Duration of smoking. Predictions of the linear, power, and log-with-baseline models; D: Age of starting to smoke. Predictions of the linear and power models. The RR is plotted against number of cigarettes per day (A and B), years smoked (C) and age of starting to smoke (D). The linear model for amount smoked is the poorest fit to the data, other models shown fitting the data similarly well.

For the 97 blocks combined the RRs predicted by the log-with-baseline model for 5, 10, 20, 30, 40 and 50 cigs/d are, respectively, 3.86, 6.30, 10.71, 14.77, 18.62 and 22.31. The RRs are similar for the power model (4.30, 6.33, 10.34, 14.61, 19.24 and 24.29), but very different for the linear model (1.36, 1.84, 3.39, 6.25, 11.50 and 21.19). See also Figure 1A which compares model predictions, and Figure 1B which shows the predictions by continent.

### Duration of smoking

Blocks<sup>[11]</sup> gives details of the 35 blocks, derived from 14 studies. CPS I and CPS II provide 20 blocks, with data by sex and age. Three further studies provide sex-specific results, the remaining nine only providing one block each. Three studies are from Europe, two South America and one Asia, the remaining eight being from North America.

Table 7 summarizes the dose-response data. A clear increase in risk with increasing duration is evident, 34 blocks (97.1%) showing a greater RR for the longest than the shortest duration group, and 22 (62.9%) showing a strictly monotonic increase in RR.

Table 8 summarizes the analyses on model suitability. The exponential model is again very poor, explaining only 31.47% of the deviance. Other models differ little, explaining 88.49% to 89.96%. The best single parameter

models are the best-fitting power model (89.96%) and log-with-baseline model (89.89%).

Fit Duration<sup>[11]</sup>, gives full details for the further analyses using the linear model and best-fitting power and log-with-baseline models, laid out as Fit Amount Smoked<sup>[11]</sup>.

Table 9 compares observed and fitted cases summed over blocks. Misfit is similar for all three models, and significant ( $P < 0.001$ ), though its extent seems relatively moderate.

Table 10 gives fitted values of  $\beta_1$  and SE and goodness-of-fit  $P$  values for the power and log-with-baseline models. We comment on six blocks where  $P$  is  $< 0.001$  for both models, and three where  $P$  is  $< 0.001$  for one model (Fit Duration<sup>[11]</sup> and Table 7 for further details). In three blocks (1: AMANDU, 29: HUMBLE, 35: PEZZO2) the RR associated with the lowest duration level is large, but the RRs associated with higher levels are not much larger (or even smaller). In another block (10: CPS I males age 65-74 years) the misfit comes from the lack of rise in risk over short duration levels, while in another (17: CPS II males age 30-44 years) it is associated with the very high risk for the longest duration. In three other blocks (11: CPS I males age 75+ years, 15: CPS I females age 65-74 years, 20: CPS II males age 65-74 years) the misfit is at least partly due to the non-monotonic dose-response.

**Table 7 Duration of smoking by current smokers (yr) - dose-response data**

Block: Study	Duration of smoking groupings (yr)	Mean values	RRs <sup>1</sup>
1: AMANDU	0-24, 25+	12.66, 41.28	5.92, 7.02 M
2: BEST	1-4, 5-9, 10-14, 15-19, 20-29, 30-39, 40+	2.64, 7.02, 12.03, 16.89, 24.03, 33.93, 51.12	1.60, 2.60, 2.30, 3.20, 4.10, 13.90, 14.20
3: BOUCOT	1-39, 40+	32.05, 51.14	42.40, 89.94 M
4: BUFFLE	1-30, 31-40, 41+	19.34, 35.56, 48.49	14.00, 14.70, 18.15 M
5: CEDERL	1-29, 30+	23.94, 41.09	1.80, 7.40 M
6: CEDERL	1-29, 30+	22.26, 39.47	1.60, 9.60 M
7: CPS I	1-29, 30+	23.70, 32.03	3.83, 6.56 M
8: CPS I	1-29, 30-34, 35-39, 40+	24.92, 31.91, 36.52, 43.07	5.58, 13.40, 16.93, 29.61 M
9: CPS I	1-29, 30-34, 35-39, 40-44, 45-49, 50+	23.57, 32.25, 37.75, 42.03, 46.80, 51.81	3.11, 6.57, 10.35, 15.21, 21.46, 33.11 M
10: CPS I	1-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60+	10.67, 30.00, 39.00, 42.21, 47.57, 52.05, 56.73, 62.87	5.46, 6.26, 5.86, 8.22, 12.48, 15.28, 18.86, 28.60
11: CPS I	1-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60+	20.00, 32.00, 37.00, 41.00, 46.25, 51.86, 57.00, 65.03	2.17, 2.27, 12.66, 3.47, 6.22, 6.31, 12.87, 13.23
12: CPS I	1-29, 30+	21.86, 31.37	6.16, 13.80 M
13: CPS I	1-29, 30-34, 35-39, 40+	23.29, 31.61, 36.72, 43.09	2.90, 6.91, 8.49, 19.48 M
14: CPS I	1-29, 30-34, 35-39, 40-44, 45-49, 50+	21.03, 31.92, 37.45, 42.21, 46.65, 52.80	1.51, 3.71, 4.73, 5.72, 7.78, 14.48 M
15: CPS I	1-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55+	21.69, 31.38, 37.19, 41.80, 47.14, 51.76, 57.56	2.30, 3.38, 3.67, 5.81, 7.01, 6.20, 3.32
16: CPS I	1-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55+	18.00, 30.00, 38.00, 41.75, 46.71, 52.25, 61.54	0.38, 3.11, 2.72, 3.22, 1.15, 1.85, 3.20
17: CPS II	1-29, 30+	19.15, 32.13	3.69, 108.27 M
18: CPS II	1-29, 30-34, 35-39, 40+	24.92, 31.91, 36.52, 43.07	7.72, 19.47, 25.01, 36.98 M
19: CPS II	1-29, 30-34, 35-39, 40-44, 45-49, 50+	23.57, 32.25, 37.75, 42.03, 46.80, 51.81	15.51, 17.86, 27.31, 44.71, 50.92, 72.07 M
20: CPS II	1-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60+	10.67, 30.00, 39.00, 42.21, 47.57, 52.05, 56.73, 62.87	9.97, 11.05, 20.45, 18.59, 24.33, 32.06, 40.43, 45.64
21: CPS II	1-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60+	20.00, 32.00, 37.00, 41.00, 46.25, 51.86, 57.00, 66.09	7.54, 4.53, 4.37, 16.30, 16.01, 13.40, 18.09, 21.79
22: CPS II	1-19, 20+	13.93, 24.23	10.98, 8.38
23: CPS II	1-29, 30-34, 35+	23.29, 31.61, 37.81	8.96, 18.34, 23.32 M
24: CPS II	1-29, 30-34, 35-39, 40-44, 45-49, 50+	21.03, 31.92, 37.45, 42.22, 46.65, 52.80	6.50, 13.04, 17.07, 20.56, 23.94, 28.45 M
25: CPS II	1-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55+	21.69, 31.38, 37.19, 41.80, 47.14, 51.76, 57.56	6.64, 5.57, 10.19, 12.96, 15.79, 18.75, 19.34
26: CPS II	1-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60+	18.00, 30.00, 38.00, 42.20, 46.71, 52.23, 57.10, 65.06	3.36, 7.92, 6.31, 8.14, 11.68, 9.45, 9.18, 19.88
27: DEAN2	1-19, 20+	15.09, 38.28	3.21, 3.84 M
28: DEAN2	1-19, 20+	13.93, 35.85	0.98, 5.88
29: HUMBLE	1-29, 30-39, 40-49, 50+	17.18, 34.07, 44.43, 56.60	15.45, 17.54, 19.61, 17.27
30: KAISE2	1-39, 40+	23.76, 51.12	4.86, 15.64 M
31: KAISE2	1-39, 40+	22.73, 48.96	9.09, 30.41 M
32: KATSOU	1-29, 30+	14.58, 42.43	1.29, 7.43 M
33: LIAW	1-20, 21-30, 31+	14.32, 26.18, 44.79	0.90, 2.60, 4.70
34: MATOS	1-24, 25-39, 40+	12.75, 31.16, 50.57	5.20, 7.40, 10.20 M
35: PEZZO2	1-35, 36+	17.15, 48.89	16.25, 26.77 M

<sup>1</sup>M indicates a strictly monotonic rise in RR with increasing duration of smoking.

In the remaining block (9: CPS I males age 55-64 years) the rise is monotonic, but the relatively small RR for the shortest duration does not fit in well with the large RRs for longer durations.

Table 11 presents results of weighted simple regression analysis of  $\beta_1$  for the power model. Significance at  $P < 0.05$  is only seen for midpoint age, with higher  $\beta_1$  values for lower ages. In forward stepwise regressions (not shown), the model included in succession midpoint age ( $P < 0.05$ ), number of cases in smokers ( $P < 0.05$ ), smoking product ( $P < 0.05$ ) and unexposed group ( $P < 0.01$ ). The final model associated increased risk with younger age, larger numbers of cases, smoking of cigarettes only or cigarettes  $\pm$  other products (compared to smoking any product) and with the unexposed being never any product (rather than never cigarettes).

For the 35 blocks combined the RRs predicted by the power model for durations of 10, 20, 30, 40 and 50 years

are, respectively, 2.21, 3.75, 5.96, 9.11 and 13.54. Figure 1C compares model predictions.

### Age of starting to smoke

Blocks<sup>[11]</sup> gives details of the 27 blocks, deriving from 15 studies. One study gives results by age and sex, two by age for males, and five by sex but not age, the remaining seven studies only providing one block each. There are similar numbers of blocks from North America (9), Europe (9) and Asia (8), only one being from elsewhere.

Table 12 summarizes the dose-response data. A relationship of risk to age of starting is not consistently seen. While 13 blocks (48.1%) show a strictly monotonic decline in risk as starting age increases, with risk often substantially higher in early starters, and 6 (22.2%) blocks show a similar non-monotonic tendency, 8 (29.6%) blocks (1, 2, 7-9, 16, 20, 21) show no such tendency.

