Effects of Maternal Periodontal Diseases on Pre-term Births

Laura Gamba
Asia Salpa
Matteo Tomasetto

Case Study

Several observational studies on pregnant women have suggested an association between pre-term births and low birthweight with respect to periodontal disease. First of all, let us define these pathological conditions:

- *Periodontal disease* is an inflammatory condition characterized by the destruction of tissue and/or bone around the teeth
- *Pre-term birth* is defined as delivery before 37 weeks of gestation; notice that this is a growing problem and it can lead to infant death or neurodevelopmental disabilities.
- *Low birthweight* is when the weight of the newborn is below 2500g

Our dataset is composed by 823 pregnant women that are affected by periodontal disease; they are subdivided into two groups based on the severity of their oral condition: 413 women are assigned to the treatment group and 410 to the control one. The patients in the treatment group received periodontal treatment, oral hygiene instruction, and tooth polishing at their follow-ups, while those assigned to control underwent only brief oral exams. Regardless of assignment, all participants attend 4 follow-up visits where periodontal data are collected.

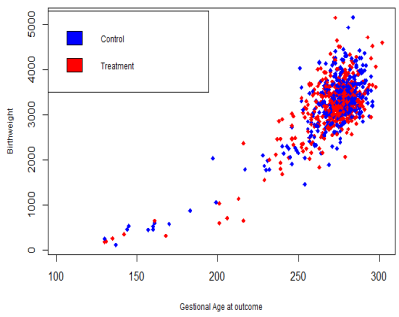
The primary outcomes of interest are *gestational age* at end of pregnancy and the *birthweight*. Moreover we have 171 variables regarding clinical measures of periodontal disease, birth outcomes and general health data such as *drug use*, *diabetes*, *BMI* and *hypertension*.

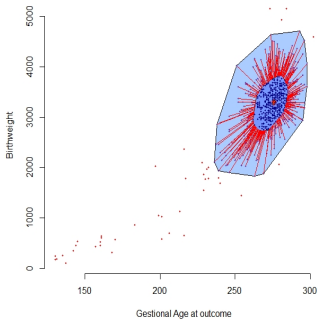
The main *goal* of our project is to determine whether treatment of maternal periodontal disease can reduce risks of pre-term birth and low birthweight.

Data exploration

First of all we focus on the scatterplot of our major outcomes, *gestational age* and *birthweight*: looking at the tail of severe cases in the left-bottom region, we believe that a nonparametric approach can be more suitable for our purpose. Morever we notice that we have a prevalence of controlled patients in these severe cases.

The bagplot shows that the bulk of the data do not present risky values of the outcomes while we have a lot of outliers corresponding to premature and underweight newborns: in the following we will consider also the sever cases only via thresholding in order not to get spoiled by the majority of the data.





Nonparametric Tests

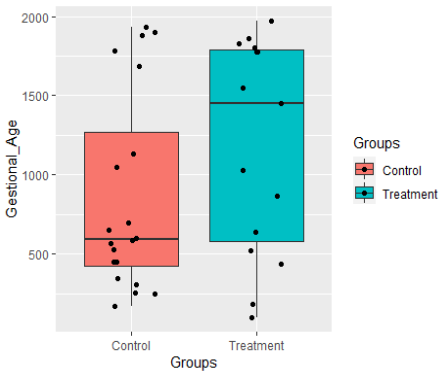
We perform two-populations nonparametric t-test to assess the differences in treatment and control groups with respect to several outcomes and variables:

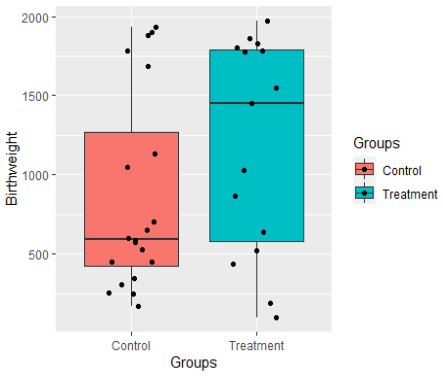
Variables	Group	Test	p-value
Gestational age	C/T	T-test median	0.061
Birthweigth	C/T	T-test median	0.07
%Bacteria at 5th visit	C/T	T-test median	0
GE at 5th visit	C/T	T-test median	0

