



Use of Transition Probabilities to Estimate the Effect of Smoking on the Duration of Episodes of Respiratory Symptoms in Diary Data

The Swiss Study on Air Pollution and Lung Diseases in Adults (SAPALDIA)

Reinhard Kaiser,¹ Christian Schindler,¹ Nino Künzli,¹ Ursula Ackermann-Lieblich,¹ Dominik Heeb,¹
Tullio C. Medici,² Jean Pierre Zellweger,³ and the SAPALDIA Team

Incompletely documented symptom episodes pose methodological problems in the analysis of diary data. The aim of this study was to develop a method of estimating the average durations of symptomatic and nonsymptomatic episodes, respectively, coping with the problem of bias due to undocumented days and censored episodes that is found in most diary studies. The authors derived their outcome variables from a Markov model using transition probabilities. To evaluate this method, the authors assessed the impact of active smoking on the duration of episodes of bronchitis symptoms and the corresponding nonsymptomatic periods, respectively, using diary data (1992–1993) obtained from 801 participants in the Swiss Study on Air Pollution and Lung Diseases in Adults. Covariate-adjusted distribution curves for the mean durations of individual episodes were estimated by Cox regression. Median values for light smokers (<10 cigarettes/day) were 60.0 symptom-free days (95% confidence interval (CI) 42.0–78.5) and 4.0 symptomatic days (95% CI 3.0–6.0), respectively, compared with medians of only 21.0 days (95% CI 16.2–29.8) for periods without bronchitis symptoms and 6.0 days (95% CI 4.9–9.0) for episodes of bronchitis symptoms in heavy smokers (≥ 30 cigarettes/day). The authors suggest that the Markov method is a feasible approach to the assessment of long term effects of smoking and environmental risk factors on the average duration of symptomatic and nonsymptomatic respiratory episodes. *Am J Epidemiol* 1998;148:600–8.

data collection; epidemiologic methods; longitudinal studies; Markov chains; respiration; smoking

Diaries in which study participants record respiratory symptoms are known to be valuable data collection instruments in air pollution epidemiology. Such data normally consist of sets of sequences of binary outcomes, one for each symptom and participant. Short to medium term effects of air pollution on health endpoints such as symptom incidence and prevalence (1–10) and, more recently, symptom duration (11, 12) have been a major focus of diary studies. However, the use of diary data for assessment of long term associations between exposure to environmental risk factors and individual respiratory health has gained less attention.

One approach is to focus on individual patterns of symptoms instead of summarizing these data over

time, as is done in time series analyses (13). The underlying hypothesis is that living in areas with high concentrations of air pollutants results in increased symptom prevalence, incidence (i.e., shorter nonsymptomatic periods), or duration of symptoms. The assessment of symptom duration poses special methodological problems because of the presence of days with missing information and the limitation of observation periods by design. If the duration of a symptomatic episode is defined as the number of consecutive days of reported symptoms (12), this implies that the episode is preceded and followed by at least one day without symptoms. Using this naive approach, a symptomatic interval cannot be counted as a single episode if there is even one day with missing information.

Generally, there are two conflicting goals in diary studies: On the one hand, study periods should be long enough to estimate the duration of symptomatic episodes accurately, and the total number of days with reported data should be large enough to detect differences in the duration of symptoms between different categories of subjects. On the other hand, the compliance of most subjects decreases with the length of the

Received for publication June 9, 1997, and accepted for publication February 19, 1998.

Abbreviations: CI, confidence interval; FEV₁, forced expiratory volume in 1 second; SAPALDIA, Swiss Study on Air Pollution and Lung Diseases in Adults.

¹ Institute for Social and Preventive Medicine, University of Basel, Basel, Switzerland.

² Division of Pneumology, University of Zürich, Zurich, Switzerland.

³ Division of Pneumology, University of Lausanne, Lausanne, Switzerland.

diary periods and the total number of diary days. In addition to the general problem of incompletely observed episodes at the beginning and end of the observation period, a study may be designed in such a way that subjects fill in the diaries during several intermittent time periods, which results in discontinuous reporting.

In the Swiss Study on Air Pollution and Lung Diseases in Adults (SAPALDIA), participants in the follow-up study were asked to fill in diaries during six periods of 4 weeks spread out over 2 years. To cope with the problems of missing data for some days and discontinuous reporting, we used transition probabilities to estimate the effects of ambient pollution on the durations of symptomatic episodes and corresponding symptom-free periods (hereafter called nonsymptomatic episodes). Here we demonstrate the advantages and disadvantages of such an approach using the effects of cigarette smoking on the individual average durations of episodes with and without bronchitis symptoms.

