ESTIMATING AGE-SPECIFIC TRANSITION RATES FOR POPULATION SUBGROUPS FROM SUCCESSIVE SURVEYS

Michael A. Stoto

August 1985
WP-85-52

This work was supported by the Division of Cancer Prevention and Control of the U.S. National Cancer Institute.

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

ABSTRACT

As an extension of a recent result in generalized stable population theory, this paper develops a new method for estimating age-specific transition rates for population subgroups based on data from two successive surveys. A test based on detailed data on smoking behavior from the 1976 and 1981 censuses of New Zealand shows that the method works well when the data are aggregated into a small number of irregular age groups and when the surveys are irregularly spaced.

ESTIMATING AGE-SPECIFIC TRANSITION RATES FOR POPULATION SUBGROUPS FROM SUCCESSIVE SURVEYS

Michael A. Stoto

Institute for the Study of Smoking
Behavior and Policy
Harvard University
Cambridge, Massachusetts 02138
USA

Since the Surgeon General's first report about the hazards of smoking in 1964, a remarkable change in smoking habits has occurred in the United States. Nearly everyone is aware in a general way of the hazards of smoking, and the proportion of adults who smoke regularly dropped from 38 percent in 1955 to 32 percent in 1983. But this statistic hides the fact that while the proportion of males who smoke declined sharply from 52 to 35 percent, the proportion of females who smoke increased from 25 to 34 percent in 1965 and has declined only slightly in recent years to 30 percent (DHEW, 1979, DHHS, 1984).

Beyond such summary statistics, health experts know very little about the characteristics of smoking change in the United States or other countries. Except in very select experimental populations, few studies have followed individuals over time and recorded changes in their smoking behavior recorded. Similarly few population-based studies include retrospectively reported changes in smoking behavior within a fixed reference period. The best information that we have about cohort changes in smoking behavior comes from a single study based on a retrospective smoking history taken in a 1978-80 survey (Harris, 1983).

Nationally representative data on smoking habits are usually in the form of sample-based estimates, in various years, of the proportion of the population in various age, sex, race, and sometimes socioeconomic subgroups that smokes. For the United States, for instance, the most consistent data series with information on changing smoking behavior is the series of studies based on the National Center for Health Statistics' regular Health Interview Survey (HIS). This survey provides estimates of the proportion of the population in the following age groups that are current or former smokers: 20-24, 25-34, 35-44, 45-64, and

65 and older. Data exist for 1965, 1970, 1976, 1978-80, and 1983 (DHEW, 1979, DHHS, 1984). Similar series of surveys exist for Australia, Canada, Great Britain, the Netherlands, and Japan. (Durch, 1985). In New Zealand, similar information, but based on a national census rather than a sample survey, is available for 1976 and 1981 (Hay and Foster, 1984).

Given the paucity of other data, these series provide valuable information about a period which has seen dramatic changes in smoking behavior, but analyzing them is a difficult demographic challenge. Demographers traditionally have estimated transition rates from one state to another - living to dead, single to married, employed to unemployed, and so on - by dividing a count of the number of transitions from a vital statistics system by a measure of exposure from a census or survey. This paper extends a recent development in mathematical demography and indirect estimation by Preston and Coale (1982) to provide a general methodology for estimating age-specific net transition rates for population subgroups from the information contained in two surveys. Although the current application and data relate to changes in smoking behavior, the method is also appropriate for other situations in which the relative sizes of population subgroups are known, with reasonable age detail, at two points in time.

If there were data on smoking behavior for n-year population age groups in two surveys taken n years apart, the demographic problem of estimating rates of change would be relatively simple. The commonly available data, however, present two difficulties to demographers. First, the data are often tabulated for irregular age groups. In the HIS data, for instance, there are 5-year, 10-year, 20-year, and open-ended age groups. This tabulation is a compromise between adequate sample sizes and the need for information about particular population groups. It is sometimes impossible to retabulate the data on a more regualar basis. Second, in most cases, the surveys have been taken at irregular intervals. Even when the data are available for 5- or 10-year age groups, the surveys are 3, 6, or 11 years apart.

The first section of this paper derives a method for using such irregular aggregate data to infer transition rates into and out of the population of smokers. It first considers the case of regular age groups, expands this to include irregular age groups, and discusses special treatment necessary for the final open interval. The second section uses the New Zealand data, which is aggregated into 5-year age groups and comes from censuses separated by 5 years, to test the method. The final section discusses the strengths and weaknesses of the new method, and the implications of the assumptions used in its derivation. A companion paper (Stoto, 1985) uses these methods to analyze in detail changes in smoking behavior in the United States.

DERIVATION

Regular Age Groups

Let us begin with the simplest case, that in which all age groups have the same width n, usually 5 or 10 years, but the two surveys are $T \neq n$ years apart.

Let ${}_nN_a$ be the population aged a to $a+n$, and ${}_nN_a^S$ be the number of smokers in the same group. Survey data reveal ${}_nP_a$, the proportion of the population aged a to $a+n$ that classify themselves as smokers. By definition,

$${}_nP_a = {}_nN_a^S / {}_nN_a \quad .$$

Furthermore, let ${}_nL_a$ be the number of person-years lived in the interval a to $a+n$ according to the population life table, and ${}_nL_a^S$ be the same quantity according to the mortality rates experienced by smokers. Similarly, let ${}_nK_a$ be the number of person-years that would be experienced in a population of smokers aged a to $a+n$ if the only increments and decrements were starting and quitting smoking. The object is to use the ${}_nP_a$ to calculate ${}_nK_a$ for each age a, and hence derive a measure of the net change in smoking behaviour.

The key variable needed to relate ${}_nK_a$ and ${}_nP_a$ in the Preston-Coale scheme is the growth rate (over time) of the population aged a to $a+n$;

$${}_nr_a = \frac{\frac{d}{dt}\,{}_nN_{a+t}}{{}_nN_a}$$

Let us also define ${}_nr_a^S$ as

$${}_nr_a^S = \frac{\frac{d}{dt}\,{}_nN_{a+t}^S}{{}_nN_a^S} \quad .$$

Furthermore, define ${}_nr_a^+ = {}_nr_a^S - {}_nr_a$.

The key relationship is that, for any population,

$$ {}_nN_a = {}_nN_0 \; exp(-\int_0^a {}_nr_y\, dy) \left[\frac{{}_nL_a}{{}_nL_0}\right] . $$

(This would be exact, except that at one point in the derivation in Preston, et al. (1982), the life table death rate ${}_nM_x$ is approximated by the population death rate ${}_nm_x$. Similarly, for the population of smokers,

$$ {}_nN_a^S = {}_nN_0^S \; exp(-\int_0^a {}_nr_y^S dy) \left[\frac{{}_nL_a^S}{{}_nL_0^S}\right]\left[\frac{{}_nK_a}{{}_nK_0}\right] . $$

Then, by definition of ${}_nP_a$,

$$ {}_nP_a = {}_nP_0 \; exp(-\int_0^a ({}_nr_y^S - {}_nr_y)dy) \left[\frac{{}_nL_a^S}{{}_nL_0^S}\right]\left[\frac{{}_nL_0}{{}_nL_a}\right]\left[\frac{{}_nK_a}{{}_nK_0}\right] . $$

Using this result for two successive age groups ${}_nP_a$ and ${}_nP_{b_0}$ yields

$$ \frac{{}_nK_b}{{}_nK_a} = \frac{{}_nP_b}{{}_nP_a} \left[\frac{{}_nL_a^S}{{}_nL_b^S} \cdot \frac{{}_nL_b}{{}_nL^a}\right] exp(\int_a^b {}_nr_y^+ dy) \quad . \qquad (1) $$

In words, this formula means that the relative numbers of smokers at ages a and b in a synthetic cohort in which changes in smoking behaviour were the only cause of increment or decrement can be estimated from the relative proportion of smokers in the population times two "correction factors". The first factor corrects for differential mortality between smokers and non-smokers. The second factors takes into account the different initial sizes of the two cohorts aged a and b at time t, and different historical changes in their smoking behaviour.