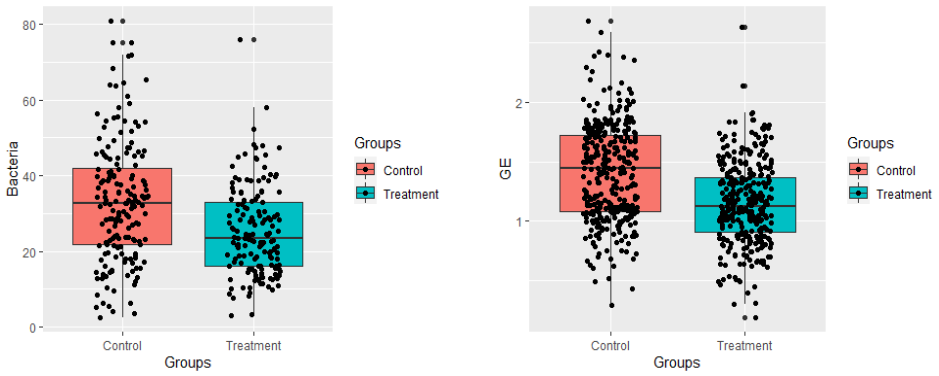
where GE is a gingival index.

Some variables of dental health are naturally different in control and treatment groups since treatment is required for severe patients and it improves dental health. Moreover we observe that *birthweight* and *gestational age* are differently distributed in the control and treatment groups; notice that we have considered severe patients only in order to achieve these results. Since treated patients show higher birthweights and higher delivery times, this is an evidence of the reduction of the risk for these pathologies due to treatment.

These results are supported also by the following plots where we can see that the treatment helps to reduce the risks of pre-term birth and low birthweight and improves the oral health





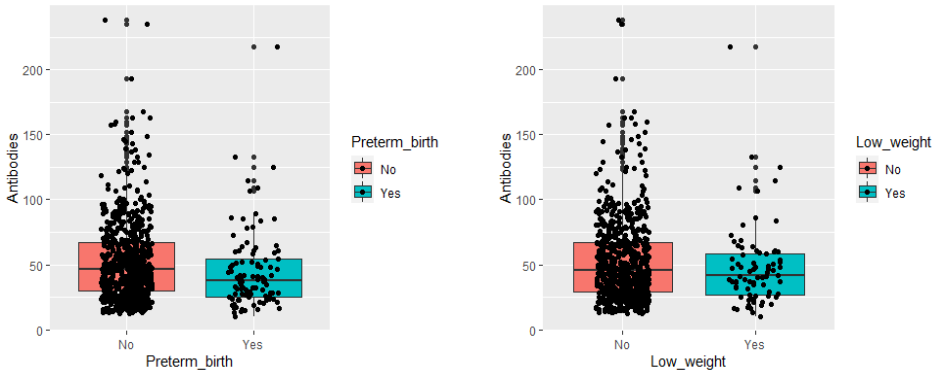


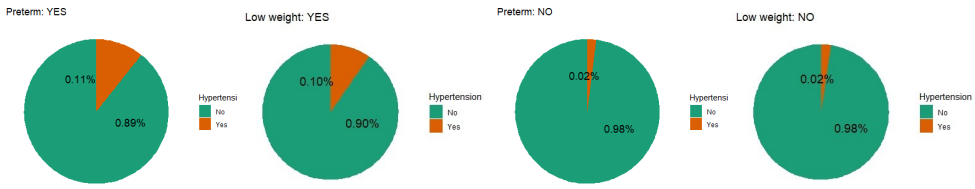
Moreover we perform two-populations nonparametric t-test to find risk factors for pre-term birth and low birthweight