MATERIALS AND METHODS

Study design

SAPALDIA consisted of a multicenter, population-based cross-sectional investigation (1991) with a diary study as a longitudinal component (1992–1993). It was designed to evaluate associations between air pollution and respiratory symptoms and diseases. A detailed description of the cross-sectional study has been given elsewhere (14). The eight study areas (Aarau, Basel, Davos, Geneva, Lugano, Montana, Payerne, and Wald) represented a wide range of urbanization, altitude, air pollution, and meteorologic conditions in Switzerland. A random sample of adults aged 18–60 years was drawn from registries of the inhabitants of these areas. Successfully recruited subjects (59 percent) were interviewed with an extended version of the European Respiratory Survey Questionnaire and underwent allergy and lung function tests (15).

In the longitudinal portion of the study, subjects with a higher probability of reporting respiratory symptoms were oversampled to enhance the study's statistical power to detect associations between air pollution and lung diseases. Accordingly, meeting one of the following criteria assessed in the cross-sectional study was sufficient for a subject to be eligible for participation in the diary study: 1) an affirmative answer to at least one of a set of 10 questions on respiratory symptoms; 2) a ratio between forced expiratory volume in 1 second (FEV_1) and forced vital capacity that was less than 80 percent of the value

predicted by European Community Coal and Steel norms; 3) a decline in FEV_1 of at least 20 percent from baseline during the bronchochallenge test; and 4) satisfying none of the above criteria and being a non-smoker. All subjects fulfilling at least one of the first three criteria were invited to participate. The control cohort, those fulfilling the fourth criterion, was restricted to a random sample of 150 nonsymptomatic nonsmokers per area. Thus, all smokers in the follow-up sample were either symptomatic or hyperreactive.

In each study area, the subjects invited to participate were randomly assigned to 16 groups of equal size. The 16 groups were enrolled in the study at intervals of 1 week to contribute diaries kept for 4 consecutive weeks. Each diary period was followed by a break of 12 weeks; thus, participants could complete up to six diary periods of 4 weeks during the 2 years of the follow-up study. Therefore, in each week of the follow-up study (except the weeks at the beginning or the end), four groups were filling in diaries. Randomization of groups was done in a stratified way in order to guarantee that proportions between cohorts varied as little as possible from week to week. The participation rate for the diary study was 61 percent (3,281 of 5,383 persons).

During the diary completion periods, participants had to record daily data on morbidity parameters (respiratory symptoms, medication use, and peak expiratory flow) and other relevant information (time spent outdoors, absence from the study area, physical activity, and number of cigarettes smoked).

We restricted our analysis to cigarette smokers. A total of 801 follow-up participants who had been identified as current smokers in the cross-sectional study were still reported to be smokers during the diary study. The sample was divided into four categories by number of cigarettes smoked per day, based on cross-sectional study data (1–9, 10–19, 20–29, and ≥ 30 cigarettes/day).

Respiratory symptoms

The 1-week diaries were designed to be filled in daily in the morning and evening from Monday through Sunday. Information on some respiratory symptoms—wheezing, dyspnea, chest tightness, and cough—was requested both in the morning (for the preceding night) and in the evening. Other symptoms—phlegm, sore throat, hoarseness, running or blocked nose, itchy or irritated eyes, and fever—had to be reported in the evening only (for the entire 24-hour period). If no symptoms were reported either in the morning or at night, a response indicating “no symptoms experienced” had to be entered.

Because bronchitis is known to be strongly associated with smoking, bronchitis symptoms were selected for the analysis. They were defined as the presence of cough and/or phlegm during the day and/or at night.

Days with missing information

An absence of more than 8 hours from the study location had to be noted in the diary for that particular day. For instance, if the absence lasted for a whole weekend, the participant was asked to fill in the diary only in the morning before leaving and at night when he or she came back the next day. For longer absences, the participant was asked to leave out the whole diary week and to restart recording after coming back. Since the present investigation was not directed at short term effects, days with absences were not generally excluded from the analysis, but they were one source of missing information.

If diurnal symptom status was not reported, the day was declared symptom-free, provided that nocturnal symptoms were explicitly denied and that evening peak expiratory flow was measured. The slight underestimation of symptom prevalence resulting from this definition outweighs the opposite bias introduced by excluding such days from analysis. All other days with incomplete symptom reports were defined as "missing."

Statistical analysis

In a first step, we estimated transition probabilities—i.e., the probabilities of becoming symptomatic or asymptomatic, respectively, from one day to another—and the durations of symptomatic and nonsymptomatic episodes for each subject. This estimation was based on a Markov model. Secondly, we assessed the impact of cigarette smoking on these previously estimated outcome measures, using covariate-adjusted Cox regression to calculate confidence intervals for the median durations. Additionally, we compared the estimates obtained from the Markov model with the average durations of completely observed symptomatic episodes (naive method).

Outcome measures. Our outcome measures—i.e., estimates of the individual average durations of nonsymptomatic and symptomatic episodes, respectively—were derived from a Markov model describing day-to-day changes in symptom status. Since only the presence or absence of symptoms was recorded, a binary model was appropriate. For simplicity, we assumed the individual series to be stationary and transition probabilities to depend only on the symptom

status of the previous day. Our model was based on the following assumptions:

1. The probability q of getting a symptom on the next day is constant (i.e., it does not depend on the number of symptom-free days prior to the current day).
2. The probability p of becoming symptom-free on the next day is constant (i.e., it does not depend on the number of symptomatic days prior to the current day).

