Micro and macro population effects in disease transmission: the case of varicella

R. SILHOL1,2,*, F. P. ALVAREZ2, C. ARENA2, J. P. AMOROS3, A. FLAHAULT2,4, T. HANSLIK2,5 AND P. Y. BOËLLE1,2

1 UPMC Université Paris 06, UMR S 707, Paris, France
2 INSERM, UMR S 707, Paris, France
3 Université de Corse, Corte, France
4 EHESP School of Public Health, Paris, France
5 APHP, Hôpital Ambroise Pareé, Department of Internal Medicine, Versailles, France

(Accepted 25 August 2009; first published online 2 October 2009)

SUMMARY

Profiles of varicella-zoster virus (VZV) seroprevalence have shown large variability in European countries in which vaccination has not been implemented. Differences in micro and macro population structures (e.g. household and municipality, respectively) may explain such variability, which is the focus of a population-based study of varicella in 12 000 children from 7800 French households in Corsica. The cumulative incidence was 89% at age 11 years, as the median age at infection was 5 years in first-born children, but decreased to 3.9 years in younger siblings. Hazard of infection in households increased as the first-born child initially enrolled in primary school. Age at infection was higher in less populated areas. Household attack rates increased with age (55% in <6 months, >90% in >3 years) and household secondary attack rate was ~70%. Several levels of population structure independently affect age at varicella infection.

Key words: Household, infectious disease epidemiology, paediatrics, transmission, varicella zoster.

INTRODUCTION

Varicella is endemic in most Western countries where vaccination has not been implemented [1]. Surprisingly, the age-specific seroprevalence of varicella-zoster virus (VZV) shows large variability in these countries, although most people are infected with VZV during childhood. For example, the median age at infection ranges between 2 years in The Netherlands and 6 years in Italy [2]. Although factors leading to differences in age at infection may be numerous, the mixing networks of children could play a central role. Indeed, it has been suggested that differences in social behaviour are associated with differences in age at infection [3–5]. For example, initial primary-school enrolment of the first-born child could be followed by the introduction of varicella into the household, but other characteristics like school or municipality size may affect the time of introduction. In 1928, the mean age of varicella infection in rural Maryland was almost 2 years higher than in urban Maryland [6], suggesting less circulation of the virus in less populated areas. Even today, such characteristics may have a smaller impact in highly connected societies.

Knowledge on how these micro and macro population properties shape the spread of infectious agents...
is necessary for infectious disease modelling. Although the increasing availability of reliable demographic and transportation data allows models to take into account more detail in their structure, such as multiple levels of sources of exposure to infection [7, 8], population density, household structures [9], or empirical contact matrices [2], there is little epidemiological data documenting several levels of exposure to infection.

Varicella constitutes an exemplary case study for assessing the impacts of population characteristics on disease spread: all newborns are susceptible or become susceptible within a few weeks, almost everyone is infected before adulthood and the disease is easily recognized. In this paper, we first describe a retrospective study of varicella infection in about 12,000 children aged ≤11 years and report the changes in hazard of varicella infection in individual, household and macro characteristics.

METHODS

Study design

We conducted a survey of 7800 Corsican households in September/October 2007. Households were contacted through children attending nursery and elementary schools (in France >99% children start nursery school at age 3 years and attend elementary school until age 11 years). Corsica is a French island located in the Mediterranean Sea, populated by 280,000 individuals. In total, there are 268 nursery and elementary schools of which 180 (67%) agreed to participate in the study. Households were included if they comprised of at least one child in a participating school, and if the parent or legal guardian of the child was willing to participate. The survey was approved by the French Commission Nationale de l’Informatique et des Libertés ethics committee (Approval Number 907091).

