

# Supplement 2

## **Estimating past smoking habits by an indirect method. An investigation into a method based on recall, with application to Great Britain**

Barbara Forey and Peter Lee

This document is an extended version of Appendix V of:

*International Smoking Statistics. A collection of historical data from 30 economically developed countries*, 2nd edition,

Barbara Forey, Jan Hamling, Peter Lee, Nicholas Wald, editors. Wolfson Institute of Preventive Medicine and Oxford University Press, London and Oxford; 2002.

Available from: P.N.Lee Statistics and Computing Ltd.,  
17 Cedar Road, Sutton, Surrey, SM2 5DA, UK.

Tel: 44 (0)20 8642 8265; Fax: 44 (0)20 8642 2135.

## Introduction

Nationally based surveys of smoking habits are now conducted regularly in many countries, providing sex- and age-specific data directly on both prevalence and amount smoked. However, such surveys existed in relatively few countries before 1970.<sup>1</sup> Several researchers have therefore attempted to estimate the prevalence of smoking for earlier periods indirectly using data on past smoking habits recalled in a recent survey. Based on current age and smoking status (never/former/current), and on age of starting and stopping smoking, each subject's smoking status can be derived for each year throughout life, allowing estimation of smoking prevalence by calendar year. Results are usually presented by sex and birth cohort. A list of references to such attempts is given in Appendix V of reference 1.

Various potential biases may affect these indirect estimates—differential mortality by smoking history (since only survivors can contribute to the survey), misreporting of current smoking habits,<sup>2</sup> inaccurate recall of events many years ago, and partial missing data. Also, since the surveys may not have been designed specifically for this purpose, they may lack appropriate detail, such as intermediate quit periods or switches between different tobacco products.

We attempt here to assess the usefulness of indirect estimates of past smoking by comparing them with equivalent data obtained directly from surveys conducted at the same time. We selected Great Britain because of the existence of a long established series of smoking surveys, conducted for the Tobacco Manufacturers Association (TMA) and its predecessor organisations, providing direct data back to 1948.<sup>3,4</sup> For indirect data we used two sources; the 1984–1985 Health and Lifestyle Study (HALS),<sup>5</sup> chosen as it was large, representative and included sufficient smoking questions, and the 1978–1981 Alderson Hospital In-Patients Study (AHIP),<sup>6</sup> chosen as it used exceptionally detailed questions on manufactured cigarette (MC) smoking history. Both HALS and AHIP provided indirect data on smoking prevalence, while AHIP also provided indirect data on amount smoked.

Apart from comparing direct and indirect estimates back to 1948, we also present indirect estimates by sex and birth cohort back to 1920.

# Subjects and Methods

## Data sources

### *HALS*

9003 subjects, representative of the British population living in private houses and aged 18+, were interviewed at home during 1984–1985. Questions covered many aspects of health and lifestyle, including smoking (Table S2.1). Questions on cigarette smoking related to MC and hand-rolled cigarettes (HRC) combined but not separately. For cigarettes, pipes and cigars subjects were categorised as regular, occasional or former smokers, and the ages of starting (as self-defined) and stopping (various definitions) smoking are given. Past regular smoking by occasional smokers is given for cigarettes only. The methods are described fully elsewhere.<sup>5</sup>

### *AHIP*

This case-control study of four major smoking-associated diseases was conducted in 10 hospital regions in England. Patients aged 35–74 and fit to be interviewed were selected on a quota basis and interviewed in hospital during 1978–1981. Our analyses include only those 4950 control subjects diagnosed as having a disease *probably or definitely not associated with smoking*. Questions related to the year of hospital admission (1975–1981) rather than year of interview (Table S2.2). Information is available on regular smoking, separately for MC, HRC, pipes and cigars. Ages of starting and stopping smoking are given, but only in age groups for HRC, pipe and cigars. Additional detail for MC gives amount smoked at up to 10 time points from age 16 to hospital admission, allowing identification of intermediate *quit periods*. The time points were no more than 5 years apart, except for subjects aged over 50 for whom the period between *age 25 and 20 years ago* was longer. The methods are described fully elsewhere.<sup>6</sup>

### *TMA*

In this annual survey of British smoking habits, about 10 000 people were interviewed annually, using a quota sampling method. Results were published initially by the TMA<sup>3</sup> and then independently.<sup>4</sup> See also reference 1, chapter 27, source 1. Current MC, HR and pipe smokers were as self-defined, and regular cigar smoking was included from 1963. Estimates were adjusted to align with known sales data; for MC and HRC until 1969, and for pipes throughout, only consumption per smoker was adjusted but for MC and HRC since 1969, both consumption and prevalence were adjusted. This adjustment had a relatively small effect, only increasing

overall prevalence of MC smoking by about 1%, and average daily consumption by half a cigarette. The methods are described fully elsewhere.<sup>4</sup>

## Estimation based on recall

### *Smoking prevalence*

A few subjects with missing data on the primary smoking question (HALS: “Do you smoke regularly?”, AHIP: “Have you ever smoked at least one manufactured cigarette a day for as long as a year?”) were excluded. Otherwise missing smoking data items were estimated (as described below). This avoided introducing a bias, as subjects with partial missing data are smokers, and to omit them would tend to reduce estimates of smoking prevalence.

For each product separately, current smokers were generally assumed to have smoked continuously from their age of starting to smoke, up to and including 1985 (HALS) or year of hospital admission (AHIP); former smokers were assumed to have smoked up to and including their year of stopping. Exceptionally, for MC smokers in AHIP, smokers were assumed to have quit temporarily for periods centred on any time point when they reported smoking no cigarettes.<sup>a</sup> In HALS, attention was restricted, as closely as the questionnaire wording allowed, to regular smoking, only about 1% of subjects in either sex being current occasional smokers.<sup>b</sup>

---

<sup>a</sup> AHIP: For MC, if the age of starting was missing, then the younger of the cohort average (Table S2.7) and the first time point for which a positive amount was given was assumed; if the age of stopping was missing, then the most recent time point with a positive amount was used, or 5 years ago if all amounts were missing. Missing data on the amounts smoked were filled in based on that for the next most recent time point (i.e. when the subject was older) with non-missing data. MC smoking status at each year between starting and stopping or admission was then established, based on whether a positive or zero amount was given for the nearest time point. Thus intermediate *quit periods* were assumed when the subject had answered zero. Each subject's history was then allocated to actual years based on age and year of hospital admission.