If this model holds true, then in the long run the percentage of symptomatic days is $q/(q + p)$, and the percentage of symptom-free days is $p/(q + p)$. Moreover, the mean length of symptom-free periods is $1/q$, and the mean length of symptomatic episodes is $1/p$. Now, q and p may be estimated as follows:

$$q' = \frac{n_{01}}{n_{01} + n_{00}}$$

and

$$p' = \frac{n_{10}}{n_{10} + n_{11}}$$

where $S = 0$ represents a symptom-free day and $S = 1$ represents a symptomatic day and

- n_{01} = number of transitions from $S = 0$ to $S = 1$;
- n_{00} = number of transitions from $S = 0$ to $S = 0$;
- n_{10} = number of transitions from $S = 1$ to $S = 0$;
- n_{11} = number of transitions from $S = 1$ to $S = 1$.

Under this model (i.e., of a first-order stationary binary Markov chain), the remaining number of episode days after the initial day of symptoms follows a geometric distribution with parameter p . Thus, the average duration of a symptomatic episode equals $1 + (1 - p)/p = 1/p = (n_{11} + n_{10})/n_{10}$ (16). Likewise, the average duration of a nonsymptomatic episode equals $1/q = (n_{00} + n_{01})/n_{01}$. Over a long and completely documented period with a sufficiently large number of symptomatic and nonsymptomatic episodes, respectively, one would almost certainly have $(n_{11} + n_{10})/n_{10} \approx$ (number of days within completely documented symptomatic episodes)/(number of completely documented symptomatic episodes). Therefore, asymptotically, our method of estimating the average duration of symptomatic episodes is equivalent to the naive method for subjects with variable symptom status in the ideal situation of complete documentation. By symmetry, the same also holds true for our estimates of the average duration of nonsymptomatic episodes. Since the naive method does not require any assumptions other than data completeness, this asymptotic

equivalence indicates that potential limitations of our method are not primarily related to the fact that it was derived from a simplistic model of symptom dynamics but rather are related to the assumption that missing days (days with missing data) are independent of symptom status.

For 56.7 percent of the subjects, transition probabilities could be estimated in this way. The remaining subjects were divided into the following categories: 1) subjects who reported only symptomatic days or only symptom-free days, 2) subjects for whom exactly one of the two transition probability estimates p' and q' was 0, although they reported both types of symptom status, and 3) subjects for whom both transition probability estimates p' and q' were 0.

If all documented days (days with documented symptom status) for a subject are symptom-free, then only a censored estimate for the average length of this subject's symptom-free periods is available—i.e., the number n_{00} of transitions between consecutive documented days. Of course, in such a case, no estimate for the average duration of symptomatic periods is available. The case in which no symptom-free days have been reported is completely symmetrical (i.e., with n_{11} replacing n_{00}). Approximately 27.7 percent of the estimates were censored in this way.

Table 1 shows the proportions of the different ways in which the estimates were obtained, stratified by smoking category. The proportions of censored estimates for symptomatic episodes were positively associated with increasing numbers of cigarettes smoked per day (test for trend: $p < 0.001$), and the proportions of censored estimates for nonsymptomatic episodes

were negatively associated with numbers of cigarettes smoked per day (test for trend: $p < 0.01$).

The following formula for the odds of reporting symptoms can be used to obtain estimates of p' and q' , respectively, in cases where missing data do not allow us to apply the original formulas:

$$f/(1-f) = \frac{q/(q+p)}{p/(q+p)} = \frac{q}{p},$$

where f denotes the individual frequency of symptoms. If $q' = 0$ and $p' > 0$, then this formula gives

$$q' = p' \frac{f}{(1-f)},$$

and if $q' > 0$ and $p' = 0$, then one obtains

$$p' = q' \frac{(1-f)}{f}.$$

Approximately 12.2 percent of the estimates were calculated in this way.

If q and p are 0 or missing and censored estimates are not available, the average durations of nonsymptomatic and symptomatic episodes cannot be calculated. This may happen if there are few changes in symptom status and none of them can be clearly located in time because of missing data. In 3.4 percent of the cases, estimates could not be computed. This percentage did not vary significantly between smoking categories (table 1).

