Breaking Down Silos in Asthma Research: The Case for an Integrated Approach

Authors: *Sadia Haider, Adnan Custovic

Department of Paediatrics, Imperial College London, London, UK
*Correspondence to s.haider@imperial.ac.uk

Disclosure: Dr Haider has declared no conflicts of interest. Prof Custovic reports personal fees from Novartis, Regeneron/Sanofi, ALK, Bayer, ThermoFisher, GlaxoSmithKline, and Boehringer Ingelheim, outside the submitted work.

Received: 07.02.18

Accepted: 31.10.18

Keywords: Asthma, big data, data-driven methods, prediction, Study Team for Early Life Asthma Research (STELAR).

Citation: EMJ Innov. 2019;3[1]:82-92.

Abstract

Asthma is a complex condition with heterogeneous patterns of symptoms underpinned by different underlying pathophysiological mechanisms and treatment responses. Analyses of data from birth cohorts and patient studies, from the subjective hypothesis-testing approach to the data-driven hypothesis-generating approach, have improved the current understanding of asthma's heterogeneity. Despite the rapid proliferation of new sources of data and increasingly sophisticated methods for data mining and revealing structure, relatively few findings have been translated into clinically actionable solutions for targeted therapeutics or improved patient care. This review focusses on why an integrated approach may be a more powerful catalyst for improved patient outcomes compared with the artificial and imposed dichotomy of hypothesis-generating versus investigator-led subjective approaches. As the factors shaping the development and control of asthma affect individuals dynamically in response to treatment or environmental factors, deeper insights can be garnered through the integration of data with human expertise and experience. The authors describe how integrative approaches may have greater power to provide a more holistic understanding of the pathophysiological mechanisms driving asthma heterogeneity, discussing some of the key methodological challenges that limit the clinical use of findings from asthma research, and highlighting how recent examples of integrative approaches are building bridges to ensure that the power of emerging sources of data, coupled with rigorous scientific scrutiny, can lead to a more nuanced understanding of asthma.

INTRODUCTION

Over the last two decades, a substantial effort has been devoted to understanding the heterogeneity of asthma.[1-7] The architecture of wheezing illness during childhood has been described based on temporal patterns of symptoms using data-driven techniques applied to longitudinal data from birth cohort studies.[3-5,7-10] As a consequence, the conceptual framework of asthma heterogeneity is now accepted within the clinical and research communities.[1] However, the main aim of discovering 'asthma endotypes'[11] and their underlying pathophysiological mechanisms for

the identification and development of novel targeted therapeutics appears as elusive as ever.

In part, progress has been stymied by the methodological and disciplinary silos. The rapid increase in the ability to generate, share, and access large amounts of data, including longitudinal clinical information and biomarkers, various 'omics' technologies, and environmental exposures, coupled with advances in data-driven techniques to analyse high-dimensional data, has made it, on occasion, challenging to discern what problems we are seeking to address, or how findings are relevant in a real-world setting.[12] Given that big data sets may contain many thousands of variables, or differ in terms of the format or level of the data (e.g., clinical history, laboratory tests, environmental and behavioural factors, various biomarkers, proteomic data, and genome-wide genotyping), it is not possible to define a priori all possible causal and associational mechanisms. An integrated approach to research may enable the power of these resources to be harnessed in ways that translate into a better understanding of causal mechanisms, more accurate diagnoses, and more personalised treatment. The integration of data, methodologies, and human expertise to understand the results can only occur through cross-disciplinary research, with the central principle that basic scientists, geneticists, clinicians, and data scientists work together to understand the clinical heterogeneity of complex diseases and the mechanisms underpinning them.

In this review, the authors set out to describe the evolution of analytical frameworks in asthma epidemiology, from the subjective hypothesis-driven to the data-driven hypothesis-generating approaches; highlight why an integrated approach may be a more powerful catalyst for improved patient outcomes; and identify the key challenges faced by healthcare professionals in adopting findings to clinical practice.

EVOLVING FRAMEWORKS OF DATA ANALYSIS IN ASTHMA RESEARCH

Long-term follow-up in birth cohort studies has allowed a shift in emphasis in temporal perspectives from the static cross-sectional approach to a more dynamic longitudinal approach. By explicitly allowing for time in the mediation of disease development, the longitudinal approach has allowed us to establish whether individuals affected by symptoms of the disease at one point in time are the same individuals who have the disease at later time points, ascertain temporal variations across individuals in terms of the timing of onset or remission and the persistence and recurrence of episodes, and identify the risk factors that discriminate these different temporal patterns.[3,10] Analytical approaches have progressed from supervised analyses testing-specific hypotheses to statistical data-driven classification techniques.[3] In the former, typologies of disease or hypotheses are proposed by investigators or clinical experts, usually based on patterns of symptoms observed in a clinical situation.[13] The Tucson Children Respiratory Study (TCRS) was one of the first studies to use longitudinal data to differentiate childhood wheezing phenotypes based on the presence of temporal patterns or the absence of symptoms.[2] Three mutually exclusive phenotypes (transient early, late-onset, and persistent wheezing) were described from data collected at two time points (aged 3 and 6 years).[2] While such studies have been instrumental in introducing and confirming the idea of heterogeneity of childhood wheezing and asthma, the subjective approach has several potential limitations. For example, there is a risk of limiting the predictive ability of a model by restricting the set of inputs, imposing a structure that does not necessarily fit the data, failing to identify groups with truly distinct patterns, and/or missing rare patterns.

