Computational Phenotyping of Obstructive Airway Diseases: Protocol for A Systematic Review

Muwada Bashir (✉ muwada.bashir@gu.se)

University of Gothenburg   https://orcid.org/0000-0003-3901-5384

Rani Basna

Goteborgs Universitet

Guo-Qiang Zhang

Goteborgs universitet Institutionen for medicin

Helena Beckman

Umea Universitet

Anne Lindberg

Umea Universitet

Linda Ekerljung

Goteborgs universitet Institutionen for medicin

Malin Axelsson

Malmo Universitet

Linnea Hedman

Umea Universitet

Lowie Vanfleteren

Goteborgs universitet Institutionen for medicin

Bo Lundbäck

Goteborgs universitet Institutionen for medicin

Eva Rönmark

Umea Universitet

Bright I. Nuwaru

Goteborgs universitet Institutionen for medicin

Abstract

Background: Over the last decade, computational sciences have contributed immensely to characterization of phenotypes of airway diseases, but it is difficult to compare derived phenotypes across studies, perhaps as a result of the different decisions that fed into these phenotyping exercises. We aim to perform a systematic review of studies using computational approaches to phenotype obstructive airway diseases in children and adults.

Methods and analysis: We will search PubMed, EMBASE, Scopus, Web of Science, Google scholar for papers published between 2010 and 2020. Conferences proceedings, reference list of included papers, and experts will form additional sources of literature. Two reviewers will independently screen the retrieved studies for eligibility, extract relevant data, and perform quality appraisal of included studies. A third reviewer will arbitrate any disagreements in these processes. Quality appraisal of the studies will be undertaken using the Effective Public Health Practice Project quality assessment tool. We will use summary tables to describe the included studies. We will narratively synthesize the generated evidence, providing critical assessment of the populations, variables, and computational approaches used in deriving the phenotypes across studies

Conclusion: As progress continues to be made in the area of computational phenotyping of chronic obstructive airway diseases, this systematic review, the first on this topic, will provide the state-of-the-art on the field and highlight important perspectives for future works.

Ethics and dissemination: No ethical approval is needed for this work is based only on the published literature and does not involve collection of any primary or human data.

Registration: The protocol of this the review process is registered in PROSPERO with the number: CRD42020164898.

Background:

Asthma and chronic obstructive pulmonary diseases (COPD) are the most common chronic respiratory diseases worldwide, largely accounting for global mortality and morbidity burden (1, 2). While one fifth of the developed world population is expected to have asthma at certain time in their life especially in Europe (3), globally around 10 % of adults currently have COPD (4). By 2030, COPD is projected to be the fourth leading cause of death globally (5, 6). Other airway diseases, such as sinusitis, allergic rhinitis, although of lesser contribution to overall mortality, collectively can affect around 10-30% of the populations of western countries (4, 5). They also account for significant loss in societal productivity due to loss of working and schooling hours and treatment expenditure (7, 8).

Over the last decade, significant progress has been made regarding improving understanding of the pathophysiological and clinical features of obstructive airway diseases. Indeed, we know today that diseases such as asthma and COPD are not single disease entities as previously thought, rather they are

heterogeneous in nature and embedded with varied underlying phenotypes (9, 10). A phenotype is 'the observable and structural and functional characteristics of an organism determined by its genotype and modulated by its environment'(11). Better understanding of the phenotypes of airway diseases will provide the opportunity for targeted, individualized, and precise management of these diseases (12).

Generally, disease phenotyping falls into two areas: hypothesis-led approach and data-driven or computational approach. The hypothesis-led phenotyping relies on classifying diseases on the basis of the characteristics of the presenting patient and the general framework has been to rely on the clinical or physiological features, based on specific triggers and pathobiology of inflammation(11, 13). As no standard exists in such classifications, the clinician relies on the current knowledge of the disease and his own experiences and presumptions; consequently, the hypothesis-led approach is said to be largely subjective and may be potentially biased (14, 15). The data-driven approach to phenotyping works through development of high-level computer algorithms that automatically learn from data and try to uncover complex patterns in a systematic and meaningful way (16). Usually, no a priori theory is employed in learning from the data, rather the computer allows the data to 'speak for itself' and uncover hidden nuances that will enhance understanding and clinical decisions; consequently the data-driven approach to phenotyping is said to be unbiased (16). The advancement in machine-led computations and novel statistical methods in human diseases has facilitated the progress now being made in data-driven phenotyping of chronic obstructive airway diseases (17). Whilst the traditional clustering technique, like hierarchical clustering and partitioning methods, has remained the most frequently used conventional approach to disease phenotyping, several emerging machine-learning approaches, such as deep learning and probabilistic modelling, are providing advanced flavor to the phenotyping exercises(18).

