# Phenotypes of symptomatic airways disease in China and New Zealand

James Fingleton[1,2,3,10], Kewu Huang[4,10], Mark Weatherall[2,5], Yanfei Guo[6], Stefan Ivanov[7], Piet Bruijnzeel[7], Hong Zhang[4], Wei Wang[4], Richard Beasley[1,2,3,11] and Chen Wang[6,8,9,11] for the NZ–China Study Group

**Affiliations**: [1]Medical Research Institute of New Zealand, Wellington, New Zealand. [2]Capital and Coast District Health Board, Wellington, New Zealand. [3]Victoria University of Wellington, Wellington, New Zealand. [4]Dept of Pulmonary and Critical Care Medicine, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, PR China. [5]University of Otago Wellington, Wellington, New Zealand. [6]Dept of Pulmonary and Critical Care Medicine, Beijing Hospital, Ministry of Health, Beijing, PR China. [7]AstraZeneca, Molndal, Sweden. [8]Dept of Pulmonary and Critical Care Medicine, China–Japan Friendship Hospital, Beijing, PR China. [9]Capital Medical University, Beijing, PR China. [10]Joint first authors. [11]Joint last authors.

**Correspondence**: James Fingleton, Medical Research Institute of New Zealand, Private Bag 7902, Wellington 6242, New Zealand. E-mail: james.fingleton@mrinz.ac.nz



**Cite this article as:** Fingleton J, Huang K, Weatherall M, *et al*. Phenotypes of symptomatic airways disease in China and New Zealand. *Eur Respir J* 2017; 50: 1700957 [https://doi.org/10.1183/13993003.00957-2017].

ABSTRACT It is uncertain whether phenotypes of asthma and chronic obstructive pulmonary disease (COPD) vary between populations with different genetic and environmental characteristics. Here, our objective was to compare the phenotypes of airways disease in two separate populations.

This was a cross-sectional observational study in adult populations from New Zealand and China. Participants aged 40–75 years who reported wheeze and breathlessness in the last 12 months were randomly selected from the general population and underwent detailed characterisation. Complete data for cluster analysis were available for 345 participants. Hierarchical cluster analysis was undertaken, based on 12 variables: forced expiratory volume in 1 s ($FEV_1$), $FEV_1$/forced vital capacity ratio, bronchodilator reversibility, peak expiratory flow variability, transfer coefficient of the lung for carbon monoxide, exhaled nitric oxide fraction, total IgE, C-reactive protein, age of symptom onset, body mass index, health status and cigarette smoke exposure.

Cluster analysis of the combined dataset described five phenotypes: "severe late-onset asthma/COPD overlap group", "moderately severe early-onset asthma/COPD overlap group", "moderate to severe asthma group with type 2 predominant disease", and two groups with minimal airflow obstruction, differentiated by age of onset. Separate analyses by country showed similar patterns; however, a distinct obese/comorbid group was observed in the New Zealand population.

Cluster analysis of adults with symptomatic airways disease suggests the presence of similar asthma/COPD overlap phenotypes within populations with different genetic and environmental characteristics, and an obese/comorbid phenotype in a Western population.

This article has supplementary material available from erj.ersjournals.com

Received: May 09 2017 | Accepted after revision: Aug 29 2017

The NZRHS is registered with the Australian New Zealand Clinical Trials Registry (www.anzctr.org.au) with identifier number ACTRN12610000666022

## Introduction

A number of studies have defined phenotypes of asthma and chronic obstructive pulmonary disease (COPD), either separately or as part of the spectrum of obstructive airways disease [1–12]. Phenotyping offers the potential to better understand differing patterns of disease and tailor an individual's treatment to their phenotype. However, interpretation of studies has been limited by differences in the specific populations studied and the varied methodology applied.

These limitations applied to the recently reported New Zealand Respiratory Health Survey (NZRHS) [13], which employed cluster analysis to describe five phenotypes of symptomatic airways disease that differed in clinical characteristics and response to inhaled medications. To validate the study findings we have performed a combined cluster analysis using data from a subset of NZRHS participants and from participants from China tested using the same methodology. The objectives were to identify and characterise phenotypes of airways disease, and explore differences between the New Zealand and China populations.

## Methods

We performed a cross-sectional observational study in two randomly selected populations in New Zealand and China.

Ethical approval was given by the Central Regional Ethics Committee of New Zealand (CEN/09/12/095), and the ethics committees of Beijing Hospital and Beijing Chao-Yang Hospital No. 8. Written informed consent was obtained from all participants prior to testing. The NZRHS was registered with the Australian New Zealand Clinical Trials Registry (www.anzctr.org.au: identifier ACTRN12610000666022).