Survey data representing ${}_nP_a$ for all ages a at two points in time, plus ancillary information on differential mortality, provide enough information to calculate the ${}_nK_a$. Suppose we have ${}_nP_{a,1}$ and ${}_nP_{a,2}$, at times t_1 and t_2, for all ages a that are multiples of n. For this time interval, we can approximate ${}_nP_a$ as ${}_nP_a = \frac{1}{2}({}_nP_{a,1} + {}_nP_{a,2})$ for all a. Letting ${}_nN_{a,i}$ and ${}_nN_{a,i}^S$ be the number in the total population and the smoking population at time t_i, we can write

$$ {}_nr_a \doteq \frac{1}{(t_2 - t_1)} \ln\left[\frac{{}_nN_{a,2}}{{}_nN_{a,1}}\right] $$

and

$$ {}_n r_a^S = \frac{1}{(t_2 - t_1)} \ln \left[\frac{{}_n N_{a,2}^S}{{}_n N_{a,1}^S} \right] . $$

Then,

$$ {}_n r_a^+ = \frac{1}{(t_2 - t_1)} \ln \left[\frac{{}_n P_{a,2}}{{}_n P_{a,1}} \right] $$

and can be calculated from the ${}_n P_{a,i}$ alone. Assuming as an approximation that ${}_n r_y^+$ is linear between a and b, the integral in equation (1) is approximately equal to

$$ \frac{(b - a)}{2} ({}_n r_a^+ + {}_n r_b^+) = \frac{(b - a)}{2(t_2 - t_1)} \ln \left[\frac{{}_n P_{b,2}}{{}_n P_{b,1}} \cdot \frac{{}_n P_{a,2}}{{}_n P_{a,1}} \right] . $$

Thus, the second correction factor is

$$ \left[\frac{{}_n P_{b,2}}{{}_n P_{b,1}} \cdot \frac{{}_n P_{a,2}}{{}_n P_{a,1}} \right]^{\frac{b-a}{2(t_2 - t_1)}} . $$

The factor to correct for differential mortality can be calculated as follows. Let ${}_n q_a$ be the proportion of the general population at exact age a who one expected to die before age $a+n$, and ${}_n q_a^S$ and ${}_n q_a^{NS}$ be similar quantities for smokers and non-smokers. Various studies have estimated R_a, the ratio of ${}_n q_a^S$ and ${}_n q_s^{NS}$ at age a. For the whole population, the identity

$$ {}_n q_a = {}_n P_a \, {}_n q_a^S + (1 - {}_n P_a) {}_n q_a^{NS} $$

must hold. Substituting $\frac{{}_n q_a^S}{R_a}$ for ${}_n q_a^{NS}$ yields

$$ {}_n q_a^S = R_a \, {}_n q_a (1 + {}_n P_a (R_a - 1))^{-1} . $$

Using standard life table calculations, the proportion of the life table population still alive at exact age is

$$ l(a+n) = l(a)(1 - {}_n q_a) , $$

and for smokers

$$ l^S(a+n) = l^S(a)(1 - {}_n q_a^S) . $$

The quantity ${}_nL_a = \int_a^{a+n} l(x)dx$ can then be approximated by $(l(a) + l(a+n))/2$, and ${}_nL_a^S$ by $n(l^S(a) + l^S(a+n))/2$, and equation (1) can be evaluated.

The result of equation (1) is a series of ratios, ${}_nK_b/{}_nK_a$, which have a definite demographic meaning, but are difficult to interpret in lay terms. Demographically, ${}_nK_b/{}_nK_a$ is the ratio of person-years lived as a smoker in the interval b to $b+n$ compared to the interval a to $a+n$ in a synthetic cohort experiencing only the age-specific net changes in smoking behavior observed in the population in the interval t, to t_2. For purposes of explication, we can calculate a constant rate of net smoking cessation over the age range or to $b+n$ as follows.

Let $k(x)$ be the number of smokers at exact age x. If the net rate of smoking cessation is constant over the range a to $b+n$, $k(x)$ has the following form

$$k(x) = Ke^{-\lambda x}$$

where λ is the net rate of cessation and K is a constant. Note that λ can be negative as well as positive; negative values of λ correspond to a $k(a)$ that increases with age. Under this assumption, we calculate

$${}_nK_a = \int_a^{a+n} k(x)dx = \frac{K}{\lambda}e^{-\lambda a}(1 - e^{-\lambda n})$$

and discover that

$$\frac{{}_nK_b}{{}_nK_a} = e^{-\lambda n}$$

so for $b = a + n$,

$$\lambda = \frac{-1}{n}\ln\left[\frac{{}_nK_b}{{}_nK_a}\right] .$$

Irregular Age Groups

If the data are aggregated into age groups of unequal width, the same approach can be used with a further approximation. Consider two adjacent age groups a to $a+n$ and b to $b+m$, where $n \neq m$. In this situation, we can apply equation (1) with two modifications.

First, we know ${}_mP_b / {}_nP_a$ rather than ${}_mP_b / {}_mP_a$ or ${}_nP_b / {}_nP_a$. With the simple assumption that $p(x)$, the proportion of the population at age x that smokes, is linear between a and $b+m$, we can approximate ${}_mP_b / {}_mP_a$ as

$$\frac{{}_mP_b}{{}_mP_a} = \frac{(m+n)\,{}_mP_b}{2n\,{}_nP_a + (m-n)\,{}_mP_b} \; .$$

Second, equation (1) requires evaluating the integral

$$\int_a^b {}_nr_y^+ dy$$

but we can calculate only ${}_nr_a^+$ and ${}_mr_b^+$. Since, *a priori*, there is no reason for ${}_nr_y^+$ to be consistently greater or less than ${}_mr_y^+$, we simply assume that they are the same and ignore the difference between n and m in calculating the integral.

Thus, with these two assumptions, we replace equation (1) with

$$\frac{{}_mK_b}{{}_mK_a} = \frac{(m+n)\,{}_mP_b}{2n\,{}_nP_a + (m-n)\,{}_mP_b} \left[\frac{{}_mL_a^S}{{}_mL_b^S} \cdot \frac{{}_mL_b}{{}_mL_a} \right] \left[\frac{{}_mP_{b,2}}{{}_mP_{b,1}} \cdot \frac{{}_nP_{a,2}}{{}_nP_{a,1}} \right]^{\frac{b-a}{2(t_2-t_1)}} \; . \qquad (2)$$

Open-ended Interval

Finally, we must deal with the open-ended interval, the proportion of the population aged b or older who smoke. As in the case of irregular age groups, we can use equation (1) with two modifications.

First, with an open-ended interval, we must approximate ${}_nP_b / {}_nP_a$, knowing only ${}_\infty P_b$ and ${}_nP_a$. In this situation the linear approximation used alone for $p(x)$ is not appropriate, so instead we assume that $p(x)$ declines exponentially. With this assumption,

$$\frac{{}_nP_b}{{}_nP_a} = \frac{{}_\infty P_b}{{}_\infty P_a}$$

so we need only calculate ${}_{\infty}P_a$.

Second, as above, we assume that ${}_n r_b^+ = {}_{\infty} r_b^+$.

Thus, for the open-ended interval, we can calculate

$$\frac{{}_nK_b}{{}_nK_a} = \frac{{}_{\infty}P_b}{{}_{\infty}P_a}\left(\frac{{}_nL_a^S}{{}_nL_b^S}\cdot\frac{{}_nL_b}{{}_nL_a}\right)\left(\frac{{}_{\infty}P_{b,2}}{{}_{\infty}P_{b,1}}\cdot\frac{{}_nP_{a,2}}{{}_nP_{a,1}}\right)^{\frac{b-a}{2(t_2-t_1)}} \quad . \tag{3}$$

Start Rates and Quit Rates

In addition to the proportion of people in various age groups who are current smokers, surveys frequently provide information on the proportions who are former smokers. This additional information can be very valuable in partially sorting out the increments and decrements in the observed net changes in smoking behavior.

Let the number of current smokers at a particular age and time be represented by C, the number of former smokers by F. The number of ever smokers, those who are either current or former smokers, is $E = C + F$.

The population of ever smokers has the special property that there are no decrements except through death--one can not go back to being a never smoker. Therefore, if the population of ever smokers is analyzed by the methods described above, we would expect λ_E to be greater than or equal to zero for all ages.

Changes in the population of current smokers reflect both transitions from never smoker to current smoker and net transitions from current smoker to former smoker. (This analysis assumes that, in a given time interval, there will be few transitions from never to current to former smoker.) Symbolically, since

$$E = C + F \quad ,$$

the increments in the three populations are related as

$$\Delta E = \Delta C + \Delta F \quad .$$

The net quit rate is the increment in the number of former smokers in a small interval divided by the number of current smokers $\Delta F/C$ which equals

$$\frac{\Delta E - \Delta C}{C} = \frac{\Delta C}{C} - \frac{\Delta E}{E} \cdot \frac{E}{C} .$$

Since $\frac{\Delta E}{E}$ is the start rate λ_E, as calculated above, and we can calculate $\lambda_C = \frac{\Delta C}{C}$ from the data for the population of current smokers, the net quit rate can be calculated as

$$\lambda_E(\frac{E}{C}) - \lambda_C .$$

Unfortunately, even multiple observations of the proportion of current and former smokers are not sufficient for estimating the number of quitters and recidivists. Repeated observations on individuals or at least retrospective reports of previous smoking behavior would be needed to estimate these gross transition rates.

