
Estimated varicella incidence on the basis of a seroprevalence survey

M. P. MUÑOZ^{1*}, A. DOMÍNGUEZ^{2,3} AND L. SALLERAS^{2,3}

¹ *Department of Statistics and Operational Research, Universitat Politècnica de Catalunya, Barcelona, Spain*

² *Department of Public Health, University of Barcelona, Barcelona, Spain*

³ *General Directorate of Public Health, Department of Health and Social Security, Generalitat de Catalunya, Barcelona, Spain*

(Accepted 25 July 2001)

SUMMARY

Varicella is a disease caused by varicella-zoster virus. It is transmitted via the respiratory route, is highly communicable and mainly affects young children. An effective vaccine is now available, whose routine use is advised by health authorities in the USA and which can prevent severe disease, although breakthrough infections do occur. In deciding whether or not to include a vaccine in the routine vaccination schedule, knowledge of the morbidity of the disease in question is fundamental. Although reporting of varicella is compulsory in Catalonia, doctors only have to report the weekly number of cases diagnosed, and not their age distribution. Given that recent data on the prevalence of the infection in Catalonia according to age groups is available, it was considered that, using these data, an estimation of age-related incidence could be made.

The objective of the present study was to estimate the incidence of varicella in Catalonia on the basis of the available seroprevalence data. A curve was fitted to the observed prevalence and point prevalence estimates for all ages were obtained. The incidence was derived by smoothed prevalence for each of these age groups. Estimated variance of the estimated incidence was obtained by the delta method. Predicted prevalence in the 0–4 years age group was calculated by the smoothed prevalence.

The model that best fitted the sample prevalence was the exponential function. The estimated number of varicella cases in this study was 46419 (95% CI 40507–52270). As the population in Catalonia in 1996 was 6090040, the previous results give an incidence rate of 762.2 per 100000 persons/year with their 95% CI (666.1–858.3).

The method described may be applied to the study of incidence rates in relation to the prevalence of diseases if we accept that the infection produces permanent immunity; the risk of mortality is the same for infected and non-infected subjects and that the disease incidence and population remain constant in time.

INTRODUCTION

Varicella is a disease caused by the varicella-zoster virus. It is highly communicable and mainly affects

young children [1]. Various studies have shown that the disease is not as benign as previously thought and causes important health and social costs [2–4]. An effective vaccine is now available, whose routine use is advised by health authorities in the United States. However, in most countries, the vaccine has not yet been incorporated into the routine vaccination sched-

* Author for correspondence: Department of Statistics and Operational Research, Universitat Politècnica de Catalunya, Pau Gargallo 5, 08028 Barcelona, Spain.

ule and is used exclusively for the protection of immunocompromised children as, in these cases, the infection may take the form of a progressive disease with a case fatality rate of between 7% (all clinical forms) and 25% (pneumonitis) before the use of antiviral drugs [5, 6].

In deciding whether or not to include a vaccine in the routine vaccination schedule, knowledge of the magnitude of the problem – that is the morbidity of the process in question – is fundamental. Statistical models predict that if vaccine coverage against varicella in children is more than 90%, this will produce a higher proportion of varicella among older subjects [7]. This has generated a debate about the age at which universal vaccination should be carried out in order to minimize cases among adults [8]. For this reason it is important to know at what ages new cases of infection in community occur.

Although reporting of varicella is compulsory in Catalonia, doctors only have to report the weekly number of cases diagnosed, and not their age distribution. Since data on the prevalence of the infection in Catalonia according to age groups has recently become available, an estimation of age-related incidence can now be made.

The objective of the present study was to estimate the incidence of varicella in Catalonia on the basis of the available seroprevalence data.

METHODS

Our analysis is based on the seroprevalence study of varicella-zoster virus infection in Catalonia (Spain) [9]. As suggested by Leske et al. [10], a curve was fitted to the observed prevalence and point prevalence estimates for all ages were obtained. The parametric models (Table 1) were adjusted as functions of age. In all of the models proposed it was decided that prevalence should be zero for age zero, as recommended by Muench [11] for the catalytic models. It was assumed that in children of 1 year of age, both the maternal antibodies still present [12, 13] and the lesser exposure of the child to the virus, due to lesser socialization, would still protect them from the infection and that the maximum incidence rates would be produced before the age of 6 [14].