Given the uncertainty of the underlying evidence and the rapid progress being made, the aim of this study is to identify, critically appraise, and synthesize data from studies that have so far used computational approaches to phenotype chronic obstructive airway diseases in children and adults. Specifically, we aim to:

1. Characterize and compare the populations included in studies of computational phenotyping of chronic airway diseases.
2. Assess and compare the criteria used to select participants included in studies of computational phenotyping of chronic airway diseases.
3. Evaluate and compare the variables used to derive phenotypes of chronic airway diseases across studies and assess the choices informing the included variables.
4. Describe and compare the computational approaches used across studies and highlight the features of each computational approach.
5. Describe the number and characteristics of phenotypes derived across studies and assess their clinical interpretation.

Methods:

Eligibility criteria:

We will include population-based studies that have used computational approaches to derive phenotypes of chronic airway diseases, whether conducted in the general population or in a clinical setting. We will exclude studies that have characterized phenotypes of chronic airway diseases based on hypothesis-based approaches.

Outcomes:

We will include studies focusing on computational phenotyping of the following chronic obstructive airway diseases:

- Asthma
- COPD
- Rhinitis
- Emphysema

Study type:

We will include observational general population-based and clinical epidemiological studies, including cohort, case-control, and cross-sectional. We do not anticipate computational phenotyping studies of airway diseases based on randomized clinical trials or other experimental study designs. Case studies and case series as well as ecological studies will be excluded.

Participants:

We will include studies conducted both in children and adults.

Years of consideration:

Studies conducted in the last ten years: (2010-2020) only will be considered for our review. The selected time window is the reported era of evolution of the use of computational approaches in phenotyping of chronic obstructive airway diseases (23).

Language:

There will be no language-based exclusions of studies, and we will endeavor to translate studies published in languages other than English.

Study identification:

To identify relevant studies for the review, we will search PubMed, EMBASE, Web of Science, Scopus and Google scholar. For unpublished materials, such as conferences proceedings, we will search databases of proceeding of conferences and databases of the grey literature, such as Open Grey. We will also contact experts in the field to request for any paper we may miss from our database searches. Finally, we will screen the reference lists of included studies to identify any additional paper.

Study selection:

Search strategy:

We have developed a preliminary search strategy to identify relevant studies for the review. The search strategy (Supplementary File 1) was developed in PubMed and will be adapted in searching the other databases.

Screening:

The search results from the different databases will be exported to Endnote for further screening. Two reviewers will independently screen the studies on the basis of the review inclusion and exclusion criteria; any discrepancies will be resolved by discussion or a third reviewer will arbitrate if a consensus is not reached. The first stage of the literature will involve removal of duplicates from the database searches; then we will perform title and abstract screening. The final stage will involve full-text screening of the studies potentially meeting the eligibility criteria not clearly identified from the titles and abstracts. We will document the screening process using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart (24) .

Data extraction:

Two reviewers will independently extract relevant data from included studies onto a data extraction form to be developed for the review; any discrepancies will be resolved by discussion or a third reviewer will arbitrate if a consensus is not reached. We will develop a data extraction form specifically designed for this review that will be used to capture relevant data from included studies. The form will initially be first piloted on two to three included studies; any amendment will be undertaken prior to using the form on all included studies.

Data items:

A minimum of the following data items will be collected from included studies onto the data extraction form: general information (authors name; publication year and study time; aim of the study and data source); information describing populations characteristics (population size, recruitment characteristics, sample size, children/adults, inclusion and exclusions criteria); type of airway disease; information about the variables selected for phenotyping (number and description of variables, rational of selection, variable measurement and definition); type and features of computational approach used; and information of the

derived phenotypes (number of phenotypes, characteristics of each phenotype, and clinical interpretation).

Quality assessment:

We will appraise the general quality of included studies using the Effective Public Health Practice Project (EPHPP), where focus of this tool will be in relation to each study's potential for selection bias; appropriateness of study design; data collection methods; withdrawals and dropouts and analysis (25). Since, to our knowledge, there are no standard tools for assessing the quality of studies on computational disease phenotyping, we will develop a preliminary checklist that will enable us to extract items related to the computational approaches used across studies and to help us compare approaches across studies.