Variables	**Group**	**Test**	**P-value**
Hypertension	Pre-term Y/N	T-test prop	0
Hypertension	Low weight Y/N	T-test prop	0.003
Diabetes	Pre-term Y/N	T-test prop	0.002
BMI	Pre-term Y/N	T-test mean	0.005
Total Bacteria at 5th	Pre-term Y/N	T-test median	0.067
Antibodies	Pre-term Y/N	T-test median	0.015
Antibodies	Low weight Y/N	T-test mean	0.0302

We can see which features contribute to have pre-term birth and low birthweight from the p-values above: for example diabetes and hypertension are more frequent in pre-term births, in addition less bacteria and more antibodies prevent these pathologies.

Even graphically we can see that *hypertension* is a risk factor and that *antibodies* reduces the risks of pre-term birth and low weight





As a final test useful for the following analysis, we prove that there is statistical evidence to assess that treatment improves the oral health and gives lower gingival indices at the follow-up visits. Moreover treated patients have also higher gingival indices at the baseline: indeed only the severe cases at the first visit will be in the treatment group.

Correlation analysis

In order to confirm the link between periodontal disease and our outcomes, we perform a correlation analysis to investigate the interactions between periodontal risk factors and other variables such as *diabetes*, *hypertension* or *BMI*.

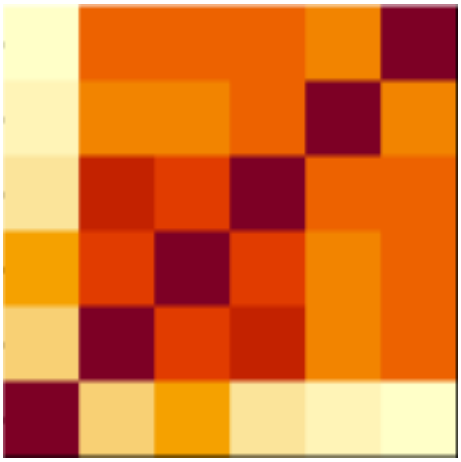
First of all we notice that correlations between these risk factors are not high: they act separately and they contribute to the problem from different points of attack. Therefore periodontal diseases are risk factors that are directly linked with low birthweights and pre-term births.

We investigate in addition the differences of the correlation matrices of risk factors in the control and treatment groups via a nonparametric permutational test

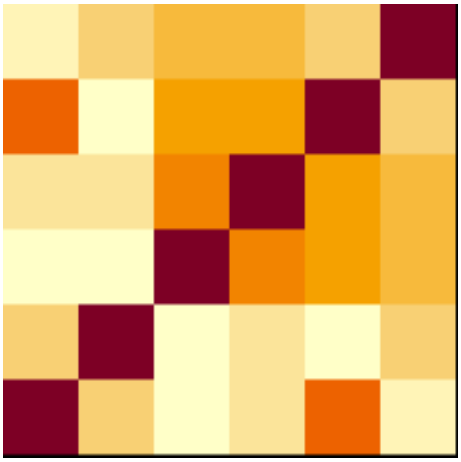
Variables	**Group**	**Test**	**P-value**
Hypertension, Diabetes, BMI, Bacteria, Antibodies	C/T	Permutational test on correlation matrices	0.05
Hypertension, Diabetes, BMI, V3 GE, V3 PD avg	C/T	Permutational test on correlation matrices	0.034

where *V3 GE* and *V3 PD avg* are gingival indices at the 3th visit.