For other products (HRC, pipes and cigars), only age of starting and age of stopping were available, and these were given only in categories. A smoker was assumed to have smoked continuously, using mid-points of the categories. If the age of start of smoking other products was missing, then the cohort average for MC smoking (Table S2.7) was assumed rather than attempting to calculate averages from the categorised data for the other products.

<sup>b</sup> HALS: For each product separately, subjects were assumed to have smoked continuously from their age of starting (or the cohort average shown in Table S2.7 if this was missing), up to and including 1985 for current smokers or, for former smokers, the age of stopping (or 5 years ago if missing). Occasional cigarette smokers were treated as former smokers if they had previously smoked regularly, otherwise they were treated as never having smoked cigarettes. The wording of the questions for cigars and pipe smoking meant that for current occasional smokers, any earlier period of regular smoking could not be assessed and they were therefore all treated as never smokers. Former occasional smokers were also treated as never smokers. Note that, although only smokers who smoked regularly at some time have been included, the question on age of starting to smoke did not specify any level, so that the smoking career as we have defined it may have included an initial period of occasional smoking.

Smoking prevalence estimates could then be calculated for any calendar year, taking as the base those subjects alive (and falling into any specified age group) at that time. Subjects were included up to 1985 for HALS, or to their year of hospital admission for AHIP. Estimates were calculated for cigarette smoking (HALS), MC smoking (AHIP) and for smoking of any product (both studies).

### ***Level of MC smoking***

This analysis was based on the AHIP questions on the number of MC smoked at up to 10 timepoints throughout life. A subject was considered a MC smoker in a particular year if he had answered positively for a time point within  $\pm 2$  years.<sup>c</sup> The average number of MC per smoker could then be calculated.

---

<sup>c</sup> The year to which each of the 10 timepoints related was calculated. For former smokers, the *latest* number smoked was taken to apply at the year when smoking stopped, but was omitted if the year stopped was missing, or for older subjects if the year stopped fell between 22 years ago and age 27. The time points were then compared with the years of interest (1950, 1955 etc), and were considered relevant if within  $\pm 2$  years. If there was more than one positive (i.e. non-zero non-missing) relevant answer then the average was taken. A subject was considered a smoker at each year of interest if there was a positive answer and the average number of cigarettes smoked per smoker could then be calculated. Note that the stated ages of starting and stopping smoking were ignored here.

## Results

Table S2.3 gives numbers of subjects by sex and birth cohort. HALS cohorts with few subjects are excluded from some analyses.

Table S2.4 gives, for HALS, numbers of cigarette smokers at time of interview and extent of missing data, while Table S2.5 shows similar data for MC smokers for AHIP. Table S2.6 shows numbers in AHIP who, based solely on their ages of starting and stopping smoking, were current smokers at selected timepoints, and the percentages who failed to give amount smoked for that timepoint. Less than 1% had missing data on age of starting in both studies. In AHIP, percentages with missing data on age of stopping or on recent amounts smoked were also very low, but were higher for amounts smoked longer ago—between 5% and 10% for 20 years ago or more. In HALS, percentages with missing data for age of stopping were higher, falling with age from about 20% for those under 35 to 4% for those over 65.

Table S2.6 also includes percentages of subjects who reported smoking a zero amount when, according to their ages of starting and stopping, they were smokers. This suggests that between 1% and 3% of smokers temporarily quit smoking at any time.

Table S2.7 shows the average age of starting smoking by sex and birth cohort for the two studies; for HALS results are shown for three products separately, although the number of female pipe smokers was very small. The age of starting was always lower for men than women, and fell for later-born cohorts, such that the gap between the sexes had virtually disappeared for the 1960–1969 cohort. The reduction with cohort was least evident for male cigarette (or MC) smokers. Age of starting was higher for cigars and pipes than for cigarettes.

Table S2.8 presents smoking prevalence estimates from the three sources for selected years, omitting estimates based on less than 20 subjects. Cigarette (HALS) or MC smoking (TMA and AHIP) is shown for both sexes, while all-product smoking is only shown for males (few women smoking products other than cigarettes). The data are also plotted on an annual basis in Figure S2.1 (males, cigarettes), Figure S2.2 (males, all products) and Figure S2.3 (females, cigarettes). We have followed the age groups most commonly used by TMA. These were altered in 1975, and estimates for that year are included in both forms. Note that the base for indirect estimates is all subjects falling within the age group at that year, but such subjects may come from a narrower age range. For instance, in 1970, AHIP subjects were aged 24–68, so that the group labelled 60+ actually only contains subjects aged 60–68, and age group 20–24 has been omitted as it contained only 4 male and 2 female 24-year-olds. TMA estimates are only

shown where indirect estimates are available for comparison. Note also that AHIP estimates after 1975 are based on progressively fewer subjects due to exclusion of subjects admitted before the given year.

Table S2.8 also shows the average difference over the period 1948–1975 between the direct TMA and the indirect estimates.

The three sets of estimates generally correspond well with the same time trends evident at all age groups for both sexes, except that AHIP estimates fall more steeply in the final five years than do the other sources.

For the central age groups (about age 20–60) HALS and AHIP estimates for cigarette smoking are generally a little higher than TMA for males, but lower for females. For older male smokers of all products, TMA estimates are consistently highest.