Table 13 summarizes the analyses on model suitability,

**Table 8** Comparing the suitability of different models relating log RR to duration of smoking by current smokers, expressed as d = years smoked/10

Model	Parameter value <sup>1</sup>	Fitted coefficient(s) (SE)	Deviance	DF	Deviance explained (%)
Null	-	-	6161.11	35	
Linear: log RR = $\beta_1 d$	-	$\beta_1 = 0.5134$ (0.0069)	672.01	34	89.09
Quadratic: log RR = $\beta_1 d + \beta_2 d^2$	-	$\beta_1 = 0.6718$ (0.0237), $\beta_2 = -0.0300$ (0.0043)	623.20	33	89.88
Cubic: log RR = $\beta_1 d + \beta_2 d^2 + \beta_3 d^3$	-	$\beta_1 = 0.6788$ (0.0607), $\beta_2 = -0.0330$ (0.0243), $\beta_3 = 0.0003$ (0.0025)	623.19	32	89.89
Power: log RR = $\beta_1 d^Y$	Y = 0.50	$\beta_1 = 1.1576$ (0.0156)	682.56	34	
	Y = 0.72	$\beta_1 = 0.8180$ (0.0110)	619.30	34	
	Y = 0.73	$\beta_1 = 0.8049$ (0.0108)	618.93	34	
	Y = 0.74	$\beta_1 = 0.7919$ (0.0106)	618.75	34	89.96
	Y = 0.75	$\beta_1 = 0.7791$ (0.0105)	618.76	34	
	Y = 0.76	$\beta_1 = 0.7665$ (0.0103)	618.95	34	
	Y = 1.00	$\beta_1 = 0.5134$ (0.0069)	672.01	34	
	Y = 2.00	$\beta_1 = 0.0863$ (0.0013)	1426.44	34	
Log: log RR = $\beta_1 \log d$	-	$\beta_1 = 1.5870$ (0.0215)	708.96	35	88.49
Exponential: log RR = $\beta_1 \exp d$	-	$\beta_1 = 0.0045$ (0.0001)	4222.01	35	31.47
Log-with-baseline: log RR = $\beta_1 \log(1 + Wd)$	W = 0.10	$\beta_1 = 6.4054$ (0.0861)	629.58	34	
	W = 0.15	$\beta_1 = 4.6503$ (0.0625)	623.99	34	
	W = 0.18	$\beta_1 = 4.0572$ (0.0545)	622.82	34	
	W = 0.19	$\beta_1 = 3.9001$ (0.0524)	622.69	34	
	W = 0.20	$\beta_1 = 3.7581$ (0.0505)	622.67	34	89.89
	W = 0.21	$\beta_1 = 3.6293$ (0.0488)	622.74	34	
	W = 0.22	$\beta_1 = 3.5117$ (0.0472)	622.89	34	

<sup>1</sup>Note that we only sought the best-fitting value of W and Y to two decimal places.

**Table 9** Duration of smoking by current smokers - observed and fitted lung cancers for the linear, best power and best log-with-baseline model, with  $\beta_1$  fitted separately for each block

Midpoint years smoked	Observed <sup>1</sup>	Fitted <sup>2</sup>		
		Linear model	Best power model	Best log-with-baseline model
< 15	109.36	89.10	93.61	91.19
15 to < 30	605.64	584.79	666.12	645.42
30 to < 45	3968.47	3917.11	4000.71	4012.22
45+	3493.56	3573.95	3499.85	3499.95
Never smoked	2533.97	2546.03	2450.69	2462.20
Total	10710.98	10710.98	10710.98	10710.98
Fit statistic <sup>3</sup>		28.14	25.70	24.53

<sup>1</sup>Observed pseudo-number of lung cancer cases, summed over blocks; <sup>2</sup>Fitted pseudo-number of lung cancer cases, summed over blocks. For each study, the fitted number of cases for each block is calculated from the numbers at risk and the fitted RRs by years smoked, derived from the fitted value of  $\beta_1$ ;

<sup>3</sup>Based on summation of (observed-fitted)<sup>2</sup>/fitted, the summation also including terms for the observed and fitted total numbers of controls (not shown). The statistic can be considered to be approximately chisquared distributed on 6 df, and is significant at  $P < 0.001$  for all three models.

taking (70-age at starting) as a duration-like measure. The exponential model again is the poorest, explaining only 60.24% of the deviance, and the log model is also poorer than the other models. Although slightly more deviance is explained by the cubic model, the best-fitting power model is the best simple model, explaining 88.29% of the deviance. Log-with-baseline models were also tried (not shown), but the best-fitting value of W was extremely low, so it became essentially identical to the linear model, having the same deviance.

Fit Age Start<sup>[11]</sup>, gives full details of the further analyses using the linear and best-fitting power models.

Table 14 compares observed and fitted cases summed over blocks. Both the linear and best-fitted power models fit well.

Table 15 gives fitted values of  $\beta_1$  and SE and goodness-

of-fit  $P$  values. For no block does either model show misfit at  $P < 0.01$ , though misfits at  $P < 0.05$  are sometimes seen. We comment on four blocks with some evidence ( $P < 0.1$ ) of misfit to both models (Table 12 and Fit Age Start<sup>[11]</sup> for further details). For block 4 (CPS I males aged 55-69) the misfit seems due to the relatively small decline in risk between age of start 1-14 and 15-19 years, compared to a greater decline subsequently. For block 5 (CPS I males aged 70-84 years), risk again declines substantially over ages of start 15-19, 20-24 and 25+ years, but risk is slightly less at age 1-14 than 15-19 years. For block 12 (ENGELA males), risk decreases from age 1-19 to 20-29 years but then falls no further. For block 17 (LIAW), the pattern is non-monotonic.

Table 16 presents results of the weighted simple regression analyses of  $\beta_1$  for the power model. Various fac-

**Table 10** Duration of smoking by current smokers - fitted values for of  $\beta_1$  and SE, and  $P$  values for goodness-of-fit tests for the best-fitting power model and log-with-baseline model

Block: Study	Log-with-baseline model <sup>1</sup> log RR = $\beta_1 [1 + (0.2 \text{ yr}/10)]$			Power model <sup>1</sup> log RR = $\beta_1 (\text{yr}/10)^{0.74}$		
	$\beta_1$	SE $\beta_1$	$P$ (fit) <sup>2</sup>	$\beta_1$	SE $\beta_1$	$P$ (fit) <sup>2</sup>
1: AMANDU	1.5114	0.5504	0.0030	0.3454	0.1210	0.0046
2: BEST	3.7163	0.4210	(0.0548)	0.7995	0.0908	0.0469
3: BOUCOT	6.7731	1.6574	NS	1.4235	0.3482	NS
4: BUFFLE	4.5141	0.4888	0.0466	0.9570	0.1029	(0.0780)
5: CEDERL	3.4044	0.7166	NS	0.7147	0.1512	NS
6: CEDERL	3.3566	0.7692	NS	0.6990	0.1609	NS
7: CPS I	3.6386	1.4531	NS	0.7530	0.3009	NS
8: CPS I	5.7248	0.4753	(0.0947)	1.2122	0.1011	(0.0518)
9: CPS I	4.8058	0.2032	0.0003	1.0130	0.0429	0.0002
10: CPS I	3.7681	0.1557	0.0002	0.7976	0.0328	0.0010
11: CPS I	3.1437	0.1768	0.0079	0.6566	0.0369	0.0073
12: CPS I	5.1328	1.5211	NS	1.0518	0.3119	NS
13: CPS I	3.9960	0.3377	0.0308	0.8342	0.0708	0.0165
14: CPS I	2.8575	0.1844	(0.0637)	0.5999	0.0388	0.0383
15: CPS I	2.6384	0.1872	0.0000	0.5544	0.0394	0.0000
16: CPS I	1.5443	0.3346	NS	0.3238	0.0702	NS
17: CPS II	9.5482	2.4662	0.0149	1.9485	0.5160	0.0076
18: CPS II	6.1160	0.5675	NS	1.2874	0.1199	NS
19: CPS II	5.9905	0.3222	(0.0749)	1.2722	0.0684	(0.0921)
20: CPS II	4.5913	0.1996	0.0058	0.9714	0.0421	0.0210
21: CPS II	3.6604	0.2069	NS	0.7645	0.0432	NS
22: CPS II	5.1073	3.0925	NS	1.1011	0.6362	NS
23: CPS II	5.8103	0.4709	NS	1.2180	0.0988	NS
24: CPS II	4.9498	0.2256	NS	1.0461	0.0476	NS
25: CPS II	4.1028	0.1646	NS	0.8656	0.0347	NS
26: CPS II	3.3310	0.1915	NS	0.6993	0.0402	NS
27: DEAN2	2.1038	0.3399	NS	0.4515	0.0723	NS
28: DEAN2	2.9155	0.6412	NS	0.5984	0.1334	NS
29: HUMBLE	4.2780	0.2807	0.0091	0.9005	0.0590	0.0148
30: KAISE2	3.8758	0.4357	NS	0.8202	0.0922	NS
31: KAISE2	4.7883	0.5320	NS	1.0198	0.1130	NS
32: KATSOU	2.9305	0.7339	NS	0.6031	0.1527	NS
33: LIAW	2.4277	0.4372	NS	0.5093	0.0921	NS
34: MATOS	3.4356	0.5336	NS	0.7276	0.1124	NS
35: PEZZO2	3.3022	0.5283	0.0002	0.7260	0.1137	0.0005

<sup>1</sup>Years smoked; <sup>2</sup>Not significant (NS) indicates  $P \geq 0.1$ .  $P$  values in the range  $0.05 \leq P < 0.1$  are shown in brackets.

tors are significant at  $P < 0.05$ , including number of lung cancer cases ( $P < 0.001$ ), continent ( $P < 0.01$ ), publication year ( $P < 0.01$ ), sex ( $P < 0.05$ ) and product smoked ( $P < 0.05$ ), though in a forward stepwise model (details not shown), only two factors were included: number of lung cancer cases ( $P < 0.001$ ) and midpoint age ( $P < 0.05$ ). However, the relationship of  $\beta_1$  to number of cases was not smooth (higher risks for 100 to  $< 200$ , and 500 to  $< 1000$  cases and lower risks for  $< 100$ , 200 to  $< 500$  and 1000+) so the result is difficult to interpret. The association with age is related to a lower  $\beta_1$  in older subjects (aged 60+ years).