Hence we have a significant difference in the correlation matrices of treatment and control group; in particular, the treatment reduces the interactions and correlations of the risk factors as we can see from the following plot



Correlation matrix
in control group



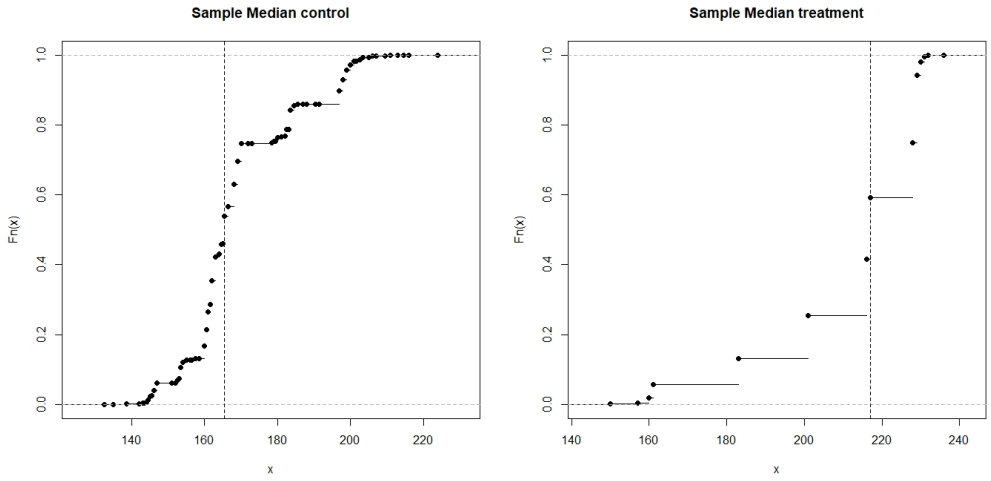
Correlation matrix
in treatment group

In conclusion, we can see again that the treatment reduces the risks of low birthweight and pre-term births giving less correlation between covariates and preventing that all the risk factors assume high values together.

Bootstrap

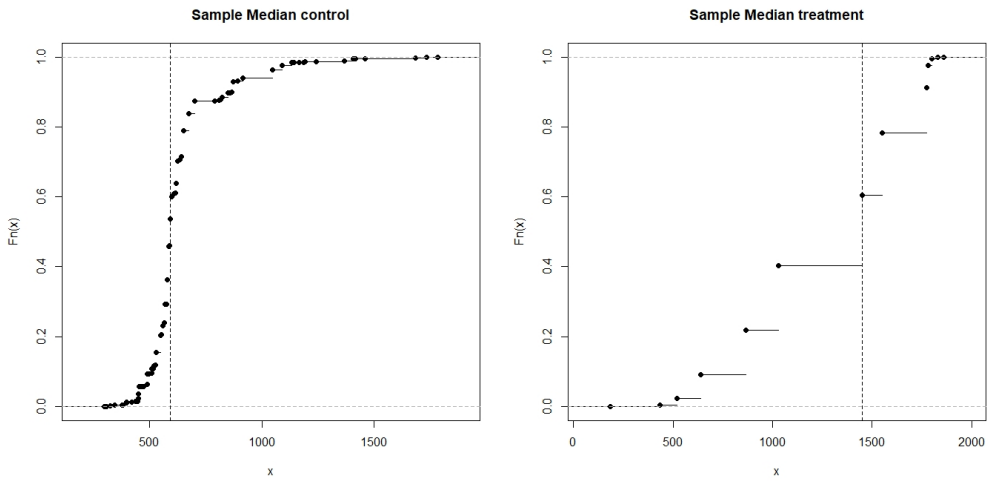
In order to understand if there are some differences between the distributions of the median in control and treatment group, we consider a bootstrap analysis. In particular, we analyze the variables *gestational age* and *birthweight* considering only the most severe cases thanks to a threshold (240 days for gestational age and $2000g$ for birthweight).

Regarding the *gestational age*, we have very different point estimates for the median in control group (165 days) and treatment group (217 days); this difference is supported by the fact that also their bootstrap distribution are concentrated on different values with smaller values in control group. Hence patients in treatment group seem to have less risk to present premature pregnancy.



Sample median distribution of severe gestational age at outcome in Control and Treatment groups

We find coherent results for *birthweight*: both the point estimates for the median and its bootstrap distribution are very different between the control group ($592g$) and the treatment group ($1450g$); in particular, the distribution of the median in control group is concentrated on smaller values. Similarly as before, this means that women in control group have typically underweight children.