In contrast, data-driven algorithms enable the analysis of large quantities of complex data for the identification of hidden patterns within such datasets. Continuous advances in computational power allow pattern discovery in high-dimensional data to take place with increasingly greater efficiency. As data-driven techniques are hypothesis-neutral, they are useful for examining heterogeneity based on distinctions that are not known a priori, and for making predictions about outcomes while remaining agnostic towards specific predictors.[14] This has allowed for the discovery of patterns that could not have been predicted in advance. Numerous data-driven algorithms

have been applied in asthma research. For example, latent class trajectory models, which are a class of probabilistic models in which repeated measurements of manifest symptoms are modelled in order to derive homogenous subtypes, have been extensively applied to derive distinct wheeze and lung function trajectories.[4,7,8,15,16] One advantage of such methods is that objective statistical criteria are used for judging whether clusters (classes or subtypes) represent true variation in the population. Clusters discovered using data-driven approaches are not observed, but hidden, and should not be referred to as 'phenotypes'; however, as this term has widely been used in the literature, the authors will continue to use this nomenclature.[6]

Discovery of wheeze phenotypes using data-driven methods is susceptible to inconsistencies with respect to the number of discovered phenotypes, the size of each class, and the labels ascribed to them.[6] For example, a review, which compared wheeze phenotypes derived from latent trajectory modelling across 28 studies, found that the number of phenotypes ranged from 3–8. Another review found considerable differences in the size of 'common' phenotypes in different cohort studies[6] (e.g., there was up to a 10-fold difference in the proportion of children classified as late-onset wheezing [3.7–35.8%]).[7] The inconsistencies between studies may arise from differences in the number of data points, the length of the intervals between data collection points, the age at follow-up, and the study's duration, sample size, population differences, and definition of the symptom[17] (e.g., parentally-reported versus doctor-diagnosed).[18]

Bayesian analysis,[9] hidden Markov models,[9] and temporal clustering[19] have been applied to challenge the paradigm of the atopic march, which assumes that there is a natural progression of symptoms from eczema to asthma and rhinitis. This paradigm is based on observations using cross-sectional data on population prevalence. However, modelling longitudinal data within individual patients revealed heterogenous patterns, with <7% of children with any of these symptoms following the atopic march trajectory.[9] Other applications of machine-learning include Bayesian networks coupled with feature selection methods for the discovery of patterns of allergic sensitisation,[20-22] principal components analysis to investigate whether syndromes of co-existing respiratory symptoms could be derived using responses to >100 questions from validated questionnaires,[23] Bayesian estimation of a mixture of Bernoulli distributions to describe the architecture of IgE responses to multiple allergenic proteins during childhood,[24] Gaussian mixture model to cluster human blood cell cytokine responses to rhinovirus-16,[25] and the use of network analysis and hierarchical clustering to explore the connectivity structure of allergen component-specific IgE, which demonstrated that the interaction patterns of IgE rather than individual 'informative' components are associated with asthma.[26]

CHALLENGES TO BRIDGING THE GAP BETWEEN BIG DATA RESEARCH AND CLINICAL USE

Currently, there is no consensus on what the best approach should be to understand asthma heterogeneity, how best to identify distinct underlying pathophysiological mechanisms, and how to implement these findings in a clinically useful way.[12] One potential flow is summarised in Figure 1.

Identification of Children at High Risk of Asthma

Prediction modelling to identify individuals at a higher risk of asthma is important and was identified as the top research priority by the European Asthma Research and Innovation Partnership (EARIP).[27] Several algorithms have been proposed for predicting persistent asthma in school age using early-life features, including the Asthma Prediction Index,[28] the Isle of Wight score,[29] the PIAMA risk score,[30] and the Leicester[31] and Manchester scores.[32] However, these tools have not been widely adopted clinically.[33] A systematic review found that these tools typically have low sensitivity and positive predictive values, making them unsuitable for the precise identification of high-risk individuals in a clinical setting.[34] Given the heterogeneity of asthma, algorithms may be required to predict different 'asthmas' instead of a one-size-fits-all tool.[35]