Testing the Method

Table 1 shows the proportion of the population that are current or ever smokers, by sex and 5-year age groups, based on identical questions in the 1976 and 1981 censuses of New Zealand. With this data, the age-specific growth rates of each of these populations can be calculated as follows.

Let ${}_5P_{a,i}$ be the proportion of the population aged a to $a+5$ at time t_i who are smokers. The ratio of two adjacent groups is simply

$$\frac{{}_5P_{a+5,2}}{{}_5P_{a,1}} = \frac{{}_5K_{a+5}}{{}_5K_a} \left[\frac{{}_5L^S_{a+5}}{{}_5L^S_a} \cdot \frac{{}_5L_a}{{}_5L_{a+5}} \right] . \tag{4}$$

The procedures described above for correcting for differential mortality and translating ${}_5K_{a+5}/{}_5K_a$ into an instantaneous rate yields the results in Table 2. The data for making the differential mortality correction, and the sensitivity of the results to the choice, will be discussed below.

Table 2 and Figures 1 and 2 present the results using the 5-year data, but employing the method for regular age groups developed in the preceeding section. The correction for differential mortality is exactly the same. The results are quite close, as they are based on essentially the same data. Note that in this and the following tables, the data for ages 20 and 75 are not plotted so that the central values can be seen more clearly. When the surveys are separated by n

Table 1. Current and ever smokers as a percentage of the population, New Zealand, 1976 and 1981.

	Male				Female			
	Current		Ever		Current		Ever	
Age	'76	'81	'76	'81	'76	'81	'76	'81
15-19	29.8	26.9	35.3	32.3	30.4	29.9	36.2	36.3
20-24	41.8	39.5	52.2	50.1	38.9	40.3	49.3	52.2
25-29	42.5	38.4	57.4	54.2	37.8	34.6	50;3	49.6
30-34	44.0	37.1	62.0	57.0	37.8	32.8	50.6	48.9
35-39	43.3	38.4	63.4	61.3	35.0	33.2	46.9	49.3
40-44	43.8	37.5	67.0	62.3	35.3	31.0	47.4	45.3
45-49	45.0	37.5	71.8	65.8	35.8	31.1	48.5	46.0
50-54	44.0	38.0	75.6	70.4	34.1	30.7	47.8	46.4
55-59	40.6	35.9	75.1	74.4	30.6	28.1	45.2	45.5
60-64	38.3	31.9	74.4	73.9	26.0	24.5	40.7	42.6
65-69	34.8	28.6	74.5	72.6	20.2	19.9	34.8	37.9
70-74	30.7	25.3	72.6	71.1	14.5	14.7	26.9	31.4
75-79	26.9	19.8	69.3	66.7	10.7	7.7	21.3	19.5
80+	20.4		61.3		5.6		13.4	

Note: The last entry in the 1981 columns represents 75+

Source: Department of Statistics, 1979, 1983

Table 2. Net quit rates from exact and 5-year regular age group methods expressed as a percentage.

	Male		Female	
Age	Exact	Regular	Exact	Regular
20	-5.64	-5.62	-5.64	-5.64
25	1.70	1.68	2.34	2.33
30	2.72	2.67	2.84	2.81
35	2.70	2.74	2.59	2.63
40	2.82	2.80	2.42	2.38
45	3.01	2.99	2.50	2.50
50	3.21	3.24	3.01	3.03
55	3.81	3.83	3.77	3.78
60	4.40	4.37	4.28	4.30
65	5.22	5.22	5.01	5.03
70	5.37	5.42	5.63	5.68
75	7.31	7.22	11.69	11.46

Fig. 1 – Five year age groups

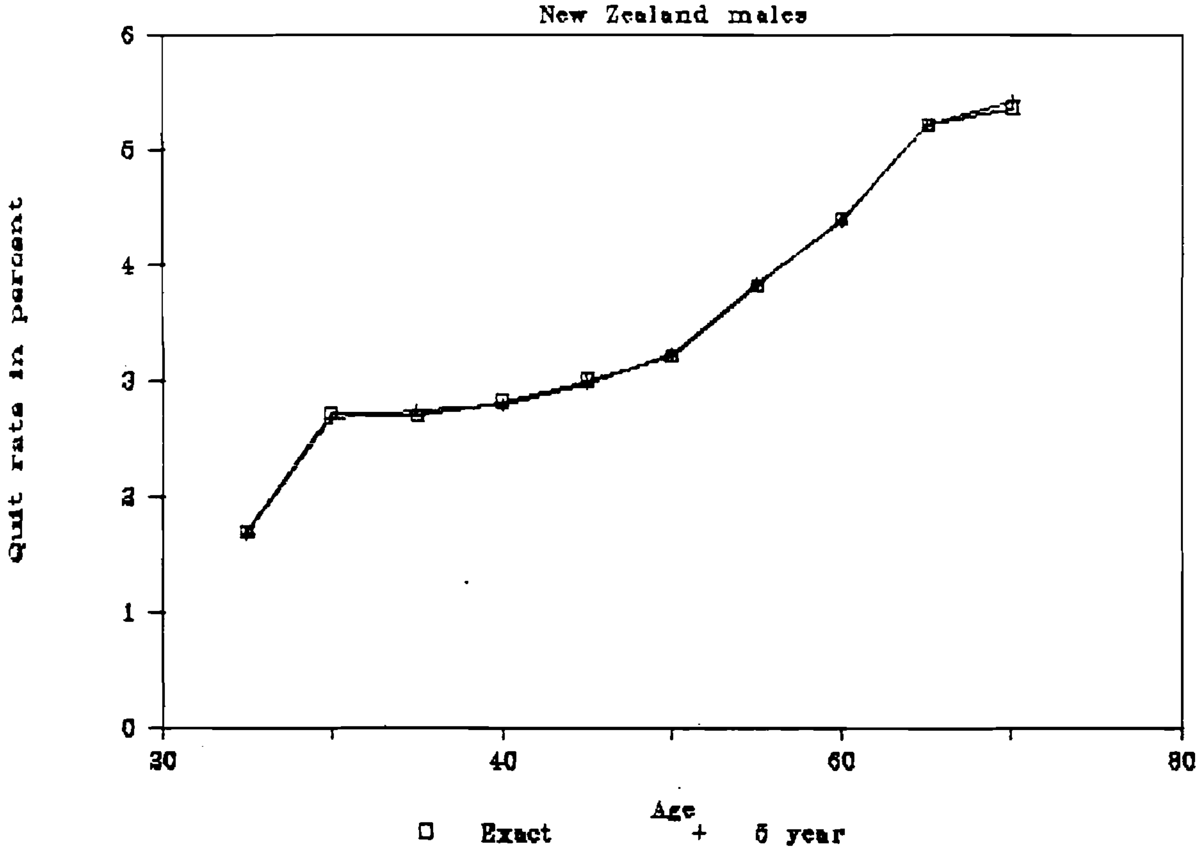
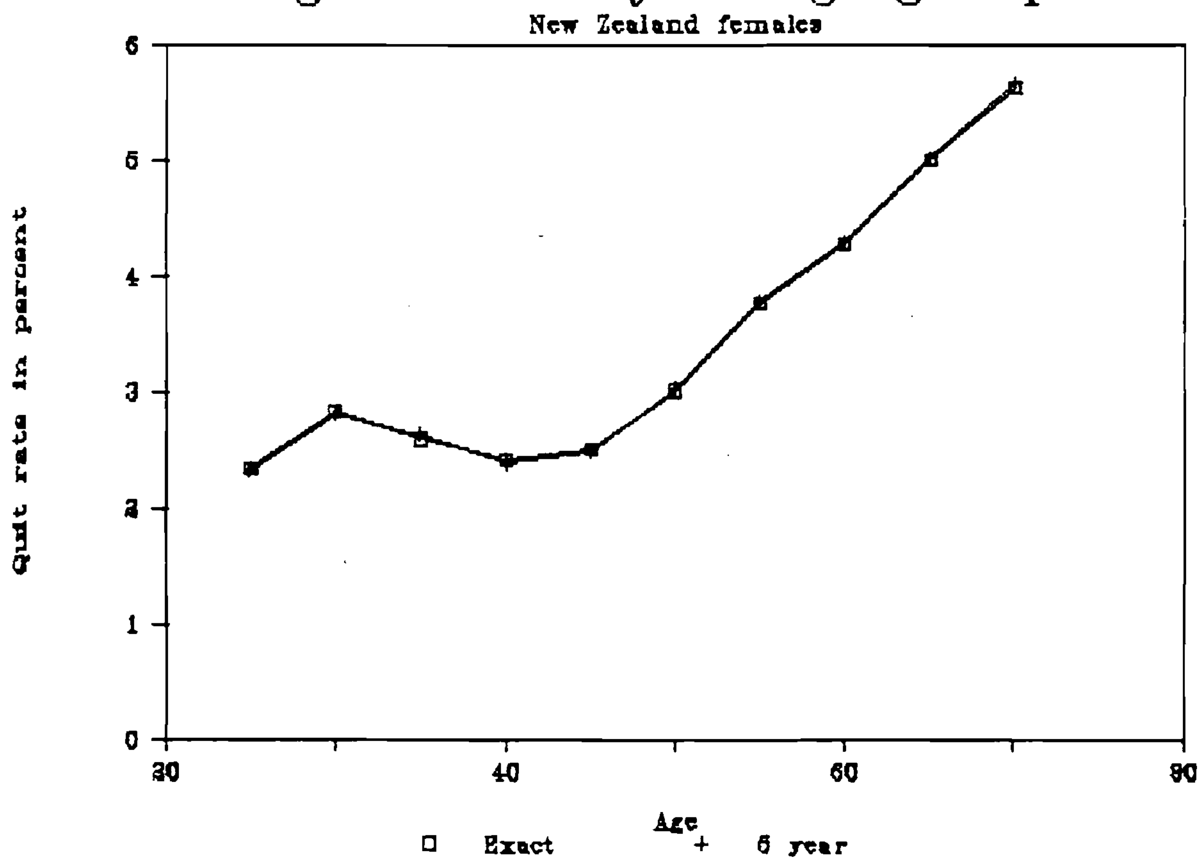