Table S2.9 compares indirect (AHIP) and direct (TMA) estimates of the average MC smoked per MC smoker for selected years. AHIP estimates almost always exceed TMA estimates, except after 1975. However, both sources show the male consumption higher than the female by a factor of 1½–2, and consumption rising with time within each sex and age group. Similar age patterns are seen at each year for both sources, with amount smoked increasing up to age 25 and then flattening off up to age 60. The AHIP estimates decline less at age 60+ than the TMA estimates, which may in part be an artefact of the different age distributions.

In Tables S2.10 and S2.11 smoking prevalence estimates from the two indirect sources are presented by sex and cohort back to 1920, well before direct TMA data are available. The AHIP estimates are generally slightly lower than HALS for the earlier-born male cohorts, but higher for the later-born males and for females. Among males both sources showed that the highest prevalence of smoking was reached by the 1900–1929 cohorts, with HALS showing the highest cigarette smoking concentrated in the 1920–1929 cohort. Both sources showed a substantial drop in prevalence between the 1920–1929 and 1930–1939 cohorts. For females, both sources showed the highest prevalence occurred in the 1920–29 cohort. Within the earlier-born female cohorts, the highest prevalence was reached about 20 years older than for males or for the late born cohorts.

Table S2.12 presents estimates of the amount of MC smoking by sex and cohort. This shows that the rise in consumption per smoker had started earlier than the period covered by Table 7. For instance, at about age 25, daily cigarette consumption rose from 16 in 1930 to 18 in 1940 for males, and from 9 to 10 for females.

## Discussion

We have compared direct estimates of smoking prevalence in Britain from TMA surveys conducted since 1948 with indirect estimates based on recall in two recent surveys (HALS and AHIP). We found a generally good level of agreement in the prevalence of smoking either of cigarettes or all products, by sex and age group, over the 35 year period studied. We also compared direct estimates of amount smoked with indirect estimates from one source (AHIP). Although indirect estimates were consistently higher than direct estimates, patterns over time were similar.

We also found quite consistent results from two indirect sources for other aspects of smoking, including prevalence in earlier decades and average age of starting to smoke.

We now consider features of the data from the three sources which might have influenced the results.

### Subjects included

The subjects in the HALS and TMA surveys were representative of Britain. Potential biases from selection of subjects in the AHIP study are difficult to assess. Limited regional data suggest that Scotland, excluded from AHIP, has higher smoking prevalence than England.<sup>4</sup> Hospital patients may have reduced their smoking due to illness, even if their condition is not smoking-associated, and this may have caused the steeper decline in AHIP prevalence estimates in the final years. AHIP estimates for 1976–1980 are also less reliable due to the diminishing subject base.

### Smoking definition

Some differences in prevalence would be expected from the differing definitions and methods used by the three sources. TMA estimates may be inflated from using self-defined rather than regular smokers, and by sales adjustment, but neither cause is likely to raise the estimate by more than about 1%. The fact that the AHIP definition of age of starting relates to regular, not self-defined, smoking may contribute to the higher average age of starting in this study, although it is only evident for males (Table S2.7).

Minor differences in the TMA definition of *all products* (exclusion of pipe smoking for women, exclusion of cigar only smokers before 1963) probably have little effect, while the less detailed questions on other products in HALS and AHIP make their *all products* estimates less reliable. The definitions of *cigarette* smoking might be expected to have had a more major



impact. For AHIP, we could construct MC prevalence estimates corresponding to the TMA estimates, but for HALS it was impossible to separate off HRC smoking. An indication of the likely size of difference can be gained from TMA data (since 1961) on prevalence of smoking of HRC *only*.<sup>3,4</sup> Among men, this rose, from 1–2% at age 16–24 to 5–9% at age 60+. It was always negligible among women. Such differences for older men are not evident in the results (Table S2.8, Figure S2.1).

### **Quit periods**

Failure to take account of intermediate quit periods would also theoretically increase the HALS estimates. Although some studies report that many smokers try unsuccessfully to give up (e.g. in HALS, 66% of male and 62% of female current smokers had seriously attempted to give up), we are unaware of published data on lengths of quit periods achieved. The reports of amounts smoked in AHIP (Table S2.6) suggest between 1% and 3% of smokers had temporarily quit at any given time point within their total smoking life. This implies that ignoring quit periods would inflate prevalence estimates by about ½%–2%.

### **Missing data**

To assess the impact of partial missing data, cigarette smoking prevalences were recalculated for HALS, omitting subjects who failed to give ages of starting or stopping, and the results compared with Table S2.8. Since most subjects omitted were former smokers (Table S2.4), assumed simplistically to have stopped five years ago, estimates for 1985 were *increased* by up to 2%, while estimates for previous years were *decreased* by up to 3%.

### **Inaccurate reporting**

Recent studies<sup>2</sup> have shown that a proportion of current smokers deny smoking at interview, which would tend to depress estimates of the current prevalence of smoking. In 1969, TMA, based on information from other surveys, altered their method of sales adjustment to include adjustment of prevalence,<sup>4</sup> suggesting that increasing social pressures against smoking may be linked to the emergence of the problem of understatement around that time. Subjects interviewed twice and giving inconsistent replies on ever having smoked are more likely to be long-term ex-smokers,<sup>2</sup> implying that, in estimates based on recall, understatement may increase with the length of the recall period. Interview setting has also been shown to influence the level of smoking denial,<sup>7</sup> perhaps relevant to the medical context of AHIP. It should also be noted the

TMA method of sales adjustment is applied to the whole sample studied though understatement may vary by sex and cohort.

### **Differential mortality**

An important potential source of bias in indirect estimates is the differential mortality suffered by smokers and nonsmokers. Only persons who survive to the date of interview can contribute to a survey and, since cigarette smokers have higher mortality than nonsmokers, survivors would be expected to have smoked less at any point in the past than the whole population alive at that time.

