Exploratory analysis of Leptospirosis cohort data

Nicolas Capian*, Avner Bar-Hen*, Peter Diggle**

*MAP5 Paris Descartes Université, 75006. **Lancaster University, UK.

Contact address : nicolas.capian.pro@gmail.com

I. Cycle of Leptospirosis infection

Leptospirosis is a zoonotic infectious disease caused by a pathogenic bacteria called a spirochet. The pathogens are maintained in nature in the renal tubules of certain mammalian species which excrete leptospiral pathogens in their urine. As illustrated in Figure 1, rodent species constitute a prime reservoir in which infection produces chronic and asymptomatic carriage. Leptospira bacteria can then infect livestock, domestic and wild animals and cause a range of disease manifestations and carrier states.

Leptospirosis is transmitted to humans by direct contact with reservoir animals or by exposure to environmental surface water or soil that is contaminated with their urine. Leptospirosis bacteria penetrate abraded skin or mucous membranes and disseminate throughout the body tissue. The incubation period lasts usually between 5 and 14 days. The infection generally first causes a febrile influenza-like illness and may progress to a severe form causing multi-system symptoms such as those displayed in Figure 1. Although the immune response eventually eliminates the pathogens, leptospira bacteria may persist for prolonged periods in immuno-privileged sites. Human beings are accidental hosts and do not shed sufficient numbers of leptospirosis bacteria to serve as reservoirs for transmission.

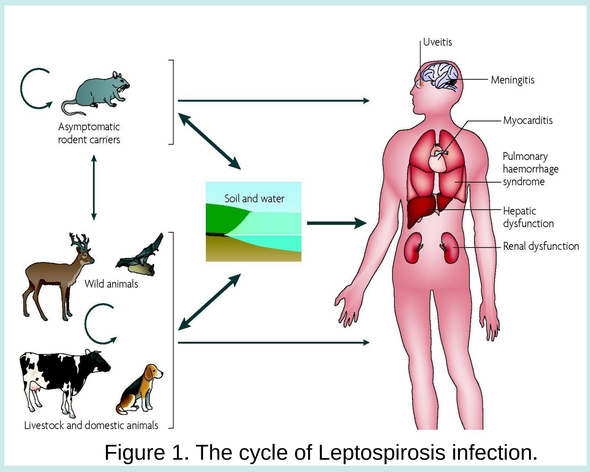

Figure 1. The cycle of Leptospirosis infection.

II. General context and problematic

Leptospirosis has a worldwide distribution both in industrialised and developing countries but the incidence of human infection is higher in the tropics and in areas with high rainfall where conditions for its transmission are particularly favourable. Over the past decade, leptospirosis has spread from its traditional rural base to become the cause of epidemics in poor urban slum communities. In Brazil, in average 3 500 cases of the severe form of the disease are reported per year with a 10-15% mortality rate [1]. Severe leptospirosis represents a small fraction of the total number of infections and the burden for milder forms of the disease is often underestimated because mild or sub-clinical leptospirosis is a significant cause of undifferentiated fever and is frequently not recognized.

Barriers to cope with leptospirosis problem have been the lack of adequate diagnostic tests and effective control measures. In order to improve the impact of prevention and intervention measures of the disease, there is a need of reliable information about the incidence rates of the disease and the risk factors which influence the progression of infection.

III. Collected data

Figure 2. Pau de Lima study site. (A) The yellow line is the boundary of the study site. (B) Typical environment at the community study. (C) Geo-referenced households.

Study site and cohort enrollment

Pau da Lima site is a 0.46 km² area made of four valleys located in the suburb of the city of Salvador de Bahia, Brasil. This area hosts a densely populated community of 14,122 inhabitants which is characterised by a high annual incidence of severe leptospirosis (35.4 cases per 100,000) observed from 1996 to 2002 and by a prevalence of leptospirosis of 15.4% (95% CI, 14-16.8%) observed in 2003 [2].

During a census in 2003, a study team from the research institution FIOCRUZ identified 12,651 residents who met the eligibility criteria of being strictly older than 5 years old. A prospective study was then set up based on a cohort of 2,003 subjects as depicted in Figure 2.