### *Participants*

In New Zealand the data for this analysis were drawn from the first 204 Caucasian subjects to complete the NZRHS that were aged 40–75 years and consented to sharing of de-identified data. In China the same methodology was used to recruit approximately 200 subjects across two sites. New Zealand participants were recruited in the greater Wellington area; Chinese participants were recruited from Chongwen, Hui Longguan and Tongzhou districts of Beijing (Site 1), and from Chanping and Chao-Yang districts of Beijing (Site 2).

The methodology of the NZRHS has been reported in detail previously [13]. Here, we summarise the methods, and highlight the standardisation and modification of methods required to replicate the NZRHS methodology in China.

### *Phase 1*

In New Zealand the participants were randomly selected from the electoral roll and received a one-page questionnaire by post. Individuals who did not return the questionnaire were sent up to two reminders to improve the response rate. Those with a publically listed telephone number who did not respond also received a telephone call and were given the option of completing the questionnaire verbally. In China the participants were randomly selected from community lists, which include details of all adults in a geographically defined district, and a single postal reminder was used to maximise the response rate. Computer-based random selection was used in both New Zealand and China. Participants for Phase 2 were drawn from potentially eligible patients who replied and were aged 40–75 years. Each site recruited as responses were received until at least 200 participants had completed Phase 2. Questionnaire responses received after recruitment was complete were recorded, but respondents were not invited to take part in Phase 2.

### *Phase 2*

Respondents with symptoms of wheeze and breathlessness in the last 12 months were invited to attend for detailed lung function testing. Spirometry (New Zealand and China site 2: Masterscreen Body; Erich

Support statement: Primary funder: AstraZeneca AB, Sweden. Secondary funding: Health Research Council of New Zealand (reference 10/174), a Government funding organisation; the Medical Research Institute of New Zealand receives funding from the Health Research Council of New Zealand through the Independent Research Organisations Capability Fund (14/1002); the Research Special Fund for Capital Health Development (Shou-Fa 2011-1004-01); Project in the National Science and Technology Pillar Program during the Twelfth Five-Year Plan Period (2012BAI05B02, 2014BAI08B04, 2013BAI09B10). This was an investigator-sponsored and -designed study. The primary funder had the opportunity to review and comment on the study protocol prior to initiation of the study. Funding information for this article has been deposited with the Crossref Funder Registry.

Conflict of interest: Disclosures can be found alongside this article at erj.ersjournals.com

Jaeger, Friedberg, Germany; China site 1: SensorMedics; Anaheim, CA, USA) was performed according to American Thoracic Society (ATS) criteria [14] pre- and post-bronchodilator on two occasions, at least 1 week apart. Ipratropium bromide 80 μg was administered at the first visit and salbutamol 400 μg at the second visit. Post-bronchodilator spirometry was performed 30 min after inhalation of the bronchodilator through a Volumatic spacer (GlaxoSmithKline, Brentford, UK). Transfer coefficient of the lung for carbon monoxide (*K*CO) was measured post-bronchodilator at the second visit and was corrected for participant haemoglobin. Locally appropriate reference ranges were used for both the New Zealand and China sites [15, 16]. Exhaled nitric oxide fraction (*F*ENO) was measured according to ATS guidelines (New Zealand: NIOX; Aerocrine, Solna, Sweden; China: NIOX MINO; Aerocrine) [17].

Details of participants' respiratory and medical symptoms and self-reported medical history were systematically collected using a questionnaire based on validated questions drawn from the European Community Respiratory Health Survey [18]. In addition, participants completed the St George's Respiratory Questionnaire (SGRQ; validated New Zealand English and Mandarin translations) at the second visit [14]. Bloods were drawn at the second visit for determination of haemoglobin (Sysmex, Mundelein, IL, USA), IgE and high-sensitivity C-reactive protein (hsCRP) (Roche, Indianapolis, IN, USA). Serum Phadiatop (Phadia, Uppsala, Sweden), a composite test for a panel of specific IgEs, with a positive test signifying atopy, was also measured.

### *Standardisation*

New Zealand and China sites used identical questions for the Phase 1 and 2 questionnaires. NZRHS questionnaires were translated into Mandarin and then back-translated to ensure consistency. Spirometry was performed according to an identical study protocol at all sites. One investigator (J.F.) visited all sites to standardise testing.