To investigate this further, we considered the life tables given by Hammond.<sup>8</sup> For a hypothetical population consisting at a given point of 50% lifelong nonsmokers and 50% current smokers of 20–39 cigarettes per day, we estimated the percentage of smokers which would have been observed in interviews conducted among survivors up to 40 years later (Table S2.13). It can be seen that the major determinant of the observed percentage is age at interview, with bias less than 1% for those age under 50, about 2% for those age 50–60 and about 5% for those age 60–70, with little difference related to the length of time to interview. For subjects interviewed at older ages, the bias becomes substantial. Although these calculations are only approximate, they suggest that this bias should not be important if attention is limited to subjects aged under 70, i.e. born since about 1910/1915 in the studies considered here.

This suggestion is supported by Harris<sup>9</sup> who presented a method of correcting for differential mortality. Using interviews in 1978–1980, he found that the main effect of correction was to increase smoking prevalence estimates for men born before 1910.

## **Conclusion**

The marked agreement between the prevalence estimates (Table S2.8 and Figures S2.1–3) suggests that most sources of potential bias have not in fact had a large influence. Indeed, it is surprising that differences in definitions of *cigarette* smoking (including HRC only for HALS) has not had more impact on prevalence estimates for older men. Conversely, none of the factors discussed provides an explanation for the higher estimates of amount smoked by indirect than direct methods. We have also demonstrated that bias due to differential mortality should be minor if attention is limited to subjects aged under 70 at interview. For both prevalence and level of smoking, the indirect method has successfully shown similar patterns to the direct estimates in respect of sex, age or cohort, and time. It therefore represents a useful approach where direct estimates are not available.

## **Acknowledgements**

We thank the Health and Lifestyle Survey at the University of Cambridge and the ESRC Data Archive for permission to use HALS data and Mrs P Wassell for preparing the manuscript. Financial support was provided by Philip Morris Europe. We alone bear responsibility for analysis and interpretation of these data.

## References

1. Forey B, Hamling J, Lee P, Wald N, editors. *International Smoking Statistics. A collection of historical data from 30 economically developed countries*, 2nd edition. London and Oxford: Wolfson Institute of Preventive Medicine and Oxford University Press; 2001.
2. 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**:109-29.
3. Lee PN, editor. *Statistics of smoking in the United Kingdom*, 7th edition. London: Tobacco Research Council; 1976. Research Paper I.
4. Wald N, Nicolaides-Bouman A, editors. *UK Smoking Statistics*, 2nd edition. London and Oxford: Wolfson Institute of Preventive Medicine and Oxford University Press; 1991.
5. Cox BD, Blaxter M, Buckle ALJ, Fenner NP, Golding JF, Gore M, *et al.* *The health and lifestyle survey. Preliminary report of a nationwide survey of the physical and mental health, attitudes and lifestyle of a random sample of 9,003 British adults*. London: Health Promotion Research Trust; 1987.
6. Alderson MR, Lee PN, Wang R. Risks of lung cancer, chronic bronchitis, ischaemic heart disease, and stroke in relation to type of cigarette smoked. *J Epidemiol Community Health* 1985;**39**:286-93.
7. Lee PN. Misclassification of smoking habits and passive smoking. A review of the evidence. *Int Arch Occup Environ Health* 1988; Suppl, 103pp. Heidelberg: Springer-Verlag
8. Hammond EC. Life expectancy of American men in relation to their smoking habits. *J Natl Cancer Inst* 1969;**43**:951-62.
9. Harris JE. Cigarette smoking among successive birth cohorts of men and women in the United States during 1900–80. *J Natl Cancer Inst* 1983;**71**:473-9.

**Table S2.1** Summary of relevant smoking questions in the HALS studyCigarettes:

Status = Current regular/current occasional/former regular for 6 months

Age started<sup>a</sup>

Current occasional smokers:

–previously smoked regularly for at least 6 months

–years since last smoked regularly

Former smokers:

–years since completely stopped

Cigars and pipe (separately):

Status = Current regular/former regular/current occasional/former occasional

Age started<sup>a</sup>

Former smokers:

–years since stopped<sup>a</sup>

Note. Regular = daily.

<sup>a</sup> Level of smoking not specified.**Table S2.2** Summary of relevant smoking questions in the AHIP studyMC, HRC, pipe and cigars (separately):Status<sup>a</sup> = Current regular for one year/former regular for one yearAge<sup>b</sup> started regular smoking

Current smokers:

–amount smoked at time of hospital admission

Former smokers:

–age<sup>b</sup> last smoked regularly

–amount last smoked

Additionally for MC only:

–amount smoked at 9 additional time points–1, 3, 5, 10, 15, 20 years before admission and at age 25, 20, 16

## Notes

Regular = daily for MC, HRC and pipe, weekly for cigars.

Time points are no more than 5 years apart, except for subjects aged over 50, for whom the period *age 25 to 20 years ago* is longer. For younger subjects, certain questions were omitted depending on age to avoid time points overlapping.<sup>a</sup> At time of hospital admission.<sup>b</sup> Actual age for MC only, otherwise categories.**Table S2.3** Number of subjects included from HALS and AHIP studies, by sex and birth cohort

Year of birth	HALS		AHIP	
	Males	Females	Males	Females
1880–1889	0	2		
1890–1899	15	48		
1900–1909	214	292	286	308
1910–1919	438	580	789	766
1920–1929	633	769	798	775
1930–1939	589	793	571	383
1940–1949	751	993	182	88
1950–1959	725	980		
1960–1969	539	637		
Total	3904	5094	2626	2320

Note. 5 HALS and 4 AHIP subjects without primary smoking data have been omitted.

