

Big asthma data: getting bigger and more beautiful?

Sarah Diver, Chris E Brightling

'Big data' is on trend and the term is used in equal measure to reflect both one of the greatest challenges and likeliest solutions to future scientific advances from fundamental understanding in astrophysics, climate change, economics, health and disease.¹ Like many trends, it means different things to different people. In medicine, it is used to describe the data derived from large populations in epidemiology studies, high fidelity multiscale 'omic datasets across spatial scales within individuals or sometimes a combination of the two. Big data will often capture information at a single time point. Typically, it does not address temporal scales of chronic disease including day-to-day variability, response to perturbations such as intercurrent infection, decompensation of the disease or response to therapeutic interventions and is rarely obtained over a life course. Observations will therefore always be limited by what is measured, when and in whom and will only ever provide estimates of what is 'real' within the larger group from which the sample is taken. Big data that includes large populations makes interpretations more robust and generalisable. Indeed, as the population studied or sample size approaches a majority, or at least a sizeable minority, of the whole population then the observations begin to no longer be estimates, but simply a description of the population.

In this issue of *Thorax*, Bloom and colleagues sought to investigate the frequency of asthma exacerbations in a general population and their association with age and asthma severity.² Asthma exacerbations are when individuals experience symptoms beyond usual day-to-day fluctuation leading to a change in treatment defined here as a requirement for a course of oral corticosteroids. Asthma exacerbations are important as they are associated with increased morbidity, mortality and significant healthcare costs.³ Bloom and colleagues used the Clinical Practice Research Datalink (CPRD)⁴ to

characterise the patients with asthma and determine the frequency of the exacerbation episodes. The CPRD is a national database contributed to by 674 general practitioner (GP) practices with a current coverage of over 11 million patients, representing ~17% of the entire UK population. In the majority of GP practices included, the database is linked to Hospital Episode Statistics data, Index of Multiple Deprivation and Office of National Statistics mortality data. The study population were representative of the UK-wide demographics in terms of age, gender and ethnicity and thus enabled the study of people with asthma from infants to old age. Importantly, the study included follow-up between 1 and 5 years to allow exacerbations to be captured.

What has the study by Bloom and colleagues taught us? Asthma is common: ~8% of the population, of which 26% require treatment at British Thoracic Society (BTS)⁵ step 4 or above, and of these 58% had exacerbations during the follow-up period. Only a small proportion were treated with regular oral corticosteroids (<2%). Conversely, a large proportion (35%) were not receiving regular therapy. An increased exacerbation frequency was associated with several comorbidities across all age groups including atopy, rhinitis, gastro-oesophageal reflux, anxiety or depression and in adults was associated with female gender, smoking history and increased body mass index. Independent of these confounders, exacerbation frequency was related to BTS treatment step in all age groups and was highest in the very young (<5 years) or older age (>55 years) groups. Lower respiratory tract infections requiring antibiotic treatment were also highest in these groups, although much less common than exacerbations treated with oral corticosteroids. Interestingly, less than 10% of the antibiotic-treated lower respiratory tract infections occurred concurrently with the asthma exacerbation and <1% preceded an exacerbation within 14 days. This suggests that in primary care, the co-administration of antibiotics and oral corticosteroids for asthma exacerbations is unusual and consistent with the BTS guidelines. This is in contrast with findings from the Azithromycin Against Placebo

for Acute Exacerbations of Asthma trial⁶ in which the majority of patients admitted to hospital for asthma exacerbations had previously received antibiotics for their exacerbation episode. This suggests that administration of antibiotics for asthma exacerbations in primary care are probably reserved for those that are most severe.

These data do largely confirm what has been previously reported. However, the scale of the UK population studied is unprecedented in asthma research. These 'big data' confirm that with asthma there remains a huge unmet need, identifies those most at risk and therefore in need of specialist care. These data are useful for healthcare providers, patients and industry. Beyond confirming earlier reports, the increased exacerbation risk in the very young and older asthmatics is especially important to understand and suggests that these high-risk groups should be targeted.

The strengths of this study is that it is 'real world' and 'big'. However, like any data, there are limitations in what is measured in whom, how it is measured and recorded introducing potential errors and bias. Notwithstanding these limitations, the authors have previously studied the positive predictive value of the asthma diagnosis in cases reported in the CPRD which was very high, suggesting a possible small underestimation of the overall prevalence of asthma but that they are likely to have captured the patients correctly in the database, particularly those with more severe disease. Thus, the limitations of using CPRD are likely to be largely due to the restriction of what data is collected; the limited quality control and assurance that can be applied to the data curation in real-world studies and the lack of data verification for self-reported outcomes such as adherence to treatment.