### *Statistical analysis*

All participants with complete data were included in the analysis. Participant characteristics and questionnaire response rates are described using simple data summaries. Hierarchical cluster analysis using the Agnes algorithm with Ward's method and Gower distance metric was performed as previously described [13]. 12 variables were chosen for cluster analysis: forced expiratory volume in 1 s ($FEV_1$) (% predicted), $FEV_1$/forced vital capacity (FVC) ratio (%), bronchodilator reversibility (% change in $FEV_1$ from baseline after salbutamol administration), peak expiratory flow (PEF) variability (expressed as % of mean), *K*CO (% predicted), log *F*ENO (ppb), log IgE ($IU \cdot L^{-1}$), log hsCRP ($mg \cdot L^{-1}$), age of onset of respiratory symptoms (years), body mass index (BMI) ($kg \cdot m^{-2}$), total SGRQ score and cigarette smoke exposure (pack-years). Functional residual capacity was used as an additional variable in the NZRHS [13], but was not utilised in this analysis as plethysmography data were unavailable at the China sites. The variables were chosen to represent multiple dimensions of airways disease, including airflow obstruction, variability and bronchodilator reversibility, parenchymal damage, symptoms, risk factors, and inflammatory biomarkers. All variables were weighted equally and the number of variables representing each dimension was limited to avoid multicollinearity, which can distort cluster structure [19, 20]. *F*ENO, IgE and hsCRP were log-transformed to ensure that extreme values did not form clusters. The number of potential clusters was determined by establishing cut-points for the dendrograms with a preference for at least 30 participants in the smallest cluster in line with the previously reported analysis and to ensure sufficient participants in each group to allow meaningful examination of cluster characteristics. Each solution was examined for group size and clinical coherence before the preferred solution was selected for phenotype description.

After completion of the combined cluster analysis two exploratory cluster analyses were performed using data from New Zealand and China separately in order to facilitate comparison of patterns seen within each population. The same methodology was used, but in view of the smaller samples clusters containing less than 10 participants were permitted.

## Results

### *Phase 1*

Participant flow through the study is shown in figure 1. Screening questionnaire response rates were ⩾69% for all sites. The proportion of respondents reporting symptoms of wheeze and breathlessness was 14.6% and 4.7% in New Zealand and China, respectively. The majority of eligible subjects who were contacted were successfully recruited into Phase 2; however, questionnaire responses continued to be received after the study was fully recruited meaning that only 30% of eligible participants were recruited to Phase 2.