Fig. 2 – Five year age groups



years, the only difference between the simple method and the regular age group method is that

$$\frac{{}_5P_{a+5,2}}{{}_5P_{a,1}} \tag{5}$$

is replaced by

$$\frac{{}_5P_{a+5,1} + {}_5P_{a+5,2}}{{}_5P_{a,1} + {}_5P_{a,2}} \sqrt{\frac{{}_5P_{a+5,2}}{{}_5P_{a+5,1}} \cdot \frac{{}_5P_{a,2}}{{}_5P_{a,1}}} \; . \tag{6}$$

If the data are smooth, as they seem to be in New Zealand, the two will be quite close.

Table 3 and Figures 3 and 4 present the results of aggregating the data into 10 year age groups (weighting by the population in the 5 year age groups) and applying the same method. The transition rates from the 10 year-regular age method apply to the middle of a 20-year age interval, so are compared to the average of the two adjacent 5-year estimates. Remembering that the 10-year calculation uses only about half as much data does the 5-year calculation, the closeness of the two estimates is remarkable. The major difference between the simple and regular age group method appears, where we would expect it, at ages 15 to 25 where smoking behavior is changing rapidly. Together, the 5- and 10-year comparisons tell us that the method for regular age groups does not introduce any serious distortions.

Table 3. Net quit rates from exact and 10-year regular age group methods expressed as a percentage.

	Male		Female	
Age	Exact	Regular	Exact	Regular
25	-1.97	0.35	-1.65	0.65
35	2.71	2.60	2.71	2.61
45	2.92	2.81	2.46	2.29
55	3.51	3.99	3.39	3.71
65	4.81	4.73	4.65	4.72
75	6.34	5.59	8.66	7.73

Table 4 and Figures 5 and 6 show the results of the irregular age group methods for the same age grouping as is used in the United States HIS tabulations. Here the results are almost as good as before, differing from the exact results in only one case by more than one percent point. Given the difference

Fig. 3 – Ten year age groups

New Zealand males

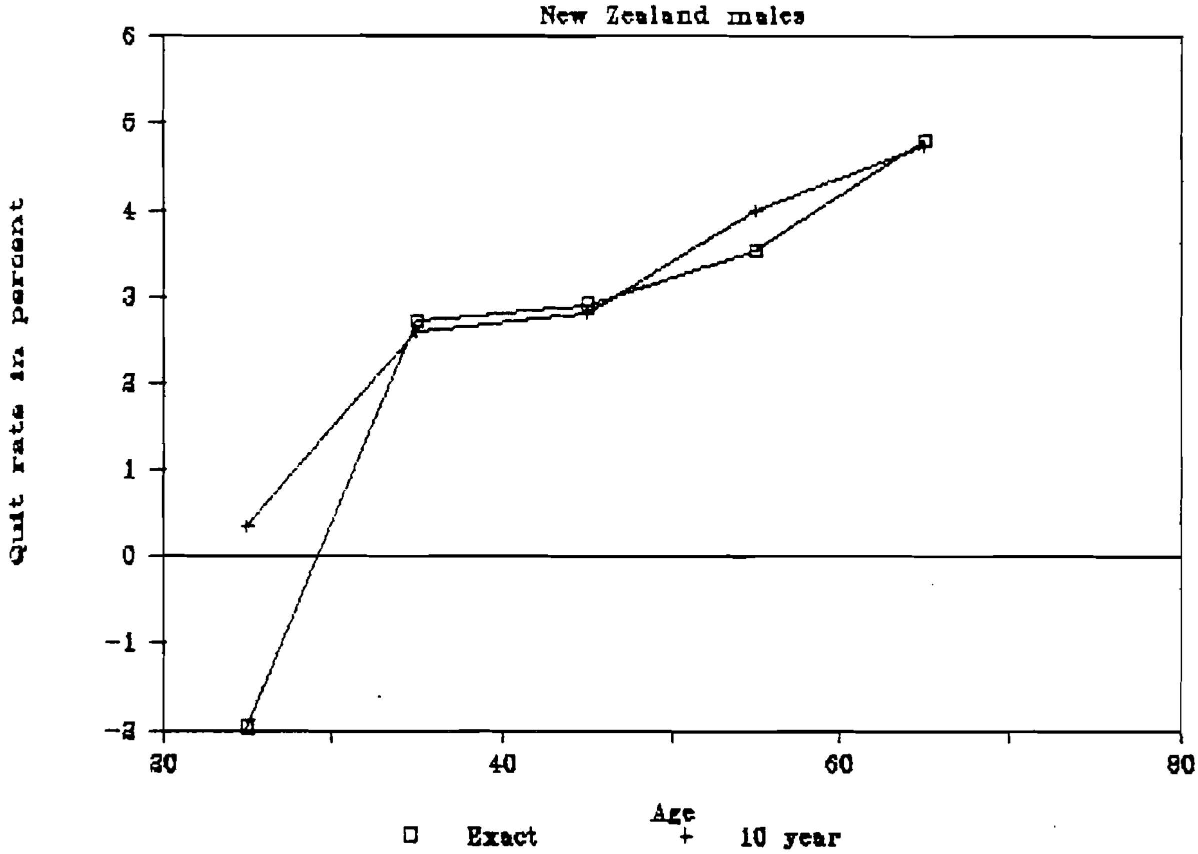
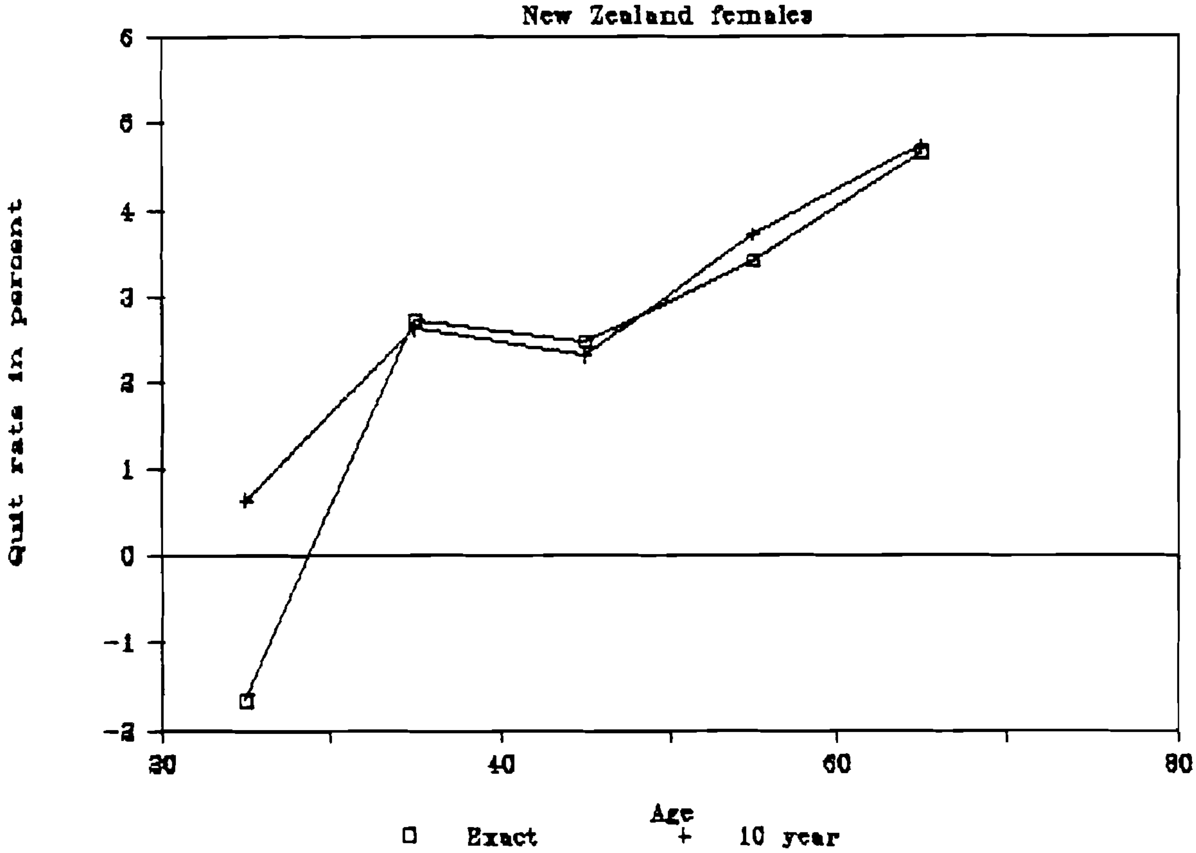