**Table S2.4** HALS study. Numbers of cigarette smokers and extent of missing data, by cigarette smoking category and sex

Cigarette smoking category <sup>a</sup>	Sex	N <sup>b</sup>	Data missing	n <sup>c</sup>	%
Current smoker	Male	1358	Age started	7	0.5
	Female	1598	Age started	12	0.8
Occasional, previously regular smoker	Male	55	Age started	0	0.0
			Age stopped	8	14.5
	Female	61	Age started	1	1.6
			Age stopped	7	11.5
Former smoker	Male	1222	Age started	14	1.1
			Age stopped	84	6.8
	Female	1002	Age started	11	1.1
			Age stopped	107	10.7

<sup>a</sup> At time of interview.<sup>b</sup> Number in smoking category.<sup>c</sup> Number with missing data.**Table S2.5** AHIP study. Numbers of MC smokers and extent of missing data, by MC smoking category and sex

MC smoking category <sup>a</sup>	Sex	N <sup>b</sup>	Data missing	n <sup>c</sup>	%
Current smoker	Male	856	Age started	0	0.0
			Current amount	2	0.2
			All amounts	2	0.2
	Female	643	Age started	2	0.3
			Current amount	0	0.0
			All amounts	0	0.0
Former smoker	Male	1188	Age started	5	0.4
			Age stopped	4	0.3
			Latest amount	14	1.2
			All amounts	7	0.6
	Female	671	Age started	5	0.7
			Age stopped	7	1.0
			Latest amount	3	0.4
			All amounts	1	0.1

MC—Manufactured cigarettes.

<sup>a</sup> At time of admission.<sup>b</sup> Number in smoking category.<sup>c</sup> Number with missing data.

**Table S2.6** AHIP study. Numbers of subjects known to have been MC smokers at selected timepoints, and percentages with missing or zero amount smoked for that timepoint

Sex	MC smoking status <sup>a</sup>		Time point			
			10 years ago	20 years ago <sup>b</sup>	Age 25 <sup>c</sup>	Age 20
Male	Current	<i>N</i>	842	569	732	706
		% with amount missing	2.2	6.5	7.5	5.7
		% with amount zero	2.5	2.5	1.4	1.4
	Former	<i>N</i>	632	676	988	957
		% with amount missing	4.4	8.6	6.8	6.9
		% with amount zero	1.3	1.2	0.8	0.9
Female	Current	<i>N</i>	625	453	497	423
		% with amount missing	2.2	8.4	7.4	6.1
		% with amount zero	1.1	1.1	1.6	0.9
	Former	<i>N</i>	428	439	460	381
		% with amount missing	4.2	9.1	6.5	6.6
		% with amount zero	1.9	2.3	2.0	1.3

MC—Manufactured cigarettes.

*N*—Number of subjects known to be a smoker at each time point, based solely on age(s) of starting and stopping smoking. Subjects with age(s) of starting or stopping missing have been omitted.

Amount—Number of cigarettes smoked at the timepoint.

<sup>a</sup> At time of admission.

<sup>b</sup> Not asked of subjects aged less than 50.

<sup>c</sup> Not asked of subjects aged less than 40.

**Table S2.7** Average age of starting smoking (number of smokers), by sex, birth cohort and smoking product

Sex	Year of birth	HALS			AHIP
		cigarettes	cigars	pipe	MC
Male	1890–1899	16.5 (8)	36.3 (4)	18.3 (4)	
	1900–1909	16.3 (171)	31.0 (66)	28.0 (93)	18.2 (219)
	1910–1919	16.6 (347)	35.0 (143)	29.7 (167)	18.5 (600)
	1920–1929	16.0 (508)	31.3 (225)	27.2 (232)	17.2 (662)
	1930–1939	16.5 (408)	28.3 (204)	27.9 (171)	17.5 (425)
	1940–1949	15.7 (536)	24.0 (308)	25.9 (213)	16.7 (133)
	1950–1959	15.1 (431)	20.8 (248)	22.5 (94)	
	1960–1969	14.7 (244)	18.2 (95)	19.8 (8)	
Female	1890–1899	28.1 (11)	(0)	(0)	
	1900–1909	26.2 (100)	43.5 (4)	(0)	27.3 (112)
	1910–1919	23.0 (266)	40.4 (10)	(0)	24.0 (421)
	1920–1929	19.7 (471)	38.1 (14)	25.3 (3)	19.7 (475)
	1930–1939	19.5 (438)	31.7 (18)	(0)	19.6 (218)
	1940–1949	17.7 (550)	25.3 (52)	19.7 (3)	17.4 (47)
	1950–1959	16.2 (543)	22.2 (44)	20.8 (4)	
	1960–1969	15.1 (300)	18.1 (10)	19.0 (2)	

MC—Manufactured cigarettes.

Note. Excluding subjects who did not state their age of starting to smoke, and 2 HALS subjects born before 1890 (both female never smokers).

**Table S2.8** Prevalence of smoking at selected years from different sources

Sex	Age group	1950			1960			1970			Average difference <sup>a</sup> 1948–1975		
		TMA	HALS	AHIP	TMA	HALS	AHIP	TMA	HALS	AHIP	HALS-TMA	AHIP-TMA	
Males	16–19	<i>N</i>	217	267	678	279	120	648	280				
		%C	51	44	49	65	56	49	55	53	-1.9	-4.1 <sup>b</sup>	
		%A	52	45	52	65	56	52	56	53	-3.2	-2.7 <sup>b</sup>	
	20–24	<i>N</i>	303	356	670	306	221	775	430				
		%C	68	73	71	67	60	63	58	61	2.1	2.4 <sup>c</sup>	
		%A	71	74	76	70	61	69	65	62	-1.1	3.2 <sup>c</sup>	
	25–34	<i>N</i>	559	793	772	588	676	1212	657	353			
		%C	70	77	72	64	69	68	60	64	60	6.5	4.0
		%A	79	81	81	73	72	75	70	67	65	0.5	1.6
	35–59	<i>N</i>	497	800	1992	1034	1593	1622	1377	1896			
		%C	66	70	64	64	67	62	55	58	57	4.2	0.6
		%A	81	78	76	78	74	76	71	65	70	-4.3	-1.7
60+	<i>N</i>			789	22		760	267	373				
	%C			46	45		46	47	39	0.6 <sup>d</sup>	-4.2 <sup>e</sup>		
	%A			69	64		68	60	57	-8.5 <sup>d</sup>	-7.6 <sup>e</sup>		
Females	16–19	<i>N</i>	333	195	638	362	64	609	407				
		%C	36	26	29	45	34	36	52	42	-5.7	-1.8 <sup>b</sup>	
	20–24	<i>N</i>	359	295	749	397	123	768	573				
		%C	48	46	55	48	45	45	54	51	-1.9	0.1 <sup>c</sup>	
	25–34	<i>N</i>	731	808	767	766	527	1238	848	192			
		%C	53	47	53	53	50	53	51	48	51	-2.4	0.6
	35–59	<i>N</i>	667	791	2099	1334	1599	1715	1767	1749			
		%C	38	34	38	47	41	46	50	45	48	-6.5	-2.0
	60+	<i>N</i>			1134	68		1028	399	377			
		%C			22	16		26	21	27	-5.8 <sup>d</sup>	0.6 <sup>e</sup>	