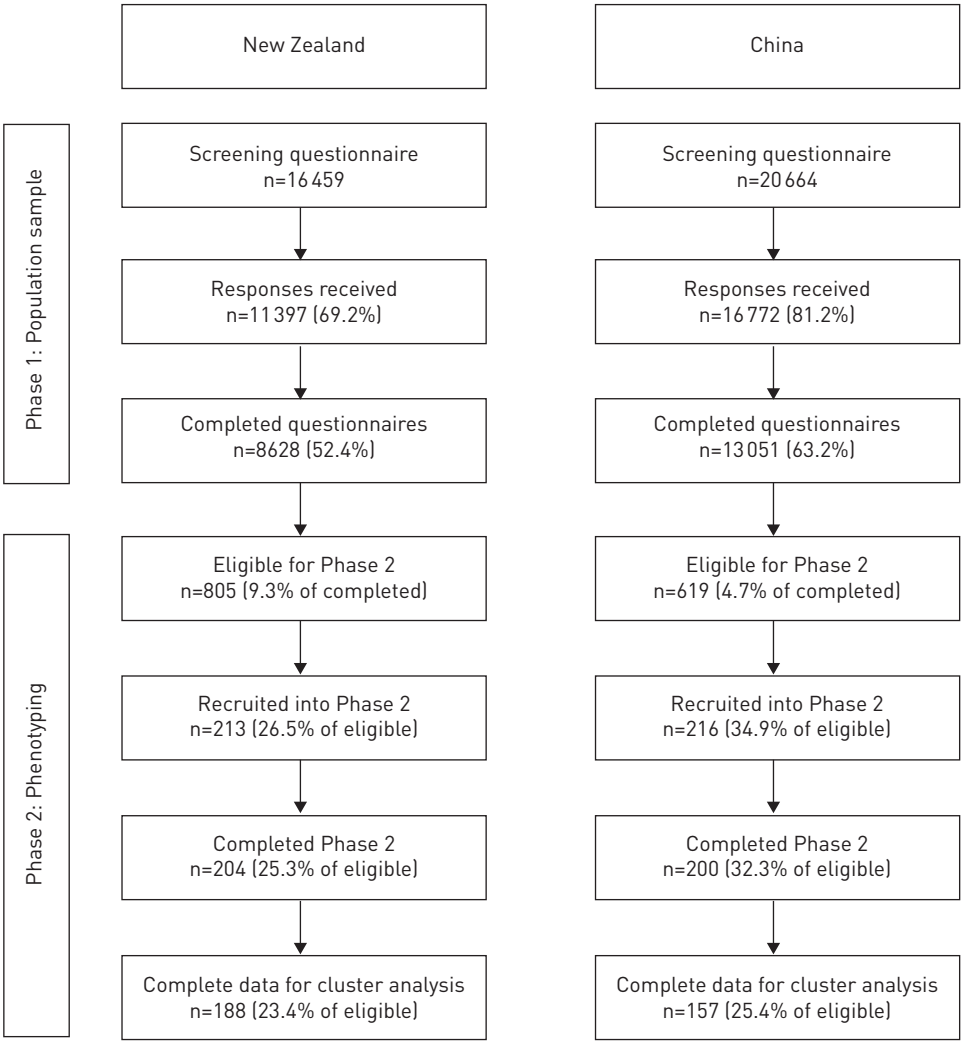


FIGURE 1 Participant flow through the study, shown separately for each study site. Phase 1 recruitment in New Zealand identified a greater number of participants eligible for Phase 2 as these numbers include those screened for the New Zealand Respiratory Health Survey but who were not eligible for inclusion in this analysis due to being aged 18–39 years.

### *Phase 2*

Participant characteristics for the 345 individuals with complete data for cluster analysis are shown in table 1. Comparison of the characteristics of participants from New Zealand and China demonstrated that participants from China had a higher mean age at symptom onset, greater cigarette smoke and biomass exposures, and lower *K*CO, but less marked systemic inflammation and lower IgE levels. Use of any inhaler in the past 12 months and of inhaled corticosteroid (ICS)-containing inhalers in the past 12 months was higher in the New Zealand population, whereas oral methylxanthine use was only seen in the participants from China. Only one participant, based in China, reported use of oral β-agonists within the past year. Around a third of participants had a post-bronchodilator $FEV_1/FVC$ ratio <0.7 and this was similar in both countries.

#### *Combined cluster analysis*

The primary analysis included all participants with complete data. The results of several clustering approaches were compared, and the one which best met pre-specified criteria relating to group size and clinical coherence was selected for phenotype description. The five-cluster solution described using the Agnes algorithm with Ward's method and Gower distance metric was considered most appropriate. This analytic method was that used for phenotype description in the NZRHS [13].

Cluster analysis described five phenotypes with differing characteristics (tables 2 and 3). Participants in Cluster A were predominantly male smokers with late-onset disease, severe obstruction, elevated hsCRP, reduced *K*CO and poor health status, marked reversibility and PEF variability, and markedly elevated total IgE. In addition to smoking, participants had the highest rates of biomass and occupational exposures.

TABLE 1 Participant characteristics

| | All participants | New Zealand | China |
|---|---|---|---|
| **Subjects** | 345 | 188 | 157 |
| **Age years** | 55.9±8.7 | 55.5±9.5 | 56.5±7.7 |
| **Male** | 157 (45.5) | 84 (44.7) | 73 (46.5) |
| **Age at onset years** | 33.2±19.8 | 28.9±20.4 | 38.4±17.7 |
| **BMI kg·m$^{-2}$** | 27.3±5.8 | 29.0±6.5 | 25.3±4.0 |
| **hsCRP mg·L$^{-1}$** | 2.5±3.7 | 2.9±4.0 | 2.0±3.2 |
| **FEV$_1$/FVC %** | 68.2±13.2 | 68.4±12.6 | 67.9±13.9 |
| **Post-bronchodilator FEV$_1$/FVC <0.7** | 125 (36.2) | 65 (34.6) | 60 (38.2) |
| **FEV$_1$ % pred** | 91.5±25.4 | 94.9±23.5 | 87.4±27.0 |
| **IgE IU·L$^{-1}$** | 243±646 | 288±822 | 189.4±321.6 |
| ***K*$_{CO}$% pred** | 94.3±20.4 | 98.7±18.3 | 89.1±21.6 |
| ***F*$_{ENO}$ ppb** | 32.0±32.6 | 32.5±32.7 | 31.5±32.6 |
| **PEF variability** | 33.2±19.2 | 25.6±14.6 | 42.3±20.2 |
| **Reversibility** | 9.8±11.6 | 10.1±11.1 | 9.5±12.2 |
| **Significant reversibility by ERS criteria** | 96 (27.8) | 55 (29.3) | 41 (26.1) |
| **SGRQ score** | 27.3±19.5 | 23.8±17.5 | 31.5±20.9 |
| **Smoking pack-years** | 12.3±21.4 | 9.8±16.6 | 15.4±25.8 |
| **Biomass exposure** | 97 (28.1) | 41 (21.8) | 56 (35.7) |
| **Occupational exposure** | 159 (46.1) | 94 (50.0) | 65 (41.4) |
| **Any inhaler use in past 12 months** | 192 (55.6) | 138 (73.4) | 54 (34.4) |
| **Any ICS-containing medication in past 12 months** | 119 (34.5) | 89 (47.3) | 30 (19.1) |
| **Oral methylxanthine use in past 12 months** | 22 (6.4) | 0 (0) | 22 (14.0) |