For the 31 blocks combined, the RRs predicted for age of start 12.5, 15, 17.5, 20, 25 and 30 years are, respectively, 8.94, 7.80, 6.83, 5.99, 4.66 and 3.66 for the power model and 8.31, 7.57, 6.91, 6.30, 5.24 and 4.36 for the linear model (Figure 1D).

## DISCUSSION

In our earlier review<sup>[1]</sup> of the evidence relating smoking

to lung cancer, we demonstrated a clear dose-response with the three measures considered, risk increasing with increasing amount smoked and duration and with decreasing age of starting. We extend this work by fitting parametric models to the dose-relationships.

We tried various models. The most useful were the “linear model” (log RR =  $\beta_1 d$ ), the “power model” (log RR =  $\beta_1 d^Y$ ) and the “log-with-baseline model” [log RR =  $\beta_1 \log(1 + Wd)$ ], where  $d$  is dose. For amount smoked, the linear model proved inadequate, but a reasonable fit was found with the other two models, the best fit being for the log-with-baseline model with  $W = 0.81$ . For duration, all three models were reasonable, the best being the power model with  $Y = 0.74$ . For age of starting, where we used a duration-like dose measure based on (70 - age of starting to smoke), the best-fitting model was again the power model, here with  $Y = 1.44$ .

Inverse-variance weighted analyses were also carried out to identify sources of heterogeneity in  $\beta_1$ . For amount smoked, as expected from our earlier review<sup>[1]</sup>, the major source was continent, the fitted slope being much less

**Table 11 Duration of smoking by current smokers - inverse-variance weighted simple regression analyses of  $\beta_1$  based on best-fitting power model**

Factor	Level	<i>n</i>	$\beta_1$ (95%CI)	<i>P</i> <sup>1</sup>	RR for 30 yr smoked
All		35	0.79 (0.72-0.86)		5.94
Sex	Male	18	0.84 (0.73-0.94)	NS	6.59
	Female	15	0.74 (0.64-0.85)		5.35
	Combined	2	0.79 (0.45-1.12)		5.89
Study type	Case-control	7	0.73 (0.50-0.97)	NS	5.24
	Prospective	28	0.80 (0.72-0.87)		6.02
Continent	North America	27	0.81 (0.73-0.88)	NS	6.19
	Europe	5	0.55 (0.20-0.90)		3.44
	Asia	1	0.51 (-0.11-1.13)		3.15
	Other	2	0.73 (0.19-1.27)		5.15
Publication year <sup>2</sup>	< 1990	8	0.76 (0.52-0.99)	NS	5.50
	1990-1994	2	0.44 (-0.20-1.09)		2.73
	1995-1999	25	0.80 (0.72-0.88)		6.08
Product <sup>3</sup>	Any product	4	0.50 (0.18-0.83)	NS	3.12
	Cigarettes +/-	10	0.86 (0.73-0.98)		6.89
	Cigarettes only	21	0.78 (0.70-0.87)		5.85
Unexposed	Never cigarettes	19	0.77 (0.69-0.86)	NS	5.71
	Never any product	16	0.85 (0.71-0.99)		6.75
Grouped midpoint age (yr)	< 50	8	1.06 (0.78-1.34)	< 0.05	11.00
	50-59	13	0.87 (0.75-0.99)		7.04
	60+	14	0.73 (0.65-0.82)		5.21
Cases in smokers	< 100	14	0.62 (0.39-0.84)	NS	4.01
	100 to < 200	9	0.73 (0.59-0.87)		5.19
	200 to < 500	6	0.78 (0.67-0.89)		5.18
	500 to < 1000	5	0.95 (0.80-1.10)		8.58
	1000+	1	0.80 (0.59-1.01)		6.04

<sup>1</sup>Probability values for factor considered, presented as < 0.001, < 0.01, < 0.05, < 0.1 or not significant (NS) ( $P \geq 0.1$ ); <sup>2</sup>Of principal publication for the study;

<sup>3</sup>Any product = Smokes cigarettes and/or pipes and/or cigars; Cigarettes +/- = Smokes cigarettes with or without other products (pipes, cigars).

**Table 12 Age of starting to smoke by current smokers (yr) - dose-response data**

Block: Study	Age of starting to smoke groupings (yr)	Mean values	RRs <sup>1</sup>
1: CEDERL	< 17, 17-18, 19+	13.83, 17.52, 22.45	6.40, 9.80, 6.50
2: CEDERL	< 17, 17-18, 19+	14.31, 17.56, 24.32	0.61, 1.84, 1.99
3: CPS I	< 15, 15-19, 20-24, 25+	11.98, 16.74, 21.29, 28.47	15.00, 9.71, 7.14, 3.43 M
4: CPS I	< 15, 15-19, 20-24, 25+	11.59, 16.56, 21.00, 30.90	18.16, 16.32, 12.00, 5.21 M
5: CPS I	< 15, 15-19, 20-24, 25+	11.19, 16.61, 21.07, 35.00	14.03, 16.60, 8.66, 1.71
6: CPS I	< 15, 15-19, 20-24, 25+	12.71, 16.88, 21.26, 31.51	9.00, 5.00, 4.00, 1.50 M
7: CPS I	< 20, 20-24, 25+	16.02, 20.95, 33.29	2.59, 3.23, 2.62
8: DEAN3	< 15, 15-19, 20-24, 25+	11.73, 16.67, 21.19, 29.95	5.67, 7.01, 6.86, 6.80
9: DEAN3	< 15, 15-19, 20-24, 25+	12.45, 16.95, 21.15, 32.19	2.41, 2.90, 2.95, 3.70
10: DORN	< 15, 15-19, 20-24, 25+	11.63, 16.43, 21.23, 30.88	23.42, 16.25, 11.06, 5.18 M
11: DORN	< 15, 15-19, 20-24, 25+	11.37, 16.81, 20.69, 32.63	14.18, 11.31, 7.94, 4.95 M
12: ENGELA	< 20, 20-29, 30+	15.08, 22.23, 35.10	7.42, 3.60, 3.79
13: ENGELA	< 20, 20-29, 30+	15.80, 22.63, 35.80	11.29, 8.15, 2.73 M
14: GAO2	< 20, 20-29, 30+	15.11, 22.17, 35.94	8.62, 6.44, 2.15 M
15: HIRAYA	< 20, 20+	14.90, 24.02	5.71, 4.35 M
16: HIRAYA	< 20, 20+	15.87, 27.40	0.78, 2.46
17: LIAW	< 21, 21-24, 25+	15.70, 21.94, 32.03	4.60, 5.90, 1.50
18: MATOS	< 15, 15-19, 20+	11.96, 16.64, 23.49	11.30, 8.60, 5.30 M
19: MIGRAN	< 16, 16-19, 20+	12.60, 17.19, 23.71	10.03, 6.93, 7.79
20: MIGRAN	< 16, 16-19, 20+	12.94, 17.31, 26.17	7.17, 8.29, 7.98
21: MRFITR	< 16, 16-17, 18-19, 20-21, 22-23, 24+	12.93, 16.45, 18.31, 20.40, 22.35, 27.74	45.91, 67.17, 50.54, 27.09, 60.06, 23.91
22: SEGI2	< 20, 20-22, 23+	15.28, 20.77, 26.60	8.21, 5.56, 1.83 M
23: SEGI2	< 20, 20-22, 23+	14.87, 20.90, 28.96	8.70, 5.68, 3.56 M
24: SEGI2	< 20, 20-22, 23+	14.37, 20.59, 30.87	3.26, 1.70, 1.52 M
25: SVENSS	< 19, 19-25, 26+	15.04, 21.31, 33.89	7.82, 13.08, 5.61
26: WAKAI	< 20, 20-29, 30+	14.91, 22.14, 37.23	3.69, 4.62, 2.08
27: WU	< 19, 19-24, 25+	15.06, 20.71, 31.44	10.32, 3.57, 1.55 M

<sup>1</sup>M indicates a strictly monotonic decline in RR with increasing age of starting to smoke.



**Table 13** Comparing the suitability of different models relating log RR to age of starting smoke by current smokers, expressed as  $d = (70 - \text{age at start})/10$ 

Model	Parameter value <sup>1</sup>	Fitted coefficient(s) (SE)	Deviance	DF	Deviance explained (%)
Null	-	-	2145.30	27	
Linear: log RR = $\beta_1 d$	-	$\beta_1 = 0.3681$ (0.0085)	276.67	26	87.10
Quadratic: log RR = $\beta_1 d + \beta_2 d^2$	-	$\beta_1 = 0.1987$ (0.0349) $\beta_2 = 0.0318$ (0.0064)	251.63	25	88.27
Cubic: log RR = $\beta_1 d + \beta_2 d^2 + \beta_3 d^3$	-	$\beta_1 = -0.0415$ (0.2143) $\beta_2 = 0.1304$ (0.0870) $\beta_3 = -0.0100$ (0.0088)	250.34	24	88.33
Power: log RR = $\beta_1 d^Y$	Y = 0.75	$\beta_1 = 0.5515$ (0.0129)	316.69	26	
	Y = 1.00	$\beta_1 = 0.3681$ (0.0085)	276.67	26	
	Y = 1.42	$\beta_1 = 0.1825$ (0.0042)	251.27	26	
	Y = 1.43	$\beta_1 = 0.1794$ (0.0041)	251.24	26	
	Y = 1.44	$\beta_1 = 0.1764$ (0.0041)	251.23	26	88.29
	Y = 1.45	$\beta_1 = 0.1734$ (0.0040)	251.25	26	
	Y = 1.46	$\beta_1 = 0.1705$ (0.0039)	251.28	26	
	Y = 1.50	$\beta_1 = 0.1592$ (0.0037)	251.67	26	
	Y = 2.00	$\beta_1 = 0.0668$ (0.0015)	284.04	26	
Log: log RR = $\beta_1 \log d$	-	$\beta_1 = 1.1460$ (0.0270)	340.23	26	84.14
Exponential: log RR = $\beta_1 \exp d$	-	$\beta_1 = 0.0058$ (0.0002)	852.85	26	60.24

<sup>1</sup>Note that we only sought the best-fitting value of Y to two decimal places.