*N*—Number of subjects. This was unknown for the TMA 1950 survey (total sample approximately 10,000). *N* for TMA for 1960 and 1970 not available, data shown are for 1961 and 1971 and are "similar".

%C—Prevalence of MC smoking (TMA and AHIP) or cigarette smoking (HALS).

%A—Prevalence of smoking any product.

<sup>a</sup> Calculated from estimates as whole numbers, no greater precision being available for TMA.

<sup>b</sup> 1948–1963.

<sup>c</sup> 1948–1968.

<sup>d</sup> 1960–1975.

<sup>e</sup> 1965–1975.



**Table S2.9** Average number of MC smoked per smoker at selected years from different sources

Sex	Age group	1950		1960		1970	
		TMA	AHIP	TMA	AHIP	TMA	AHIP
Males	16–19	9.7	14.6 (150)	13.0	15.7 (65)		
	20–24	14.1	19.1 (239)	17.9	18.1 (134)		
	25–29	15.9	19.9 (235)	17.7	19.6 (168)	19.5	24.9 (84)
	30–34			18.6	23.2 (196)	19.9	22.4 (134)
	35–49			21.2	21.9 (780)	20.8	24.3 (670)
	50–59			21.0	21.3 (205)	19.9	22.9 (426)
	60+					15.8	21.1 (152)
Females	16–19	5.2	10.0 (59)	7.3	10.3 (26)		
	20–24	7.8	10.6 (160)	9.6	12.3 (56)		
	25–29	7.4	11.8 (160)	9.5	14.0 (81)	14.5	15.3 (36)
	30–34			11.9	13.0 (134)	13.7	15.3 (60)
	35–49			12.2	13.9 (585)	14.7	15.7 (495)
	50–59			10.7	12.4 (115)	14.2	15.3 (351)
	60+					10.1	13.9 (110)

MC—Manufactured cigarettes.

Note. The bases (number of MC smokers) for the AHIP estimates are shown in brackets. The bases for the TMA estimates are unknown, but can be derived approximately from Table S2.8.

**Table S2.10** HALS Study. Prevalence (%) of smoking at selected years, by sex and birth cohort

Sex	Smoking category	Year of birth	1920	1925	1930	1935	1940	1945	1950	1955	1960	1965	1970	1975	1980	1985
Male	Cigarettes	1890–1899	47	47	47	47	47	40	40	40	40	33	13	13	13	7
		1900–1909	33	62	75	76	75	73	68	65	61	55	48	38	29	19
		1910–1919	1 <sup>a</sup>	13	41	63	75	76	74	69	64	60	53	48	38	28
		1920–1929			1	11	40	71	76	76	73	70	61	53	46	32
		1930–1939					1	10	31	58	63	63	57	52	46	36
		1940–1949							1	9	36	63	66	58	49	40
		1950–1959									1	10	36	53	50	39
1960–1969											1	10	36	53	50	39
Male	All	1890–1899	73	67	67	67	67	60	60	60	53	47	33	33	33	20
		1900–1909	35	67	80	81	82	81	77	75	71	65	60	50	42	31
		1910–1919	1	13	42	65	77	81	79	75	71	69	64	60	49	37
		1920–1929			1	12	41	73	79	79	77	74	67	59	54	40
		1930–1939					2	11	32	60	65	66	62	58	53	44
		1940–1949							1	9	36	64	69	63	55	46
		1950–1959									1	10	36	54	51	41
1960–1969											1	11	35	36		
Female	Cigarettes	1890–1899	15	19	19	19	19	23	21	21	17	15	13	13	13	8
		1900–1909	3	10	19	23	28	29	30	28	26	24	20	17	14	10
		1910–1919	0	1	8	23	34	38	38	39	39	39	36	32	26	21
		1920–1929			0	2	19	46	51	53	52	52	50	49	44	36
		1930–1939					0	3	16	36	47	48	48	46	43	34
		1940–1949							0	2	19	44	49	45	39	33
		1950–1959										0	4	28	46	45
1960–1969												0	9	35	36	

Note. Bases are given in Table S2.3.

<sup>a</sup> This group of smokers comprises 2 subjects born in 1910 and starting at ages 8 and 9, 1 subject born in 1911 starting at age 7 and 2 subjects born in 1912 starting at ages 6 and 8.