Questionnaires

In September 2007, participating schools distributed the survey questionnaires to all attending children. Households receiving more than one questionnaire were instructed to return only one. The following data were obtained for each participating household: place of residence (municipality and zip code), and the number and age of household members. Individual-level information concerning varicella was obtained for all children aged ≤11 years living in the household. This included date of birth, current school or playgroup attendance, and the name of school or playgroup. History of varicella infection (yes/no) and age at infection were also obtained. Questionnaires were returned to the school at most 1 week after distribution, and mailed back to the study investigator. Data were double entered and measures to assure data quality were applied.

Missing data treatment

The following children were excluded from analysis: age >11 years or missing information on date of birth (n=2122, 15%), missing varicella history (107 children with unknown history of varicella and 298 with positive reported history of varicella but unknown age at infection, 0.7% and 2.1%). For geographical analyses we also excluded those with missing municipality of residence (n=107, 0.7%). For attack rate (AR) and secondary attack rate (SAR) calculations, we excluded households with any missing varicella history (307 households, 3.9% of all households, corresponding to 520 children, 3.7% of all children).

Spatial data

We used data from both the French National Institute for Statistics and Economical Studies (INSEE) and the French Ministry of National Education in order to evaluate demographic characteristics of participating schools, households and children. The total size of Corsican youth population was calculated by summing the annual number of births. The distance between Corsican municipalities was calculated as the Euclidean distance of their centroids.

Statistical analysis

We considered the age at which varicella infection occurred as a time-to-event variable, with possible censoring at the time of observation. A smooth estimate of the hazard of infection was obtained using a linear log-spline with nodes 6 months apart [10]. We computed the cumulative incidence of varicella according to age, i.e. the percentage of individuals who had acquired varicella at a defined age [11]. In order to account for rounding in age at infection, we considered ages reported in completed years or half-years to be actually interval-censored in the following
or current year (e.g. varicella at 3 years or 3 years 6 months was computed to occur between 3 and 4 years). First-born children and younger siblings were analysed separately. Confidence intervals (CI) for hazards and cumulative incidences were calculated using 200 bootstrap samples.

Methods used to determine household characteristics associated with varicella incidence must account for the changing structure of households over time (births, school enrolments, infections). We adjusted time-dependent Cox proportional hazards models to study associations between hazard of varicella infection and two sibship characteristics: (1) the current number of children in the household, and (2) the current number of children in the household aged \( \geq 3 \) years.

In order to evaluate intra-household varicella transmissibility, we calculated the calendar dates of infections in households using information about each subject’s current age and age at infection. Within household \( i \), let \( (t_{i,1}, t_{i,2}, \ldots, t_{i,m}) \) be the sorted calendar dates (in months) of the observed varicella cases in ascending order, with \( t_{i,s} = +\infty \) if the case of children \((i, s)\) was not observed at the time of the study. A 10-month period \([t_{i,c}, t_{i,c} + 10]\) defines a period of varicella exposure if there was a case at \( t_{i,c} \) and if \((t_{i,c} - t_{i,c-1} > 10)\) (i.e. no case in the previous 10 months). In each period of varicella exposure, we counted the number of initially susceptible children \([t_{i,r} \geq t_{i,c} ; (i, r) \text{ born before } t(i, c)]\) and the number of cases \((t_{i,r} \leq t_{i,c} + 10)\) (see Fig. 1). Each varicella case was included in only one period. Periods of varicella exposure with at least two susceptible subjects at first, including the proband varicella case, were selected for analysis. Household ARs were calculated as the percentage of susceptible children at the beginning of each period of varicella exposure (counting the first case) infected by the end of the period. Household age-specific ARs in children were calculated in the <6 months, 6 months to 3 years, and \( >3 \) years age groups, as well as in periods including at least one susceptible child aged <6 months and in those including at least one susceptible child aged \( >3 \) years. Finally, we computed the household SAR in each period as the percentage of cases in those initially susceptible, excluding the first case.