Cox regression. In the present analysis, we assessed the effect of cigarette smoking on the average

TABLE 1. Availability of transition probabilities for bronchitis symptoms, SAPALDIA† study, Switzerland, 1992–1993

| | Smoking category (cigarettes/day) | | | | | | | |
|--------------------------------|-----------------------------------|-----------|--------------------|-----------|--------------------|-----------|------------------|-----------|
| | 1–9 (n = 136) | | 10–19 (n = 189) | | 20–29 (n = 285) | | ≥30 (n = 191) | |
| | % | 95% CI† | % | 95% CI | % | 95% CI | % | 95% CI |
| Censored estimates | | | | | | | | |
| Symptoms on all days*** | 0 | | 2.1 | 0.1–4.2 | 3.9 | 1.6–6.1 | 11.0 | 6.6–15.4 |
| No symptoms** | 27.2 | 19.7–34.7 | 26.5 | 20.2–32.7 | 23.9 | 18.9–28.8 | 16.2 | 11.0–21.5 |
| Noncensored estimates | | | | | | | | |
| q defined; p defined‡ | 55.1 | 46.8–63.5 | 55.6 | 48.5–62.6 | 56.1 | 50.4–61.9 | 59.7 | 52.7–66.6 |
| q defined; $p = 0$ § | 8.1 | 3.5–12.7 | 7.4 | 3.5–12.7 | 7.4 | 3.7–11.1 | 4.2 | 1.3–7.0 |
| $q = 0$; p defined¶ | 4.4 | 1.0–7.9 | 6.3 | 2.9–9.8 | 5.3 | 2.7–7.9 | 5.8 | 2.5–9.1 |
| Estimates not available | | | | | | | | |
| q missing; p missing# | 5.2 | 1.4–8.9 | 2.1 | 0.1–4.2 | 3.5 | 1.4–5.6 | 3.1 | 0.7–5.6 |

** $p < 0.01$; *** $p < 0.001$ (test for trend).

† SAPALDIA, Swiss Study on Air Pollution and Lung Diseases in Adults; CI, confidence interval.

‡ q = the probability of getting a symptom on the next day; p = the probability of losing a symptom on the next day.

§ p can be calculated by using q and the frequency of symptomatic days.

¶ q can be calculated by using p and the frequency of symptomatic days.

Both q and p cannot be calculated.

durations of symptomatic and nonsymptomatic episodes. Because of the presence of censored data (in subjects without changes in symptom status), Cox regression was used for this purpose. Multivariate analysis controlled for sex, age, age², body mass index (weight (kg)/height (m)²), study area, season (the proportion of winter days among all days with documented symptom status), maternal asthma history, paternal asthma history, sibling asthma history, and atopy (defined as a positive skin test). To obtain covariate-adjusted distribution curves of individual average durations for each of the four smoking categories (1–9, 10–19, 20–29, and ≥30 cigarettes/day), we used this latter factor as a stratum variable in the Cox regression model (i.e., a different baseline distribution function was estimated for each of these categories).

RESULTS

Table 2 shows the distributions of age and body mass index among males and females in the four smoking categories. Table 2 also gives the length of participation for each group and the prevalence of the factors for which we controlled in the analysis. Heavy smokers (≥30 cigarettes/day) tended to be older, to have a higher mean body mass index, and to have fewer documented days than light smokers (<10 cigarettes/day). However, all of these trends were relatively weak. A sibling history of asthma and a positive skin test were most frequent in light smokers. The high prevalence of these known predictors of respira-

tory symptoms suggests that they should be controlled for in a multivariate analysis.

The estimates of episode duration were highly skewed in all exposure categories. When compared with the durations of completely observed episodes (naive method) (figure 1), the estimated durations were considerably longer (median of 4.0 days vs. 2.3 days) and had a wider distribution (interquartile range of 5.8 days vs. 2.5 days) and a long tail, with a maximum of 82 days, as compared with 21 days when symptomatic episodes could not exceed the 28-day observation periods by definition. Most of the long durations of symptomatic episodes were censored estimates. Noncensored estimates correlated only weakly with length of participation (Pearson correlation coefficient = 0.09).

Figure 2 shows covariate-adjusted distribution curves for the individual average duration of episodes with bronchitis symptoms in the four smoking categories; figure 3 shows the corresponding distribution curves for the individual average duration of intervals without bronchitis symptoms. As expected, these figures display opposite trends with increasing numbers of cigarettes smoked. Periods of bronchitis symptoms tended to last longer and nonsymptomatic episodes tended to be shorter in heavy smokers. There was little difference between the categories of 10–19 cigarettes per day and 20–29 cigarettes per day regarding the average duration of symptomatic episodes, or between the categories of 1–9 cigarettes per day and 10–19 cigarettes per day regarding the average duration of