**Table 14** Age of starting to smoke by current smokers - observed and fitted lung cancers for the linear and best power model, with  $\beta_1$  fitted separately for each block

Age of starting (yr)	Observed <sup>1</sup>	Fitted <sup>2</sup>	
		Linear model	Best power model
< 16	1304.92	1294.48	1387.73
16 to < 20	1964.76	1906.55	1921.20
20 to < 24	1227.48	1266.99	1216.48
25+	2173.90	2219.26	2122.14
Never smoked	894.41	878.20	917.92
Total	7565.47	7565.47	7565.47
Fit statistic <sup>3</sup>		4.77	9.75

<sup>1</sup>Observed pseudo-number of lung cancer cases, summed over blocks;

<sup>2</sup>Fitted pseudo-number of lung cancer cases, summed over blocks. For each study, the fitted number of cases for each block is calculated from the numbers at risk and the fitted relative risks for age of starting to smoke, derived from the fitted value of  $\beta_1$ ; <sup>3</sup>Based on summation of (observed-fitted)<sup>2</sup>/fitted, the summation also including terms for the observed and fitted total numbers of controls (not shown). The statistic can be considered to be approximately chi-squared distributed on 6 df and is not significant ( $P > 0.1$ ) for both models.

steep for Asian than European or North American studies. However, it proved more difficult to identify meaningful major sources for the other measures.

We discuss various issues relating to interpretation of these findings.

### Adequacy of literature search and publication bias

All the data used came from the IESLC database. The source paper<sup>[1]</sup> demonstrated that the search was comprehensive, though limited to papers published before 2000 and studies of 100+ cases. Publication bias was discussed earlier<sup>[1]</sup>, evidence for its existence being considered not strong. The probability of dose-response results being published might depend on the strength of the overall

relationship seen. While clearly demonstrated for passive smoking and lung cancer<sup>[3]</sup>, this seems less relevant here, the association with active smoking being so strong. Nevertheless, some publication bias may exist.

There are various reasons why the fitted dose-relationships may not accurately reflect the true relationships.

### Misclassification of smoking status

It is well-documented (*e.g.*,<sup>[12,13]</sup>) that some subjects deny current or past smoking, so increasing the apparent lung cancer risk in reported never smokers and biasing downwards the estimated smoking RR. Such misclassification is difficult to adjust for, as it varies by aspects of study design, the questions asked, and also by sex, age, location and other demographics. Indeed, higher denial rates in Asian populations<sup>[14]</sup> may contribute to the markedly weaker observed associations seen in Asia.

In prospective studies there is an additional problem, especially in studies with long-term follow-up with no re-interviews to update smoking status. In particular, some subjects classified at baseline as current smokers may quit during follow-up. Also some never smokers may start, though this is less likely given the subjects' age at baseline in many studies.

### Misclassification of amount smoked

Similar problems arise. Subjects may understate (or overstate) the amount they smoke, and during follow-up in prospective studies, may reduce or increase the amount smoked. Although some studies, particularly case-control, may ask questions on habits at various times during the subject's smoking career, the data reported may relate to average consumption. Someone smoking, say, 30 cigarettes/d for 20 years, then 10 cigarettes/d for 20 years, may not have the same risk as someone smoking 20 cigarettes/d for the whole 40 years period. Difficulties in re-

**Table 15** Age of starting to smoke by current smokers - fitted values of  $\beta_1$  and SE, and  $P$  values for goodness-of-fit tests for the linear model and for the best-fitting power model

Block: Study	Linear model <sup>1</sup> log RR = $\beta_1(70 - a)/10$			Power model <sup>1</sup> log RR = $\beta_1(70 - a)/10^{1.44}$		
	$\beta_1$	SE $\beta_1$	$P$ (fit) <sup>2</sup>	$\beta_1$	SE $\beta_1$	$P$ (fit) <sup>2</sup>
1: CEDERL	0.0385	0.0082	NS	0.1828	0.0394	NS
2: CEDERL	0.0140	0.0089	NS	0.0699	0.0445	NS
3: CPS I	0.0461	0.0041	0.0490	0.2167	0.0188	NS
4: CPS I	0.0524	0.0029	(0.0959)	0.2399	0.0135	(0.0579)
5: CPS I	0.0512	0.0047	0.0195	0.2408	0.0223	0.0307
6: CPS I	0.0301	0.0036	(0.0510)	0.1480	0.0175	NS
7: CPS I	0.0237	0.0046	NS	0.1168	0.0238	0.0828
8: DEAN3	0.0332	0.0041	NS	0.1482	0.0189	0.0342
9: DEAN3	0.0209	0.0039	NS	0.0960	0.0187	NS
10: DORN	0.0549	0.0037	(0.0874)	0.2527	0.0166	NS
11: DORN	0.0455	0.0027	NS	0.2086	0.0125	NS
12: ENGELA	0.0360	0.0037	0.0109	0.1682	0.0171	0.0269
13: ENGELA	0.0442	0.0045	NS	0.2186	0.0221	NS
14: GAO2	0.0394	0.0066	NS	0.1909	0.0318	NS
15: HIRAYA	0.0317	0.0022	NS	0.1497	0.0104	0.0288
16: HIRAYA	0.0207	0.0029	NS	0.1086	0.0152	NS
17: LIAW	0.0296	0.0050	0.0396	0.1448	0.0242	(0.0661)
18: MATOS	0.0411	0.0063	NS	0.1933	0.0296	NS
19: MIGRAN	0.0373	0.0087	NS	0.1612	0.0385	NS
20: MIGRAN	0.0400	0.0108	NS	0.1859	0.0519	NS
21: MRFITR	0.0686	0.0237	NS	0.2931	0.1055	NS
22: SEGI2	0.0468	0.0140	NS	0.2325	0.0639	NS
23: SEGI2	0.0440	0.0128	NS	0.1974	0.0557	NS
24: SEGI2	0.0192	0.0109	NS	0.0945	0.0500	NS
25: SVENSS	0.0432	0.0051	NS	0.2064	0.0246	NS
26: WAKAI	0.0275	0.0068	NS	0.1244	0.0318	NS
27: WU	0.0348	0.0063	NS	0.1741	0.0305	NS

<sup>1</sup>a = Age of starting to smoke; <sup>2</sup>Not significant (NS) indicates  $P \geq 0.1$ .  $P$  values in the range  $0.05 \leq P < 0.1$  are shown in brackets.

**Table 16** Age of starting to smoke by current smokers - inverse-variance weighted simple regression analyses of  $\beta_1$  based on best-fitting power model

Factor	Level	$n$	$\beta_1$ (95%CI)	$P^1$	RR for 15 yr start
All		27	0.18 (0.16-0.20)		7.80
Sex	Male	17	0.19 (0.17-0.21)	< 0.05	9.49
	Female	9	0.14 (0.11-0.17)		5.14
	Combined	1	0.14 (0.04-0.25)		5.40
Study type	Case-control	10	0.15 (0.11-0.20)	NS	5.94
	Prospective	17	0.18 (0.16-0.20)		8.37
Continent	North America	9	0.21 (0.19-0.23)	< 0.01	11.49
	Europe	9	0.16 (0.13-0.19)		6.38
	Asia	8	0.14 (0.11-0.17)		5.17
	Other	1	0.19 (0.08-0.31)		9.50
Publication year <sup>2</sup>	< 1990	11	0.15 (0.11-0.19)	< 0.01	5.59
	1990-1994	4	0.14 (0.11-0.17)		5.17
	1995-1999	12	0.20 (0.18-0.22)		10.37
Product <sup>3</sup>	Any product	3	0.17 (0.11-0.24)	< 0.05	7.65
	Cigarettes +/-	16	0.19 (0.17-0.21)		9.09
	Cigarettes only	8	0.13 (0.09-0.17)		4.69
Unexposed	Never cigarettes	4	0.19 (0.13-0.25)	NS	9.05
	Never any product	23	0.17 (0.15-0.20)		7.66
Grouped midpoint age (yr)	< 50	7	0.18 (0.13-0.23)	NS	7.73
	50-59	9	0.21 (0.17-0.25)		10.98
	60+	11	0.17 (0.14-0.19)		6.88
Cases in smokers	< 100	12	0.14 (0.11-0.17)	< 0.001	5.09
	100 to < 200	7	0.20 (0.16-0.24)		9.88
	200 to < 500	4	0.17 (0.23-0.20)		7.02
	500 to < 1000	3	0.23 (0.20-0.26)		14.57
	1000+	1	0.15 (0.11-0.19)		5.71

<sup>1</sup>Probability values for factor considered, presented as < 0.001, < 0.01, < 0.05, < 0.1 or not significant (NS) ( $P \geq 0.1$ ); <sup>2</sup>Of principal publication for the study;

<sup>3</sup>Any product = Smokes cigarettes and/or pipes and/or cigars; Cigarettes +/- = Smokes cigarettes with or without other products (pipes, cigars).

membering smoking history also form part of the problem. Also the dose of smoke constituents received may not be directly proportional to the amount smoked<sup>[15]</sup>.

### **Misclassification of duration and age of starting**

Subjects may not remember the exact age of starting, and indeed there may be differences in definition between studies - age of first trying a cigarette, or age of starting to smoke regularly? Also duration may not represent a continuous period. Risk may be affected by intermediate quit periods, which may be asked about differently in different studies.

### **Estimating midpoints of ranges**

The statistical methods used require estimates of midpoints of ranges used. We have not attempted sensitivity analyses based on alternative procedures for defining midpoints.

### **Use of pseudo-numbers**

Our methodology requires knowledge, for each block, of the numbers of cases and controls (or at risk) in each smoking group. As such data are not always provided, and indeed for covariate-adjusted data are only hypothetical, we used the method of Hamling *et al.*<sup>[4]</sup> to estimate pseudo-numbers corresponding exactly to the reported RRs and CIs. These pseudo-numbers have been shown<sup>[16]</sup> to allow accurate estimation of RRs and CIs relative to a different base group from that used originally, and should be adequate for model fitting. This issue seems less important than others considered so far.