The geographical variation in varicella hazard was determined by a Cox proportional hazards model with frailty [12]. More precisely, the hazard rate of varicella infection at age \( a \) in municipality \( j \) was formulated as \( k(a, j) = k_0(a) \exp(u_j) \) with \( u_j \sim N(0, \sigma_u^2) \), \( k_0(a) \) the spatial-averaged hazard depending only on age and \( u \) a zero-mean normally distributed random variable (or frailty) taking the same values for all children residing in the same municipality. The random effect is applied in areas where cases tend to be either younger (\( u_j \) positive) or older (\( u_j \) negative) than average. The variability was assessed by comparing the variance of the area-level random effect (\( \sigma_u^2 \)) to 0.

In order to explore geographical clustering of the hazards of varicella infection, the spatial autocorrelation of age at varicella infection was computed using Moran’s \( I \) on the estimated frailties. Moran’s \( I \) computes the ratio of the covariance between contiguous municipalities and the total variance [13, 14]. It measures whether neighbouring areas preferentially

---

**Fig. 1.** Periods of exposure are defined as 10-month time periods starting with one varicella case in a household with at least another susceptible sibling. In household H1, there is one single period (P1: three susceptible children and two cases) while there are two periods in H2 (P2: two susceptible children and one case; P3: two susceptible children and two cases). B, Birth; V, varicella.
display changes from average in the same direction (in this case $0 < I \leq 1$) or opposite direction ($-1 \leq I < 0$), or if there is no association at all between the characteristic and the geographical component ($I = 0$).

Neighbouring areas were described by a first-order contiguity matrix $W$: $W_{ij}$ is 1 if municipalities $i$ and $j$ share geographical borders, and 0 otherwise. $I$ was calculated as

$$I = \frac{n}{\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}} \times \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} u_i u_j / \sum_{i=1}^{n} u_i^2.$$

Moran’s $I$ statistic may be disaggregated to yield a series of local indices. Local indicator for spatial autocorrelation (LISA) is the result of such a disaggregation, and can be mapped and tested to provide an indication of clustering patterns [15, 16], here the similarity in age at varicella infection.

Finally, we determined how population and school size affected age at infection. The number of inhabitants for each municipal residential zone was determined from the 1999 French national census database. We also computed the total number of inhabitants in municipalities within a 5 km radius of each household. A proportional hazards Cox model was fit where school or population size was cut into quartiles. Multivariable models, including individual, household and spatial information were also fit.

All models and estimations were computed using R [17]. We used GeoDa [18] software and the R package spdep [19] to create the contiguity matrix and compute Moran’s $I$ and LISA.

**RESULTS**

Overall, 7934 households including 14,118 individuals participated in the survey. Smaller schools (<60 children) were less likely to participate than larger schools; however, parents of children attending smaller schools agreed more often to participate (64% vs. 55%, respectively). After the exclusions detailed in the Methods section, we included 11,501 individuals aged 0–11 years which corresponds to 36% of Corsican children in this age range.

**Individual characteristics**

The cumulative incidence of varicella in the study population increased with age, reaching 89% (95% CI 88–90) at 11 years. The median age at infection was 4.7 years (95% CI 4.6–4.8).

Figure 2a shows that the hazard of varicella infection increased from birth to age 5 years and decreased thereafter. The hazard was shifted to younger ages in the younger children of the sibship. A change in slope was noticeable in the hazard of infection of first-born children at age 3 years which corresponds to enrolment in nursery school, but was not present for younger siblings. Figure 2b reports the corresponding cumulative incidences. Median age at varicella was 5-0 years (95% CI 4.9–5.1) in first-born children, yet significantly decreased to 3-9 years (95% CI 3.8–4.0) in younger children. However, both cohorts eventually reached a cumulative incidence of ~90% at age 11 years. The hazard ratios (HR) from the Cox proportional hazards model confirmed that the hazard of
varicella infection increased monotonically with sibship rank (Table 1).