TABLE 2. Clinical characteristics, participation, and prevalence of control variables, SAPALDIA* study, Switzerland, 1992–1993

| | Smoking category (cigarettes/day) | | | | | | | | | | | |
|--------------------------------|-----------------------------------|-----------------|----------------------|--------------------|-----------------|--------------------|--------------------|-----------------|--------------------|------------------|-----------------|--------------------|
| | 1–9 (n = 136) | | | 10–19 (n = 189) | | | 20–29 (n = 285) | | | ≥30 (n = 191) | | |
| | No. | Mean or % | SD* or 95% CI* | No. | Mean or % | SD or 95% CI | No. | Mean or % | SD or 95% CI | No. | Mean or % | SD or 95% CI |
| Mean age (years) | | | | | | | | | | | | |
| Males | 50 | 39.4 | 11.4 | 80 | 41.1 | 11.8 | 159 | 42.0 | 10.0 | 140 | 42.3 | 9.4 |
| Females | 86 | 39.3 | 10.2 | 109 | 41.3 | 9.7 | 126 | 40.2 | 9.7 | 51 | 41.2 | 6.8 |
| Mean body mass index† | | | | | | | | | | | | |
| Males | | 23.7 | 2.6 | | 24.4 | 3.1 | | 24.8 | 3.6 | | 25.4 | 3.9 |
| Females | | 22.7 | 4.2 | | 22.1 | 2.9 | | 23.6 | 4.4 | | 23.2 | 3.9 |
| Participation‡ | | | | | | | | | | | | |
| Days | | 114.3 | 54.6 | | 116.4 | 54.9 | | 107.1 | 56.1 | | 107.6 | 57.0 |
| Periods | | 4.0 | 1.9 | | 4.1 | 1.9 | | 3.8 | 2.0 | | 3.8 | 2.0 |
| Season§ | | 0.51 | 0.23 | | 0.48 | 0.23 | | 0.50 | 0.25 | | 0.50 | 0.25 |
| History of maternal asthma (%) | | 4.4 | 1.0–7.9 | | 5.3 | 2.1–8.5 | | 3.9 | 1.6–6.1 | | 3.2 | 0.7–5.6 |
| History of paternal asthma (%) | | 7.4 | 3.0–11.7 | | 7.4 | 3.7–11.1 | | 6.7 | 3.8–9.6 | | 5.8 | 2.5–9.1 |
| Sibling with asthma (%) | | 17.6 | 11.2–24.1 | | 15.9 | 10.7–21.1 | | 9.2 | 5.8–12.5 | | 11.8 | 7.2–16.3 |
| Positive skin test (%) | | 33.1 | 25.2–41.0 | | 18.4 | 12.9–24.0 | | 18.8 | 14.3–23.4 | | 25.0 | 18.9–31.1 |

* SAPALDIA, Swiss Study on Air Pollution and Lung Diseases in Adults; SD, standard deviation; CI, confidence interval.

† Weight (kg)/height (m)².

‡ Participants were asked to fill in diaries for six periods of 4 weeks each (168 days).

§ Proportion of winter days among all days with symptom status reported.

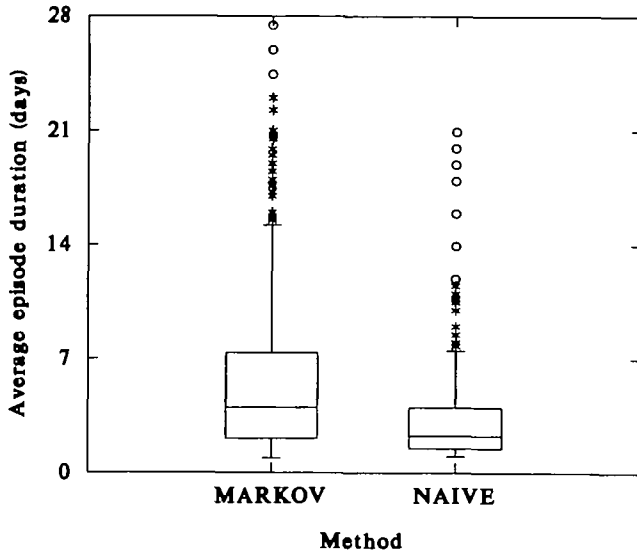


FIGURE 1. Estimated individual average durations of bronchitis symptom episodes (Markov method vs. naive method) among 454 subjects: Swiss Study on Air Pollution and Lung Diseases in Adults, 1992–1993. The 25th, 50th, and 75th percentiles define the lower edge, inner line, and upper edge of each box, respectively. Values exceeding the 75th percentile by more than 1.5-fold the interquartile range are plotted individually. The minimum value defines the end of the lower whisker.

nonsymptomatic episodes. However, the overall differences between light (<10 cigarettes/day) and heavy (≥30 cigarettes/day) smokers were considerable (table 3). We estimated medians of 60.0 symptom-free days (95 percent confidence interval (CI) 42.0–78.5) and

4.0 symptomatic days (95 percent CI 3.0–6.0), respectively, for light smokers compared with medians of only 21.0 symptom-free days (95 percent CI 16.2–29.8) and 6.0 symptomatic days (95 percent CI 4.9–9.0), respectively, for heavy smokers.

DISCUSSION

In this paper, we have provided a simple and efficient method of dealing with diary data from a 2-year study in which subjects recorded information during several separate 4-week periods. This type of participation was less demanding for participants than a single 6-month diary period. A design with multiple intermittent diary periods results in an increased number of symptomatic and nonsymptomatic episodes exceeding the limits of the documented periods, adding to the problem of missing data encountered in diary studies. Since the symptom duration data tend to be highly skewed, with a minimum of 1 day and a long tail (12), a naive approach considering only complete episodes not only results in a substantial loss of data but introduces bias towards shorter episodes.