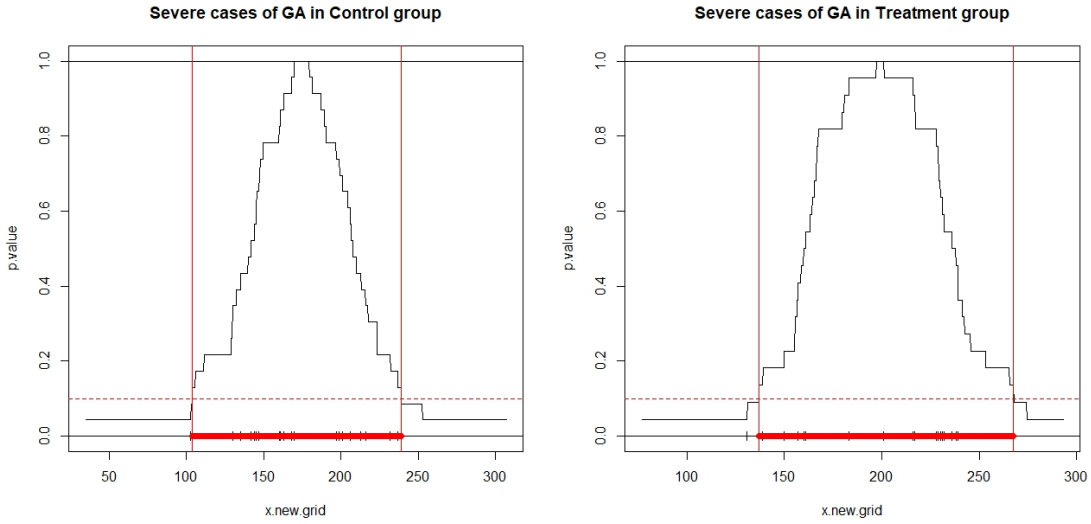


Sample median distribution of severe birthweight in Control and Treatment groups

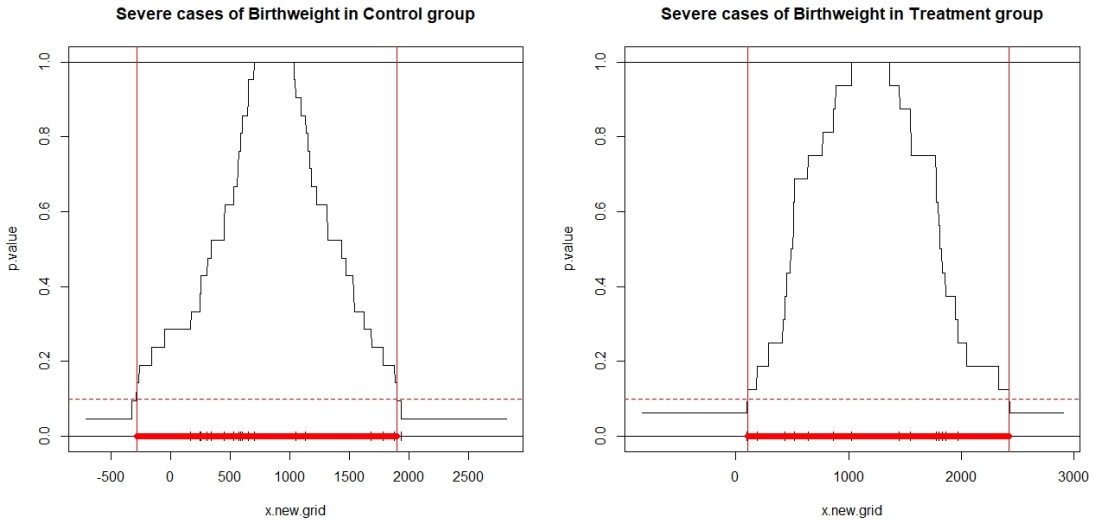
Conformal Prediction

In order to predict numerically the outcomes in control and treatment groups, we perform also a conformal prediction analysis taking into account the most severe cases only.