Data are presented as n, mean±SD or n (%). BMI: body mass index; hsCRP: high-sensitivity C-reactive protein; FEV$_1$: forced expiratory volume in 1 s; FVC: forced vital capacity; *K*$_{CO}$: transfer coefficient for carbon monoxide adjusted for lung volume, corrected for haemoglobin; *F*$_{ENO}$: exhaled nitric oxide fraction; PEF: peak expiratory flow; ERS: European Respiratory Society; SGRQ: St George's Respiratory Questionnaire; ICS: inhaled corticosteroid.

This group may represent a severe late-onset atopic asthma/chronic bronchitis/emphysema overlap group in smokers with systemic inflammation.

Cluster B contained individuals with moderate partially reversible airflow obstruction, significant cigarette smoke exposure, elevated hsCRP and poor health status, but significant variability, preserved *K*$_{CO}$ and early onset of symptoms. This group may represent a composite phenotype of childhood asthma with concomitant COPD, predominantly chronic bronchitis.

Cluster C was characterised by mild to moderate airflow obstruction with significant variability and reversibility, and elevated *F*$_{ENO}$, total IgE and blood eosinophils. Participants in this cluster had the highest prevalence of eczema, rhinitis, atopy and diagnosed asthma, and had minimal cigarette exposure. This group appears to represent an asthma phenotype with type 2 inflammation.

Clusters D and E were the largest and demonstrated minimal airflow obstruction or reversibility. Cluster D comprised participants with late-onset disease and minimal atopy, whereas Cluster E comprised participants with early-onset disease and raised serum IgE levels, and most had a diagnosis of asthma. Participants in both groups reported symptoms of wheeze and breathlessness in the previous 12 months, and had similar rates of antibiotic or oral steroid use to the other groups, but little evidence of active disease at the time of testing. Cluster D was notable for a similar health status to Cluster C despite minimal evidence of airflow obstruction at the time of testing. We propose Cluster D may represent a nonatopic adult-onset phenotype and Cluster E a mild/intermittent atopic childhood-onset phenotype.

The proportion of participants from New Zealand and China differed between the phenotypes (table 3). A greater proportion of New Zealand participants was allocated to the atopic asthma and mild/intermittent early-onset atopic phenotypes; a greater proportion of China participants was allocated to the two asthma/COPD overlap and mild/intermittent late-onset nonatopic phenotypes.

### *Exploratory cluster analyses*

#### New Zealand site

Five disease phenotypes were identified (supplementary tables S1 and S2). Four phenotypes were similar to the groups identified in the combined New Zealand/China analysis: a severe late-onset asthma/chronic