We have shown that the use of a simple Markov model allows investigators to avoid the problems described. A MEDLINE® search suggested that Markov models are increasingly being used in scientific papers (37 papers in 1966–1975, 57 in 1976–1983, 178 in 1984–1990, and 258 in 1991–1996). The application of this method in epidemiology, however, has been relatively scarce. With regard to environmental epi-

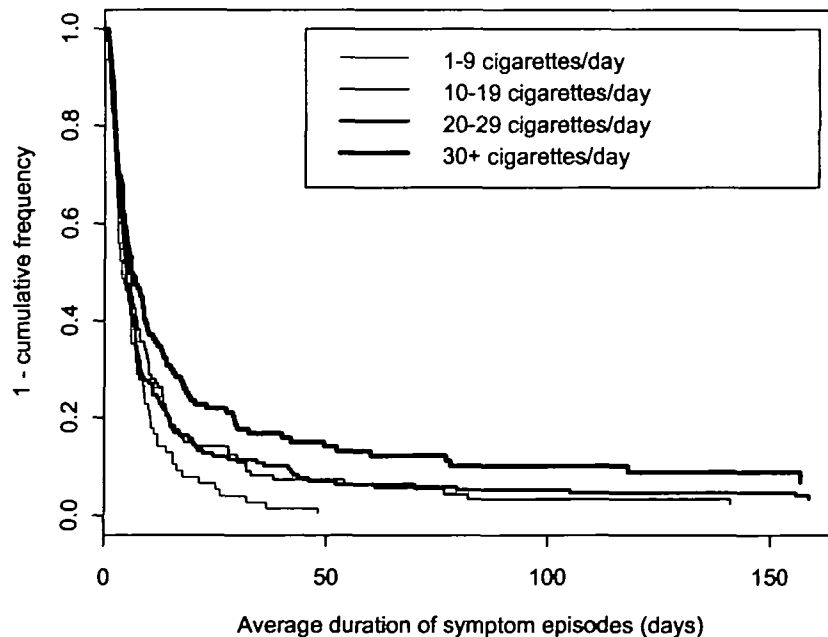


FIGURE 2. Distribution of the individual average durations of bronchitis symptom episodes (adjusted for covariates listed in text), by smoking category: Swiss Study on Air Pollution and Lung Diseases in Adults, 1992–1993.

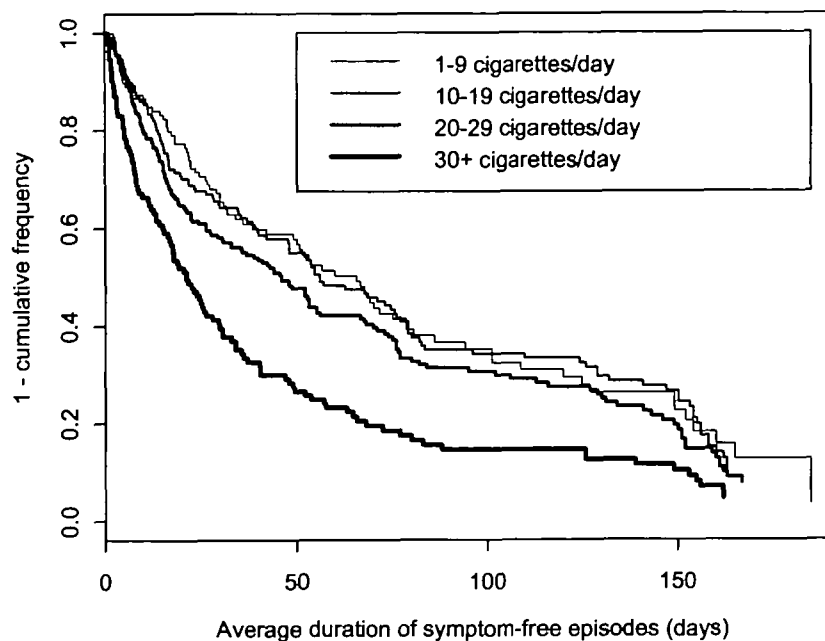


FIGURE 3. Distribution of the individual average durations of episodes without bronchitis symptoms (adjusted for covariates listed in text), by smoking category: Swiss Study on Air Pollution and Lung Diseases In Adults, 1992–1993.

miology, the model was first applied by Korn and Whittemore for analyzing the relation between a binary health response and short term air pollution exposure (17) as a way of avoiding the problems of previous approaches, such as autocorrelation and missing values. However, as far as we know, two-stage approaches, using parameters of a Markov model as outcomes to be further analyzed, have not been employed so far. The specific assumptions about this model do not seem to represent a serious methodological limitation, since our method of estimating the average duration of symptomatic episodes is asymptotically equivalent to the naive method in the case of complete data over a long observation period with a large number of documented symptomatic episodes. In the case of shorter observation periods, however, the estimates may exhibit some seasonal dependency. We therefore used the percentage of

documented days in the winter season as a model covariate.