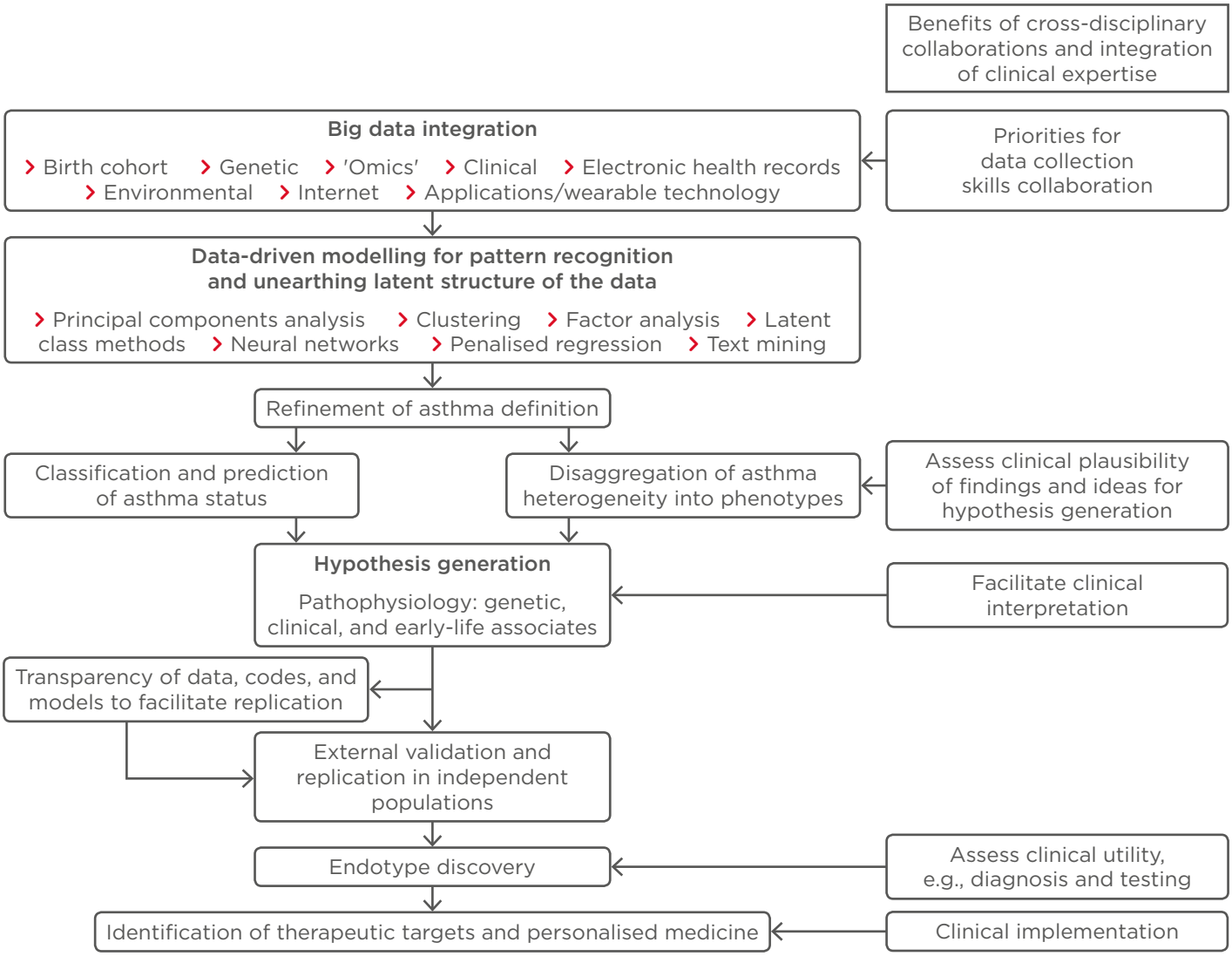

Figure 1: From asthma research to clinical implementation.

Lack of Uniformity in Defining the Dependent Variable

Comparison of prediction tools and adoption in practice is complicated by the fact there is no uniform operational definition of asthma. This creates challenges in identifying consistent early-life predictors, genetic and environmental associates, and pathophysiologic mechanisms.[34,36] A number of studies have indicated that the choice of case definition has a large impact on the estimate of asthma prevalence, as well as performance measures of predictive models. van Wonderen et al.[37] found 60 different definitions of childhood asthma in cohort studies in 122 published articles.[37] Applying four common definitions to a single cohort, the authors found that prevalence estimates varied from 15.1–51.1%.[37] These finding have implications for comparing studies that use different definitions of asthma and suggest the importance of conducting sensitivity analyses to assess the impact of heterogeneous definitions.

Transparency of Replication of Algorithms

Clinicians require access to their patients' data in an absorbable and reliable way that integrates seamlessly with their clinical workflow and does not detract from their key priority of providing quality care during a short patient visit. Without interpretive tools that can be readily incorporated in daily practice, there may be a risk of valuable research findings being overlooked, as actions for decision-making may not be

obvious. The statistical literacy of the clinical community is not keeping pace with the proliferation of new data-driven techniques and the associated terms (e.g., negative matrix factorisation,[19] probabilistic causal network analysis,[38] decision trees,[39] and least absolute shrinkage and selection operator [LASSO]-penalised logistic regression[31]). Computational transparency and reproducibility of research findings are increasingly complicated by the density and complexity of the code underlying models implemented using a variety of programming languages.[40] Such issues are increasingly being recognised, with organisations, including Fairness, Accountability, and Transparency in Machine Learning, calling for greater awareness, debate, and research on such issues. Recently, practical solutions have been proposed, such as a toolkit for enhanced transparency, which includes the use of open-source software, documentation of analyses steps, data archiving, and version control of code using web-based hosting services, such as GitHub, Inc., San Francisco, California, USA.[41] Timely syntheses of findings from the growing research output can help clinicians to understand research with a potential for clinical application. For example, Pecak et al.[42] recently developed a catalogue of 190 potential asthma biomarkers from 73 studies covering 13 omics platforms (including genomics, epigenomics, transcriptomics, and proteomics).[42] They identified 10 candidate genes linked to asthma that were present in at least two omics levels, thus demonstrating the potential for prioritising specific biomarker research and the development of targeted therapeutics.

