Where do we stand with asthma phenotypes derived from data-driven methods? A systematic review

Rita Amaral, Tiago Jacinto, Bernardo Sousa-Pinto, João Fonseca

Abstract

Asthma phenotypes can be refined using methods without a priori assumptions (data-driven). We aimed to describe asthma phenotypes derived with data-driven methods, using variables easily measurable in a clinical setting, and to summarize their consistency.

Systematic review of the literature using PubMed, Scopus and Web of Science, with no date/language restrictions. We included studies reporting adult asthma phenotypes applying data-driven methods with variables that can be easily collected in a clinical setting. Exclusion criteria were studies focusing exclusively on children and/or genotyping and theory-driven methods. Studies were assessed by two independent reviewers.

Of the 3,757 articles identified, 52 were included. The most frequent number of phenotypes identified per study was 4 and 5 phenotypes. Most of the studies focused on patients from specialized centers(n=41;79%). Variables used for phenotyping substantially differ according to the samples' characteristics, data availability, study aim and data sources. Variables were matched and categorized into six groups: clinical(n=38), functional(n=32), socio-demographic(n=25), inflammation(n=23), atopy(n=19), and other(n=9). Studies evaluating the consistency of phenotypes(n=15;29%) followed ≥1 of the criteria: longitudinal stability, cluster repeatability, reproducibility, and validity.

Data-driven methods have been increasingly used to derive asthma phenotypes, particularly in specialized centers. However, there's significant heterogeneity that requires clinical expertise and

statistical know-how for their evaluation. Studies with population-based samples and reporting longitudinal consistency of data-driven phenotypes are scarce.