In our example, estimates were available for more than 96 percent of the subjects. To increase this percentage, we considered additional steps such as imputing data for single missing days, but less than half of the remaining cases would have been gained for the analysis. A more sophisticated approach that would be generally applicable might consist of deriving maximum likelihood estimates under the assumption that missing days occur without any relation to symptom status.

The present model assumes that missing days are uninformative—i.e., that the mechanisms leading to days with missing information are unrelated to the underlying dynamics of symptom status (18). As regards undocumented day-to-day transitions of symptom status at the beginning or end of a diary period,

TABLE 3. Covariate-adjusted estimates* of the individual average durations of bronchitis symptom episodes and corresponding nonsymptomatic periods (in days), SAPALDIA† study, Switzerland, 1992–1993

| | Smoking category (cigarettes/day) | | | | | | | |
|-------------------------------------|-----------------------------------|-----------|--------------------|-----------|--------------------|-----------|------------------|-----------|
| | 1–9 (n = 136) | | 10–19 (n = 189) | | 20–29 (n = 285) | | ≥30 (n = 191) | |
| | Median | 95% CI† | Median | 95% CI | Median | 95% CI | Median | 95% CI |
| Episodes of bronchitis symptoms | 4.0 | 3.0–6.0 | 6.0 | 4.2–7.5 | 5.0 | 4.3–6.0 | 6.0 | 4.9–9.0 |
| Periods without bronchitis symptoms | 60.0 | 42.0–78.5 | 56.0 | 47.7–79.0 | 46.0 | 35.5–56.0 | 21.0 | 16.2–29.8 |

* Adjusted for sex, age, age², body mass index, study area, season, family history of asthma, and atopy (see "Materials and Methods").

† SAPALDIA, Swiss Study on Air Pollution and Lung Diseases in Adults; CI, confidence interval.

this assumption is certainly valid; but it does not match reality where undocumented day-to-day transitions within a diary period are concerned. In our example, the vast majority of subjects (90 percent) had very few (<5 percent) unexplained missing transitions (i.e., transitions for which at least one of the days was without documented symptom status and without indicated absence from the community). As a consequence of these nonignorable missing days, duration estimates may be biased to a certain extent. However, as long as reporting behavior varies little across exposure categories, differences in duration estimates between these categories may be less affected when being considered on a relative scale rather than an absolute scale. To assess the potential magnitude of this bias, we performed a sensitivity analysis using the individual percentage of missing transitions as an additional covariate. While this had almost no effect on the association between duration of symptom episodes and smoking intensity, the association between symptom-free episodes and smoking intensity decreased slightly. Moreover, upon analyzing associations between baseline symptoms (cross-sectional examination) and individual follow-up symptom prevalences, we found some suggestion that missing days were positively associated with symptom-free status. Further assessment of the potential impact of nonignorable missing data on duration estimates is beyond the scope of this analysis.

We tested our method using diary data from SAPALDIA and obtained plausible estimates for the effects of smoking on the individual average durations of bronchitis symptom episodes and corresponding nonsymptomatic episodes, respectively. Although the clinical feature of smoking as a risk factor for respiratory ill health is well known, it has not been previously quantified with regard to symptomatic and non-symptomatic episodes, to our knowledge. This study sample consisted of smokers who were either symptomatic or hyperreactive. The true effects of smoking might even have been underestimated, because of the fact that heavy smokers may reduce their cigarette consumption or quit smoking if they become symptomatic.

We suggest that the method of transition probabilities is a feasible approach to the assessment of long term effects of smoking and environmental risk factors on the average durations of symptomatic respiratory episodes and intervals without such symptoms.

ACKNOWLEDGMENTS

The Swiss Study on Air Pollution and Lung Diseases in Adults (SAPALDIA) is part of Swiss National Research

Program 26A, which is supported by the Swiss National Science Foundation (grant 4026-28099) and the Federal Office of Education and Science. SAPALDIA Basel is part of the European Respiratory Survey.