FUTURE DIRECTIONS: TAKING AN INTEGRATIVE APPROACH

Integrating Data

The proliferation of new data types coupled with advances in computational power may offer new opportunities for integrating different data sources to understand common complex diseases more holistically. Recent advances in molecular techniques offer promising opportunities to disentangle phenotypic characteristics that reflect underlying pathological mechanisms.[43] In this context, systems biology is an approach that investigates organisms as integrated systems comprising dynamic and interrelated genetic, protein, metabolic, and cellular components. Combined with mathematical, bioinformatic, and computational techniques, systems biology can help to elucidate the directionality of relationships between variables at a more holistic level, thereby moving away from associative to more causal analyses.[38,44] In the longer term, findings from such data have the potential for the development of non-invasive and quick diagnostic assessments for use in clinics.[45-47] With the birth of genome-wide association studies (GWAS), researchers are able to investigate the relationship between hundreds of thousands of genetic markers with a phenotype.[48] However, most large GWAS in the field of asthma use the broadest possible definition of the primary outcome (e.g., 'doctor-diagnosed asthma'). In contrast, using deep phenotyping, a recent comparatively small GWAS discovered the association of a specific asthma phenotype (early-life onset with severe exacerbations) with a functional variant in a novel susceptibility gene *CDHR3* (rs6967330).[49] This SNP was not associated with doctor-diagnosed asthma in any of the large-scale GWAS. Subsequent *in silico* studies have shown that rs6967330 mediates rhinovirus-C binding and replication, and that a coding SNP in *CDHR3* mediates enhanced rhinovirus-C binding and increased progeny yields.[50] Several companies are currently pursuing this as a therapeutic target. This example shows the potential of moving from much better phenotyping to genetic association studies, discovery of mechanisms through functional studies, and the identification of therapeutic targets for tailored clinical treatment. Figure 2 summarises this desired sequence.

New possibilities for asthma research are also emerging from personally tracked data from the ubiquitous use of digital devices. Data from Google, Twitter, and Facebook have made real-time information about daily behaviours, health status, and geographical locations widely accessible on an unprecedented scale. The potential for using web-based data for surveillance of trends has been demonstrated in other diseases, such as flu,[51] lupus,[52] and multiple sclerosis.[53] In contrast, traditional sources of surveillance data are based on a time lag, which makes prompt responses infeasible. Real-time models could help healthcare

facilities anticipate asthma-related visits and hospitalisations, and plan staffing and resource management in areas of high risk. A recent study has capitalised on the use of online data to demonstrate the potential for asthma surveillance.[54] Text mining was used to link asthma-related tweets with electronic medical records using geolocation data, along with near real-time environmental data from an air quality sensor. When the number of asthma-related tweets increased in a particular week, the number of asthma emergency department visits or hospitalisations increased proportionally during the following week. The predictive model suggested patterns of accident and emergency visits with around 75% accuracy.

Individually generated data are also emerging from synergies between medical technology and smartphones. Bluetooth-enabled smart inhalers and peak flow meters[55,56] allow individuals to monitor lung function, medication use, and severity of symptoms. myAirCoach, which is a pan-European Union (EU) consortium comprising patient groups, academic institutions, and technology and pharmaceutical companies, aims to provide an evidence base for the benefits of integrating sensor technology with computational modelling to provide personalised feedback to patients on how to manage their condition daily.[57,58] The use of such data may provide clinicians with warnings on exacerbations, which would allow them to tailor medication accordingly.

Table 1 summarises the strengths and limitations of different sources of data.[3,10,38,42,56-64] These types of data have the potential to uncover different aspects of asthma heterogeneity with greater granularity and certainty, but they are a complement to, rather than a substitute for, traditional or other forms of data.

Integrating Multidisciplinary Expertise

One potential risk of 'allowing the data to speak for itself' is that data analysis may become divorced from rigorous scientific scrutiny and meaningful clinical interpretation.[12]