**Table S2.11** AHIP Study. Prevalence (%) of smoking at selected years, by sex and birth cohort

Sex	Smoking category	Year of birth	1920	1925	1930	1935	1940	1945	1950	1955	1960	1965	1970	1975	1980 <sup>a</sup>
Male	MC	1900–1909	13	48	63	66	67	67	64	59	53	45	38	30	16 (109)
		1910–1919	0	7	33	55	65	68	64	63	60	55	50	43	28 (398)
		1920–1929			1	8	39	69	74	72	69	65	60	51	39 (387)
		1930–1939					1	8	36	60	66	64	61	53	43 (326)
		1940–1949							2	13	54	64	63	48	39 (116)
Male	All	1900–1909	14	53	72	77	78	80	77	77	72	56	47	47	32 (109)
		1910–1919	0	7	36	61	74	77	75	75	73	70	65	57	40 (398)
		1920–1929			1	9	43	76	82	81	79	76	72	67	51 (387)
		1930–1939					1	8	38	64	72	71	69	64	53 (326)
		1940–1949							2	14	57	69	69	57	50 (116)
Female	MC	1900–1909	1	9	17	20	25	30	31	30	30	29	24	19	13 (115)
		1910–1919	0	1	8	26	37	45	45	45	45	45	41	35	23 (408)
		1920–1929			0	3	22	47	55	57	57	56	54	49	37 (420)
		1930–1939					0	4	22	43	47	46	49	45	39 (213)
		1940–1949							1	1	34	50	52	47	35 (55)

MC—Manufactured cigarettes.

<sup>a</sup> Estimates for 1980 exclude about half the subjects who were hospitalised before 1980, and the bases are given in brackets. Bases for other years are given in Table S2.3.**Table S2.12** AHIP Study. Average number of MC per smoker at selected years, by sex and birth cohort

Sex	Year of birth		1920	1925	1930	1935	1940	1945	1950	1955	1960	1965	1970	1975	1980
Male	1900–1909	N	60	154	180	117					160	140	116	96	71
		nMC	9.2	12.0	16.4	17.6					21.2	20.7	21.5	20.3	17.7
	1910–1919	N		101	319	449	398	148			472	431	401	368	291
		nMC		7.9	11.7	15.1	17.7	18.7			22.1	22.2	22.2	21.1	18.8
	1920–1929	N				113	375	550	437	202	529	510	487	441	365
nMC					10.4	12.9	17.1	20.0	20.4	22.1	23.5	24.1	25.1	23.0	
1930–1939	N						76	246	346	294	312	348	323	270	
	nMC						12.1	13.9	17.5	19.7	23.5	24.4	25.5	24.6	
1940–1949	N								42	105	73	118	98	84	
	nMC								10.4	15.1	19.2	23.6	25.0	24.1	
Female	1900–1909	N	7	38	53	37					86	83	80	63	48
		nMC	5.9	7.0	9.2	10.0					12.8	14.1	14.4	13.3	12.4
	1910–1919	N		12	96	216	221	98			326	323	322	295	220
		nMC		5.3	7.3	8.6	10.5	12.1			13.8	14.7	15.0	15.0	14.1
	1920–1929	N				53	230	381	301	134	417	421	422	407	351
nMC					7.8	9.5	10.9	11.3	11.1	13.5	14.8	15.3	16.3	15.9	
1930–1939	N						32	100	163	136	147	184	189	165	
	nMC						6.8	9.6	12.0	13.4	15.5	16.2	17.9	16.8	
1940–1949	N								10	35	31	46	44	39	
	nMC								7.6	9.8	13.8	15.7	18.7	17.5	

MC—Manufactured cigarettes.

N—The base is the number of subjects giving a positive amount smoked within 2 years of the selected year. This will not correspond exactly to the prevalence estimate given in Table S2.11—see Methods.

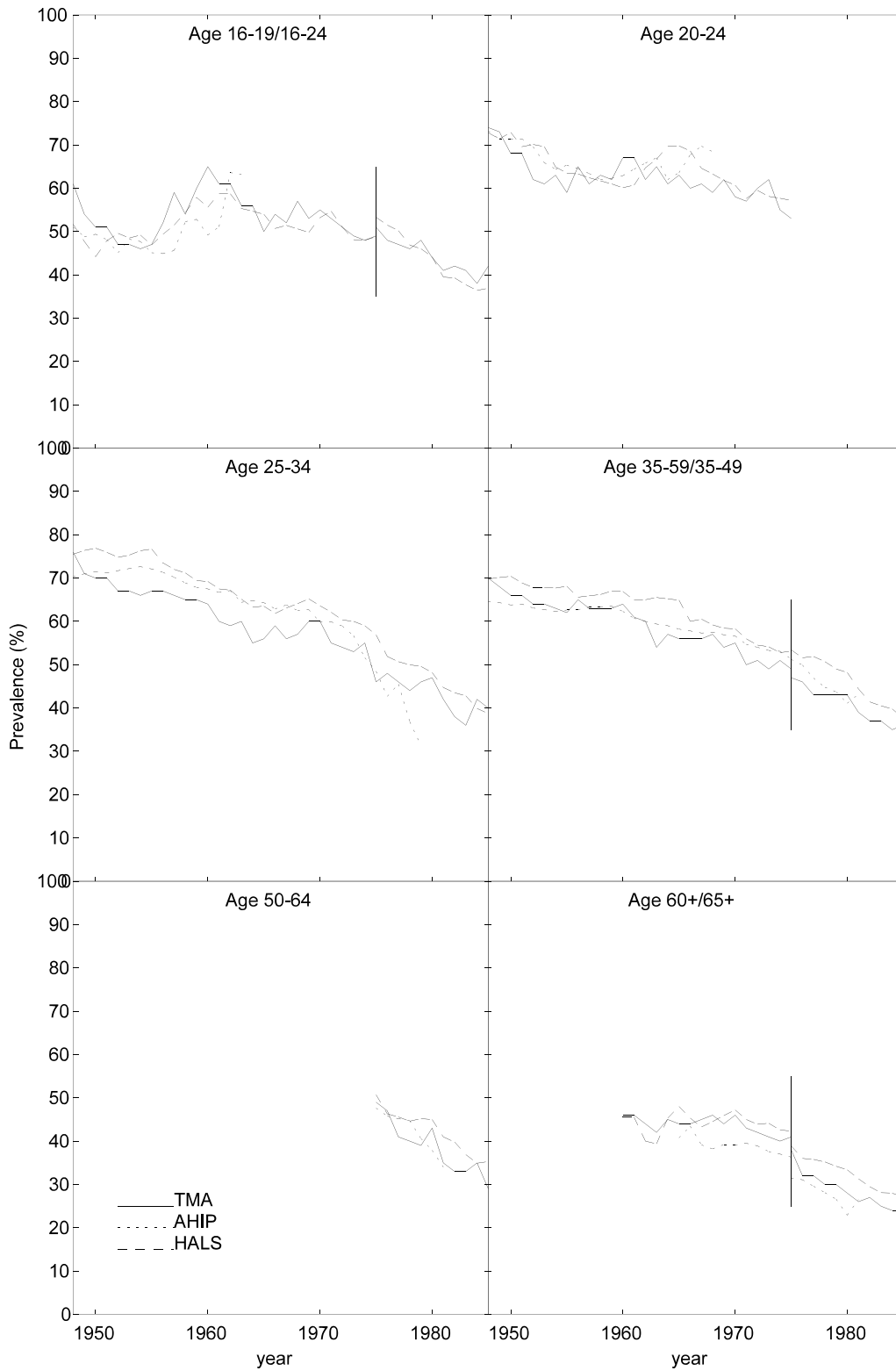
nMC—Average number of MC per MC smoker.