In contrast, well-conducted large multicentre, multinational observational and interventional trials have advantages of using standardised procedures, quality control and assurance measures with careful data curation and monitoring. However, they often have restrictive inclusion and exclusion criteria introducing acquisition bias of participants and, in spite of their complexity and often large expense, might be restricted to simple clinical and physiological measures with perhaps consideration of single 'omic technologies such as genomics. Major international academic-led asthma consortia, private-public initiatives and industry-led programmes have provided insights in asthma pathogenesis, disease heterogeneity and led to

Department of Infection, Immunity and Inflammation, Leicester NIHR BRC, Institute for Lung Health, University of Leicester, Leicester, UK

Correspondence to Professor Chris E Brightling, Department of Infection, Immunity and Inflammation, University of Leicester, Glenfield Hospital, LE3 9QP, Leicester, UK; ceb17@le.ac.uk

the emergence of new treatment paradigms, eg, GABRIEL,⁷ Severe Asthma Research Programme,⁸ Airway Disease Predicting Outcomes through Patient Specific Computational Modelling,⁹ Unbiased Biomarkers for the Prediction of Respiratory Disease Outcomes¹⁰ and others. However, to capitalise on these discoveries, there needs to be a greater integration between routine healthcare records, research databases and biosamples through change in culture to better embed research into routine clinical care. This requires streamlined consent procedures with appropriate ethics and research governance which would need to be achieved within existing and future legal frameworks. For example, the proposed introduction of the revised General Data Protection Regulation¹¹ will automatically come into force on a harmonised basis across the European Union (EU) from May 2018 and, coupled with the future complexities of the relationship between the UK and the EU, will bring challenges to extend research programmes in the UK and to wider Europe. Several initiatives including UK Biobank¹² and the National Institute for Health Research Bioresource¹³ are at the forefront of these strategies and respiratory medicine needs to remain at the cutting edge of these opportunities and work towards new federated and linked data and biosample repositories for respiratory disease research.

Bloom and colleagues should be congratulated on the success of their 'big data' study, which has demonstrated the

potential of accessing real-world data to inform our understanding of risk for asthma exacerbations. The challenge now is to integrate better the knowledge from primary and secondary care databases with research focused biological data including 'omic technologies to further inform our understanding of asthma and other respiratory disease. This, together with emerging technologies for monitoring adherence, providing feedback on inhaler technique and home monitoring of inflammation, will provide greater insights on asthma control and future risk in individual patients scaled-up to population studies.

Thus, it is almost inevitable that the need to continue and expand the use of 'big data' in a real-world setting means the research fashion will remain 'big' is beautiful for some time to come.

Competing interests None declared.

Provenance and peer review Commissioned; internally peer reviewed.

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.



To cite Diver S, Brightling CE. *Thorax* 2018;**73**:311–312.

Published Online First 29 December 2017



► <http://dx.doi.org/10.1136/thoraxjnl-2017-210650>

Thorax 2018;**73**:311–312.
doi:10.1136/thoraxjnl-2017-211148

REFERENCES

- 1 Marx V. Biology: The big challenges of big data. *Nature* 2013;498:255–60.
- 2 Bloom CI, Nissen F, Douglas IJ, *et al.* Exacerbation risk and characterisation of the UK's asthma population from infants to old age. *Thorax* 2018;**73**:313–20.
- 3 Custovic A, Johnston SL, Pavord I, *et al.* EAACI position statement on asthma exacerbations and severe asthma. *Allergy* 2013;**68**:1520–31.
- 4 Herrett E, Gallagher AM, Bhaskaran K, *et al.* Data resource profile: clinical practice research datalink (CPRD). *Int J Epidemiol* 2015;**44**:827–36.
- 5 British Thoracic Society/Scottish Intercollegiate Guidelines Network. *British guideline on the management of asthma*, 2016.
- 6 Johnston SL, Szigeti M, Cross M, *et al.* Azithromycin for acute exacerbations of asthma: The AZALEA randomized clinical trial. *JAMA Intern Med* 2016;**176**:1630–7.
- 7 Moffatt MF, Gut IG, Demenais F, *et al.* A large-scale, consortium-based genome-wide association study of asthma. *N Engl J Med* 2010;**363**:1211–21.
- 8 Severe Asthma Research Programme (SARP). <http://www.severeasthma.org/home.html> (accessed 4 Dec 2017).
- 9 Airway Disease Predicting Outcomes through Patient Specific Computational Modelling (AirPROM). <http://www.europeanlung.org/en/projects-and-research/projects/airprom/home> (accessed 4 Dec 2017).
- 10 Unbiased Biomarkers in Prediction of respiratory disease outcomes (U-BIOPRED). <http://www.europeanlung.org/projects-and-research/projects/u-biopred/home> (accessed 4 Dec 2017).
- 11 EU General Data Protection Regulation (GDPR). <https://www.eugdpr.org/> (accessed 4 Dec 2017).
- 12 Wain LV, Shrine N, Artigas MS, *et al.* Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet* 2017;**49**:416–25.
- 13 National Institute for Health Research (NIHR). BioResource | Bringing people together and leading research. <https://bioresource.nihr.ac.uk/> (accessed 11 Dec 2017).