Fig. 4 – Ten year age groups

New Zealand females



in the number of data points used--5 versus 13 age groups-- and the lack of obvious bias in the results, the methods seems to perform reasonably well.

Table 4. Net quit rates from exact and irregular age group methods expressed as a percentage.

	Male		Female	
Age	Exact	Irregular	Exact	Irregular
27.5	2.21	1.98	2.59	2.18
35.0	2.70	2.66	2.59	2.64
50.0	3.21	3.50	3.01	2.72
65.0	4.81	5.76	4.65	4.47

Table 5 and Figures 7 and 8 show the sensitivity of the estimation procedure to the assumptions used in the correction for differential mortality. For the differential mortality calculations, I used Coale and Demeny (1966) West model life tables with male and female life expectancies of 75 and 71 years respectively (level 23) to obtain the ${}_5q_a$ values for the differential mortality correction. Figures 7 and 8 shows the effect of assumming mortality to be about 2.5 years higher or lower (one level in the Coale and Demeny tables), denoted by "1-22", "1-23", and "1-24". The change is hardly noticeable at the lower ages and not serious at the older ages. Since the effect of changing the level of mortality is so slight, the simpler use of the easily available Coale-Demeny lifetables rather than the actual lifetable seems justified.

For the previous calculations, I used the following age-specific differential mortality ratios for smokers compared to non-smokers, based on the American Cancer Society's twelve-year follow-up study of more than one million volunteers classified as smokers or non-smokers at the beginning fo the study (Hammond, 1966, cited in DHEW, 1979, 1980):

	Mortality Ratio	
Age	Male	Female
20-35	1.00	1.00
35-44	1.82	1.12
45-54	2.20	1.31
55-64	1.86	1.27
65-74	1.58	1.31
75-84	1.35	1.14