Table 1. Varicella hazard ratios (HR) according to individual, sibship, municipality and school characteristics

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>HR</th>
<th>95% CI</th>
<th>Score (logrank) test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children sibship rank</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>7512</td>
<td>Ref.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3274</td>
<td>1.52</td>
<td>(1.44–1.61)</td>
<td>P &lt; 0.00001</td>
</tr>
<tr>
<td>3</td>
<td>521</td>
<td>1.75</td>
<td>(1.53–2.01)</td>
<td></td>
</tr>
<tr>
<td>≥4</td>
<td>62</td>
<td>2.59</td>
<td>(1.75–3.84)</td>
<td></td>
</tr>
<tr>
<td>Sibship</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of children in the sibship*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>7277</td>
<td>Ref.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5646</td>
<td>1.40</td>
<td>(1.33–1.47)</td>
<td>P &lt; 0.0001</td>
</tr>
<tr>
<td>3</td>
<td>1166</td>
<td>1.46</td>
<td>(1.34–1.60)</td>
<td></td>
</tr>
<tr>
<td>≥4</td>
<td>155</td>
<td>2.01</td>
<td>(1.60–2.54)</td>
<td></td>
</tr>
<tr>
<td>Number of school-aged children (age ≥3 years)*†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>8954</td>
<td>Ref.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥1</td>
<td>5108</td>
<td>1.55</td>
<td>(1.47–1.62)</td>
<td>P &lt; 0.00001</td>
</tr>
<tr>
<td>Municipality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of inhabitants in the municipality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;106</td>
<td>183</td>
<td>Ref.</td>
<td></td>
<td>P &lt; 0.0001</td>
</tr>
<tr>
<td>106–261</td>
<td>409</td>
<td>1.24</td>
<td>(0.98–1.56)</td>
<td></td>
</tr>
<tr>
<td>261–719</td>
<td>1012</td>
<td>1.29</td>
<td>(1.05–1.60)</td>
<td></td>
</tr>
<tr>
<td>&gt;719</td>
<td>9897</td>
<td>1.43</td>
<td>(1.17–1.75)</td>
<td></td>
</tr>
<tr>
<td>Number of inhabitants in a 5 km radius of the individual municipality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;760</td>
<td>476</td>
<td>Ref.</td>
<td></td>
<td>P = 0.007</td>
</tr>
<tr>
<td>760–1458</td>
<td>730</td>
<td>1.11</td>
<td>(0.96–1.29)</td>
<td></td>
</tr>
<tr>
<td>1458–3686</td>
<td>2050</td>
<td>1.10</td>
<td>(0.96–1.25)</td>
<td></td>
</tr>
<tr>
<td>&gt;3686</td>
<td>8245</td>
<td>1.18</td>
<td>(1.04–1.33)</td>
<td></td>
</tr>
<tr>
<td>School</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School size quartile‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;43</td>
<td>284</td>
<td>Ref.</td>
<td></td>
<td>P &lt; 0.0001</td>
</tr>
<tr>
<td>43–98</td>
<td>948</td>
<td>1.64</td>
<td>(1.27–2.11)</td>
<td></td>
</tr>
<tr>
<td>98–164</td>
<td>1102</td>
<td>1.82</td>
<td>(1.42–2.33)</td>
<td></td>
</tr>
<tr>
<td>&gt;164</td>
<td>701</td>
<td>1.57</td>
<td>(1.23–2.02)</td>
<td></td>
</tr>
</tbody>
</table>

CI, confidence interval.
* Time-dependent variables: for those variables, n corresponds to the number of at-risk periods entered in the model.
† Model adjusted for sibship rank.
‡ Model estimated on children aged from 3 to 7 years.

Household characteristics

As shown in Table 1, the hazard of varicella infection consistently increased with the number of children in the sibship and increased when there was at least one child aged ≥3 years. This effect was present in the first-born child (HR 1.29, 95% CI 1.17–1.41) as well as in those born afterwards (HR 1.68, 95% CI 1.34–2.10) (data not shown).