The SAPALDIA Team—*Study Director*: Dr. P. Leuenberger (Division of Pneumology, University Hospital Lausanne, Lausanne); *Program Director*: Dr. U. Ackermann-Liebrich (Institute for Social and Preventive Medicine, University of Basel, Basel); *Principal Investigators*: Drs. J. P. Bongard (private practitioner, Geneva), O. Brändli (Zürcher Höhenklinik, Wald), G. Domenighetti (Ospedale Regionale, Locarno), R. Keller (Klinik Barmelweid, Aarau), A. P. Perruchoud (University Hospital Basel, Basel), M. H. Schöni (Alpine Kinderklinik, Davos), J. M. Tschopp (Centre Valaisan de Pneumologie, Montana), and J. P. Zellweger (Division of Pneumology, University Hospital Lausanne, and Hôpital de Zone, Payerne); *co-Principal Investigators*: Drs. G. Bolognini (Ospedale della Beata Vergine, Mendrisio), P. Braun (Zürcher Hochgebirgsklinik, Davos), W. Karrer (Luzerner Höhenklinik, Montana), H. Keller-Wossidlo (Klinik Barmelweid), T. C. Medici (University Hospital Zurich, Zurich), G. Solari (private practitioner, Lugano), and B. Villiger (Thurgauisch-Schaffhausische Höhenklinik, Davos); *Central Unit*: Drs. N. Künzli, L. Grize, B. W. Martin, C. Schindler, and E. Zemp (Institute for Social and Preventive Medicine, University of Basel) and Dr. J. Schwartz (Harvard School of Public Health, Boston, Massachusetts); *Environmental Unit*: Drs. C. Monn and P. Alean (Federal Institute of Technology, Zurich) and Drs. C. Defila and A. G. Peeters (Swiss Meteorological Institute, Zurich); *Allergology Unit*: Drs. B. Wüthrich (University Hospital Zurich) and K. Blaser (Swiss Institute of Allergy and Asthma Research, Davos); *local medical field-workers*: Drs. C. Bron (Alpine Kinderklinik), M. Brutsche (Centre Valaisan de Pneumologie), S. Elsasser (University Hospital), P. Guldemann (Zürcher Höhenklinik), P. Hufschmid (Klinik Barmelweid), J. C. Luthi (Hôpital de Zone, Payerne), and A. Radaelli (Ospedale Regionale).

REFERENCES

1. Hammer DI, Hasselblad V, Portnoy B, et al. Los Angeles student nurse study: daily symptom reporting and photochemical oxidants. *Arch Environ Health* 1974;28:255-60.
2. Hoek G, Brunekreef B. Acute effects of a winter air pollution episode on pulmonary function and respiratory symptoms of children. *Arch Environ Health* 1993;48:328-35.
3. Hoek G, Brunekreef B. Effects of low level winter air pollution concentrations on respiratory health of Dutch children. *Environ Res* 1994;64:136-50.
4. Ostro BD, Lipsett MJ, Mann JK, et al. Air pollution and respiratory morbidity among adults in Southern California. *Am J Epidemiol* 1993;137:691-700.
5. Neas LM, Dockery DW, Koutrakis P, et al. The association of ambient air pollution with twice daily peak expiratory flow rate measurements in children. *Am J Epidemiol* 1995;141:111-22.
6. Robertson G, Lebowitz MD. Analysis of relationships be-

- tween symptoms and environmental factors over time. *Environ Res* 1984;33:130-43.
7. Roemer W, Hoek G, Brunekreef B. Effect of ambient winter air pollution on respiratory health of children with chronic respiratory symptoms. *Am Rev Respir Dis* 1993;147:118-24.
 8. Schwartz J, Hasselblad V, Pitcher H. Air pollution and morbidity: a further analysis of the Los Angeles student nurses data. *J Air Pollut Control Assoc* 1988;38:158-62.
 9. Schwartz J, Zeger S. Passive smoking, air pollution, and acute respiratory symptoms in a diary study of student nurses. *Am Rev Respir Dis* 1990;141:62-7.
 10. Schwartz J, Dockery DW, Neas LM, et al. Acute effects of summer air pollution on respiratory symptom reporting in children. *Am J Respir Crit Care Med* 1994;150:1234-42.
 11. Braun-Fahrländer C, Ackermann-Lieblich U, Schwartz J, et al. Air pollution and respiratory symptoms in preschool children. *Am Rev Respir Dis* 1992;145:42-7.
 12. Schwartz J. Air pollution and the duration of acute respiratory symptoms. *Arch Environ Health* 1992;47:116-22.
 13. Pope CA III, Schwartz J. Time series for the analysis of pulmonary health data. *Am J Respir Crit Care Med* 1996;154(suppl):S229-33.
 14. Martin BW, Ackermann-Lieblich U, Leuenberger P, et al. SAPALDIA: methods and participation in the cross-sectional part of the Swiss Study on Air Pollution and Lung Diseases in Adults. *Soz Präventivmed* 1997;42:67-84.
 15. Burney PG, Luczynska C, Chinn S, et al. The European Community Respiratory Health Survey. *Eur Respir J* 1994;7:954-60.
 16. Feller W. An introduction to probability theory and its applications. New York, NY: John Wiley and Sons, Inc, 1968.
 17. Korn EL, Whittemore AS. Methods for analyzing panel studies of acute health effects of air pollution. *Biometrics* 1979;35:795-802.
 18. Little RJ, Rubin DB. Statistical analysis with missing data. (Wiley series in probability and mathematical statistics). New York, NY: John Wiley and Sons, Inc, 1987.