Outcome definition

Detection of antibodies is not always a proof of current infection as some antibodies may persist for long periods after an infection. Fast rise in titre from consecutive serum samples is considered to be diagnostic proof of recent infection and it was taken as the outcome of interest in the cohort study. The titre of leptospira antibodies was determined using the microscopic agglutination test (MAT). An infection between two time points was reported either when the MAT titres changed from negative to ≥1:50 (sero-conversion) or when a four-fold rise in MAT titres was observed. The case of interactions of antibodies associated with different Leptospira serovars was also taken into account in the definition of infection.

Temporal aspect and nature of data

The collecting of cohort data was planned as four successive annual follow-ups and lasted from year 2003 to 2008.

III. (continued) Five annual serum surveys of the cohort were performed during the summer season (October-January) so as to be able to prospectively identify up to four recent leptospirosis infections per individual.

The study team also performed four interviews of the subjects and surveys of their household during the winter epidemic season for leptospirosis (April-August) so as to evaluate exposures. The distribution of the time points of the collected data are displayed in Figure 4 and the middle plot reveals that the exposures data were not regularly collected as originally planned.

Various types of data were collected such as demographical factors, socio-economical factors, health factors, household environment, individual behaviour at home and at work and animal reservoirs near household and workplace. Climatic data and environmental data such as garbage accumulation and sewerage system locations were also collected at different time points.

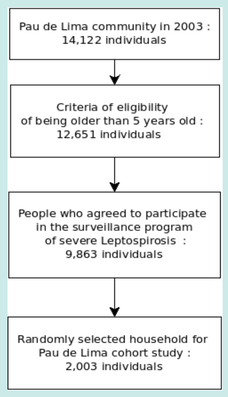

Figure 3. Selection of the cohort of individuals.

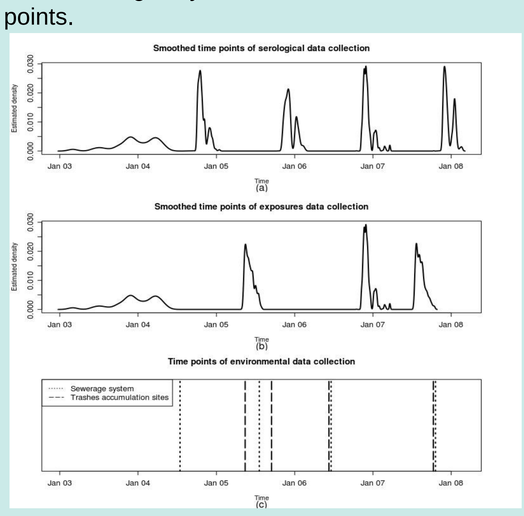

Figure 4. Temporal points of collect of the cohort data from 2003 to 2008.

IV. Preliminary exploration of the data

Missing data occurred and were mainly due to individuals who moved outside the study region. 1,251 individuals (62%) had complete serological and exposures data for the three first follow-ups.

The quality of spatial representation of the study site population by the cohort was checked as followed. A random sample of households from Pau de Lima was drawn and compared to the locations of households of the cohort. The null hypothesis of absence of spatial clustering between the two sets of locations was tested using the statistic D(s) [3]. Figure 5 reveals that there is no significant evidence of spatial clustering and no significant evidence of wrong spatial representation of the study site population by the cohort.

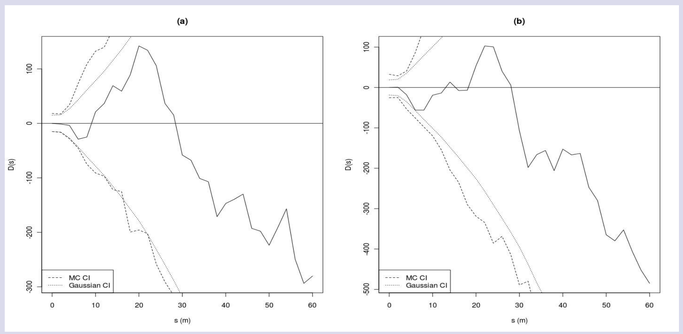

Figure 5. Observed test statistic D(s) for distance s (in metres) and confidence intervals (CI) under the null hypothesis of non spatial clustering for a random sample of Pau De Lima households against (a) the households of the cohort and (b) the households of the cohort for individuals with complete data.

All the following results are based on the data for the three first follow-ups and for the individuals with complete data. The exposures information here below are based on the data from the first follow-up.

Demographical and health factors : the individuals were young with a median age of 22 years old and a median weight of 57kg. There were more women (59%) and, 58% and 36% of people had brown or black skin color respectively. 70% had not completed primary school and 82% were able of reading. The average daily income per person was 0.8$, 65% had a social security number and 14% had a property title for their household.

