**Environmental data do not improve a clinical asthma prediction tool for children**

Anina M. Pescatore, MSc1, Ben D. Spycher, PhD1, Maja Jurca, MD1, Erol A. Gaillard, MD, PhD2, Claudia E. Kuehni, MD, MSc1

1Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

2Division of Child Health, Department of Infection, Immunity and Inflammation, University of Leicester, Leicester, UK

**ONLINE REPOSITORY**

**Methods**

We used the R package glmnet to fit the penalized logistic regression. The parameter alpha was set to 1 so that only a LASSO (least absolute shrinkage and selection operator) type penalty was included. This tends to retain only the most influential predictors. The parameter λ, which determines the magnitude of the penalty, was set to a value that maximized the area under the receiver operating characteristic curve of resulting predictions in 10-fold cross-validation.1 If λ=0, this is equal to a conventional logistic regression including all potential predictors.

All potential predictors with more than 2 response categories were ordinal variables. We coded them as multiple dichotomous variables that represented all possible cut-off points, separating lower from higher categories. For instance, the number of cigarettes /day that a mother smoked (<1, 1-10, >10) was coded into two dichotomous variables indicating ≥1cigarette/ day and >10 cigarettes/day. This procedure resulted in 30 binary variables that entered variable selection in addition to the risk score of the Childhood Asthma Risk Assessment Tool (CARAT). Missing values in potential predictor variables did not exceed 5.5% (except for parental education; 11%) and were interpreted as the absence of the respective risk factor where possible, or were recoded with the most common category of the variable. Data were prepared using Stata 12.0 and analysed using R version 2.15.2.

