The Leicester Cohorts

Phenotype definition in childhood wheezing illness: a data-driven approach

BD Spycher[1], M Silverman[2], CE Minder[1], CE Kuehni[1]

1) Dept. of Social and Preventive Medicine, University of Berne, Switzerland

2) Division of Child Health, University of Leicester, UK

Introduction

There is strong evidence suggesting that disease conditions classified as childhood asthma actually consist of a range of separate disorders. Attempts to distinguish between phenotypes have typically been based on few pre-selected criteria such as the occurrence of wheeze in various age windows ("early transient wheeze", "persistent wheeze", "late onset wheeze")[1] or triggers of wheeze attacks ("viral wheeze", "multiple trigger wheeze").[2]

Aims

To identify and define phenotypes of childhood wheeze using a data-driven approach.

Methods

From a **population-based cohort study**, we included all preschool children for whom parents reported wheeze or chronic cough in at least one of two surveys (N=323). The surveys were carried out two years apart when the children were aged 1-5 and 4-8 years respectively.

We employed a **probabilistic clustering approach (finite mixture modelling)** which allowed for the simultaneous modelling of both qualitative and quantitative data.[3] In this approach disease phenotypes are modelled as distinct probability distributions (components of the model) for a given set of observed features. Each component of the model is taken to represent a phenotype. Once the model has been fitted the estimated parameters for each component can be interpreted in terms of probabilities of observing a particular combination of features in a child given that it belongs to the respective phenotype.

We included the following set of features:

- age, gender
- symptoms reported in both surveys (repeated measures): wheeze ever, frequency of attacks, shortness of breath, triggers of wheeze (colds/other), seasonal variation (winter/summer), diurnal variation (day/night), awake at night with cough, triggers of cough (colds/other)
- physiological examinations: skin-prick tests, lung function and airway responsiveness (AR)

Steps of analysis included:

1. Presetting the number of model components
2. Fitting the model for the given number of components
3. Assigning each child to the component (phenotype) which it was most likely to belong to, given its observed features and the estimated parameters
4. Repeating steps 1 to 3 for varying numbers of groups

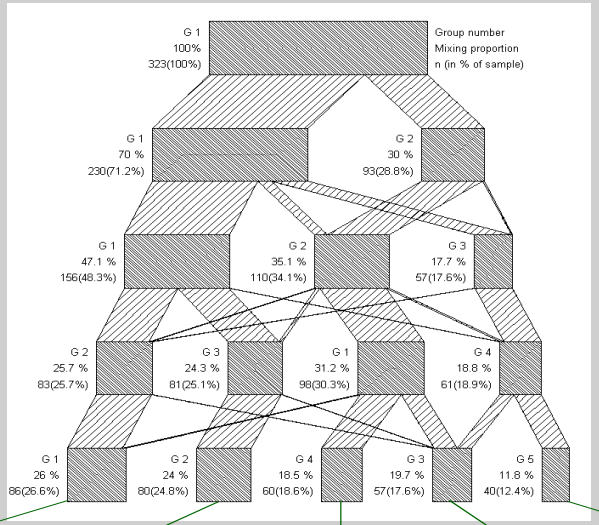
Contact: Claudia Kuehni, kuehni@ispm.unibe.ch

Results

Using a stepwise approach the model was fitted with the number of components varying from 2 to 5.

A distinction between "coughers" (children with cough without colds but no wheeze) and "wheezers" (children with wheeze ever) began at an early stage. At the 3-component level the model identified a group of "coughers only" with no wheeze ever (47% of the sample), a group of "persistent wheezers" with a high rate of atopy, persistent symptoms and multiple triggers (35%), and a group of "transient wheezers" with wheeze ever (usually only with colds) but no or only infrequent current attacks in the second survey (18%).

Increasing the number of components in the model led to a division of the coughers into a persistent and a transient group (4-component model) and in a next step (5-component model) a new group of wheezers was formed which was somewhat intermediate between the persistent and the transient group (below some typical features of these phenotypes).



Clustering flow chart

The figure illustrates the formation of clusters as the number of components is increased. Dark shaded boxes in a given layer represent clusters identified by a model with the corresponding number of components (box width proportional to cluster size). Light shaded bars connecting the layers represent the proportion of cluster members at a given level who were recruited from the clusters of the previous level.

"persistent coughers"	"transient coughers"	"atopic wheezers"	"non-atopic wheezers"	"transient wheezers"
Never wheeze, cough without colds in both surveys, waking with cough in survey 1, reduced lung function and medium AR	Never wheeze, waking with cough and cough without colds in survey 1 but not in survey 2, good lung function and low AR	Frequent attacks in both surveys, wheeze also apart from colds, for some wheeze attacks mainly in summer months, waking with cough and cough without colds in both surveys, high prevalence of atopy, poor lung function, high AR	Younger children, current wheeze in survey 1, if current wheeze in survey 2 infrequent attacks and shortness of breath, wheeze mainly in winter, low prevalence of atopy , slightly reduced lung function & slightly increased AR	Older children, mainly boys, wheeze ever, infrequent wheeze in survey 1, no current wheeze in survey 2, no night cough or cough without colds, good lung function and low AR.

Conclusion – Discussion

- The findings provide **further evidence for the existence of different phenotypes of wheezing illness and chronic cough**, confirming and partly reconciling existing classifications.
- **Probabilistic clustering methods might be a useful and relatively objective tool for phenotype definitions** in childhood asthma and a source for new aetiological hypotheses.
- **The set of variables included and model specification need further consideration**
  - A model including important risk factors (not reported here) revealed phenotypes suggesting distinct aetiologies, e.g. a phenotype of persistent cough associated with parental smoking.
  - Models could be extended to allow for a severity gradient.
- **There is a need for tools to validate identified clusters**. Validation techniques may include:
  - Statistical goodness of fit measures and cluster validation techniques
  - Comparing results from independent cohorts
  - Comparing clusters with respect to clinically relevant endpoints such as prognosis or response to treatment

References

1 Martinez FD, Wright AL, Taussig LM, Holberg CJ, Halonen M, Morgan WJ. Asthma and wheezing in the first six years of life. New England Journal of Medicine 1995;332:133-8.

2 Kuehni CE, Davis A, Brooke AM, Silverman M. Are all wheezing disorders in very young (preschool) children increasing in prevalence? Lancet 2001;357:1821-5.

3 Hunt L, Jorgensen M. Mixture model clustering using the MULTIMIX program. Australian & New Zealand Journal of Statistics 1999;41:153-71.