Registration and reporting:

The full protocol for this systematic review is registered in the International Prospective Register of Systematic Reviews with the number CRD42020164898 according to the requirements of the PRISMA-P guideline(27, 28).

Results:

We will tabulate all data items extracted from studies, where a detailed descriptive narrative summary for each included study will be synthesized and presented. We do not aim to perform any quantitative summary (meta-analysis) for included studies as this is not the goal of the current work. However, we will employ narrative synthesis of the underlying evidence, focusing at least on the following aspects: strengths and limitations of included studies and computational approaches used; description of derived phenotypes across studies and their clinical relevance; issues of reproducibility of each phenotyping exercises; etc. (26).

Discussion:

The findings derived till date from studies using computational methods to phenotype chronic airway diseases have highlighted the importance of using these methods in delineating the heterogeneous nature of these diseases (14, 22, 29-31). Still, the question about the reproducibility and clinical relevance of derived phenotypes remains a valid one. Factors of population characteristics, variables used to derive disease phenotypes, computational approaches used, and characteristics of derived phenotypes and their comparability across studies are issues that demand further scrutiny.

The current review, the first on the topic, to our knowledge, is an attempt to address these overarching issues. Findings from the review will therefore contribute in advancing the field of computational phenotyping of chronic obstructive airway diseases.

Conclusion:

As progress continues to be made in the area of computational phenotyping of chronic obstructive airway diseases, systematically surveying the field and appraising the evidence so far generated will help identify potential research gaps and how to fill them. The evidence to be generated from the current systematic review will therefore provide the current state-of-the-art on the field and will highlight important perspectives for future works. This synthesis will give researchers in the area an accessible summary to guide their works in the use of computational approaches to phenotype chronic airway diseases.

Abbreviations:

COPD: chronic obstructive pulmonary disease

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

EPHPP: Effective Public Health Practice Project

Declarations:

Ethical approval:

For the purpose of this review, no primary patients or human data will be collected or retrieved, so there will be no need for ethics approval.

Consent for publication:

As no primary data is collected form human subjects, participants' consent for publication is not needed for this study. All authors participating in this study have thoroughly reviewed and agreed on publishing the content of this protocol manuscript.

Data availability:

The data and articles used in this review, along with the analysis codes will be availed through a repository sets that will be generated during the current study.

Funding:

Authors contribution:

M.B, R.B, G.Z, B.N significantly contributed to writing and drafting of this protocol manuscript. Other coauthors: H.B, A.L, L.E, M.A, L.H, L.V, B.L and E.R significantly contributed to reviewing, revising and final drafting of this article file. All authors contributed to draft versions of the manuscript, read and approved the final manuscript, and are accountable for the accuracy and integrity of this work.

References

1. https://ginasthma.org/. Global Strategy for Asthma Management and Prevention: Gobal initative for asthma; 2019 [report].
2. Global, regional, and national age-sex-specific mortality and life expectancy, 1950-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392(10159):1684-735.
3. Network GA. The Global Asthma Report 2014. Auckland, New Zealand: Global Asthma Network, 2014. 2019.
4. Halbert R, Natoli J, Gano A, Badamgarav E, Buist AS, Mannino D. Global burden of COPD: systematic review and meta-analysis. European Respiratory Journal. 2006;28(3):523-32.
5. disease(GOLD) Gifcol. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease.
6. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. 2006;3(11):e442.
7. Dykewicz MS, Hamilos DL. Rhinitis and sinusitis. J Allergy Clin Immunol. 2010;125(2):S103-S15.
8. Ray NF, Baraniuk JN, Thamer M, Rinehart CS, Gergen PJ, Kaliner M, et al. Healthcare expenditures for sinusitis in 1996: contributions of asthma, rhinitis, and other airway disorders. J Allergy Clin Immunol. 1999;103(3):408-14.
9. Wardlaw A, Silverman M, Siva R, Pavord I, Green R. Multi-dimensional phenotyping: towards a new taxonomy for airway disease. Clinical & Experimental Allergy. 2005;35(10):1254-62.
10. Weatherall M, Travers J, Shirtcliffe P, Marsh S, Williams M, Nowitz M, et al. Distinct clinical phenotypes of airways disease defined by cluster analysis. European Respiratory Journal. 2009;34(4):812-8.
11. Rice JP, Saccone NL, Rasmussen E. Definition of the phenotype. Adv Genet. 2001;42:69-76.