Using SPSS 7.5 software [15], the parameters β_0 , β_1 , their standard error, the asymptotic correlation matrix of the parameter estimates and the R^2 statistic for all cases were estimated by means of non-linear regression. The criteria for accepting a proposed model

Table 1. *Adjusted models for prevalence as function of age*

Model	Expression
EXPO1	$b_0 \cdot (1 - \exp(-b_1 \times \text{age}))$
LOGIS1	$\frac{1}{(1 + \exp(-b_0 - b_1 \times \text{age}))} - \frac{1}{(1 + \exp(-b_0))}$

was that the parameters were significant at 95% CI, with low correlation between them and a high R^2 statistic.

In accordance with Leske et al. [10], we assumed that:

- (1) varicella-zoster virus infection confers permanent immunity;
- (2) the mortality risk is the same in infected and non-infected individuals;
- (3) varicella is a stable disease in a stable population, i.e. disease incidence and population composition remain constant over time.

Under assumptions (1) to (3), as suggested by Beutels et al. [16], the incidence can be derived by smoothed prevalence for each of these age groups by using the following formula:

$$I_{\text{age}(i)} = 1 - \left[\frac{1 - P_{\text{age}(i+h_i)}}{1 - P_{\text{age}(i)}} \right]^{1/h_i}, \quad (1)$$

where $I_{\text{age}(i)}$ is the incidence at age i , $P_{\text{age}(i)}$ is the smoothed prevalence at age i and h_i designates the age interval.

The statistical methodology to estimate the annual number of new infected cases and their asymptotic confidence interval is explained in the Appendix.

Farrington [17] and Grenfell and Anderson [18] have suggested approaches to estimating incidence rates from seroprevalence data that are more intuitive, but applying these methodologies to our data, we obtained results not consistent with our prevalence. Maybe this was because we did not have the estimated prevalence in the 0–4 years group and the sample was selected from the age groups specified in Table 3.

RESULTS

We adjusted the two possible models specified in Table 1 to the sample data (see Fig. 1). We chose the EXPO1 model because it fitted the data significantly better than the other model.

The criteria used to choose the EXPO1 model instead of the LOGIS1 model took into account the

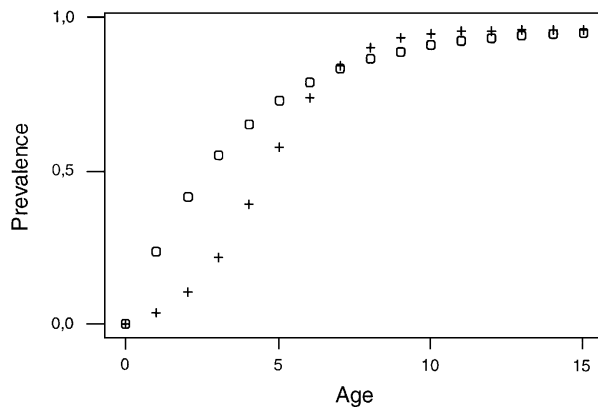


Fig. 1. Smoothed prevalence. (oooooo EXPO1 model, + + + + + LOGIS1 model)

level of curvature of the sample prevalence function and the plausibility of the epidemiological results. It should be pointed out that the LOGIS1 model assumes that the 0–4 years age group has considerably lower prevalence than that obtained with the EXPO1 model. The differences between these two estimated prevalence functions are significant up to the age of 6 years. This means that the results for incidence obtained using the LOGIS1 model are not epidemiologically plausible.

The way in which this study was carried out meant that it was not necessary to introduce any restriction in order to obtain non-negative incidence values. This could be avoided thanks to the choice of the non-decreasing monotonous prevalence function.

The estimated parameters are shown in Table 2. Other models tested gave similar results.

Prediction of the prevalence in the 0–4 years age group was made by means of the equation (A 1). Age-specific varicella incidence rates, with standard errors, are shown in Table 3. We obtained these results by applying equations (A 2) and (A 3) to the sample specified in the same table. In these cases the expressions needed to calculate the estimated variance are also in the Appendix.

From equations (A 4) and (A 5) we calculated the estimated annual number of incident cases, which was 46419. The results are also in Table 3. The 95% confidence intervals, obtained from equations (A 6) to (A 8) were 40507 and 52270.

As the population in Catalonia in 1996 was 6090040, these results give an incidence rate of 762.2 per 100000 persons/year with their 95% CI 666.1 and 858.3.