Even though there are not very relevant differences with the intervals that overlap, we can notice that the confidence interval of *gestional age* in treatment group recovers larger values suggesting that we have some hope to avoid pre-term births.



Also looking at *birthweight*, we have similar results with confidence interval presenting larger values in treatment group.

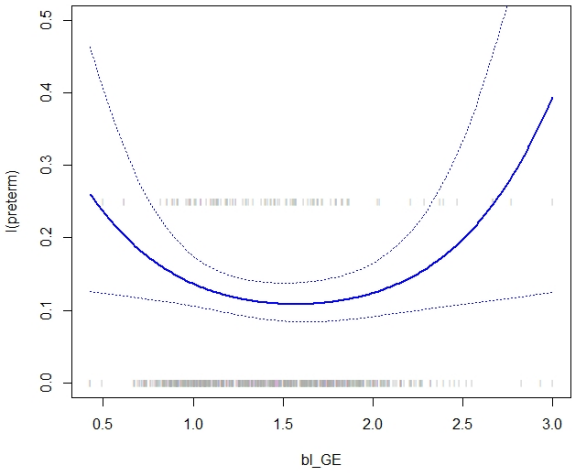


Logistic regression

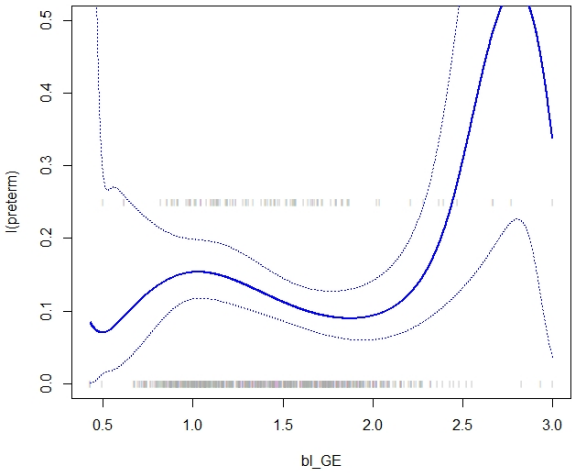
We try to build a model for the dichotomous variable *pre-term* which indicates whether the pregnancy ended before 259 days; in order to do that, we used the periodontal variables at the baseline as covariates. In particular we have

- *BL GE*: the Silness-Löe Gingival Index that has higher values in more severe inflammations
- *BL %BOP*: fraction of oral sites bleeding
- *BL PD avg*: whole-mouth average pocket depth
- *BL CAL avg*: whole-mouth average clinical attachment level
- *BL Calc I*: whole-mouth average calculus index
- *BL Pl I*: whole-mouth average plaque index
- *BL DNA*: Total amount of bacterial DNA extracted from plaque as a measure of total bacterial concentration

After many trials, we focus on a model which uses *BL GE* as covariates and we select a 2-degree polynomial looking at the p-value of the χ^2 test.

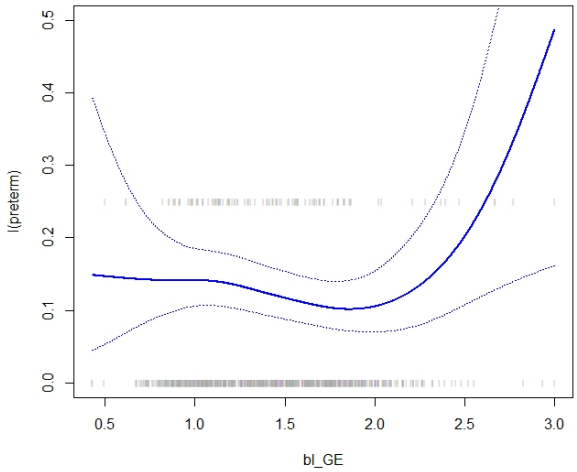


We can see that the probability of having a pre-term pregnacy increases for extreme values of the index; since in these regions we have very few data, we consider a 3-degree B-Spline regression using knots where we have less data



We notice again that the probability of having a pre-term birth increases a lot for high values of *BL GE*. Moreover we have an unexpected oscillatory behaviour in the middle; this could be caused by the control-treatment procedure: indeed the most severe cases at the baseline are treated and show an improvement of their periodontal disease, while the less severe patients are in control group and do not get better oral health. Therefore patients with *BL GE*$\simeq 2.0$ present a smaller probability of pre-term birth thanks to the treatment effect.