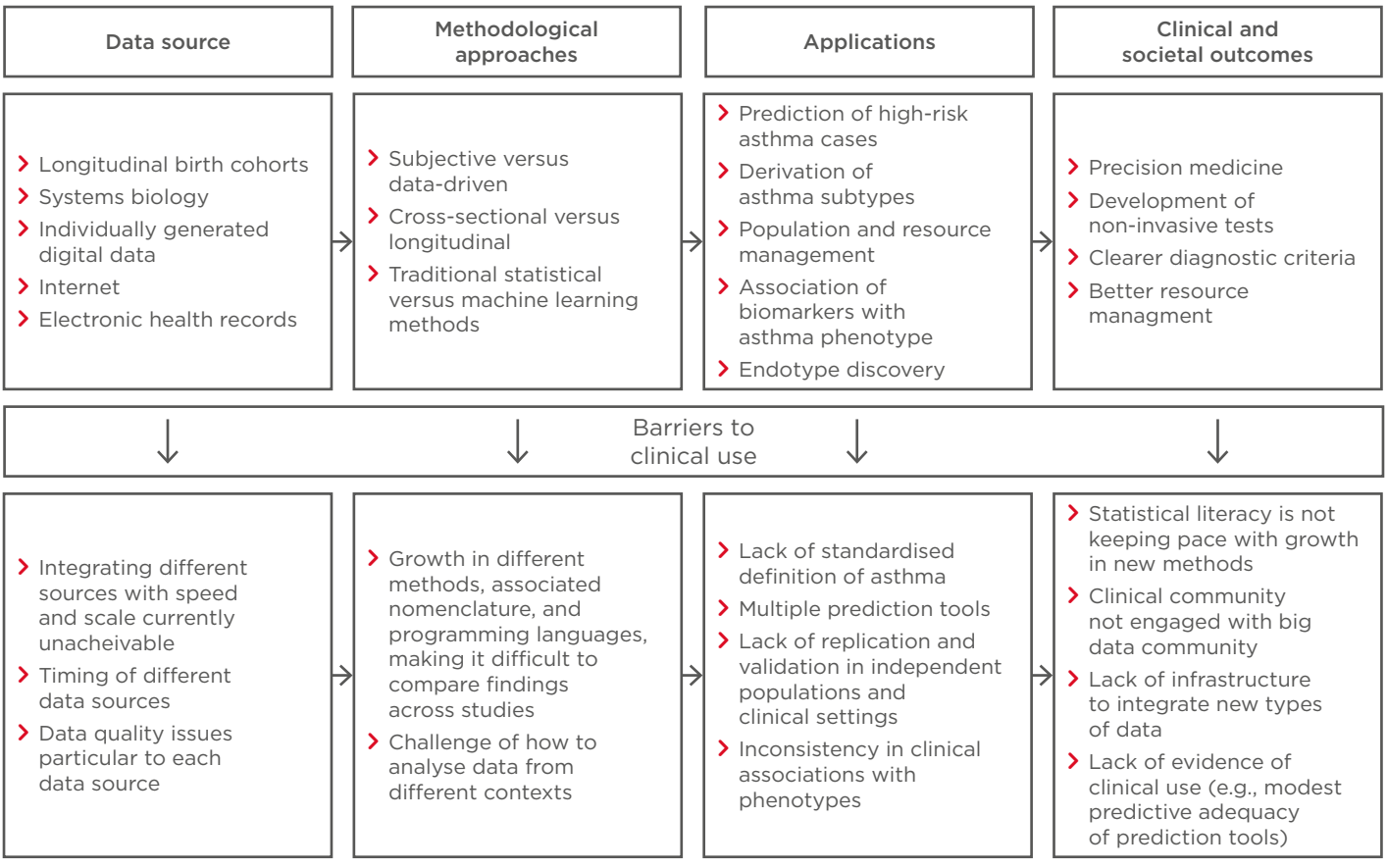

Figure 2: Barriers to clinical implementation of asthma research.

The use of modern techniques, such as machine learning, does not and should not preclude the use of more traditional statistical hypotheses-testing approaches.[14,20,25] The patterns can be discovered in large and heterogeneous data, yet clinical and basic science domain experts can guide formulation of new hypotheses and provide interpretation to findings.[59] For example, a recent study, which applied latent profile analysis to the Tasmanian Longitudinal Health Study, identified six discrete lung function trajectories,[16] five of which were remarkably similar to trajectories from pre-school age to early adulthood in two UK birth cohorts.[15] Using logistic regression, the study found that three of these trajectories were associated with childhood asthma, and the same trajectories were also associated with chronic obstructive pulmonary disease in later life, suggesting that early-life risk factors could lead to poorer lung growth and adult risk factors could accelerate lung function decline.

Table 1: Examples of asthma data sources with associated strengths and limitations.

Data source	Strengths	Limitations
Longitudinal birth cohorts	Explicitly includes the dimension of time, which allows the natural history of disease development to be studied.	Sample attrition and differential loss to follow-up can introduce bias.
	Can collect data on multiple outcomes simultaneously.	Labour and cost-intensive due to the need for a large sample size and the potentially long follow-up duration of the study.
	Questions can be tailored to a specific theme or disease.	Maintaining follow-ups can be challenging.
	Can look at associations of early-life risk factors and exposures with disease outcomes later in life.	Risk of recall bias.
	Systematic observations before the onset of disease.	Change in question wording over time.
	Incorporation of validated standardised questionnaires (e.g., ISAAC).	Large sample sizes and long duration required for discovering rare subtypes of disease.
	Potential for pooling different birth cohorts with similar questions and time points.	Not practical for rare outcomes.
Individually generated data	Data can be collected in real-time through the use of digital devices, wearable technologies, or medical devices, such as electronic inhalers.	Low long-term adoption of apps or technologies.
	Potential to collect data on multiple domains (e.g., health behaviours, symptoms, and environment).	Potential of low-quality data due to incorrect use of technology or malfunction.
	Improved self-management.	Technologies are not aligned with public health computer systems, meaning limited benefit for clinical management.
	Improved patient–clinician dialogue.	Risk of devices malfunctioning.
	Monitoring of severity of symptoms.	Missing data arising from an adverse health event.
	Data can be captured passively for some wearable devices.	
	Data collected outside of the clinical setting – greater patient insight into their own health.	
Internet text data: social media, such as Twitter, Google searches, and Facebook	Real-time data.	Risk of false predictions if models rely on historical search terms.
	Geolocation information combined with real-time data collection can reveal dynamic changes in disease over time.	Challenge of distinguishing 'noise' from genuine health episodes due to spam or searches not linked to health episodes.
	Readily available.	Recalibration of models to reflect changing search terms or unanticipated events (e.g., pandemics).
	Elucidate differential risks due to geolocation tagging.	The unstructured nature of the data makes it challenging to link to other sources of data.
	Potentially useful for forecasting.	
	Large sample size.	