The results obtained for the estimated of incidence and its standard error are similar, the age groups for

which sample values are available, to those obtained by applying the method described by Maschner [19] where it is necessary to work with the adjusted prevalence function, as this method assumes that it will always be constant and not decreasing. In the opposite case, negative incidence values are obtained. This hypothesis cannot always be assumed when working with sample prevalence values. In addition, the methodology we propose allows an estimate of the incidence and its standard error in the 0–4 years age group, values which cannot be obtained from the equations (2) and (3) proposed by Maschner [19].

DISCUSSION

The estimated number of varicella cases in this study was 46419, which supposes an incidence rate of 762.2 per 100000 persons/year. It has been generally accepted that the varicella incidence rate should approximate that of the birth rate [20, 21].

In 1996, in Catalonia, there were 53400 births and the population was 6090040, giving a birth rate of 876 per 100000. The difference between this theoretical rate and the estimated incidence rate of varicella cases is small and may be due to the age distribution of the population and the percentage of individuals that acquire varicella in their lifetime (98% in the seroprevalence study considered).

Our results show that 35391 cases (76% of the total) occurred in the 0–4 years age group. Thus, routine vaccination at 15 months, coinciding with the MMR vaccine, would reduce the incidence of varicella and the circulation of the wild virus in children and the probable exposure of susceptible adults.

In addition, as some cases (404 in our model) occur in the 15–34 years age group, and given that varicella is more severe and complications more frequent in adults, a temporary programme of preadolescent vaccination could help to reduce the proportion of susceptible young adults. With a high level of vaccination coverage, this strategy would, in 10 years, immunize all the population under 21 years of age [22]. Our experience with the hepatitis B vaccination programme of preadolescents allows us to predict the success of this strategy in avoiding any increase in the proportion of susceptible adolescents and young adults [10].

The problem of not having information about prevalence in the 0–4 years age group was solved by considering data from nearby European countries [23–27]. This information was very important in the

Table 2. EXPO1 parameter estimates

Parameter	All
b_0	0.9786
S.E.*	0.0105
b_1	0.2761
S.E.*	0.0286
r_{b_0, b_1}^\dagger	-0.4257
R^{2*}	0.7420

* R^2 : R^2 statistic.

† r_{b_0, b_1} : Sample correlation coefficient.

choice of the function of prevalence to be adjusted. Of the different expressions for this function, those that give results similar to those obtained by the cited authors, were chosen.

Fornaro et al. [24] remark that the majority of varicella cases occur between 3 and 5 years of age and Farley and Miller [28] point out that the majority of cases diagnosed by sentinel doctors occur in children under 4. In the United States [29], in the 6–10 years age group, the prevalence of susceptible subjects is 18%, while in Catalonia, in the age group 5–9 years old, this figure is only 14.9%, indicating that infections occur a little earlier than in the United States. Different factors could explain these differences between countries [30–33].

Weller [30] and Sinha [31] found that in semitropical and tropical countries varicella occurs at an older age

that in temperate ones. Sinha suggests that this may be due to epidemiological interference by other prevalent viruses in these areas, especially herpes simplex virus. Mandal et al. [32] showed that people living in rural areas in West Bengal do not contract varicella until adulthood. Alvarez y Muñoz et al. [33] found that low educational level was a risk factor for susceptibility to varicella-zoster virus infection. Dworkin [34] suggested that racial differences could influence the susceptibility of adults.

It is worth pointing out that with the methodology we have applied in this study, it has been possible to predict the number of susceptible children in the 0–4 years age group without information in the studied sample.

We were not able to apply methods described by authors such as Ades and Nokes [35] or Maschner [36] to model the incidence in function of age and time, as we only had one seroepidemiological survey to work with. It will be necessary to review the results obtained when we have more seroepidemiological studies, which will allow us to contrast these estimates as well as analyse the evolution of the incidence rate of this disease over time.

In conclusion, using this formula, annual incidence can be estimated from the age-specific prevalence alone, but the problem as suggested by Keiding [37], lies in finding the correct function to smooth out the observed prevalence.