In addition, we can see a leverage effect on the right side due to the extreme values of *BL GE*; in order to fix it we fit another model using Smoothing Splines penalized by the second derivative penalty, obtaining a p-Value of the χ^2 test of 0.094.

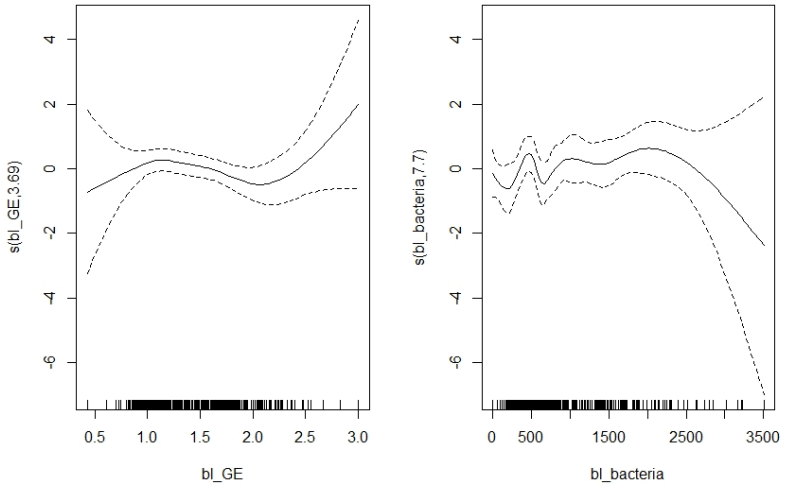


.

As a final step we fit different Generalized Additive Models using *BL GE* and other baseline features; then we compare the models using the Akaike information criterion (AIC).

Variables	**DF**	**AIC**
BL GE, BL PD avg	5.027	618.109
BL GE, BL Pl I	4.975	617.830
BL GE, BL Calc I	4.948	619.086
BL GE, BL %BOP	5.272	617.786
BL GE, BL DNA	12.394	413.845
BL GE	3.946	617.110

The best model seems to be the one which uses the *Silness-Löe Gingival Index* and the bacterial concentration.



Survival analysis

As final attempt, since we have the delivery times and right-censored data, we consider a survival analysis to assess and quantify the risk and protective factors in our dataset.

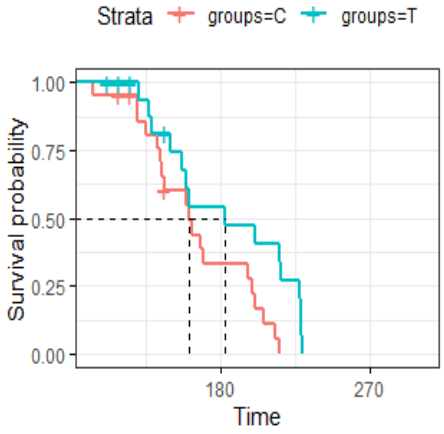
First of all, we consider categorical factors and we perform log-rank tests obtaining the following results

Group	**Threshold**	**p-value**	**Hazard Ratio**
Treatment/Control	≤ 240	0.03	0.5
Hispanic Y/N	//	0.02	0.83
Pubblic assistance Y/N	//	0.04	1.16
Hypertension Y/N	//	6e-06	2.36
Diabetes Y/N	//	2e-04	2.067
Use tobacco Y/N	//	0.008	1.32

