99_paper

randy

2022-04-19

Introduction

The motivation for applying the broken stick model was to facilitate the statistical analysis and testing of critical ages in the onset of childhood development with extensions to multiple imputation (van Buuren 2018b).

The broken stick model fits in the latter tradition, and features an intuitive parametrisation of each individual growth curves by a series of connected straight lines.

Methods Study population

General descriptions:

(why it is reasonable and useful)

- **what type of study**: The EPIC Observational Study is a prospective, multicenter, observational longitudinal study

- **multiple center**: The data was collected through the Cystic Fibrosis Foundation Patient Registry (CFFPR)

- **already in use**: The design of the EPIC study has been previously reported [28,29]. Including 59 centers

- **exclusive informations**: Include demographic and risk factor data, respiratory cultures, clinical encounter information, and clinical outcomes, such as bacterial infections or pulmonary exacerbations

- **what is the goal**:

- **our goal**:

Subjects

Variable	testing, N = 457[1]	training, N = 913[1]
Genotype		
Two alleles F508del	233 (51%)	510 (56%)
One allele F508del	158 (35%)	319 (35%)
Others or Unknown	66 (14%)	84 (9.2%)
Gender		
Female	240 (53%)	453 (50%)
Male	217 (47%)	460 (50%)
Race		
Other	16 (3.5%)	46 (5.0%)
White	441 (96%)	867 (95%)
Ethnicity		
Hispanic	19 (4.3%)	24 (2.7%)
Non-Hispanic	421 (96%)	853 (97%)
Unknown	17	36
Visit number	45 (31, 62)	45 (32, 59)
Age mean	8.16 (6.68, 9.93)	8.11 (6.69, 9.98)
Age baseline	3.14 (3.06, 3.23)	3.13 (3.05, 3.22)
Age final	12.7 (10.2, 16.0)	12.9 (10.3, 15.9)
Follow up years	9.6 (7.0, 12.8)	9.7 (7.1, 12.7)
Height mean	125 (117, 134)	125 (117, 135)
Height baseline	93.7 (91.8, 97.0)	94.0 (91.1, 96.8)
Weight mean	27 (22, 34)	27 (22, 34)
Weight baseline	14.10 (13.20, 15.40)	14.10 (13.00, 15.30)

[1]n (%); Median (IQR)

- **check for time!!**: 1772 participants enrolled between *2004 and 2006*. The data cut-off for the current analysis was *December 2013*.

- The enrolled children with a confirmed diagnosis of CF before the age of 12 without exposure of *Pseudomonas aeruginosa*

- Based on either never infected or culture negative for at least 2 years); with respiratory cultures negative for *Pseudomonas aeruginosa*

- Either no prior isolation of Pseudomonas aeruginosa, or at least a two-year period with no isolation of P. aeruginosa (Treggiari et al., 2009).

- The beginning of follow-up was defined as an individual's earliest registry entry or pulmonary exacerbation

- a registry entry could have occurred before their enrollment in the EPIC Observational Study

- End of follow-up in the study was defined as a patient's latest CF registry encounter or pulmonary exacerbation.

- **sample size**: Of the **1772** children enrolled, **76497** visit observations in the EPIC OBS study

- we identified **1325** individuals with usable data

- 1325 for the final data analysis

- **how many** missing or lost to-follow-up with time-varying covariates and outcomes

- **%** had one survey record, **%** had a second, and **%** had a third, and the remaining **%** had **#th** surveys.

- Total of **# obs** in **# subject** individuals were identified after eliminating registration age over 5

- Total of **# obs** in **# subject** individuals were identified after eliminating follow up time less than 5 years

- Total of **# obs** in **# subject** individuals were identified after eliminating avaiable visit times less than 10

- **testing/training**: randomly

Table1 Information

- half of the cohort was female (49%)

- most were Caucasian (95%)

- Our genotype of interest was defined by the number of confirmed $\Delta F508$ (`F508del`) mutations (`0`: without mutation, `1`: with one allele mutated, or `2`: with both alells mutated), where `0` included both those with both alleles missing or those with unknown status with respect to this mutation.

- Homozygosity of $\Delta F508$ ($\Delta F508/\Delta F508$)

- the most common CF-causing mutation (Cystic Fibrosis Foundation, 2016) is associated with increased risk of infection and lower lung function in CF (Rosenfeld et al., 2012, Sanders et al., 2010).

- the majority were F508del homozygous (**%**),

- the median weight percentile at enrollment was 36.7 **(IQR:)**.

- The median age at enrollment was 5.8 **(IQR:)** years

- median length of follow-up 7.8 **(IQR:)** years

Tools

the **brokenstick v 2.0.0** package. The package contains tools to fit the broken stick model to data, export the fitted model's parameters, create imputed values of the model, and predict broken stick estimates for new data.

Results

Substantive researchers often favour repeated measures over the use of linear mixed models because of their simplicity.

For example, we can easily fit a subject-level model to predict future outcomes conditional on earlier data with repeated measures data.

While such simple regression models may be less efficient than modelling the complete data (Diggle, Heagerty, Liang, and Zeger 2002, Sec. 6.1),

The broken stick model requires a specification of a sensible set of time points at which the measurements ideally should have been taken.

- For each subject, the model predicts or imputes hypothetical observations at those times

- so the substantive analysis applies to the repeated measures instead of the irregular data.

applications for brokenstick model:

• to approximate individual trajectories by a series of connected straight lines; • to align irregularly observed curves to a joint age grid; • to impute realisations of individual trajectories; • to estimate the time-to-time correlation matrix; • to predict future observations.

Discussion

Reference

Supplementary

Variable	N = 1,370[1]
Genotype	
Two alleles F508del	743 (54%)
One allele F508del	477 (35%)
Others or Unknown	150 (11%)
Gender	
Female	693 (51%)
Male	677 (49%)
Race	
Other	62 (4.5%)
White	1,308 (95%)
Ethnicity	
Hispanic	43 (3.3%)
Non-Hispanic	1,274 (97%)
Unknown	53
Visit number	45 (32, 60)
Age mean	8.12 (6.68, 9.97)
Age baseline	3.13 (3.06, 3.22)
Age final	12.8 (10.2, 16.0)
Follow up years	9.6 (7.0, 12.7)
Height mean	125 (117, 134)
Height baseline	94.0 (91.4, 97.0)
Weight mean	27 (22, 34)
Weight baseline	14.10 (13.10, 15.30)

[1]n (%); Median (IQR)