12. Vanfleteren LE, Kocks JW, Stone IS, Breyer-Kohansal R, Greulich T, Lacedonia D, et al. Moving from the Oslerian paradigm to the post-genomic era: are asthma and COPD outdated terms? Thorax. 2014;69(1):72-9.

13. Basile AO, Ritchie MD. Informatics and machine learning to define the phenotype. Expert Rev Mol Diagn. 2018;18(3):219-26.

14. Pinto LM, Alghamdi M, Benedetti A, Zaihra T, Landry T, Bourbeau J. Derivation and validation of clinical phenotypes for COPD: a systematic review. Respiratory Research. 2015;16(1):50.

15. Banda JM, Seneviratne M, Hernandez-Boussard T, Shah NH. Advances in Electronic Phenotyping: From Rule-Based Definitions to Machine Learning Models. Annu Rev Biomed Data Sci. 2018;1:53-68.

16. Pathak J, Kho AN, Denny JC. Electronic health records-driven phenotyping: challenges, recent advances, and perspectives. J Am Med Inform Assoc. 2013;20(e2):e206-11.

17. Che Z, Kale D, Li W, Bahadori MT, Liu Y. Deep Computational Phenotyping. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15; Sydney, NSW, Australia. 2783365: ACM; 2015. p. 507-16.

18. Basile AO, Ritchie MD. Informatics and machine learning to define the phenotype. Expert Rev Mol Diagn. 2018;18(3):219-26.

19. Weatherall M, Shirtcliffe P, Travers J, Beasley R. Use of cluster analysis to define COPD phenotypes. Eur Respir J. 36. England2010. p. 472-4.

20. Vazquez Guillamet R, Ursu O, Iwamoto G, Moseley PL, Oprea T. Chronic obstructive pulmonary disease phenotypes using cluster analysis of electronic medical records. Health Informatics J. 2018;24(4):394-409.

21. Burgel PR, Paillasseur J, Caillaud D, Tillie-Leblond I, Chanez P, Escamilla R, et al. Clinical COPD phenotypes: a novel approach using principal component and cluster analyses. Eur Respir J. 2010;36(3):531-9.

22. Castaldi PJ, Benet M, Petersen H, Rafaels N, Finigan J, Paoletti M, et al. Do COPD subtypes really exist? COPD heterogeneity and clustering in 10 independent cohorts. Thorax. 2017;72(11):998-1006.

23. Jaimini U, Thirunarayan K, Kalra M, Venkataraman R, Kadariya D, Sheth A. "How Is My Child's Asthma?" Digital Phenotype and Actionable Insights for Pediatric Asthma. JMIR Pediatr Parent. 2018;1(2).

24. Simons M, Busch K, Avolio A, Kiat H, Davidson A. Improving the quality of the evidence–The necessity to lead by example. J Clin Neurosci. 2017;46:165-6.

25. Yost J, Dobbins M, Traynor R, DeCorby K, Workentine S, Greco L. Tools to support evidence-informed public health decision making. BMC Public Health. 2014;14:728-.

26. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme Version. 2006;1:b92.

27. Muwada Bashir Awad Bashir, Rani Basna, Guo-Qiang Zhang, Helena Backman, Anne Lindberg,Linda Ekerljung, Malin Axelsson, Linnea Hedman, Lowie Vanfleteren, Bo Lundbäck, Eva Rönmark,Bright I Nwaru. Computational phenotyping of obstructive airway diseases: a systematic review.PROSPERO 2020 CRD42020164898 Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020164898.

28. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Systematic reviews. 2015;4(1):1.

29. Garcia-Aymerich J, Benet M, Saeys Y, Pinart M, Basagana X, Smit HA, et al. Phenotyping asthma, rhinitis and eczema in M e DALL population-based birth cohorts: an allergic comorbidity cluster. Allergy. 2015;70(8):973-84.

30. Halpern Y, Horng S, Choi Y, Sontag D. Electronic medical record phenotyping using the anchor and learn framework. J Am Med Inform Assoc. 2016;23(4):731-40.

31. Burgel P-R, Paillasseur J-L, Roche N. Identification of clinical phenotypes using cluster analyses in COPD patients with multiple comorbidities. BioMed research international. 2014;2014.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- PRISMAPchecklistfilled.docx
- Searchstrategysup.complete.pdf