Regarding varicella transmissibility, 1788 periods of exposure were identified, comprising of 1615 periods with two susceptible children at the beginning of the period of exposure (resulting in one case 503, two cases 1112), 159 with three susceptible children (one case 19, two cases 37, three cases 103) and 14 with four susceptible children (one case 2, two cases 2, three cases 3, four cases 7). Table 2 describes ARs and SARs. One child aged >3 years was present in 91% (95% CI 90–92) of the periods, in which the AR was ~93% (95% CI 92–94). Conversely, this rate was 76% (95% CI 74–79) in the 6 months to 3 years age group and only 55% (95% CI 50–60) in the <6 months age group. The AR was smaller when children in the household were younger: 74% (95% CI 70–78) when all children were <3 years and 87% (95% CI 86–88) with at least one child aged >3 years. Overall, the SAR in household children was 70% (95% CI 67–72), and was markedly lower for households with at least one susceptible children aged <6 months compared to none (47% vs. 72%, $P < 0.0001$).

Area characteristics

Hazard of varicella changed with location ($\sigma^2 = 0.038$, $P < 10^{-5}$). Neighbouring areas were more similar than others (Moran’s index $I = 0.16$, $P = 0.0008$). Figure 3 shows that the areas presenting local spatial
correlation for varicella hazard corresponded with the most populated areas. Investigating local spatial autocorrelation, we found that children residing in the two largest Corsican cities and neighbouring municipalities had a lower age at infection. Figure 4a shows lower hazard of varicella, and at an older age, in children residing in the less populated areas, with Figure 4b showing a corresponding cumulative incidence shifted to the right compared to the most populated quartile.

In the Cox proportional hazard model, the HR for population size in the municipality showed an increasing trend with decreasing population (Table 1), and this effect was maintained even after adjusting for sibship rank. Small school size was also associated with infection at older age (Table 1), and this hazard did not change when the model was adjusted on population size (HR for Q4 1.54, 95% CI 1.17–2.02, data not shown).

Table 2. Age-specific household attack rates and secondary attack rates of varicella

<table>
<thead>
<tr>
<th></th>
<th>All periods (n = 1788)</th>
<th>All susceptible children aged ≥6 months</th>
<th>All susceptible children ≤3 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All periods (n = 1485)</td>
<td>Yes (n = 1485)</td>
<td>No (n = 303)</td>
</tr>
<tr>
<td>Attack rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>84% (83–85)</td>
<td>87% (86–88)</td>
<td>73% (70–76)</td>
</tr>
<tr>
<td>&lt;6 months</td>
<td>55% (50–60)</td>
<td>n.a.</td>
<td>55% (49–60)</td>
</tr>
<tr>
<td>6 months to 3 years</td>
<td>76% (74–79)</td>
<td>76% (73–78)</td>
<td>84% (76–91)</td>
</tr>
<tr>
<td>&gt;3 years</td>
<td>93% (92–94)</td>
<td>93% (92–94)</td>
<td>93% (60–96)</td>
</tr>
<tr>
<td>Secondary attack rate</td>
<td>70% (67–72)</td>
<td>74% (72–76)</td>
<td>47% (42–53)</td>
</tr>
</tbody>
</table>

n.a., Not available.
Values in parentheses are 95% confidence intervals.

Fig. 3. Local spatial autocorrelation in age at infection. (a) Locations having significant (P < 0.01) spatial autocorrelation in varicella infection hazard. (b) Corsican children population. LISA, Local indicator for spatial autocorrelation.
DISCUSSION

Our study highlights how multiple levels of population structure shape the hazard of varicella infection. At the household level, this hazard was greater for second and later-born children compared to the first; at the community level, the age of varicella infection decreased as the population and school sizes increased.