**References**

1. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010; 33:1-22.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **TableS1. Associations of environmental and socioeconomic factors at age 1-3 years with asthma at age 6-8** | | | | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  | Unadjusted models | | | Score-adjusted models | | | Full model | | |
| Potential predictor |  | OR | 95% CI | p-value | OR | 95% CI | p-value | OR | 95% CI | p-value |
|  |  |  |  |  |  |  |  |  |  |  |
| **Environmental exposures** |  |  |  |  |  |  |  |  |  |  |
| Ethnicity | South Asian | 0.79 | (0.59,1.06) | 0.11 | 1.25 | (0.90,1.75) | 0.19 | 1.55 | (0.97,2.46) | 0.07 |
| Nursery care |  | 0.86 | (0.67,1.11) | 0.25 | 0.69 | (0.52,0.92) | 0.01 | 0.66 | (0.49,0.89) | 0.01 |
| Older siblings | ≥1 | 1.06 | (0.81,1.38) | 0.69 | 0.95 | (0.70,1.29) | 0.74 | 0.95 | (0.68,1.32) | 0.76 |
|  | >2 | 1.25 | (0.83,1.90) | 0.28 | 1.03 | (0.64,1.65) | 0.91 | 1.15 | (0.67,1.96) | 0.61 |
| Heating | gas, coal, other (vs. central heating only) | 1.07 | (0.81,1.41) | 0.62 | 1.13 | (0.83,1.54) | 0.45 | 1.15 | (0.83,1.60) | 0.40 |
| Cooking fuel | gas, other (vs. electrical stove only) | 0.69 | (0.52,0.91) | 0.01 | 0.91 | (0.66,1.25) | 0.55 | 0.82 | (0.58,1.16) | 0.27 |
| Pet ownership | cat | 1.00 | (0.72,1.38) | 1.00 | 0.90 | (0.62,1.30) | 0.57 | 0.91 | (0.62,1.35) | 0.65 |
|  | dog | 1.13 | (0.82,1.55) | 0.47 | 1.05 | (0.73,1.50) | 0.80 | 1.05 | (0.71,1.58) | 0.80 |
|  | other furry pet | 1.47 | (0.99,2.18) | 0.06 | 1.12 | (0.71,1.77) | 0.63 | 1.19 | (0.73,1.96) | 0.48 |
|  | bird | 0.87 | (0.46,1.65) | 0.67 | 0.80 | (0.38,1.67) | 0.55 | 0.74 | (0.34,1.61) | 0.45 |
| Mother smoking during pregnancy |  | 1.14 | (0.80,1.62) | 0.46 | 0.97 | (0.65,1.45) | 0.90 | 0.70 | (0.37,1.30) | 0.25 |
| Mother smoking (number of cigarettes /day) | ≥1 | 1.39 | (1.03,1.89) | 0.03 | 1.15 | (0.81,1.64) | 0.42 | 1.33 | (0.74,2.38) | 0.35 |
|  | >10 | 1.65 | (1.09,2.49) | 0.02 | 1.57 | (0.97,2.53) | 0.07 | 1.70 | (0.85,3.39) | 0.13 |
| Other person smoking in household (number of cigarettes /day) | ≥1 | 0.84 | (0.63,1.14) | 0.27 | 0.91 | (0.65,1.27) | 0.57 | 0.76 | (0.47,1.21) | 0.25 |
|  | >10 | 1.10 | (0.73,1.65) | 0.66 | 1.13 | (0.71,1.78) | 0.61 | 1.39 | (0.73,2.63) | 0.32 |
| Breastfed (months) | any duration (vs. no breastfeeding) | 0.79 | (0.62,1.02) | 0.07 | 0.92 | (0.70,1.29) | 0.55 | 1.09 | (0.66,1.80) | 0.74 |
|  | ≥1 | 0.76 | (0.59,0.98) | 0.03 | 0.85 | (0.64,1.13) | 0.25 | 0.70 | (0.40,1.24) | 0.22 |
|  | ≥4 | 0.83 | (0.63,1.09) | 0.19 | 0.95 | (0.70,1.30) | 0.75 | 1.13 | (0.65,1.95) | 0.66 |
|  | >6 | 0.88 | (0.63,1.22) | 0.44 | 1.01 | (0.70,1.46) | 0.95 | 1.06 | (0.62,1.82) | 0.82 |
| Self-reported traffic density (at home) | at least moderate | 0.91 | (0.71,1.17) | 0.47 | 0.86 | (0.64,1.14) | 0.30 | 0.85 | (0.62,1.17) | 0.31 |
|  | high | 0.74 | (0.47,1.15) | 0.18 | 0.87 | (0.53,1.43) | 0.59 | 0.93 | (0.55,1.59) | 0.79 |
|  |  |  |  |  |  |  |  |  |  |  |
| **Socioeconomic factors** |  |  |  |  |  |  |  |  |  |  |
| Crowding (persons/room) | > 1 | 0.81 | (0.60,1.10) | 0.18 | 0.77 | (0.55,1.09) | 0.15 | 0.67 | (0.43,1.04) | 0.08 |
|  | > 1.5 | 0.71 | (0.39,1.28) | 0.25 | 0.88 | (0.46,1.69) | 0.70 | 1.04 | (0.49,2.19) | 0.92 |
| Single parents |  | 1.32 | (0.89,1.95) | 0.17 | 0.87 | (0.55,1.36) | 0.53 | 0.90 | (0.54,1.51) | 0.70 |
| High parental education |  | 1.02 | (0.79,1.32) | 0.86 | 1.13 | (0.85,1.51) | 0.40 | 1.15 | (0.84,1.58) | 0.39 |
| Townsend deprivation index\* | more affluent | 0.94 | (0.69,1.28) | 0.68 | 0.92 | (0.65,1.31) | 0.65 | 1.01 | (0.64,1.60) | 0.97 |
|  | affluent | 0.88 | (0.68,1.13) | 0.31 | 0.88 | (0.66,1.17) | 0.36 | 0.97 | (0.62,1.52) | 0.90 |
|  | deprived | 1.15 | (0.89,1.49) | 0.28 | 1.21 | (0.90,1.62) | 0.21 | 1.48 | (0.92,2.40) | 0.11 |
|  | more deprived | 1.00 | (0.73,1.36) | 0.98 | 0.93 | (0.65,1.33) | 0.71 | 0.76 | (0.46,1.24) | 0.26 |
| Living in an urban area† |  | 0.97 | (0.76,1.25) | 0.82 | 1.10 | (0.83,1.46) | 0.51 | 1.20 | (0.83,1.73) | 0.33 |
| Range: 0 to 15 points, 0 represents low risk for having asthma 5 years later, 15 high risk1 | | | | | | | | | | |
| \*The categories are cut-offs between the following Townsend Deprivation Index intervals: [-5.522, -2.981], [-2.886, -1.264], [-1.250, 0.908], [0.909, 4.403], [4.418, 11.072] | | | | | | | | | | |
| †Living in Leicester post code areas LE1 to LE5 | | | | | | | | | | |