### **Adjustment for other smoking variables**

Our analyses compare risk relative to never smokers, all the RRs in any block being adjusted for the same variables. As RRs relative to never smokers cannot be adjusted for other smoking variables, we necessarily restricted attention to estimates adjusted for age and non-smoking characteristics. This is possibly unfortunate as, for example, later starters may smoke less than earlier starters. In theory one could study the extent of such bias based on studies presenting risk (compared to never smokers) jointly by more than one dose measure. However, few studies present such data and we did not investigate this.

### **Use of simple models based on published results**

We restricted attention to models of a relatively simple functional form, partly as it is much easier to explain results and conduct tests of heterogeneity where differences between blocks can be expressed in terms of one parameter ( $\beta_1$ ). Also, the numerous data uncertainties may not justify a more complex approach. Such an approach is better pursued using individual person data from large studies. This would allow fitting of models simultaneously accounting for amount smoked and duration, and allow a more precise risk estimation. In the context of a systematic review and meta-analysis, involving many studies conducted years ago with the data unlikely

to be accessible, we made no attempt to obtain individual data sets.

### **Model fit**

Goodness-of-fit has been studied in various ways. First, we used the “pool-first” approach<sup>[9,10]</sup> to compare the deviance of models with a common  $\beta_1$  per block but a different functional form of the dose-relationship. The exponential ( $\log RR = \beta_1 \exp d$ ) and the log model ( $\log RR = \beta_1 \log d$ ) clearly fitted substantially worse than other models, and were not pursued further. Also, the power model ( $\log RR = \beta_1 d^y$ ) and the log-with-baseline model [ $\log RR = \beta_1 \log (1 + Wd)$ ] generally fitted better than the linear model ( $\log RR = \beta_1 d$ ), though for age of starting the best-fitting log-with-baseline model had such a low estimate of  $W$  that it became equivalent to the linear model. While the deviance of the linear model was reduced by adding quadratic and cubic terms this advantage was small. We concentrated most on the power and/or log-with-baseline models, given their greater simplicity, and the fact that the cubic model fitted worse than these alternatives for amount smoked and not materially better for duration or age of start.

We then restricted attention to the linear, power and, except for age of start, the log-with-baseline model, fitting separate  $\beta_1$  values to each block. We investigated goodness-of-fit by studying plots of observed and predicted RRs (not shown), and by comparing observed and predicted numbers, both within block (Fit Amount Smoked<sup>[11]</sup>, Fit Duration, and Fit Age Start<sup>[11]</sup>) and summed over block (Tables 4, 9 and 14). This allowed two general conclusions. First, the best models (log-with-baseline for amount smoked, power for duration and age of start) fitted the shape of the dose relationship well. Given the large number of cases analyzed (54245 for amount smoked, 10711 for duration and 7575 for age of start) it is unsurprising that formal misfit existed for amount smoked and duration, but this seems relatively unimportant. Second, there were significant misfits for some blocks. The results section comments on the worst cases. Sometimes these are due to unusual response patterns, difficult to fit by any plausible model, sometimes to differing response patterns in different blocks. Thus, for amount smoked, there are some blocks where the slope flattens off at high consumption, but others where the reverse is true. The explanation for this is unclear, but attempting to account for it by more complex models seems unattractive, as compared to the models selected, which involve a common shape and variation only in slope ( $\beta_1$ ).

### **Sources of heterogeneity**

We carried out weighted regression analyses to investigate sources of heterogeneity. While some factors (*e.g.*, age and sex) could be better evaluated using pooled analyses based on individual person data, and problems arise from correlations between variables studied, these analyses should detect major sources.

For amount smoked, these analyses only identified

continent as a significant factor, other associations seen in the simple analyses becoming non-significant once continent was accounted for. The smaller  $\beta_1$  for Asian studies is consistent with our earlier analyses<sup>[16]</sup>, and may relate to higher denial rates of smoking in Asia.

For duration and age of starting, the regression analyses showed a tendency for  $\beta_1$  to be greater in studies involving more lung cancer cases and studies of younger people. Higher values in males than females and lower values in Asian studies were not independently significant. There was also some evidence for duration of higher  $\beta_1$  values for smoking cigarettes, than smoking any product.

### Comparison with some previous work

Attempts have been made before to model the relationship of lung cancer to amount smoked and duration. For example, Doll *et al.*<sup>[17]</sup>, in a much cited paper, based on data for British doctors who started smoking at ages 16-25 and smoked 40 or less per day, modelled the annual lung cancer incidence at age 40-79 by the expression

$$0.273 \times 10^{12} \times (\text{cigarettes/d} + 6)^2 \times (\text{age} - 22.5)^{4.5}.$$

They noted “significant ( $P < 0.01$ ) upward curvature of the dose-response relationship in the range 0-40 cigarettes/d, which is what might be expected if more than one of the ‘stages’ (in the multistage genesis of bronchial carcinoma) was strongly affected by smoking.” They also noted a drop off in response above 40 cigarettes/d, though based on few cases, and discussed various explanations for it. Our analyses show little evidence of upward curvature with amount smoked. However, this does not rule out smoking affecting more than one stage of a multistage process; indeed there is strong evidence this is true<sup>[18]</sup>.

Taking (age - 22.5 years) as an approximate indicator of duration, the model of Doll *et al.*<sup>[17]</sup> suggests risk rises steeply with increasing duration, according to a fourth or fifth power relationship. At first sight, this appears to conflict with our findings, where the power relationship we fitted was only somewhat above linear (Figure 1C). However, whereas Doll and Peto’s analysis compares risk by age for people of a similar age of start, our modelling compares risk by age of start for people of a given age. Here, the relationship of risk to duration will be much less steep. This can be illustrated by applying formulae for a form of the multistage model where risk affects the first and penultimate stages, the effect on the penultimate stage being twice as strong as for the first stage, a form known to fit smoking and lung cancer relationships quite well<sup>[18,19]</sup>. The RR for a 70-year-old starting at age 15 is estimated as 1.66 times higher than for a 70-year-old starting at age 30. This ratio somewhat exceeds the ratio of durations ( $55/40 = 1.38$ ), but much less than predicted by a fourth or fifth power relationship ( $1.38^{4.5} = 4.26$ ).

### Summing up

Based on 71 studies described in 87 publications<sup>[20-106]</sup> we demonstrated that for all three measures of dose studied (amount smoked, duration and age of start), the shape of their relationship with lung cancer can be described

quite accurately using simple models. Though, for all dose measures, there is evidence of misfit for some data blocks, these seem mainly due to unusual response patterns difficult to fit with plausible models, or to different blocks showing differing shapes of the dose-relationship. The main limitations of the models relate to the data they were fitted to. Misclassification of smoking status and of dose may produce bias, as may failure to update smoking habits during follow-up in prospective studies, and failure to adjust for other indices of dose. Nevertheless, the models presented characterize the observed relationships of lung cancer to amount smoked, duration, and age of start more fully than previously attempted.

## ACKNOWLEDGMENTS

We thank Philip Morris Products S.A. for supporting this research. The opinions and conclusions of the authors are their own, and do not necessarily reflect the position of Philip Morris Products S.A. We also thank Pauline Wasell, Diana Morris and Yvonne Cooper for assistance in typing various drafts of the paper and obtaining relevant literature.

## COMMENTS

### Background

No previous meta-analysis has used parametric models in order to quantify more precisely the relationship between smoking and lung cancer. Using a database of all epidemiological studies of 100 or more lung cancer cases published before 2000, models are fitted relating lung cancer risk to amount smoked, duration of smoking and age of starting to smoke.

### Research frontiers

Based on all the studies providing relevant data, the models fitted show that the risk, relative to never smokers, rises from 3.86 to 22.31 between 10 and 50 cigarettes/d, from 2.21 to 13.54 between 10 and 50 years smoked, and from 3.66 to 8.94 between age of starting 30 and 12.5 years. There is little heterogeneity between studies for duration of smoking or age started, but there is clear heterogeneity for amount smoked, with RRs for 50 cigarettes/d being 7.23 for studies in Asia, as compared to 26.36 for North American and 22.16 for European studies.

### Innovations and breakthroughs

The new feature of this paper is the comprehensive assessment of the shape of the dose-responses studied, with a number of alternative functional forms studied, and the best-fitting one selected. The fitted models, which describe the relationships well, are each quite simple in form, allowing ready meta-analysis of individual study estimates.

### Applications

The fitted models allow more precise quantification of the hazards of smoking than previously reported, and will assist smoking and health researchers.

### Terminology

Linear model: The logarithm of the RR is linearly related to dose. In the power model it is related to dose raised to a power. In the log-with-baseline model, it is related to the logarithm of dose with an offset for background risk. Pseudo-numbers are estimates of numbers of cases and controls, by dose level, derived from published RRs, which allow fitting of the models.

### Peer review

The authors meta-analyzed smoking/lung cancer relationships using parametric modelling according to the IESLC database. They found that the models describe the dose-relationship well and concluded that they can be used to more precisely estimate the lung cancer risk from smoking. The limitation has been fully discussed in the discussion part. This study provides some interesting results for further research into smoking and lung cancer.