TABLE 2 Cluster analysis variables by cluster allocation

| | Cluster A | Cluster B | Cluster C | Cluster D | Cluster E |
|---|---|---|---|---|---|
| **Subjects** | 30 | 43 | 38 | 166 | 68 |
| **Pre-bronchodilator $FEV_1$ % pred** | 42.9±15.1 | 78.7±19.8 | 81.7±15.6 | 100.7±19.4 | 104.0±18.0 |
| **Pre-bronchodilator $FEV_1$/FVC %** | 41.6±11.3 | 62.1±11.9 | 60.6±8.2 | 73.4±8.8 | 75.3±5.7 |
| **Bronchodilator reversibility % $FEV_1$ change** | 26.0±16.1 | 12.6±11.8 | 21.0±10.6 | 5.5±7.7 | 5.2±4.8 |
| **PEF variability % of mean** | 44.6±17.6 | 47.3±22.5 | 39.7±17.6 | 31.1±17.7 | 21.0±11.3 |
| **Post-bronchodilator $K_{CO}$% pred** | 63.8±20.7 | 95.5±17.4 | 91.8±19.7 | 95.0±17.4 | 106.8±14.8 |
| **$F_{ENO}$ ppb** | 31.8±44.4 | 16.7±9.9 | 82.4±42.4 | 23.6±15.7 | 34.4±34.3 |
| **Log $F_{ENO}$** | 3.05±0.83 | 2.66±0.56 | 4.27±0.55 | 2.97±0.63 | 3.21±0.80 |
| **IgE IU·L$^{-1}$** | 617±1643 | 146±183 | 466±686 | 136±297 | 277±590 |
| **Log IgE** | 4.6±2.0 | 4.0±1.7 | 5.4±1.3 | 3.8±1.4 | 4.5±1.5 |
| **hsCRP mg·L$^{-1}$** | 4.8±8.9 | 3.3±4.7 | 1.8±1.6 | 2.2±2.4 | 2.0±1.6 |
| **Log hsCRP** | 0.66±1.25 | 0.61±1.0 | 0.18±0.92 | 0.32±0.97 | 0.38±0.84 |
| **Age at onset years** | 43.9±18.4 | 14.0±13.6 | 24.6±14.3 | 46.7±11.4 | 12.5±9.8 |
| **BMI kg·m$^{-2}$** | 23.4±4.0 | 27.2±4.7 | 24.4±3.7 | 28.0±6.2 | 28.9±5.9 |
| **SGRQ score** | 41.8±18.7 | 44.6±20.6 | 23.4±19.3 | 25.7±17.7 | 16.2±11.4 |
| **Smoking pack-years** | 47.5±36.4 | 16.3±19.3 | 3.6±8.0 | 9.5±16.3 | 6.2±13.7 |
| **Comparison[#]** | | | | | |
| Pre-bronchodilator $FEV_1$ % pred | −− | −− | − | + | + |
| Pre-bronchodilator $FEV_1$/FVC % | −− | ± | − | ± | + |
| Bronchodilator reversibility % $FEV_1$ change | ++ | ++ | ++ | −− | −− |
| PEF variability % of mean | ++ | ++ | + | ± | −− |
| Post-bronchodilator $K_{CO}$ % pred | −− | ± | ± | ± | + |
| Log $F_{ENO}$ | ± | − | ++ | ± | ± |
| Log IgE | ± | ± | ++ | ± | ± |
| Log hsCRP | ++ | ++ | −− | − | − |
| Age of onset years | ++ | −− | −− | ++ | −− |
| BMI kg·m$^{-2}$ | − | ± | − | ± | ± |
| SGRQ score | ++ | ++ | − | ± | −− |
| Smoking pack-years | ++ | ++ | −− | −− | −− |

Data are presented as n or mean±SD. $FEV_1$: forced expiratory volume in 1 s; FVC: forced vital capacity; PEF: peak expiratory flow; $K_{CO}$: transfer coefficient for carbon monoxide adjusted for lung volume, corrected for haemoglobin; $F_{ENO}$: exhaled nitric oxide fraction; hsCRP: high-sensitivity C-reactive protein; BMI: body mass index; SGRQ: St George's Respiratory Questionnaire; COPD: chronic obstructive pulmonary disease. Cluster A: severe late-onset asthma/COPD overlap; Cluster B: early-onset asthma/COPD overlap; Cluster C: atopic asthma; Cluster D: adult-onset nonatopic; Cluster E: early-onset atopic mild/intermittent. [#]: compared with the overall mean value of all participants as reference (++: >20% above; +: 10–20% above; ±: within 10%; −: 10–20% below; −−: >20% below).

bronchitis/emphysema overlap group in smokers with raised IgE and systemic inflammation (Cluster 1), a mild late-onset asthma/COPD overlap group in smokers (Cluster 3), a severe type 2 predominant early-onset asthma group (Cluster 2) and a mild/intermittent atopic group (Cluster 5). A fifth cluster was characterised by late-onset of symptoms, obesity and poor respiratory health status despite relatively preserved lung function (Cluster 4). This group had evidence of marked systemic inflammation and high rates of comorbidities, and represents an obese/comorbid phenotype previously described [13].

### China sites

Five clusters were identified. Review of the dendrogram (supplementary figure S2) and cluster characteristics suggested it was preferable to report the first two clusters as a single group. Combination of the first two clusters led to four clinically coherent groups, similar to the groups identified in the combined New Zealand/China analysis (supplementary tables S3 and S4). There was a severe late-onset asthma/chronic bronchitis/emphysema overlap group in smokers with atopy and systemic inflammation (Clusters i and ii), a mild early-onset asthma/chronic bronchitis overlap group in smokers (Cluster iii), and two groups represented by mild/intermittent late-onset disease separated by the degree of atopy (Clusters iv and v). There was no evidence of an obese/comorbid phenotype with systemic inflammation among participants from China.