Table 1 continued.

Data source	Strengths	Limitations
Systems biology data: multiple types of 'omics' data (e.g., genomics, proteomics, and metabolomics)	A more holistic approach for investigating causal biological pathways that might inform endotype discovery and targeted therapies.	Large sample sizes required to have sufficient power to detect associations.
	Data can be used to model complex interdependencies between multiple dimensions (e.g., genome, transcriptome, epigenome, microbiome, and metabolome).	Replication in independent populations required for validation.
		Risk of false-positive associations.
		Tends to be captured at single time points.
		Data collection is small-scale compared with other data types.
		Difficult to externally validate findings due to cost and complexity of data collection.
		Data is not readily accessible unlike other sources.
Electronic health records	High granularity of clinical information: diagnoses, medication, test results, comorbidities, and demographics.	Potentially important information not routinely collected and requires replication in independent populations.
	Real-world population.	Data on medication adherence or asthma control not recorded, which is a potentially modifiable risk factor.
	Large sample size.	Useful for association analyses but of limited benefit for causal analyses.
	Curation of patient cohorts for epidemiological investigations, population management, and resource planning.	Inconsistent data quality.
		Confounding factors not recorded in the database (for example, environmental).
		Variability in data types (structured and unstructured).

ISAAC: The International Study of Asthma and Allergies in Childhood.

As the number of relationships being tested increases, there is a risk of identifying false-positive associations in the absence of previous guidance about the clinical plausibility of such findings.[65] Big data can only explain part of the picture, and clinicians can provide a more contextualised understanding through their experience, knowledge of detailed clinical histories, and being able to explain variations across their patients.[66] Experts can review the findings from big data studies, which may generate promising leads for further enquiry.

An integrated approach to big data may enable us to harness the power of big data in ways that translate into a better understanding of causal mechanisms, more accurate diagnoses, and more personalised treatment. Integration can occur at different levels through cross-disciplinary research (for example, the Study Team for Early Life Asthma Research [STELAR] consortium,[60] MedALL,[67] U-BIOPRED,[68] Breathing Together consortium[69]), wherein basic scientists, geneticists, clinicians, and data scientists work together to understand the mechanisms of relevance to clinical heterogeneity of asthma. Another way of bridging the divide between the clinical and big data communities is to understand the tools clinicians need to improve outcomes for their patients by taking a 'team science' approach. As an example, a recent pan-EU consensus exercise led by the EARIP sought to identify key areas for research funding that would, most likely, improve asthma diagnosis and patient care.[27] Experts comprised patients, patient organisations, healthcare professionals, researchers, industry representatives, and policy influencers.

The prediction of asthma in preschool children with reasonable accuracy, how to integrate new biomarkers (such as genomics, proteomics, and metabolomics) in the diagnosis and monitoring of asthma, and the measurement of exhaled volatile organic compounds were identified as priority areas for research. This demonstrates how integrating multidisciplinary expertise has the potential to inform research, and for findings to be translated into improved outcomes for patient care.

CONCLUSION

One of the goals of asthma research is to understand disease heterogeneity with the aim of providing personalised treatment. There needs to be a shift away from the artificial dichotomy of data-driven hypothesis-generating versus more traditional hypothesis testing approaches towards a more integrated one, whereby cross-disciplinary collaborations can facilitate rigorous scientific scrutiny and interpretation of findings. No single source of data can uncover the complex dynamics driving asthma heterogeneity, and triangulation (integration of evidence from several different approaches with differing and unrelated sources of bias) is critically important to fill current knowledge gaps and improve causal inference.[70,71] With the advent of new and exciting sources of data, there is huge potential for integrating these to provide a more holistic understanding of the disease at a very personal level. As the factors shaping the development and control of asthma affect individuals dynamically in response to treatment or environmental factors, deeper insights can be garnered through integration. Knowing in real-time when and where symptoms are exacerbated, in combination with refined subtypes and environmental data, may help identify personal triggers and inform a personally tailored care plan.

Research needs to take greater steps to demonstrate clinical utility, or it risks being consigned to research for research's sake. Tools need to be developed that clinicians can integrate into daily practice to make decision-making more efficient and personalised. Steps need to be taken to improve the statistical literacy of healthcare professionals through greater education to bridge the divide with the big data industry. It is essential that clinicians engage in debates surrounding big data and healthcare as a step towards breaking down the siloed approach.