APPENDIX

Statistical methodology proposed to estimate the annual number of new infected subjects and their asymptotic confidence interval

If we suppose that the smoothed prevalence is exponential, for example EXPO1 from Table 2,

$$P_{\text{age}(i)} = b_0 * (1 - \exp(-b_1 \times \text{age}(i))). \quad (\text{A } 1)$$

Combining the expressions (1) and (A 1), the estimated incidence is

$$I_{\text{age}(i)} = 1 - \left(\frac{1 - b_0 \times (1 - \exp(-b_1 \times \text{age}(i + h_i)))}{1 - b_0 \times (1 - \exp(-b_1 \times \text{age}(i)))} \right)^{1/h_i}. \quad (\text{A } 2)$$

Obviously, $I_{\text{age}(i)}$ is a function of age and the random variables b_0 and b_1 , estimators of the parameters β_0, β_1 . Thus, by the delta method [10] we estimated the variance of the estimated incidence as follows

$$\text{Var}(I_{\text{age}(i)}) = \left[\frac{\partial I_{\text{age}(i)}}{\partial b_0} \right]^2 \text{Var}(b_0) + 2 \left[\frac{\partial I_{\text{age}(i)}}{\partial b_0} \right] \left[\frac{\partial I_{\text{age}(i)}}{\partial b_1} \right] \text{Cov}(b_0, b_1) + \left[\frac{\partial I_{\text{age}(i)}}{\partial b_1} \right]^2 \text{Var}(b_1) \quad (\text{A } 3)$$

Values for estimated variances of b_0 and b_1 , $\text{Var}(b_0)$ and $\text{Var}(b_1)$, and covariances, $\text{Cov}(b_0, b_1)$ were obtained by non-linear regressions as suggested above.

The calculations necessary for the expressions used in the equation, which allows us to find the estimated

Table 3. *Smoothed age-specific prevalence, estimated age rate incidence and the number of cases of chickenpox in Catalonia, 1995*

Age (years)	Sample size	Prevalence	P^*_{age}	I_{age}	s.e.†	N_{age}	$Susc_{age}$	$Cases_{age}$
0–4			0.415	0.225	0.018	268 425	156 961	35 391
5–9	261	0.851	0.837	0.190	0.015	282 603	46 060	8 740
10–14	622	0.918	0.943	0.097	0.042	341 403	19 430	1 882
15–24	117	0.940	0.975	0.016	0.012	970 872	24 455	382
25–34	216	0.935	0.978	0.001	0.001	957 829	20 613	22
35–44	221	0.995	0.979	< 0.000	0.000	860 303	18 315	1
45–54	220	0.991	0.979	< 0.000	0.000	763 495	16 243	0
55–64	224	0.996	0.979	< 0.000	0.000	653 404	13 900	0
> 64	255	1	0.979			991 706		
All	2136					6 090 040	315 976	46 419

* P^*_{age} : smoothed prevalence proportion for the middle of age interval, determined from the EXPO1 model.

† s.e., Standard Error.

variance of the incidence (equation (A 3)), may be performed analytically or by the use of an algebraic manipulator such as MAPLE 5 [38].

The expressions necessary in order to calculate the estimated variance of the incidence of the EXPO1 model are:

$$\begin{aligned} \text{Var}(I_{age(i)}) &= \left[\frac{\partial I_{age(i)}}{\partial b_0} \right]^2 \text{Var}(b_0) + 2 \left[\frac{\partial I_{age(i)}}{\partial b_0} \right] \left[\frac{\partial I_{age(i)}}{\partial b_1} \right] \text{Cov}(b_0, b_1) + \left[\frac{\partial I_{age(i)}}{\partial b_1} \right]^2 \text{Var}(b_1) \\ \left[\frac{\partial I_{age(i)}}{\partial b_0} \right] &= - \frac{(\exp(-b_1(\text{age}(i+h_i)) - \exp(-b_1 \text{age}(i))) \left(\frac{1-b_0+b_0 \exp(-b_1(\text{age}(i+h_i)))}{1-b_0+b_0 \exp(-b_1(\text{age}(i)))} \right)^{\frac{1}{h_i}}}{(1-b_0+b_0 \exp(-b_1(\text{age}(i+h_i))))(1-b_0+b_0 \exp(-b_1(\text{age}(i))))h_i} \\ \left[\frac{\partial I_{age(i)}}{\partial b_1} \right] &= \frac{(\exp(-b_1(\text{age}(i+h_i))(1-b_0)\text{age}(i+h_i) + \exp(-b_1 \text{age}(2i+h_i))h_i b_0)}{(1-b_0+b_0 \exp(-b_1(\text{age}(i+h_i))))(1-b_0+b_0 \exp(-b_1(\text{age}(i))))h_i} \\ &\quad + \frac{\text{age}(i)\exp(-b_1 \text{age}(i))(1-b_0)b_0 \left(\frac{1-b_0+b_0 \exp(-b_1(\text{age}(i+h_i)))}{1-b_0+b_0 \exp(-b_1(\text{age}(i)))} \right)^{\frac{1}{h_i}}}{(1-b_0+b_0 \exp(-b_1(\text{age}(i+h_i))))(1-b_0+b_0 \exp(-b_1(\text{age}(i))))h_i} \end{aligned}$$