Household environment : around a third of the individuals lived in a household located less than 10 m from open sewers. 11% of the individuals have already experienced a flooding. 69% of individuals had some vegetation or bushes and 24% had banana trees less than 10 m from their household. 64% had some accumulated material into disuse near their household.

Individual behavior near household : 40% of the individuals claimed to have contact with water flooding, 26% with sewage, 37% with mud and 22% with garbage close to home. 8% of the individuals had dug into the sewer close to their household and 11% had removed trashes into the sewer.

Reservoirs near household: 23% individuals declared the sight of rats near their household, 44% had a dog in their household, 17% had a cat and 38% individuals had chickens.

Observed infections : The number of recent Leptospirosis infections (and respective 100 person-year rate of infection) over the three first follow-ups are decreasing with 46 (3.7), 30 (2.3) and 26 (2) infections as displayed in Figure 6. 10 individuals have had two recent infections over the three follow-ups.

To estimate the spatial variation of the risk of infection, the density ratio method [4] was used. The estimated smooth spatial intensity of the locations of infected individuals was divided (or in another word adjusted) by the estimates of the spatial intensity of the locations for all individuals with complete data. Indeed, it can happen that a high spatial intensity of infections only results from the fact that the area is densely populated. This adjustment is used to remove this factor. The bottom-left plot of Figure 6 reveals that areas with high adjusted intensity of infection (in red) are located in the north-est part of the valleys except for valley 3 where it is also found in the souther part.

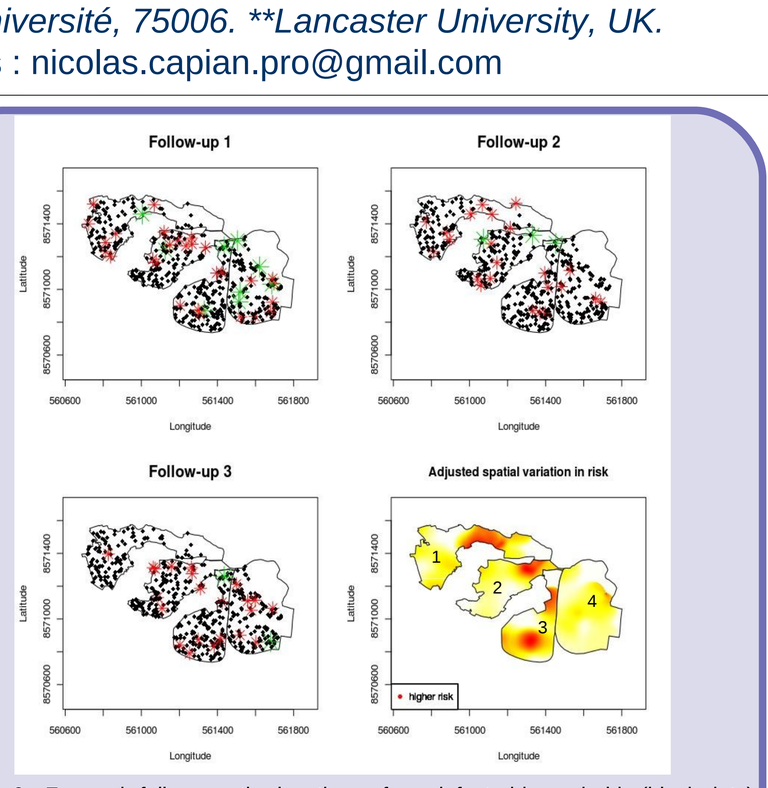

Figure 6. For each follow-up, the locations of non-infected households (black dots) and of infected households (red asterisk for 1 infection, green asterisk for 2 infections). Bottom-right : adjusted spatial variation in risk computed as the ratio of the intensity of infected individuals' locations out of the intensity of all individuals' locations. The four valleys are numbered in this plot.

V. Modelling of the factors-to-infection relationship

An initial list of 29 categorical and numerical covariates was constituted in order to model the risk of recent infection. The covariate of age was recoded as a categorical variable because its relationship with the risk of infection was not found to be monotonic.

The relationship between the probability of infection and the factors was modelled by a generalised linear model with Bernouilli distribution and logit link function. Based on a step-wise method of selection at the 5% level, the 9 first covariates were kept to give the final model (Table 1).