Note. For earlier born cohorts, questions on smoking at age 25 and 20 years ago leave a gap in the data.

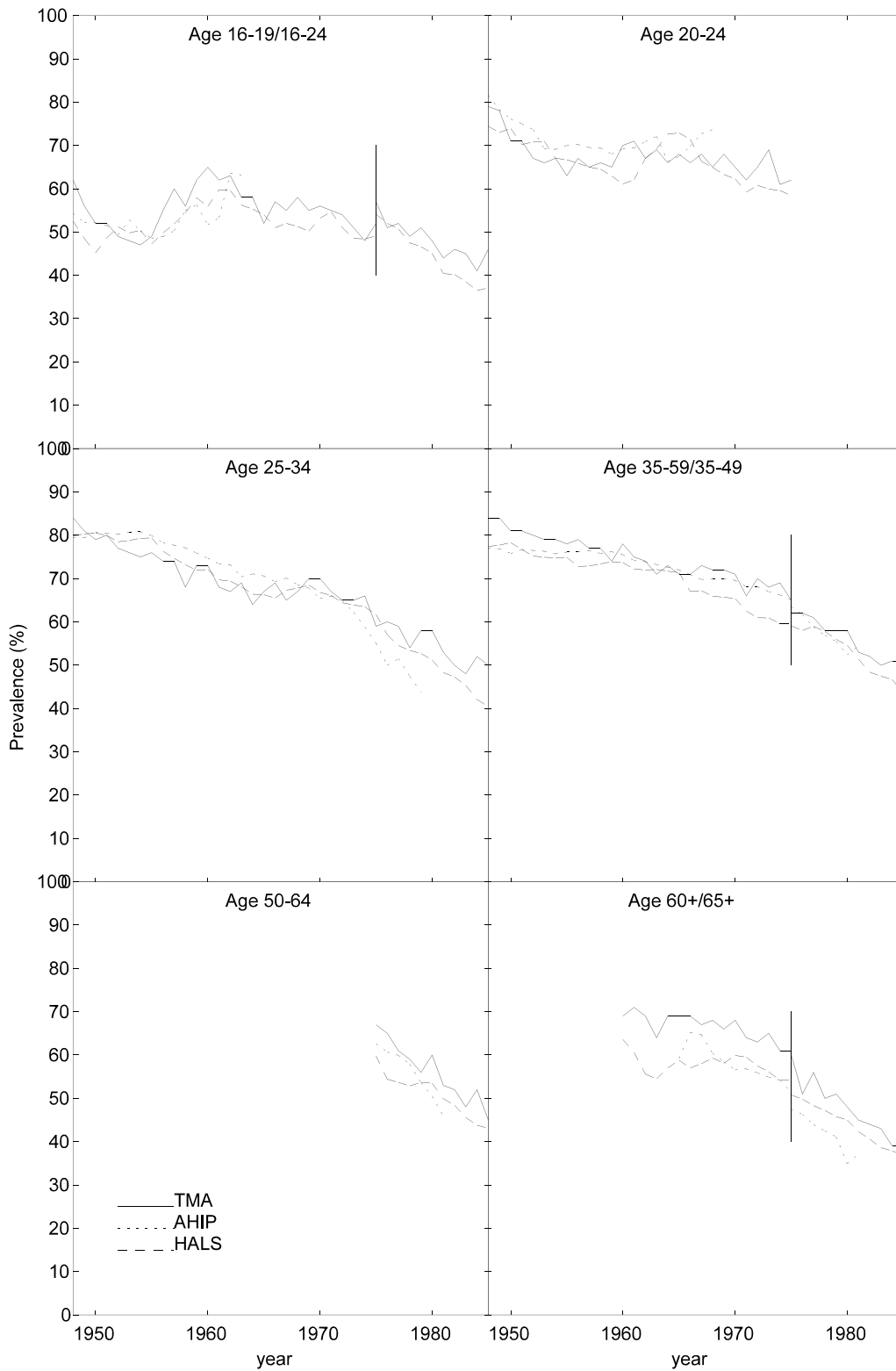
**Table S2.13** Percentage of smokers among survivors in a hypothetical population starting with 50% never smokers and 50% current smokers of 20–39 cigarettes per day

Age at start	Age at interview					
	35	45	55	65	75	85
25	49.8					
35		49.3				
45		49.5			36.8	
55			47.6		44.2	
65			47.8		44.7	27.9
75			48.3	46.4	39.0	29.2
					42.4	32.3
						39.3

**Figure S2.1** Prevalence of MC smoking (TMA and AHIP) or cigarette smoking (HALS) among men, from different sources, 1948–1985.



**Figure S2.2** Prevalence of smoking (any product) among men, from different sources, 1948–1985.



**Figure S2.3** Prevalence of MC smoking (TMA and AHIP) or cigarette smoking (HALS) among women, from different sources, 1948–1985.