To estimate the annual infection rate, we firstly need to estimate the susceptible population in each age group $Susc_{age(i)}$ as

$$Susc_{age(i)} = N_{age(i)} \times (1 - P_{age(i)}), \tag{A 4}$$

where $N_{age(i)}$ is the population at age group i .

Finally, the estimated annual number of new infected cases $Cases_{age(i)}$ occurring during the age interval, will be

$$Cases_{age(i)} = I_{age(i)} \times Susc_{age(i)} \tag{A 5}$$

To calculate the variance of the $Cases_{age(i)}$, we take the estimated susceptible population as real susceptible population. If not, by the delta method we could estimate this variance but the results obtained in this case were not essentially different.

So,

$$\text{Var}(Cases_{age(i)}) = \text{Var}(I_{age(i)}) \times (Susc_{age(i)})^2. \tag{A 6}$$

The estimate of the total annual number of incident cases, $\text{Case}_{\text{total}}$ will be

$$\text{Case}_{\text{total}} = \sum_i \text{Cases}_{\text{age}(i)} \quad (\text{A } 7)$$

and the estimated variance will be

$$\text{Var}(\text{Case}_{\text{total}}) = \sum_i \text{Var}(\text{Cases}_{\text{age}(i)}). \quad (\text{A } 8)$$

This allows us to calculate the asymptotic confidence interval for the total annual number of the incident cases.