	All follow-ups		First follow-up		Second follow-up		Third follow-up	
Risk factor	OR	pvalue	OR	pvalue	OR	pvalue	OR	pvalue
Age between 15 and 24 y.o	3.607	< 0.001	5.13	0.001	1.65	0.57	4.6	0.05
Age between 25 and 34 y.o	6.953	< 0.001	6.76	0.001	3.66	0.18	16.77	< 0.001
Age between 35 and 44 y.o	3.391	0.006	4.46	0.02	3.13	0.25	1.67	0.68
Age more than 44 y.o	3.262	0.014	0.76	0.81	2.84	0.30	10.51	0.01
Is a man	2.069	0.001	2.47	0.005	2.84	0.009	1.22	0.64
Household Income	0.999	0.022	0.998	0.003	1	0.69	1	0.74
Has social security number	0.479	0.01	0.34	0.01	1.07	0.93	0.37	0.08
Residence time	1.027	0.065	1.02	0.42	1.06	0.02	1.01	0.66
Vegetation near home	2.061	0.012	1.26	0.56	3.01	0.03	6.03	0.08
Contact with garbage near home	1.677	0.032	1.65	0.16	0.73	0.59	2.45	0.05
Rises chickens	1.501	0.074	1.3	0.42	0.8	0.73	1.81	0.23
Contact with mud close to home	1.498	0.081	1.56	0.22	0.96	0.92	2.8	0.045

Table 1. Estimates (Odds-Ratio) for the parameters of the GLM obtained by aggregating all the data over the three follow-ups and comparison with the estimates of the same model obtained for each follow-up.

The part of the variation in risk of infection which was not explained by the factors of the previous model was fitted to the spatial coordinates using a generalised additive model. The estimated spatial smoothed term was found to be significant at the 1% level, suggesting that there exist other significant factors of spatial variation in the risk of Leptopirosis infection which were not taken into account by the model. The map of the remaining spatial variation is displayed in Figure 7.

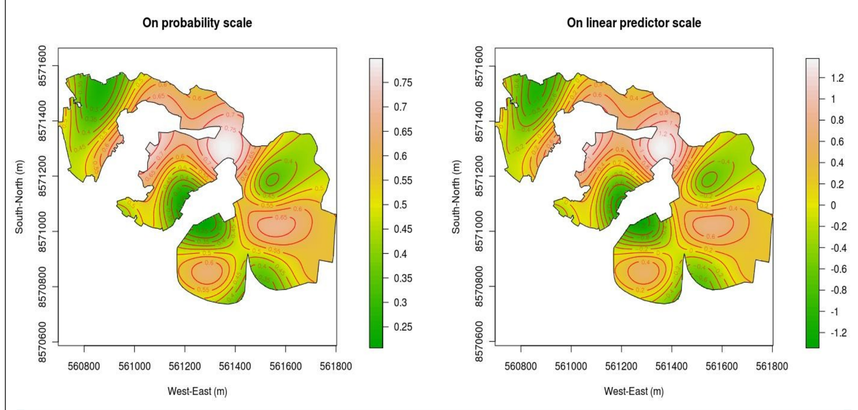

Figure 7. Using the data from all individuals with complete data and from the three first follow-ups, the risk of recent infection was adjusted by the effects of the risk factors displayed in Table 1. These plots represent the remaining spatial variation in risk of infection on probability scale (left) and on linear predictor scale (right) obtained by fitting a smooth spatial term of a generalised additive model.

VI. Conclusion and further work

The final model for data aggregated over time reveal that the risk of recent of infection is higher for young men with a low income, without social covering, living near vegetation and with contact with garbage near home.

As shown in Figure 4, exposures data were not always collected between the successive sero-surveys. The possible effects of this bias on the results should be investigated.

The particular and heterogeneous landscape of Pau de Lima suggests to carry out valley-specific analyses which will be the object of a further work.

In addition, as the number of infections decreases over time, a temporal effect will be included in the model. The change of the estimates of Table 1 for each follow-up also motivates to evaluate the significance of the effect of interaction between time and risk factors on the risk of recent infection.

Other covariates will be looked at such as climatic factors, sewage system locations and garbage accumulation sites locations.

Finally, over the follow-up period, ongoing control measures and interventions have taken place and their effect on the outcome of the cohort study should be considered.

Note: The bibliographical references marked between brackets are not displayed here. If interested, please feel free to ask for them by contacting the author.