

Microti data analysis

Material and methods

Two outcome variables of interest were identified

- **Micro** microbiological status of the lymph nodes (dichotomous variable Pos / Neg)
- **Histo** histological grade of granulomas (ordinal variable with four levels: g1 / g2 / g3 / g4)

On the basis of the nature of the outcome variable, we adapted specific regression models with the aim of identifying which variables among those measured significantly influence the outcomes. In this sense, the variables take on the meaning of predictors, therefore able to predict the outcome value with some degree of error.

Micro

For this analysis, the statistical unit is the single lymph node. The microbiological status of the lymph nodes (Micro: Positive / Negative) was modeled using a Bayesian logistic regression model using the following predictors:

Size of the lymph node: Diameter calculated from the area and then standardized in z-score *NAF*: presence of acid fast bacteria in lymph node granulomas. This variable was obtained by calculating for each single lymph node the average number of bacteria present in the various granulomas and then categorizing the lymph node itself as $NAF = 1$ when the average was greater than 0 and $NAF = 0$ (i.e. absence of acid-resistant bacteria) in the case of averages equal to 0. *MNC*: similarly to what was done for *NAF* was done for the categorization of lymph nodes on the basis of the presence or absence of cells in granulomas. *NGR*: total number of granulomas per lymph node used in the model after logarithmic transformation. *sG1*, *sG2*, *sG3*, *sG4*: to grasp the influence that the histological grade of granulomas has on the microbiological state of a lymph node, we constructed four new variables, one for each histological grade, which identify the proportions of granulomas of the different grades for each single lymph node. For each lymph node we calculated the number of granulomas of the different histological grades and then we divided it by the total number of granulomas (*NGR*) per lymph node, thus obtaining a value between 0 and 1 which corresponds to the proportion of granulomas of the different grades. For example, 16 granulomas were observed in lymph node # 8, of which 0 of Grade 1 and 2, 13 of grade 3 and 3 of grade 4, so for lymph nodes 8 we will have this profile $sG1 = 0$, $sG2 = 0$, $sG3 = 13/16$ (0.81), $sG4 = 3/16$ (0.19). By way of example we report in the following table lymph nodes and their profile in relation to the microbiological status to the proportion of granulomas of different histological grades and the total number of granulomas.

Idlinf	Micro	sG1	sG2	sG3	sG4	NGR
76	1	0	0.05	0.74	0.21	19
598	0	0	0.48	0.36	0.15	33
601	0	0	0.29	0.29	0.43	7
619	0	0	0.00	0.50	0.50	6
620	0	0	0.00	0.70	0.30	10