REFERENCES

- Gershon AA, Takahashi M, White CJ. Varicella vaccine. In: Plotkin SA, Orenstein WA, eds. *Vaccines* 3rd ed. Philadelphia: WB Saunders Company, 1999: 475–507.
- Choo PW, Donahue JG, Manson JE, Platt R. The epidemiology of varicella and its complications. *J Infect Dis* 1995; **172**: 706–12.
- Lieu TA, Black SB, Riejer N, Ray PP, Lewis EM, Shinefield HR. The cost of childhood chickenpox parents' perspective. *Pediatr Infect Dis J* 1994; **13**: 173–7.
- Weber DJ, Rutala NA, Parham C. Impact and cost of varicella prevention in a University hospital. *Am J Public Health* 1988; **78**: 19–23.
- Feldman MD, Lott L. Varicella in children with cancer: impact of antiviral therapy and prophylaxis. *Pediatrics* 1987; **80**: 465–72.
- Whitley RJ. Varicella-zoster virus. In: Mandell GL, Bennet JE, Dolin R, eds. *Principles and practice of infectious diseases*, 5th ed. Philadelphia: Churchill Livingstone, 2000: 1580–6.
- Halloran ME, Cochi SL, Lieu TA, Wharton M, Fehrs L. Theoretical epidemiologic and morbidity effects of routine varicella immunization of preschool children in the United States. *Am J Epidemiol* 1994; **140**: 81–104.
- American Academy of Pediatrics. Varicella vaccine update. *Pediatrics* 2000; **105**: 136–41.
- Salleras L, Domínguez A, Vidal J, Plans P, Salleras M, Taberner JL. Seroepidemiology of varicella-zoster virus infection in Catalonia (Spain). Rationale for universal vaccination programmes. *Vaccine* 2000; **19**: 183–8.
- Leske MC, Ederer F, Podgor MJ. Estimating incidence from age-specific prevalence in glaucoma. *Am J Epidemiol* 1981; **113**: 606–13.
- Muench H. *Catalytic models in epidemiology*. Cambridge: Harvard University Press, 1959.
- Baba K, Yabunchi H, Takahashi M, Ogra P. Immunologic and epidemiologic aspects of varicella infection acquired during infancy and early childhood. *J Pediatr* 1982; **100**: 881–5.
- White E. Chickenpox in Kerale, India. *Indian J Publ Hlth* 1950; **22**: 141–51.
- Gershon AA, Steinberg SP, Gelb L, the National Institute of Allergy and Infectious Diseases Collaborative Varicella Vaccine Study Group. Clinical reinfection with varicella-zoster virus. *J Infect Dis* 1984; **149**: 137–42.
- SPSS for Windows, Release 7.5 (Computer program). SPSS inc., 1997.
- Beutels P, Bonanni P, Tormans G, Canale F, Cuneo Crovari P. An economic evaluation of universal pertussis vaccination in Italy. *Vaccine* 1999; **17**: 2400–9.
- Farrington CP. Modeling forces of infection for measles, mumps and rubella. *Stat Med* 1990; **9**: 953–67.
- Grenfell BT, Anderson RM. The estimation of age-related rates of infection from case notifications and serological data. *J Hyg* 1985; **95**: 419–36.
- Maschner IC. Determining the size of a cross-sectional sample to estimate the age-specific incidence of an irreversible disease. *Stat Med* 1994; **13**: 2369–81.
- Preblud SR, D'Angelo LJ. Chickenpox in the United States, 1972–1977. *J Infect Dis* 1979; **140**: 257–60.
- CDC. Summary of notifiable diseases, United States, 1991. *MMWR*, 1992; **40**, **53**: 1–64.
- Salleras L, Plans P, Vidal J, et al. Serological evaluation of the universal hepatitis B vaccination programme of preadolescents in Catalonia (Spain). *Vacunas Invest Pract* 2000; **1**: 3–6.
- Miller E, Vurdien J, Farrington P. Shift in age in chickenpox. *Lancet* 1993; **1**: 308–9.
- Fornaro P, Gandini F, Marin M, et al. Epidemiology and cost analysis of varicella in Italy: results of a sentinel study in the pediatric practice. *Pediatr Infect Dis J* 1999; **18**: 414–9.
- Paul E, Thiel T. Epidemiology of varicella zoster infection. Results of a prospective study in the Ansbach area. *Hautarzt* 1996; **47**: 604–9.
- Deguen S, Chan NP, Flahault A. Epidemiology of chickenpox in France. *J Epidemiol Commun Hlth* 1985; **52** suppl 1: 465–95.
- Trlifajova J, Svandova E, Pokorny J, Pokorni J. A laboratory study of age-related varicella incidence and prevalence in the Czech Socialist Republic. *Acta Virol* 1989; **33**: 183–7.
- Farley CK, Miller E. Varicella-zoster virus epidemiology a changing scene? *J Infect Dis* 1996; **174** suppl 3: S314–9.
- Van Loon F, Markowitz L, McQuillan G, et al. Varicella seroprevalence in US population. In: Program and abstracts of the 33rd Interscience Conference on Antimicrobial Agents and Chemotherapy (New Orleans). Washington DC: American Society for Microbiology, 1993: 359.
- Weller TH. Varicella and herpes zoster. Changing concepts of the natural history, control and importance

- of a not so benign virus. *N Engl J Med* 1983; **309**: 1362–8.
31. Sinha DP. Chickenpox: a disease predominantly affecting adults in West Bengal, India. *Int J Epidemiol* 1976; **5**: 367–74.
 32. Mandal BK, Mukherjee PP, Murphy C, Mukherjee R, Naik T. Adult susceptibility to varicella in the tropics is a rural phenomenon due to lack of previous exposure. *J Infect Dis* 1998; **178** Suppl 1: S52–4.
 33. Alvarez y Muñoz MT, Torres J, Damasio-Santana L, et al. Susceptibility to varicella-zoster infection in individuals 1 to 29 years of age in Mexico. *Arch Med Res* 1999; **30**: 60–3.
 34. Dworkin RH. Racial differences in herpes zoster and age at onset of varicella. *J Infect Dis* 1996; **174**: 239–40.
 35. Ades AE, Nokes DJ. Modeling age- and time-specific incidence from seroprevalence: Toxoplasmosis. *Am J Epidemiol* 1993; **137**: 1022–34.
 36. Maschner IC. Fitting a multiplicative incidence model to age and time-specific prevalence data. *Biometrics* 1996; **52**: 492–9.
 37. Keiding N. Incidence-prevalence relationships. In: Armitage P, Colton T, eds. *Encyclopedia of biostatistics*. Chichester: John Wiley & Sons, 1998, 2004–7.
 38. Maple V, Release 5.1 (Computer program). Waterloo Maple Inc., 1998.