# Big Asthma Data - getting bigger and more beautiful?

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Thorax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>thoraxjnl-2017-211148</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Editorial</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>12-Dec-2017</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Diver, Sarah; Institute of Lung Health, University of Leicester Brightling, Chris; ILH</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Asthma, Asthma Epidemiology, Asthma in primary care, Clinical Epidemiology, Paediatric asthma</td>
</tr>
</tbody>
</table>
Big Asthma Data - getting bigger and more beautiful?

S. Diver, C. Brightling

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

Corresponding author

Professor C.E. Brightling, Institute For Lung Health, Department of Infection, Immunity and Inflammation, University of Leicester, Glenfield Hospital, LE3 9QP, Leicester, United Kingdom. E-mail: ceb17@le.ac.uk Tel: +44 116 2563998, Fax: +44 116 2502787.
'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. 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 multi-scale ‘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 inter-current 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 generalizable. 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. 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. Bloom and colleagues used the Clinical Practice Research Datalink (CPRD) to characterise the patients with asthma and determine the frequency of the exacerbation episodes. The CPRD is a national database contributed to
by 674 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 from between 1-5 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 BTS\textsuperscript{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 co-morbidities 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 AZALEA trial\textsuperscript{6} 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 health care 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 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 under-estimation 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 multi-centre, 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 the emergence of new treatment paradigms e.g. GABRIEL\textsuperscript{7}, SARP\textsuperscript{8}, AirPROM\textsuperscript{9}, UBIOPRED\textsuperscript{10} and others. However, to capitalise upon these discoveries there needs to be a greater integration between routine health care records, research data bases 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 Regulation\textsuperscript{11} will automatically come into force on a harmonised basis across the 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 Biobank\textsuperscript{12} and the NIHR Bioresource\textsuperscript{13} 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 focussed 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.
References