References

1. Pavord ID et al. After asthma: Redefining airways diseases. Lancet. 2017;391(10118):350-400.
2. Martinez FD et al. Asthma and wheezing in the first six years of life. The Group Health Medical Associates. N Engl J Med. 1995;332(3):133-8.
3. Howard R et al. Distinguishing asthma phenotypes using machine learning approaches. Curr Allergy Asthma Rep. 2015;15(7):38.
4. Deliu M et al. Asthma phenotypes in childhood. Expert Rev Clin Immunol. 2017;13(7):705-13.
5. Deliu M et al. Identification of asthma subtypes using clustering methodologies. Pulm Ther. 2016;2:19-41.
6. Oksel C et al. Classification of pediatric asthma: From phenotype discovery to clinical practice. Front Pediatr. 2018;6:258.
7. Belgrave DC et al. Characterizing wheeze phenotypes to identify endotypes of childhood asthma, and the implications for future management. Expert Rev Clin Immunol. 2013;9(10):921-36.
8. Belgrave DC et al. Trajectories of lung function during childhood. Am J Respir Crit Care Med. 2014;189(9):1101-9.
9. Belgrave DC et al. Developmental profiles of eczema, wheeze, and rhinitis: Two population-based birth cohort studies. PLoS Med. 2014;11(10):e1001748.
10. Prosperi MC et al. Predicting phenotypes of asthma and eczema with machine learning. BMC Med Genomics. 2014;7 Suppl 1:S7.
11. Lotvall J et al. Asthma endotypes: A new approach to classification of disease entities within the asthma syndrome. J Allergy Clin Immunol. 2011;127(2):355-60.
12. Belgrave D et al. Disaggregating asthma: Big investigation versus big data. J Allergy Clin Immunol. 2017;139(2):400-7.
13. Wilson NM. Wheezy bronchitis revisited. Arch Dis Child. 1989;64(8):1194-9.
14. Belgrave D, Custovic A. The importance of being earnest in epidemiology. Acta paediatrica. 2016;105(12):1384-6.
15. Belgrave DCM et al. Lung function trajectories from pre-school age to adulthood and their associations with early life factors: A retrospective analysis of three population-based birth cohort studies. Lancet Resp Med. 2018;6(7):526-34.
16. Bui DS et al. Childhood predictors of lung function trajectories and future COPD risk: A prospective cohort study from the first to the sixth decade of life. Lancet Resp Med. 2018;6(7):535-44.
17. Oksel C et al. Causes of variability in latent phenotypes of childhood wheeze. J Allergy Clin Immunol. 2018. pii:S0091-6749(18)31723-8. [Epub ahead of print].
18. Belgrave DCM et al. Joint modeling

of parentally reported and physician-confirmed wheeze identifies children with persistent troublesome wheezing. J Allergy Clin Immunol. 2013;132(3):575-83 e12.