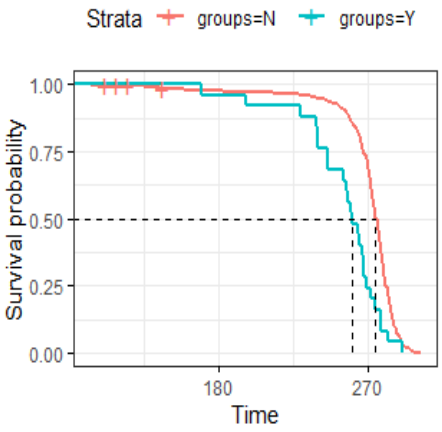
Therefore, looking at the earliest pre-term births only, we have a significance difference for the survivals curves in control and treatment group: furthermore we notice that treatment avoids severe cases and it is a protective factor with Hazard ratio equal to 0.5.

Moreover other categorical variables as *hypertension*, *diabetes*, *use of tobacco* are risk factors for the delivery times. Even *Pubblic assistance*, which indicates if the delivery is paid by the government, is a risk factor: indeed, with pubblic assistance, even poor people with bad health conditions can afford a delivery; while, without this facilitation, we are considering mainly

rich people that are usually more healthy since they can afford extra medical expenses.



Kaplan-Meier estimator for treatment and control groups

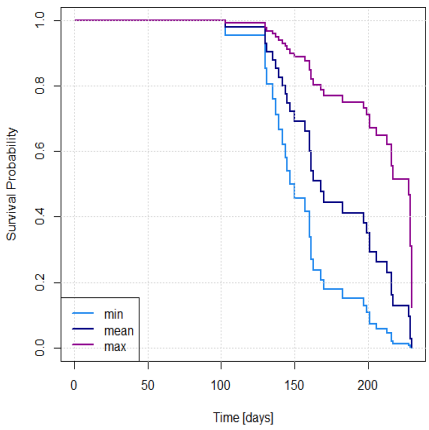


Kaplan-Meier estimator for hypertension

Let us consider now numerical covariates via Cox model

Variable	**Threshold**	**p-value**	**Hazard Ratio**
Age	//	0.934	1.01
BMI	//	0.0362	1.1
BL GE	≤ 240	0.046	0.436
BL PD avg	≤ 240	0.038	0.4965
V3 Calc I	≤ 240	0.0796	3.22

We confirm that *BMI* is a risk factor while *Age* is not significant. Moreover we observe the counter-intuitive behaviour already encountered in logistic regression: gingival indices at the baseline such as *BL GE* and *BL PD avg* are protective factors with respect to pre-term births if we consider severe cases only; instead they become risky when measured at the following visits.



Survival curve for
different values of BL GE

This is coherent with the treatment effect: higher gingival indices at the baseline imply the treatment for that patient which improves the oral health and ends up with a lower risk of pre-term births. Hence having a bad oral condition at the baseline results in a protective factor thanks to the treatment. Instead, if we look at the gingival indices at the following visits, they become risk factors since treatment is already in action and it has already improved the oral health.

Conclusions

In conclusion, thanks to nonparametric techniques, we have seen that periodontal diseases are quite directly linked with delivery problems and we have several arguments in favour of the reduction of the pre-term birth and low birthweight risks due to the treatment.

Finding a proper relationship between periodontal indices and our outcomes is a tough problem due to the problem complexity, to the treatment procedure that changes these periodontal variables and to the presence of other risk factors; as next step, it could be useful to make a further analysis using a dataset of patients without treatment effects.

References

[1] https://higgi13425.github.io/medicaldata/reference/opt.html

[2] Michalowicz BS, Hodges JS, DiAngelis AJ, Lupo VR, Novak MJ, Ferguson JE, Buchanan W, Bofill J, Papapanou PN, Mitchell DA, Matseoane S, Tschida PA. "Treatment of periodontal disease and the risk of preterm birth." New England Journal of Medicine. 2006 Nov 2. 355 (18): 1885-94.