Fig. 5 – Irregular age groups

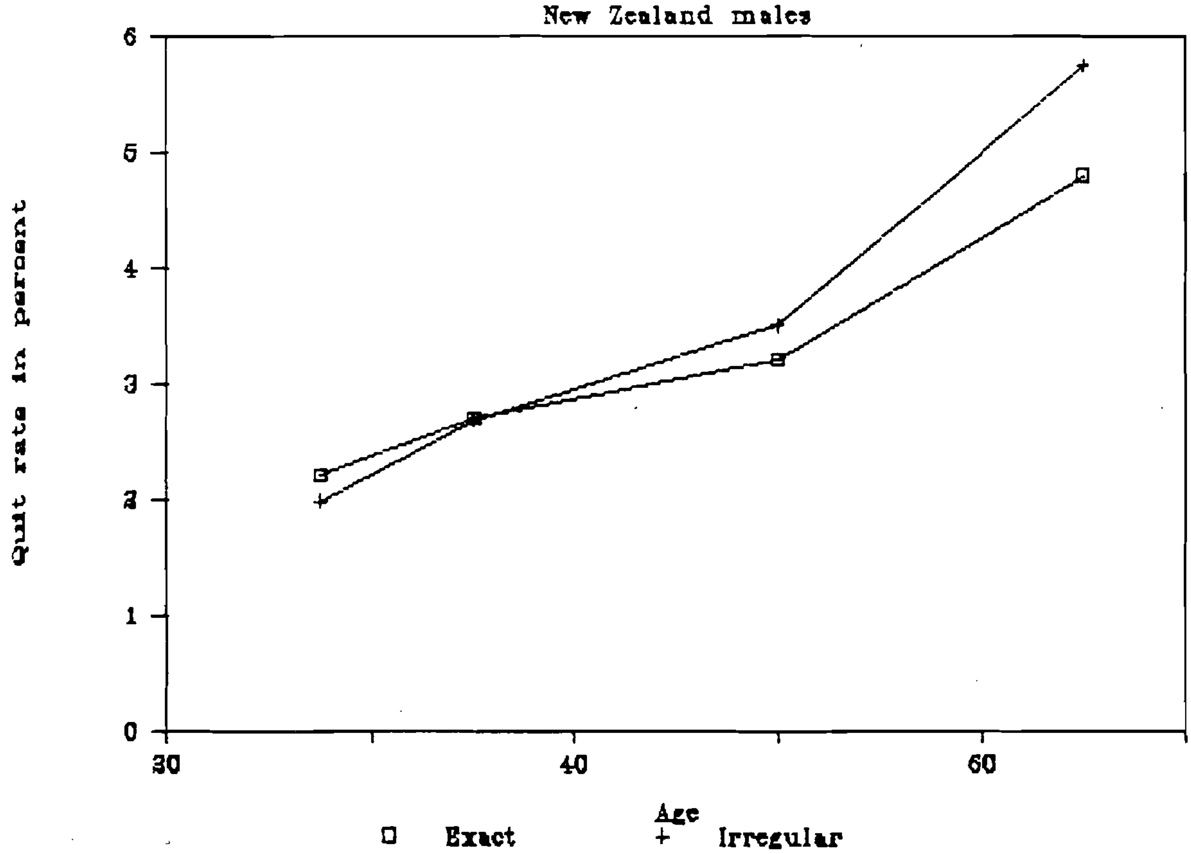


Fig. 6 – Irregular age groups

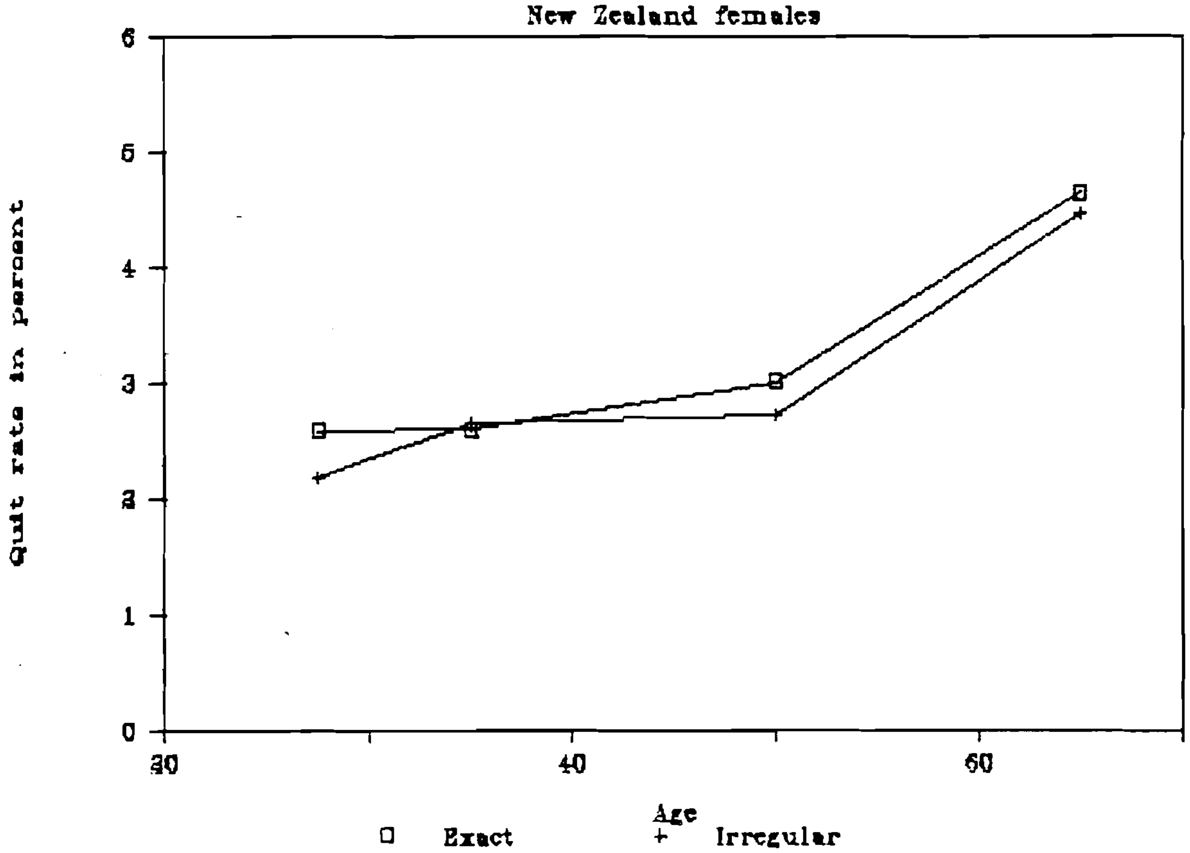


Table 5. Sensitivity to differential mortality assumptions. Net quit rate expressed as a percentage.

	Male: Model life table level			Female: Model life table level		
Age	22	23	24	22	23	24
27.5	1.99	1.98	1.98	2.18	2.18	2.18
35.0	2.68	2.66	2.65	2.64	2.64	2.64
50.0	3.56	3.50	3.44	2.74	2.72	2.71
62.5	5.86	5.76	5.64	4.52	4.47	4.41

	Male: Differential mortality			Female: Differential mortality		
Age	1	2	3	1	2	3
27.5	1.98	2.01	2.06	2.18	2.24	2.24
35.0	2.66	2.68	2.69	2.64	2.68	2.66
50.0	3.50	3.43	3.50	2.72	2.76	2.72
62.5	5.76	5.46	5.76	4.47	4.27	4.47

Other studies, however, yield different mortality ratios. The same reports of the Surgeon General (DHEW, 1979, 1980), for instance, cite studies of U.S. veterans (Rogot, 1974) and Swedish women (Cederlof et al., 1975) with the following mortality ratios:

Male Age	Mortality Ratio	Female Age	Mortality Ratio
20-35	1.00	1.00	
30-34	1.52	18-39	1.8
35-44	1.95	40-49	1.9
45-54	1.83	50-59	1.3
55-64	1.53	60-69	1.1
65-74	1.32		

Figures 7 and 8 show the results for the New Zealand data under three different assumptions about mortality ratios. The first calculation, "1-23", uses the ACS estimates that have been used all along. The second calculation, "2-23", uses the alternatives shown above. The third calculation, "3-23", uses the ACS estimates for ages 35 and above, but assumes that the mortality ratio is 2.0 for ages 15 to 34. Such a high mortality ratio could not be caused by smoking, but as the Swedish study and others suggest, would simply reflect differences between the smoking and non-smoking population with regard to other risky behaviors. The results show that the specific correction for differential mortali-

Fig. 7 – Sensitivity analysis

New Zealand males

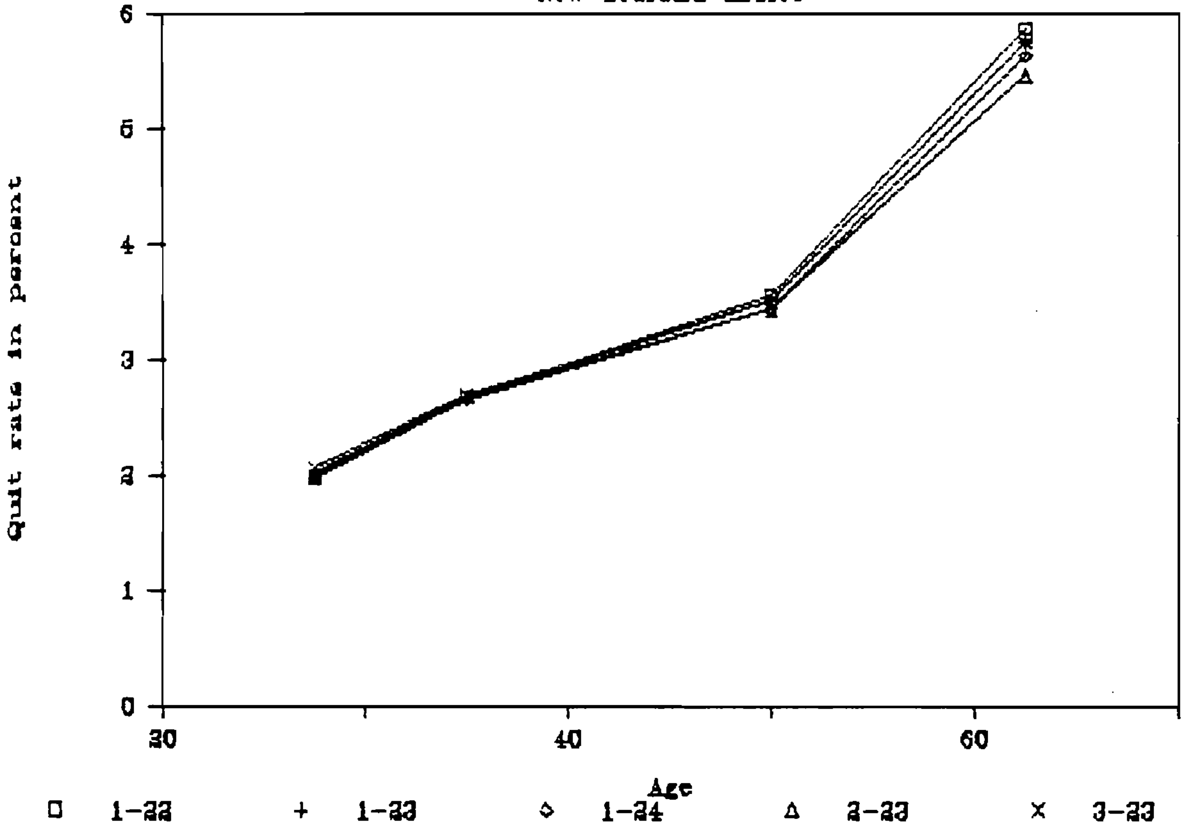
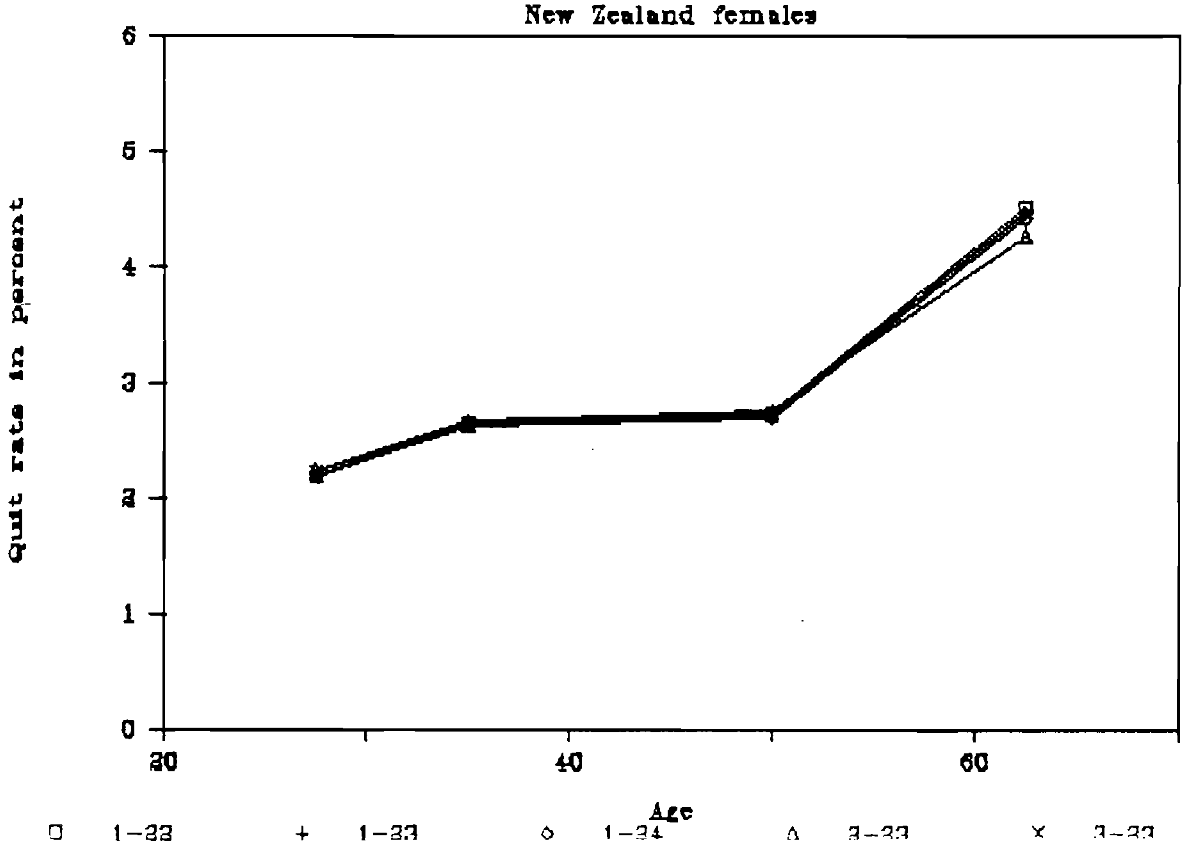