## REFERENCES

- 1 Lee PN, Forey BA, Coombs KJ. Systematic review with meta-analysis of the epidemiological evidence in the 1900s relating smoking to lung cancer. *BMC Cancer* 2012; **12**: 385 [PMID: 22943444 DOI: 10.1186/1471-2407-12-385]
- 2 Fry JS, Lee PN, Forey BA, Coombs KJ. How rapidly does the excess risk of lung cancer decline following quitting smoking? A quantitative review using the negative exponential model. *Regul Toxicol Pharmacol* 2013; **67**: 13-26 [PMID: 23764305 DOI: 10.1016/j.yrtph.2013.06.001]
- 3 Fry JS, Lee PN. Revisiting the association between environmental tobacco smoke exposure and lung cancer risk. I. The dose-response relationship with amount and duration of smoking by the husband. *Indoor Built Environ* 2000; **9**: 303-316 [DOI: 10.1177/1420326X0000900602]
- 4 Hamling J, Lee P, Weitkunat R, Ambühl M. Facilitating meta-analyses by deriving relative effect and precision estimates for alternative comparisons from a set of estimates presented by exposure level or disease category. *Stat Med* 2008; **27**: 954-970 [PMID: 17676579 DOI: 10.1002/sim.3013]
- 5 Thun MJ, Day-Lally C, Myers DG, Calle EE, Flanders WD, Zhu BP, Namboodiri MM, Heath CW Jr. Trends in tobacco smoking and mortality from cigarette use in cancer prevention studies I (1959 through 1965) and II (1982 through 1988). In: Shopland DR, Burns DM, Garfinkel L, Samet JM, editors. Changes in cigarette-related disease risks and their implications for prevention and control. Rockville, Maryland: US Department of Health and Human Services, National Institutes of Health, National Cancer Institute, 1997: 305-382. Available from: URL: [http://cancercontrol.cancer.gov/tcrb/monographs/8/m8\\_4.pdf](http://cancercontrol.cancer.gov/tcrb/monographs/8/m8_4.pdf)
- 6 Forey B, Hamling J, Lee P, Wald N. International Smoking Statistics: A collection of historical data from 30 economically developed countries. 2nd ed. London and Oxford: Wolfson Institute of Preventive Medicine and Oxford University Press, 2002 [DOI: 10.1093/acprof:oso/9780198508564.001.0001]
- 7 Forey B, Hamling J, Lee P. International Smoking Statistics. A collection of worldwide historical data. 2nd ed. Sutton: P N Lee Statistics and Computing Ltd., 2006-2012. Available from: URL: <http://www.pnlee.co.uk/iss.htm>
- 8 US Department of Health and Human Services. National health and nutrition examination survey (NHANES). National Center for Health Statistics. Available from: URL: <http://www.cdc.gov/nchs/nhanes.htm>
- 9 Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *Am J Epidemiol* 1992; **135**: 1301-1309 [PMID: 1626547]
- 10 Berlin JA, Longnecker MP, Greenland S. Meta-analysis of epidemiologic dose-response data. *Epidemiology* 1993; **4**: 218-228 [PMID: 8512986 DOI: 10.1097/00001648-199305000-00005]
- 11 [http://www.pnlee.co.uk/downloads/ieslc3/lc\\_dose\\_response\\_paper\\_additionalfiles.htm](http://www.pnlee.co.uk/downloads/ieslc3/lc_dose_response_paper_additionalfiles.htm)
- 12 Lee PN. Misclassification of smoking habits and passive smoking. A review of the evidence. Heidelberg: Springer-Verlag, 1988 [DOI: 10.1007/978-3-642-73822-7]
- 13 Lee PN, Forey BA. Misclassification of smoking habits as determined by cotinine or by repeated self-report - a summary of evidence from 42 studies. *J Smoking-Related Dis* 1995; **6**: 109-129
- 14 Lee PN, Forey B, Fry JS. Revisiting the association between environmental tobacco smoke exposure and lung cancer risk. III. Adjustment for the biasing effect of misclassification of smoking habits. *Indoor Built Environ* 2001; **10**: 384-398 [DOI: 10.1177/1420326X0101000605]
- 15 Baker RR, Dixon M, Mariner DC, Shepperd CJ, Scherer G, Ogden MW, Robinson JH, Sinclair NM, Sherwood N, Akimura Y, Sakamoto K, Röper W, Tricker AR, Marchand V, Varignon B, Lionetti G. Terms used for exposure to smoke. *Beitr Tabakforsch Int* 2004; **21**: 250
- 16 Orsini N, Li R, Wolk A, Khudyakov P, Spiegelman D. Meta-analysis for linear and nonlinear dose-response relations: examples, an evaluation of approximations, and software. *Am J Epidemiol* 2012; **175**: 66-73 [PMID: 22135359]
- 17 Doll R, Peto R. Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. *J Epidemiol Community Health* 1978; **32**: 303-313 [PMID: 744822 DOI: 10.1136/jech.32.4.303]
- 18 Lee PN. Studying the relationship of smoking to lung cancer using the multistage model of carcinogenesis. A review. Sutton, Surrey: P N Lee Statistics and Computing Ltd., 1995. Available from: URL: <http://www.pnlee.co.uk/Reports.htm>
- 19 Day NE, Brown CC. Multistage models and primary prevention of cancer. *J Natl Cancer Inst* 1980; **64**: 977-989 [PMID: 6929006]
- 20 Akiba S. Analysis of cancer risk related to longitudinal information on smoking habits. *Environ Health Perspect* 1994; **102** Suppl 8: 15-19 [PMID: 7851325 DOI: 10.1289/ehp.94102s815]
- 21 Amandus H, Costello J. Silicosis and lung cancer in U.S. metal miners. *Arch Environ Health* 1991; **46**: 82-89 [PMID: 2006898 DOI: 10.1080/00039896.1991.9937433]
- 22 Archer VE, Gillam JD, Wagoner JK. Respiratory disease mortality among uranium miners. *Ann NY Acad Sci* 1976; **271**: 280-293 [DOI: 10.1111/j.1749-6632.1976.tb23123.x]
- 23 Rylander R, Axelsson G, Andersson L, Liljequist T, Bergman B. Lung cancer, smoking and diet among Swedish men. *Lung Cancer* 1996; **14** Suppl 1: S75-S83 [PMID: 8785669 DOI: 10.1016/S0169-5002(96)90212-3]
- 24 Modigh C, Axelsson G, Alavanja M, Andersson L, Rylander R. Pet birds and risk of lung cancer in Sweden: a case-control study. *BMJ* 1996; **313**: 1236-1238 [PMID: 8939112 DOI: 10.1136/bmj.313.7067.1236]
- 25 Ben-Shlomo Y, Smith GD, Shipley MJ, Marmot MG. What determines mortality risk in male former cigarette smokers? *Am J Public Health* 1994; **84**: 1235-1242 [PMID: 8059878 DOI: 10.2105/AJPH.84.8.1235]
- 26 Marmot MG, Shipley MJ, Rose G. Inequalities in death-specific explanations of a general pattern? *Lancet* 1984; **1**: 1003-1006 [PMID: 6143919]
- 27 Department of National Health and Welfare Canada. A Canadian study of smoking and health. Canada: Department of National Health and Welfare, 1966
- 28 Boucot KR, Weiss W, Seidman H, Carnahan WJ, Cooper DA. The Philadelphia pulmonary neoplasm research project: basic risk factors of lung cancer in older men. *Am J Epidemiol* 1972; **95**: 4-16 [PMID: 5007365]
- 29 Brett GZ, Benjamin B. Smoking habits of men employed in industry, and mortality. *Br Med J* 1968; **3**: 82-85 [PMID: 5662965 DOI: 10.1136/bmj.3.5610.82]
- 30 Bross ID, Gibson R. Risks of lung cancer in smokers who switch to filter cigarettes. *Am J Public Health Nations Health* 1968; **58**: 1396-1403 [PMID: 5691372 DOI: 10.2105/AJPH.58.8.1396]
- 31 Buffler PA, Pickle LW, Mason TJ, Contant C. The causes of lung cancer in Texas. In: Mizell M, Correa P, editors. Lung cancer: causes and prevention, Proceedings of the International Lung Cancer Update Conference, New Orleans, Louisiana, March 3-5, 1983. Deerfield Beach, FL: Verlag Chemie International Inc., 1984: 83-99
- 32 Ives JC. Environmental exposures and lung cancer risk among women in Harris County, Texas, 1977-1980 [Thesis]. Houston, TX: University of Texas, Health Science Centre, 1984
- 33 Cederlöf R, Friberg L, Hrubec Z, Loric U. The relationship of smoking and some social covariables to mortality and cancer morbidity. A ten year follow-up in a probability sample of 55,000 Swedish subjects age 18-69, Part 1/2. Stockholm:



- Karolinska Institute, Dept of Environmental Hygiene, 1975
- 34 **Nordlund LA**, Carstensen JM, Pershagen G. Cancer incidence in female smokers: a 26-year follow-up. *Int J Cancer* 1997; **73**: 625-628 [PMID: 9398036]
  - 35 **Chang AK**, Barrett-Connor E, Edelstein S. Low plasma cholesterol predicts an increased risk of lung cancer in elderly women. *Prev Med* 1995; **24**: 557-562 [PMID: 8610078 DOI: 10.1006/pmed.1995.1089]
  - 36 **Chow WH**, Schuman LM, McLaughlin JK, Bjelke E, Gridley G, Wacholder S, Chien HT, Blot WJ. A cohort study of tobacco use, diet, occupation, and lung cancer mortality. *Cancer Causes Control* 1992; **3**: 247-254 [PMID: 1610971 DOI: 10.1007/BF00124258]
  - 37 **Comstock GW**, Alberg AJ, Huang HY, Wu K, Burke AE, Hoffman SC, Norkus EP, Gross M, Cutler RG, Morris JS, Spate VL, Helzlsouer KJ. The risk of developing lung cancer associated with antioxidants in the blood: ascorbic acid, carotenoids, alpha-tocopherol, selenium, and total peroxyl radical absorbing capacity. *Cancer Epidemiol Biomarkers Prev* 1997; **6**: 907-916 [PMID: 9367064]
  - 38 **Correa P**, Pickle LW, Fontham E, Dalager N, Lin Y, Haenszel W, Johnson WD. The causes of lung cancer in Louisiana. In: Mizell M, Correa P, editors. Lung cancer: causes and prevention. New York: Verlag Chemie International Inc., 1984: 73-82
  - 39 **Fontham ET**, Pickle LW, Haenszel W, Correa P, Lin YP, Falk RT. Dietary vitamins A and C and lung cancer risk in Louisiana. *Cancer* 1988; **62**: 2267-2273 [PMID: 3179940]
  - 40 **Burns DM**, Shanks TG, Choi W, Thun MJ, Heath CW Jr, Garfinkel L. The American Cancer Society cancer prevention study I: 12-year follow-up of 1 million men and women. In: Changes in cigarette-related disease risks and their implications for prevention and control. Rockville, MD: US Department of Health and Human Services, National Institutes of Health, National Cancer Institute, 1997: 113-304. Available from: URL: [http://cancercontrol.cancer.gov/tcrb/monographs/8/m8\\_3.pdf](http://cancercontrol.cancer.gov/tcrb/monographs/8/m8_3.pdf)
  - 41 **Hammond EC**. Smoking habits and air pollution in relation to lung cancer. In: Lee HK, editor. Environmental factors in respiratory diseases. New York: Academic Press Inc., 1972: 177-198
  - 42 **Thun MJ**, Myers DG, Day-Lally C, Namboodiri MM, Calle EE, Flanders WD, Adams SL, Heath CW Jr. Age and the exposure-response relationships between cigarette smoking and premature death in Cancer Prevention Study II. In: Changes in cigarette-related disease risks and their implications for prevention and control. Rockville, MD: US Department of Health and Human Services, National Institutes of Health, National Cancer Institute, 1997: 383-475. Available from: URL: [http://cancercontrol.cancer.gov/tcrb/monographs/8/m8\\_5.pdf](http://cancercontrol.cancer.gov/tcrb/monographs/8/m8_5.pdf)
  - 43 **Darby S**, Whitley E, Silcocks P, Thakrar B, Green M, Lomas P, Miles J, Reeves G, Fearn T, Doll R. Risk of lung cancer associated with residential radon exposure in south-west England: a case-control study. *Br J Cancer* 1998; **78**: 394-408 [PMID: 9703290 DOI: 10.1038/bjc.1998.506]
  - 44 **Wicken AJ**. Environmental and personal factors in lung cancer and bronchitis mortality in Northern Ireland, 1960-1962. London: Tobacco Research Council, 1966
  - 45 **Dean G**, Lee PN, Todd GF, Wicken AJ. Report on a second retrospective mortality study in North-East England - Part I. Factors related to mortality from lung cancer, bronchitis, heart disease and stroke in Cleveland County, with particular emphasis on the relative risks associated with smoking filter and plain cigarettes. London: Tobacco Research Council, 1977
  - 46 **de Klerk NH**, Musk AW. Silica, compensated silicosis, and lung cancer in Western Australian goldminers. *Occup Environ Med* 1998; **55**: 243-248 [PMID: 9624278 DOI: 10.1136/oem.55.4.243]
  - 47 **Doll R**, Peto R, Wheatley K, Gray R, Sutherland I. Mortality in relation to smoking: 40 years' observations on male British doctors. *BMJ* 1994; **309**: 901-911 [PMID: 7755693 DOI: 10.1136/bmj.309.6959.901]
  - 48 **Doll R**, Peto R. Mortality in relation to smoking: 20 years' observations on male British doctors. *Br Med J* 1976; **2**: 1525-1536 [PMID: 1009386 DOI: 10.1136/bmj.2.6051.1525]
  - 49 **Doll R**, Gray R, Hafner B, Peto R. Mortality in relation to smoking: 22 years' observations on female British doctors. *Br Med J* 1980; **280**: 967-971 [PMID: 7417764 DOI: 10.1136/bmj.280.6219.967]
  - 50 **Dorant E**, van den Brandt PA, Goldbohm RA. A prospective cohort study on Allium vegetable consumption, garlic supplement use, and the risk of lung carcinoma in The Netherlands. *Cancer Res* 1994; **54**: 6148-6153 [PMID: 7954460]
  - 51 **Dorgan JF**, Ziegler RG, Schoenberg JB, Hartge P, McAdams MJ, Falk RT, Wilcox HB, Shaw GL. Race and sex differences in associations of vegetables, fruits, and carotenoids with lung cancer risk in New Jersey (United States). *Cancer Causes Control* 1993; **4**: 273-281 [PMID: 8318643]
  - 52 **McLaughlin JK**, Hrubec Z, Blot WJ, Fraumeni JF. Smoking and cancer mortality among U.S. veterans: a 26-year follow-up. *Int J Cancer* 1995; **60**: 190-193 [PMID: 7829214 DOI: 10.1002/ijc.2910600210]
  - 53 **Enstrom JE**. Smoking cessation and mortality trends among two United States populations. *J Clin Epidemiol* 1999; **52**: 813-825 [PMID: 10529023]
  - 54 **Kahn HA**. The Dorn study of smoking and mortality among U.S. veterans: report on eight and one-half years of observation. In: Haenszel W, editor. Epidemiological approaches to the study of cancer and other chronic diseases. Bethesda, MD: U.S. Department of Health, Education, and Welfare. Public Health Service National Cancer Institute, 1966: 1-125
  - 55 **Engeland A**, Haldorsen T, Andersen A, Tretli S. The impact of smoking habits on lung cancer risk: 28 years' observation of 26,000 Norwegian men and women. *Cancer Causes Control* 1996; **7**: 366-376 [PMID: 8734831 DOI: 10.1007/BF00052943]
  - 56 **Enstrom JE**, Heath CW. Smoking cessation and mortality trends among 118,000 Californians, 1960-1997. *Epidemiology* 1999; **10**: 500-512 [PMID: 10468422 DOI: 10.1097/00001648-199909000-00007]
  - 57 **Gao CM**, Tajima K, Kuroishi T, Hirose K, Inoue M. Protective effects of raw vegetables and fruit against lung cancer among smokers and ex-smokers: a case-control study in the Tokai area of Japan. *Jpn J Cancer Res* 1993; **84**: 594-600 [PMID: 8340248 DOI: 10.1111/j.1349-7006.1993.tb02018.x]
  - 58 **Gillis CR**, Hole DJ, Boyle P. Cigarette smoking and male lung cancer in an area of very high incidence. I. Report of a case-control study in the West of Scotland. *J Epidemiol Community Health* 1988; **42**: 38-43 [PMID: 3418284 DOI: 10.1136/jech.42.1.38]
  - 59 **Haenszel W**, Shimkin MB, Mantel N. A retrospective study of lung cancer in women. *J Natl Cancer Inst* 1958; **21**: 825-842 [PMID: 13599015]
  - 60 **Hammond EC**, Selikoff IJ, Seidman H. Asbestos exposure, cigarette smoking and death rates. *Ann N Y Acad Sci* 1979; **330**: 473-490 [PMID: 294198 DOI: 10.1111/j.1749-6632.1979.tb18749.x]
  - 61 **Hammond EC**, Horn D. Smoking and death rates; report on forty-four months of follow-up of 187,783 men. II. Death rates by cause. *J Am Med Assoc* 1958; **166**: 1294-1308 [PMID: 13513355]
  - 62 **Hirayama T**. Life-style and mortality: A large scale census based cohort study in Japan. In: Wahrendorf J, editor. Contributions to epidemiology and biostatistics. Basle: Karger, 1990: 6
  - 63 **Hitosugi M**. Epidemiological study of lung cancer with special reference to the effect of air pollution and smoking habit. *Bull Inst Public Health* 1968; **17**: 237-256
  - 64 **Gillis CR**, Hole DJ, Hawthorne VM. Cigarette smoking and