## Discussion

This study is, to the best of our knowledge, the first to use standardised data from populations in different countries to explore phenotypes of symptomatic airways disease. Combined cluster analysis described five phenotypes of airways disease, with individuals from New Zealand and China represented in each

TABLE 3 Phenotype description using additional analysis variables

| | Cluster A | Cluster B | Cluster C | Cluster D | Cluster E |
|---|---|---|---|---|---|
| **Subjects** | 30 | 43 | 38 | 166 | 68 |
| **Demographics** | | | | | |
| Age years | 62.0±6.4 | 57.2±8.1 | 52.8±8.7 | 57.3±8.1 | 50.7±8.4 |
| Male | 24 (80.0) | 18 (41.9) | 13 (34.2) | 70 (42.2) | 32 (47.1) |
| New Zealand | 8 (26.7) | 19 (44.2) | 24 (63.2) | 79 (47.6) | 58 (85.3) |
| **Risk factors/symptoms** | | | | | |
| Smoking status | | | | | |
| Current smoker | 14 (46.7) | 14 (32.6) | 2 (5.3) | 25 (15.1) | 3 (4.4) |
| Ex-smoker | 14 (46.7) | 13 (30.2) | 11 (29.0) | 49 (29.5) | 23 (33.8) |
| Never-smoker | 2 (6.7) | 16 (37.2) | 25 (65.8) | 92 (55.4) | 42 (61.8) |
| Biomass exposure | 13 (43.3) | 12 (27.9) | 11 (29.0) | 40 (24.1) | 21 (30.9) |
| Occupational exposure | 17 (56.7) | 20 (46.5) | 13 (34.2) | 75 (45.2) | 34 (50.0) |
| Chronic bronchitis symptoms | 16 (53.3) | 17 (39.5) | 6 (15.8) | 54 (32.5) | 9 (13.2) |
| **Previous respiratory diagnoses** | | | | | |
| Asthma | 10 (33.3) | 33 (76.7) | 32 (84.2) | 68 (41.0) | 57 (83.8) |
| Chronic bronchitis | 15 (50) | 25 (58.1) | 9 (23.7) | 53 (31.9) | 17 (25.0) |
| COPD | 8 (26.7) | 8 (18.6) | 2 (5.3) | 6 (3.6) | 1 (1.5) |
| Emphysema | 10 (33.3) | 6 (14.0) | 0 (0) | 8 (4.8) | 2 (2.9) |
| **Atopy** | | | | | |
| Phadiatop positive | 7 (23.3) | 21 (48.8) | 31 (81.6) | 48 (28.9) | 45 (66.2) |
| Eczema diagnosis | 12 (40.0) | 20 (46.5) | 24 (63.2) | 86 (51.8) | 42 (61.8) |
| Rhinitis | 9 (30.0) | 35 (81.4) | 34 (89.5) | 103 (62.1) | 54 (79.4) |
| Eosinophil count $\times 10^9$ $L^{-1}$ | 0.2±0.2 | 0.2±0.1 | 0.9±2.2 | 0.2±0.4 | 0.2±0.2 |
| **Comorbidities** | | | | | |
| CVD | 9 (30.0) | 12 (27.9) | 4 (10.5) | 37 (22.3) | 12 (17.7) |
| GORD | 8 (26.7) | 18 (41.9) | 10 (26.3) | 70 (42.2) | 25 (36.8) |
| Hypertension | 12 (40) | 20 (46.5) | 7 (18.4) | 58 (34.9) | 18 (26.5) |
| Depression and/or anxiety | 2 (6.7) | 6 (14.3) | 7 (18.4) | 29 (17.5) | 15 (22.1) |
| **Medication use in last 12 months** | | | | | |
| Any inhaler[#] | 17 (56.7) | 26 (60.5) | 32 (84.2) | 70 (42.2) | 47 (69.1) |
| ICSs | 4 (13.3) | 11 (25.6) | 11 (28.9) | 28 (16.9) | 20 (29.4) |
| Combination ICS/LABA | 5 (16.7) | 10 (23.3) | 6 (15.8) | 26 (15.7) | 8 (11.8) |
| **Healthcare use in last 12 months** | | | | | |
| Urgent emergency department/hospital visit | 6 (20.0) | 2 (4.7) | 7 (18.4) | 9 (5.4) | 3 (5.9) |
| One or more courses of antibiotic | 9 (30.0) | 11 (25.6) | 12 (31.6) | 41 (30.7) | 22 (35.3) |
| Oral steroids | 2 (6.7) | 5 (11.6) | 5 (13.2) | 11 (6.6) | 7 (10.3) |
| **Lung function** | | | | | |
| Post-bronchodilator $FEV_1$/FVC % | 45.1±12.4 | 65.4±11.6 | 68.0±8.6 | 75.9±8.9 | 78.7±6.0 |