Fig. 8 – Sensitivity analysis

New Zealand females



ty matters only in the older ages, and even at that, not very much.

More detailed data on differential mortality ratios by amount of cigarettes smoked is available, but is not easily applied to the data at hand. A more sophisticated analysis might take this into account, but as the results are not especially sensitive to the differential mortality assumptions, this is not likely to matter very much.

Tables 6 and 7 differentiate between net changes in the proportion of current smokers and changes in the proportion of ever smokers, using the methods described in the previous section. The columns labeled dC/C represent the net rate of change in the population of smokers and are the same numbers that have been discussed up to this point. The columns labeled dE/E are calculated by applying the same method to the population of ever smokers. For the calculations in Table 6, I assumed that the differential mortality between former smokers and non-smokers was simply one-half that for smokers and non-smokers of the same age and sex. For Table 7, I made the simpler but nearly equivalent assumption that there was no differential mortality between ever smokers and the population at large.

Table 6. Adjusted quit rates from 5-year regular age group methods expressed as a percentage.

	Male			Female		
Age	dC/C	dE/E	Quit rate	dC/C	dE/E	Quit rate
20	-5.62	-6.99	2.71	-5.64	-7.33	3.17
25	1.68	-0.76	2.63	2.33	-0.11	2.47
30	2.67	0.13	2.67	2.81	0.56	2.81
35	2.74	0.23	2.74	2.63	0.51	2.63
40	2.80	0.31	2.80	2.38	0.69	2.38
45	2.99	0.30	2.99	2.50	0.57	2.50
50	3.24	0.31	3.24	3.03	0.84	3.03
55	3.83	0.22	3.83	3.78	0.92	3.78
60	4.37	0.15	4.37	4.30	1.07	4.30
65	5.22	0.24	5.22	5.03	1.19	5.03
70	5.42	0.56	5.42	5.68	1.55	5.68
75	7.22	1.14	7.22	11.46	5.84	11.46

Analysis of the dE/E columns show that the only substantial increments to the ever smoker category occur around age 20 and, to a smaller extent, age 25. At other ages, the increments are generally small but negative. The negative results could be due to an improper differential mortality correction or to former smokers gradually forgetting or refusing to say that they were once smokers.

Table 7. Adjusted quit rates from irregular age group methods expressed as a percentage.

Age	Male			Female		
	dC/C	dE/E	Quit rate	dC/C	dE/E	Quit rate
27.5	-1.98	0.46	2.59	-2.18	0.09	2.31
35.0	-2.66	-0.21	2.66	-2.64	-0.70	2.64
50.0	-3.50	-0.09	3.50	-2.72	-0.32	2.72
62.5	-5.76	-0.67	5.76	-4.47	-0.79	4.47

Since these negative values are small compared to the changes in the current smoker category they are not of much consequence to the final results, and independent studies suggest that smokers start before age 25 or 30, I calculated the final "adjusted quit rates" by using dC/C for ages 30 and above, and both dC/C and dE/E for younger ages.

Figure 9 shows the final adjusted quit rates for New Zealand males and females between 1976 and 1981. For ages 35 and above, the rates are quite similar, and increase substantially with age. At age 27.5, the net rate of change in the number of smokers is about the same for men and women, but men are still starting to smoke and women are not. The result is that the adjusted quit rate for young men is substantially higher than that for young women.

For purposes of comparison, Figure 10 shows the same rates for the United States in roughly the same period: 1976 to 1983. The smoking data are from *Health U.S.* (DHHS, 1984), and I used level 23 West model life table for the men and level 24 West for the women. The adjusted quit rates for U.S. males are roughly similar to those for New Zealand males, except that they are lower in the oldest age interval. Adjusted quit rates for U.S. females are similar at older ages but substantially lower at younger ages.

DISCUSSION

Although I developed it for and tested it on smoking data, the technique in this paper has general applications in demographic work. It applies to data on proportions of the population in two or more subgroups by regular or irregular age intervals at two points in time. It is especially valuable when the age groups are irregular and the spacing between the surveys in different that the age interval in the data. The technique can be used, for instance, to study changes in the proportions of married and single, employed and unemployed, healthy and sick, and so on. It can also be used to study more than two popula-