Participation bias may be present in the data, as we used voluntary reports by parents. If a negative history of varicella was strongly related to non-response, one would have expected increasing participation with age; but the participation rate according to age did not show a trend. Even through this rules out a strong age-related bias, we could not further verify if non-response was ignorable. Classification bias is unlikely, as true subclinical varicella is rare and clinical diagnosis easy in children presenting lesions [20]. Recall bias should also be limited, given the large positive predictive and negative predictive values of varicella history for school-aged children (95% and 90%, respectively) [21–23]. Furthermore, our estimates of cumulative incidence of varicella at age 11 years are similar to that of a recent independent French seroprevalence study (90% at age 11 years [24]) as well as earlier figures in France [25]. However, the cumulative incidence in Corsica appeared to be shifted towards higher ages, by <1 year, compared to those reported in continental France [24]. This may be due to our correct handling of age in our models, expressed in completed years using methods for interval-censored data; but it could also be the consequence of a lower population density in Corsica in relation to other French regions. In fact, such an effect was noticeable within Corsica, where smaller municipality size was associated with delayed acquisition of the disease. Differences in household size did not explain this effect, as these were the same in the less and more populated areas. This was not a consequence of age at first school enrolment as >99% of 3-year-old children attend school in Corsica (INSEE data), with little variation according to the population size of municipalities, and day-care options are sparse for younger children.

An outstanding feature of our dataset was the joint documentation of population characteristics at several levels: individual, household, school and municipality. This allowed us to apply time-to-event methods in order to investigate how individual, household and macro characteristics, as well as their interplay, impact the hazard of varicella infection. The most marked finding was a noticeable change in slope around age 3 years of the hazard of varicella infection in first-born children (Fig. 2a), probably linked to the increased exposure to varicella due to initial school enrolment. A second – more predictable – finding was the quantified impact of children attending school on
their siblings: the hazard increased by 57% in second and later-born children when the first-born child was aged $\geq 3$ years. This result strongly indicates that first-born children introduce the disease into their households. Moreover, the age at which infection occurred in second or later-born children was younger than in first-born children and almost all periods of exposure to varicella included one child aged $\geq 3$ years.

The differences in age at infection linked to population size may indicate spatial hierarchy in the spread of varicella, as described for measles in the UK [26]. Indeed, sustained varicella circulation in populated areas will lead to infectious contacts at a younger age, although these contacts will be less frequent in less populated areas which will subsequently delay disease occurrence. Interestingly, the association between population size and varicella infection weakened when population density (as defined by the population within a 5 km radius of the municipal residential zones), rather than population size, was considered. Since school and leisure activities tend to be organized within municipalities, this may indeed limit true geographical or density-dependent effects.

The AR and SAR calculations were dependent on the chosen duration for the period of exposure. In households where age at varicella was rounded to the nearest completed year, the dates of varicella cases were only computable to a 1 year range, so that a large time-window was necessary to identify grouped cases. As expected, increasing the duration of the period of exposure led to an increase in AR and SAR – in a sensitivity analysis, the AR was 78% (95% CI 77–80) using a 6-month period instead of 84% using a 10-month period, and the SAR was 58% (95% CI 56–61) (instead of 70%) for the respective time periods. All comparisons remained qualitatively valid. The SAR calculated with a 10-month period of exposure was similar to that reported in unvaccinated children (71% [27]); the smaller AR in children aged $<6$ months was consistent with immune protection by maternal antibodies [28, 29].

Finally, the household SAR calculation will be biased if cases may be infected in the community, or if tertiary cases are present [30]. The high transmissibility of varicella and small size of households makes tertiary cases unlikely; however, common exposure at school or elsewhere could be possible. Statistical methods are available to partition the sources of infection [31]; however, they could not be applied here, since timing of varicella exposure outside the household was not known.

Our study helps elucidate several of the components which may explain the widely different varicella seroprevalence profiles reported in Europe [2]. The decision to vaccinate the population remains a major issue in these countries [32]. Our findings suggest that further mathematical models and cost-utility analysis in determining the benefit of mass vaccination may be enhanced by including a more detailed population structure.

ACKNOWLEDGEMENTS

The survey was conducted through the Sentinelles Network (for more information visit: http://www.sentiweb.org). The authors thank the French Education Nationale, l’Université de Corse and l’URML Corse for their technical support. We thank Anders Boyd for useful comments. Partial financial support was provided by Sanofi Pasteur MSD.

DECLARATION OF INTEREST

None.

REFERENCES