- male lung cancer in an area of very high incidence. II. Report of a general population cohort study in the West of Scotland. *J Epidemiol Community Health* 1988; **42**: 44-48 [PMID: 3418285 DOI: 10.1136/jech.42.1.44]
- 65 **Humble CG**, Samet JM, Pathak DR, Skipper BJ. Cigarette smoking and lung cancer in 'Hispanic' whites and other whites in New Mexico. *Am J Public Health* 1985; **75**: 145-148 [PMID: 3966619 DOI: 10.2105/AJPH.75.2.145]
  - 66 **Pathak DR**, Samet JM, Humble CG, Skipper BJ. Determinants of lung cancer risk in cigarette smokers in New Mexico. *J Natl Cancer Inst* 1986; **76**: 597-604 [PMID: 3457198]
  - 67 **Friedman GD**, Tekawa I, Sadler M, Sidney S. Smoking and mortality: the Kaiser Permanente experience. In: *Changes in cigarette-related disease risks and their implications for prevention and control*. Rockville, MD: US Department of Health and Human Services, National Institutes of Health, National Cancer Institute, 1997: 477-499. Available from: URL: [http://cancercontrol.cancer.gov/tcrb/monographs/8/m8\\_6.pdf](http://cancercontrol.cancer.gov/tcrb/monographs/8/m8_6.pdf)
  - 68 **Selby JV**, Friedman GD. Epidemiologic evidence of an association between body iron stores and risk of cancer. *Int J Cancer* 1988; **41**: 677-682 [PMID: 3366489 DOI: 10.1002/ijc.2910410507]
  - 69 **Kanellakis A**, Trichopoulos D, Michalakopoulos N, Margoudakis S, Kanellaki K, Xirouchaki E, Kalapothaki V. The relationship between smoking of Greek cigarettes and the development of lung cancer. *Mater Med Greca* 1976; **4**: 351-355
  - 70 **Trichopoulos D**, Kalandidi A, Tzonou A. Incidence and distribution of lung cancer in Greece. *Excerpta Med Int Congr Ser* 1982; **558**: 10-17
  - 71 **Katsouyanni K**, Trichopoulos D, Kalandidi A, Tomos P, Riboli E. A case-control study of air pollution and tobacco smoking in lung cancer among women in Athens. *Prev Med* 1991; **20**: 271-278 [PMID: 2057473 DOI: 10.1016/0091-7435(91)90026-Z]
  - 72 **Kaufman DW**, Palmer JR, Rosenberg L, Stolley P, Warshauer E, Shapiro S. Tar content of cigarettes in relation to lung cancer. *Am J Epidemiol* 1989; **129**: 703-711 [PMID: 2923118]
  - 73 **Kinlen LJ**, Willows AN, Goldblatt P, Yudkin J. Tea consumption and cancer. *Br J Cancer* 1988; **58**: 397-401 [PMID: 3179194 DOI: 10.1038/bjc.1988.227]
  - 74 **Knekt P**. Vitamin E and smoking and the risk of lung cancer. *Ann N Y Acad Sci* 1993; **686**: 280-27; discussion 280-27; [PMID: 8512253 DOI: 10.1111/j.1749-6632.1993.tb39187.x]
  - 75 **Knekt P**, Seppänen R, Järvinen R, Virtamo J, Hyvönen L, Pukkala E, Teppo L. Dietary cholesterol, fatty acids, and the risk of lung cancer among men. *Nutr Cancer* 1991; **16**: 267-275 [PMID: 1775388 DOI: 10.1080/01635589109514165]
  - 76 **Koo LC**, Ho JHC, Saw D. Is passive smoking an added risk factor for lung cancer in Chinese women? *J Exp Clin Cancer Res* 1984; **3**: 277-283
  - 77 **Koo LC**, Ho JHC, Saw D. Active and passive smoking among female lung cancer patients and controls in Hong Kong. *J Exp Clin Cancer Res* 1983; **2**: 367-375
  - 78 **Liaw KM**, Chen CJ. Mortality attributable to cigarette smoking in Taiwan: a 12-year follow-up study. *Tob Control* 1998; **7**: 141-148 [PMID: 9789932 DOI: 10.1136/tc.7.2.141]
  - 79 **McDonald JC**, Liddell FD, Dufresne A, McDonald AD. The 1891-1920 birth cohort of Quebec chrysotile miners and millers: mortality 1976-88. *Br J Ind Med* 1993; **50**: 1073-1081 [PMID: 8280638 DOI: 10.1136/oem.50.12.1073]
  - 80 **MacLennan R**, Da Costa J, Day NE, Law CH, Ng YK, Shanmugaratnam K. Risk factors for lung cancer in Singapore Chinese, a population with high female incidence rates. *Int J Cancer* 1977; **20**: 854-860 [PMID: 591126 DOI: 10.1002/ijc.2910200606]
  - 81 **Matos E**, Vilensky M, Boffetta P, Kogevinas M. Lung cancer and smoking: a case-control study in Buenos Aires, Argentina. *Lung Cancer* 1998; **21**: 155-163 [PMID: 9857993 DOI: 10.1016/S0169-5002(98)00055-5]
  - 82 **Lee PN**. Mortality from smoking-associated diseases in Great Britain. A statistical analysis of British data from the U.S.A.-U.K.-Norway migrant study. Sutton, Surrey: P N Lee Statistics and Computing Ltd., 1979. Available from: URL: <http://www.pnlee.co.uk/Reports.htm>
  - 83 **Kuller LH**, Ockene JK, Meilahn E, Wentworth DN, Svendsen KH, Neaton JD. Cigarette smoking and mortality. MRFIT Research Group. *Prev Med* 1991; **20**: 638-654 [PMID: 1758843 DOI: 10.1016/0091-7435(91)90060-H]
  - 84 **Kuller LH**, Ockene J, Meilahn E, Svendsen KH. Relation of forced expiratory volume in one second (FEV1) to lung cancer mortality in the Multiple Risk Factor Intervention Trial (MRFIT). *Am J Epidemiol* 1990; **132**: 265-274 [PMID: 2372006]
  - 85 **Nam CB**, Hummer RA, Rogers RG. Underlying and multiple causes of death related to smoking. *Popul Res Policy Rev* 1994; **13**: 305-325 [DOI: 10.1007/BF01074340]
  - 86 **Parkin DM**, Vizcaino AP, Skinner ME, Ndhlovu A. Cancer patterns and risk factors in the African population of south-western Zimbabwe, 1963-1977. *Cancer Epidemiol Biomarkers Prev* 1994; **3**: 537-547 [PMID: 7827583]
  - 87 **Pershagen G**, Akerblom G, Axelsson O, Clavensjö B, Damber L, Desai G, Enflo A, Lagarde F, Mellander H, Svartengren M. Residential radon exposure and lung cancer in Sweden. *N Engl J Med* 1994; **330**: 159-164 [PMID: 8264737 DOI: 10.1056/NEJM199401203300302]
  - 88 **Peto R**, Speizer FE, Cochrane AL, Moore F, Fletcher CM, Tinker CM, Higgins IT, Gray RG, Richards SM, Gilliland J, Norman-Smith B. The relevance in adults of air-flow obstruction, but not of mucus hypersecretion, to mortality from chronic lung disease. Results from 20 years of prospective observation. *Am Rev Respir Dis* 1983; **128**: 491-500 [PMID: 6614643]
  - 89 **Pezzotto SM**, Poletto L. Occupation and histopathology of lung cancer: A case-control study in Rosario, Argentina. *Am J Ind Med* 1999; **36**: 437-443 [PMID: 10470008]
  - 90 **Pezzotto SM**, Mahuad R, Bay ML, Morini JC, Poletto L. Variation in smoking-related lung cancer risk factors by cell type among men in Argentina: a case-control study. *Cancer Causes Control* 1993; **4**: 231-237 [PMID: 8391337]
  - 91 **Prescott E**, Osler M, Hein HO, Borch-Johnsen K, Lange P, Schnohr P, Vestbo J. Gender and smoking-related risk of lung cancer. The Copenhagen Center for Prospective Population Studies. *Epidemiology* 1998; **9**: 79-83 [PMID: 9430273 DOI: 10.1097/00001648-199801000-00016]
  - 92 **Segi M**, Kurihara M, Ishikawa S, Haenszel W. Epidemiological survey on lung cancer and smoking. *Lung Cancer* 1979; **19**: 157-165 [DOI: 10.2482/haigan.19.157]
  - 93 **Shaw GL**, Falk RT, Deslauriers J, Frame JN, Nesbitt JC, Pass HI, Issaq HJ, Hoover RN, Tucker MA. Debrisoquine metabolism and lung cancer risk. *Cancer Epidemiol Biomarkers Prev* 1995; **4**: 41-48 [PMID: 7894323]
  - 94 **Sobue T**, Suzuki T, Fujimoto I, Matsuda M, Doi O, Mori T, Furuse K, Fukuoka M, Yasumitsu T, Kuwahara O. Case-control study for lung cancer and cigarette smoking in Osaka, Japan: comparison with the results from Western Europe. *Jpn J Cancer Res* 1994; **85**: 464-473 [PMID: 8014103 DOI: 10.1111/j.1349-7006.1994.tb02381.x]
  - 95 **Speizer FE**, Colditz GA, Hunter DJ, Rosner B, Hennekens C. Prospective study of smoking, antioxidant intake, and lung cancer in middle-aged women (USA). *Cancer Causes Control* 1999; **10**: 475-482 [PMID: 10530619 DOI: 10.1023/A:1008931526525]
  - 96 **Stockwell HG**, Lyman GH, Waltz J, Peters JT. Lung cancer in Florida. Risks associated with residence in the central Florida phosphate mining region. *Am J Epidemiol* 1988; **128**: 78-84 [PMID: 2837899]
  - 97 **Svensson C**, Pershagen G, Klominek J. Smoking and passive smoking in relation to lung cancer in women. *Acta Oncol* 1989; **28**: 623-629 [PMID: 2590538 DOI: 10.3109/02841868909092282]

- 98 **Tenkanen L**, Hakulinen T, Teppo L. The joint effect of smoking and respiratory symptoms on risk of lung cancer. *Int J Epidemiol* 1987; **16**: 509-515 [PMID: 3440661 DOI: 10.1093/ije/16.4.509]
- 99 **Tenkanen L**, Hakulinen T, Hakama M, Saxén E. Sauna, dust and migration as risk factors in lung cancer among smoking and non-smoking males in Finland. *Int J Cancer* 1985; **35**: 637-642 [PMID: 3997283]
- 100 **Tsugane S**, Watanabe S, Sugimura H, Arimoto H, Shimosato Y, Suemasu K. Smoking, occupation and family history in lung cancer patients under fifty years of age. *Jpn J Clin Oncol* 1987; **17**: 309-317 [PMID: 2826845]
- 101 **Tulinius H**, Sigfússon N, Sigvaldason H, Bjarnadóttir K, Tryggvadóttir L. Risk factors for malignant diseases: a cohort study on a population of 22,946 Icelanders. *Cancer Epidemiol Biomarkers Prev* 1997; **6**: 863-873 [PMID: 9367058]
- 102 **Tverdal A**, Thelle D, Stensvold I, Leren P, Bjartveit K. Mortality in relation to smoking history: 13 years' follow-up of 68,000 Norwegian men and women 35-49 years. *J Clin Epidemiol* 1993; **46**: 475-487 [PMID: 8501474 DOI: 10.1016/0895-4356(93)90025-V]
- 103 **Wakai K**, Ohno Y, Genka K, Ohmine K, Kawamura T, Tamakoshi A, Aoki R, Kojima M, Lin Y, Aoki K, Fukuma S. Smoking habits, local brand cigarettes and lung cancer risk in Okinawa, Japan. *J Epidemiol* 1997; **7**: 99-105 [PMID: 9255031 DOI: 10.2188/jea.7.99]
- 104 **Wu AH**, Henderson BE, Pike MC, Yu MC. Smoking and other risk factors for lung cancer in women. *J Natl Cancer Inst* 1985; **74**: 747-751 [PMID: 3857370]
- 105 **Harris RE**, Zang EA, Anderson JI, Wynder EL. Race and sex differences in lung cancer risk associated with cigarette smoking. *Int J Epidemiol* 1993; **22**: 592-599 [PMID: 8225730 DOI: 10.1093/ije/22.4.592]
- 106 **Yamaguchi N**, Kido M, Hoshuyama T, Manabe H, Kikuchi Y, Nishio T, Ohshima LH, Watanabe S. A case-control study on occupational lung cancer risks in an industrialized city of Japan. *Jpn J Cancer Res* 1992; **83**: 134-140 [PMID: 1555994 DOI: 10.1111/j.1349-7006.1992.tb00077.x]

**P- Reviewers** Iftikhar IH, Shen XC, Zhang J **S- Editor** Gou SX  
**L- Editor** A **E- Editor** Zheng XM





Published by **Baishideng Publishing Group Co., Limited**  
Flat C, 23/F., Lucky Plaza,  
315-321 Lockhart Road, Wan Chai, Hong Kong, China  
Fax: +852-65557188  
Telephone: +852-31779906  
E-mail: [bpgoffice@wjgnet.com](mailto:bpgoffice@wjgnet.com)  
<http://www.wjgnet.com>