Fig. 9 - Adjusted quit rates

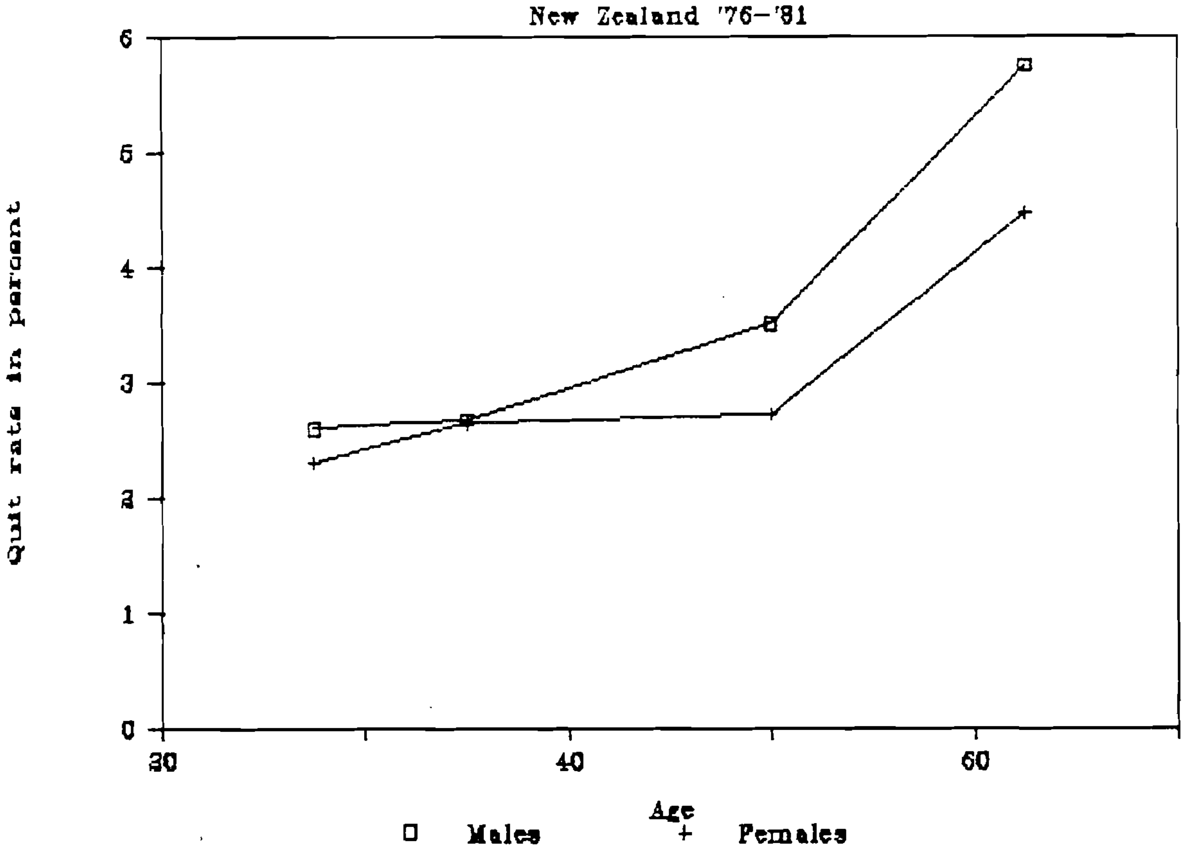
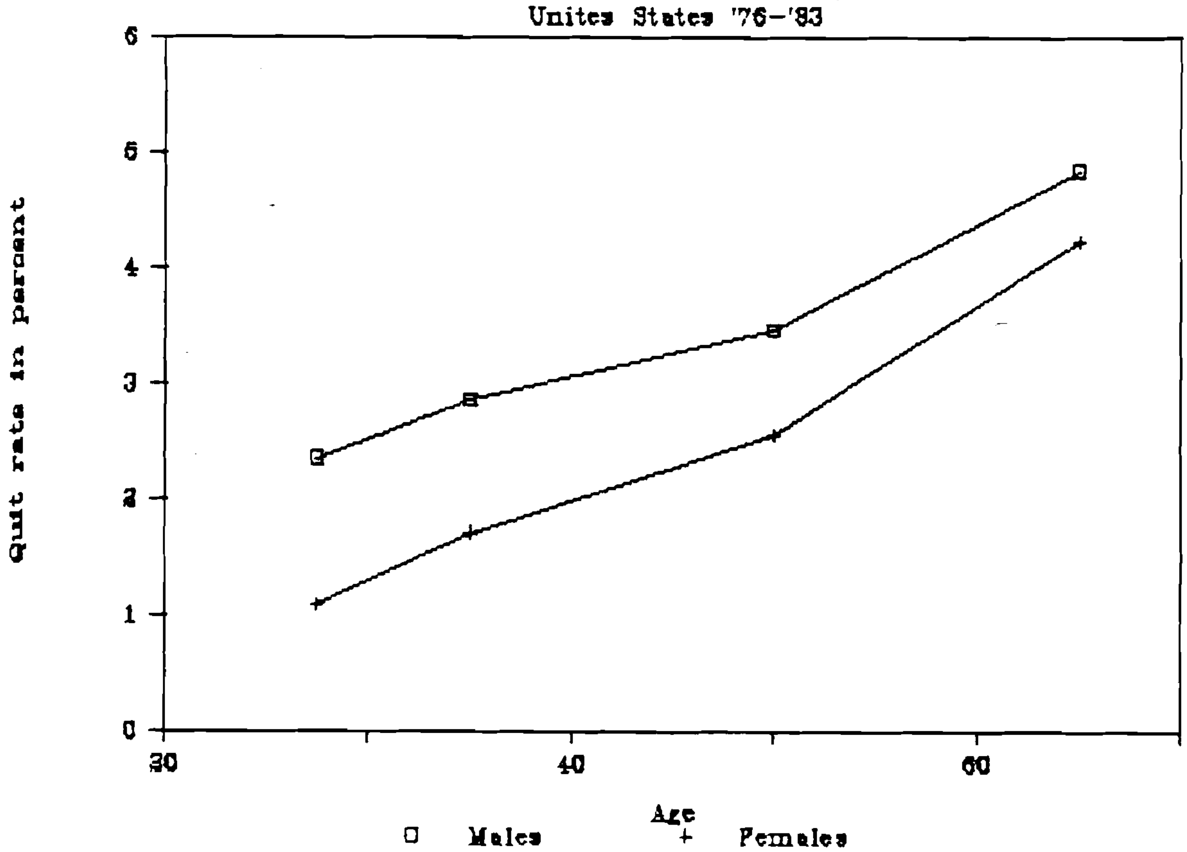


Fig. 10 - Adjusted quit rates



tion groups, such as the proportion of light, medium, and heavy smokers (Stoto, 1985). But because there is no information on the number of transitions in particular directions, the method can only be used to estimate net transition rates. In special circumstances, like the transition to being an ever smoker, more information can be obtained.

The main assumption of the method is that the age-specific growth rates of the population subgroups are approximately constant within the age intervals dictated by the data. As the tests on the New Zealand data show, this is often a reasonable assumption. Of course, the smaller the age intervals, the better the approximation. The technique also relies on independent estimates of differential mortality, but the results are not overly sensitive to these estimates.

Because equations (1), (2) and (3) involve only relative proportions, the technique avoids certain problems. If smoking is systematically under-reported by the same proportion at all ages and at both points in time, for instance, the estimates are still correct. If there is emigration or immigration, but the migrants and residents have similar smoking habits, the estimates are correct. As long as the proportion of smokers and non- smokers is consistent, the surveys can even have different levels of population coverage. But if smoking is more likely to be reported in one survey than another, for instance, the results will be systematically biased.

Furthermore, since the method relies on rates of change rather than absolute levels, it may be possible to compare data from different sources that use different definitions of smoking status. If, for instance, one country's definition of a smoker includes occasional smokers but another's does not, and occasional smokers are a constant but unknown fraction of all smokers, the results will be comparable.

The method developed here can be generalized in a variety of ways. The Preston and Coale (1982) result applies to period data, but Arthur and Vaupel (1984) show that similar results can be obtained for cohorts and other data gathering frames. These may enable us to combine information about growth rates over time from one study with information about relative proportions smoking at different ages from another study and yield a more accurate result than would be possible from either study alone. In the United States, for instance, there are three series of surveys with information on the smoking behavior of teenagers, each with different sampling schemes, age categories, and definitions of smoking (Massey et al., 1984). Combining information about growth rates for one age group from an annual survey of high school seniors with an accurate estimate of the relative proportions smoking at various ages from another survey of all children, for instance, might provide more accurate results than could be obtained from either survey alone.

REFERENCES

Arthur, W. Brian and James W. Vaupel (1984) Some General Relationships in Population Dynamics. *Population Index* 50(2):214-226.

Cederlof, R., L. Freiberg, Z. Hrubec, and U. Lorich (1975) *The Relationship of Smoking and Some Social Covariates to Mortality and Cancer Mortality. A Ten Year Follow-up in a Probability Sample of 55,000 Swedish Subjects Age 18 to 69.* Part I and II. Stockholm: Karolinska Institute.

Durch, Jane (1985) National Surveys on Smoking Behavior with Demographic Characteristics. Institute for the Study of Smoking Behavior and Policy, Harvard University. September.

Hammond, E. C. (1966) Smoking in Relation to Death Rates of One Million Men and Women. Pages 127-204 in *Epidemiological Approaches to the Study of Cancer and Other Chronic Diseases*, edited by W. Haenszel. National Cancer Institute Monograph 19, U.S. Department of Health, Education, and Welfare, National Cancer Institute.

Harris, Jeffrey E. (1983) Cigarette Smoking Among Successive Birth Cohorts of Men and Women in the United States During 1900-80. *Journal of the National Cancer Institute* 71(3):473-479.

Hay, D.R. Hay and F.H. Foster (1984) Intercensal Trends in Cigarette Smoking in New Zealand 1: Age, Sex, and Ethnic Status. *The New Zealand Medical Journal* 97(752):283-5 (May).

Massey, Marilyn M., Gayle Boyd, Margaret Mattson, and Marcia R. Feinleib (1984) Inventory of Surveys on Smoking. Draft, National Cancer Institute, October.

Preston, Samuel H. and Ansley J. Coale (1982) Age Structure, Growth, and Accession: A New Synthesis. *Population Index* 48(2):217-59.

Rogot, E. (1974) Smoking and Mortality Among U.S. Veterans. *Journal of Chronic Diseases* 27:189-203.

Stoto, Michael A. (1985) *Changes in Adult Smoking Behavior in the United States: 1955 to 1983.* Laxenburg, Austria: International Institute for Applied Systems Analysis. (Forthcoming Working Paper)

U.S. Department of Health Education and Welfare (DHEW) (1979) *Smoking and Health: A Report of the Surgeon General.* Office of the Assistant Secretary for Health, Washington D.C.

U.S. Department of Health Education and Welfare (DHEW) (1980) *The Health Consequences of Smoking for Women: A Report of the Surgeon General.* Office of the Assistant Secretary for Health, Washington D.C.

U.S. Department of Health and Human Services (DHHS) (1984) *Health U. S. 1984.* National Center for Health Statistics, Washington, D.C.