19. Schoos AMM et al. Atopic endotype in childhood. J Allergy Clin Immun. 2016;137(3):844-51.
20. Lazic N et al. Multiple atopy phenotypes and their associations with asthma: Similar findings from two birth cohorts. Allergy. 2013;68(6):764-70.
21. Simpson A et al. Beyond atopy: Multiple patterns of sensitization in relation to asthma in a birth cohort study. Am J Respir Crit Care Med. 2010;181(11):1200-6.
22. Oksel C, Custovic A. Development of allergic sensitization and its relevance to paediatric asthma. Curr Opin Allergy Clin Immunol. 2018;18(2): 109-16.
23. Smith JA et al. Dimensions of respiratory symptoms in preschool children: Population-based birth cohort study. Am J Respir Crit Care Med. 2008;177(12):1358-63.
24. Howard R et al. Evolution of IgE responses to multiple allergen components throughout childhood. J Allergy Clin Immunol. 2018;142(4):1322-30.
25. Custovic A et al. Cytokine responses to rhinovirus and development of asthma, allergic sensitization, and respiratory infections during childhood. Am J Respir Crit Care Med. 2018;197(10):1265-74.
26. Fontanella S et al. Machine learning to identify pairwise interactions between specific IgE antibodies and their association with asthma: A cross-sectional analysis within a population-based birth cohort. PLoS medicine. 2018;15(11):e1002691.
27. Garcia-Marcos L et al. Priorities for future research into asthma diagnostic tools: A PAN-EU Consensus exercise from the European asthma research innovation partnership (EARIP). Clin Exp Allergy. 2018;48(2):104-20.
28. Castro-Rodriguez JA. The Asthma Predictive Index: A very useful tool for predicting asthma in young children. J Allergy Clin Immunol. 2010;126(2):212-6.
29. Kurukulaaratchy RJ et al. Predicting persistent disease among children who wheeze during early life. Eur Respir J. 2003;22(5):767-71.
30. Caudri D et al. Predicting the long-term prognosis of children with symptoms suggestive of asthma at preschool age. J Allergy Clin Immunol. 2009;124(5):903-10.e1-7.
31. Pescatore AM et al. A simple asthma prediction tool for preschool children with wheeze or cough. J Allergy Clin Immunol. 2014;133(1):111-8.e1-13.
32. Wang R et al. Individual risk assessment tool for school age asthma prediction in UK birth cohort. Clinical and experimental allergy. 2018. [Epub ahead of print].
33. Brand PL. The Asthma Predictive Index: Not a useful tool in clinical practice. J Allergy Clin Immunol. 2011;127(1):293-4.
34. Luo G et al. A systematic review of predictive models for asthma development in children. BMC Med Inform Decis Mak. 2015;15:99.
35. Matricardi PM et al. Predicting persistence of wheezing: One algorithm does not fit all. Eur Respir J. 2010;35(3):701-3.
36. Rodriguez-Martinez CE et al. Factors predicting persistence of early wheezing through childhood and adolescence: A systematic review of the literature. J Asthma Allergy. 2017;10:83-98.
37. van Wonderen KE et al. Different definitions in childhood asthma: How dependable is the dependent variable? Eur Respir J. 2010;36(1): 48-56.
38. Bunyavanich S, Schadt EE. Systems biology of asthma and allergic diseases: A multiscale approach. J Allergy Clin Immunol. 2015;135(1): 31-42.
39. Prosperi MC et al. Challenges in interpreting allergen microarrays in relation to clinical symptoms: A machine learning approach. Pediatr Allergy Immunol. 2014;25(1):71-9.
40. Groeneveld PW, Rumsfeld JS. Can Big data fulfill its promise? Circ Cardiovasc Qual Outcomes. 2016;9(6):679-82.
41. Perkel JM. A toolkit for data transparency takes shape. Nature. 2018;560(7719):513-5.
42. Pecak M et al. Multiomics data triangulation for asthma candidate biomarkers and precision medicine. Omics. 2018;22(6):392-409.
43. Tang HH et al. Trajectories of childhood immune development and respiratory health relevant to asthma and allergy. eLIFE 2018;7.
44. Greene CS, Troyanskaya OG. Integrative systems biology for data-driven knowledge discovery. Seminars in Nephrology. 2010;30(5):443-54.
45. Custovic A et al. Evolution pathways of IgE responses to grass and mite allergens throughout childhood. J Allergy Clin Immunol. 2015;136(6):1645-52 e8.
46. Holt PG et al. Distinguishing benign from pathologic TH2 immunity in atopic children. J Allergy Clin Immunol. 2016;137(2):379-87.
47. Simpson A et al. Patterns of IgE responses to multiple allergen components and clinical symptoms at age 11 years. J Allergy Clin Immunol. 2015;136(5):1224-31.
48. March M et al. Genome-wide association studies in asthma: Progress and pitfalls. Adv Genomics Genet. 2015;5:107-19.
49. Bonnelykke K et al. A genome-wide association study identifies *CDHR3* as a susceptibility locus for early childhood asthma with severe exacerbations. Nat Genet. 2014;46(1):51-5.
50. Bochkov YA et al. Cadherin-related family member 3, a childhood asthma susceptibility gene product, mediates rhinovirus C binding and replication. Proc Natl Acad Sci U S A. 2015;112(17):5485-90.
51. Butler D. When Google got flu wrong. Nature. 2013;494(7436):155.
52. Radin M, Sciascia S. Infodemiology and seasonality of systemic lupus erythematous using google trends. Ann Rheum Dis. 2017;26(8):886-9.
53. Moccia M et al. Google Trends: New evidence for seasonality of multiple sclerosis. J Neurol Neurosur Ps. 2016;87(9):1028-9.
54. Ram S et al. Predicting asthma-related emergency department visits using big data. IEEE J Biomed Health Inform. 2015;19(4):1216-23.
55. Asthma MD. Features. Available at: http://www.asthmamd.org/features/. Last accessed: 25 January 2018.
56. Adherium Ltd. Smartinhaler medication sensors. Available at: http://www.smartinhaler.com/devices/. Last accessed: 25 January 2018.
57. Ryan D et al. Use of electronic medical records and biomarkers to manage risk and resource efficiencies. Eur Clin Respir J. 2017;4(1):1293386.
58. My Air Coach. Available at: http://www.myaircoach.eu/content/what-myaircoach-project. Last accessed: 25 January 2018.
59. Deliu M et al. Features of asthma which provide meaningful insights for understanding the disease heterogeneity. Clin Exp Allergy. 2018;48(1):39-47.
60. Custovic A et al. The Study Team for Early Life Asthma Research (STELAR) consortium 'Asthma e-lab': Team science bringing data, methods and investigators together. Thorax 2015;70(8):799-801.
61. Dai H et al. Predicting asthma prevalence by linking social media data and traditional surveys. Ann Am Acad Political Soc Sci. 2017;669(1): 75-92.
62. Sircar G et al. Allergic asthma biomarkers using systems approaches. Front Genet. 2014;4:308.
63. Bloom CI et al. Exacerbation risk and characterisation of the UK's asthma population from infants to old age.

Thorax. 2017;73(4):313-20.

64. Turner SW et al. Applying UK real-world primary care data to predict asthma attacks in 3776 well-characterised children: A retrospective cohort study. NPJ Prim Care Respir Med. 2018;28.

65. Rumsfeld JS et al. Big data analytics to improve cardiovascular care: Promise and challenges. Nat Rev Cardiol. 2016;13(6):350-9.

66. Neff G. Why big data won't cure us. Big data. 2013;1(3):117-23.

67. Bousquet J et al. Birth cohorts in asthma and allergic diseases: Report of a NIAID/NHLBI/MeDALL joint workshop. J Allergy Clin Immunol. 2014;133(6):1535-46.

68. Fleming L et al. The burden of severe asthma in childhood and adolescence: Results from the paediatric U-BIOPRED cohorts. Eur Respir J. 2015;46(5):1322-33.

69. Turner S et al. Pulmonary epithelial barrier and immunological functions at birth and in early life - Key determinants of the development of asthma? A description of the protocol for the Breathing Together study. Wellcome Open Res. 2018;3:60.

70. Lawlor DA et al. Triangulation in aetiological epidemiology. Int J Epidemiol. 2016;45(6):1866-86.

71. Munafo MR et al. Robust research needs many lines of evidence. Nature. 2018;553(7689):399-401.