Data are presented as n, mean±SD or n (%). COPD: chronic obstructive pulmonary disease; CVD: cardiovascular disease; GORD: gastro-oesophageal reflux disease; ICS: inhaled corticosteroid; LABA: long-acting β-agonist; $FEV_1$: forced expiratory volume in 1 s; FVC: forced vital capacity; COPD: chronic obstructive pulmonary disease. Cluster A: severe late-onset asthma/COPD overlap; Cluster B: early-onset asthma/COPD overlap; Cluster C: atopic asthma; Cluster D: adult-onset nonatopic; Cluster E: early-onset atopic mild/intermittent. [#]: denotes a positive answer to the question "Have you used any inhalers to help your breathing at any time in the last 12 months".

phenotype in differing proportions. The study provides evidence of asthma/COPD overlap phenotypes in two distinct populations with considerable differences in genetic heritage and environmental exposures. Asthma/COPD overlap phenotypes have been identified in previous cluster analyses [1, 4, 8], and the consistent identification of asthma/COPD overlap groups in both the primary combined analysis and the separate exploratory analyses provides important validation of their existence. We have also demonstrated that the candidate obese/comorbid phenotype identified in previous cluster analyses [3, 6] was present in the Caucasian New Zealand population, but not the China population.

We identified two asthma/COPD overlap phenotypes. One was characterised by severe airflow obstruction, with evidence of asthma and emphysema, coexistent chronic bronchitis, and markedly raised serum IgE and hsCRP, indicating systemic inflammation. The heavy smoking histories, together with the highest rates of biomass and occupational exposures and the later age of onset, suggest that these exposures are likely to be relevant to disease pathogenesis. Cigarette and biomass exposure, rather than atopy, may be contributing to the elevated IgE in this group given the low proportion of participants with a positive Phadiatop. In contrast, the other overlap group had less severe airflow obstruction and early onset of

symptom presentation, significant but less pronounced tobacco smoke, biomass and occupational exposures, lower IgE levels, and preserved *K*CO. This group may represent childhood asthma with concomitant COPD, predominantly chronic bronchitis. While two distinct phenotypic groups were identified it is notable that we did not describe a separate COPD cluster and it is probable that clinically relevant subphenotypes may also be present, *e.g.* "pure emphysema in smokers" that was identified in the original New Zealand cluster analysis [4].

The presence of these overlap groups is broadly consistent with both the Dutch hypothesis and/or an interaction between separate disease processes, *i.e.* asthma and COPD [21–23]. Whatever the underlying cause(s), at least one in five adults with symptomatic airflow obstruction exhibited one of these overlap phenotypes, and yet our knowledge of their natural histories and the most appropriate management is currently limited [24]. This is in part because there are no universally accepted definitions and therefore different studies may be characterising different populations under the same broad label [25]. It is possible that progress will accelerate as we move away from trying to assign each patient a particular label and come to focus on the "treatable traits" each individual possesses [26].

Patients with the overlap phenotypes are not represented in the majority of major randomised controlled trials performed to date as their tobacco exposure and prior diagnosis of COPD would have led to their exclusion from most asthma trials, and the presence of reversible airflow obstruction, atopy and prior diagnosis of asthma their exclusion from most COPD trials [27, 28]. There is an urgent need for further research to determine the natural history, longitudinal stability and treatment responsiveness of patients with the overlap phenotypes to guide evidence-based recommendations [29]. These studies should attempt to control for the confounding effect of differences in treatments between participants seen in all cross-sectional studies. The longitudinal data provide the ultimate test of the clinical relevance of a phenotype [30] and do not always demonstrate clinically important differences in outcome when such studies are performed [31].

Another candidate phenotype identified in this study is the obese/comorbid phenotype seen in the New Zealand cohort. Participants in this group had multiple comorbidities, evidence of systemic inflammation and disproportionately poor health status relative to severity of airflow obstruction. This is consistent with both the obesity-related asthma phenotypes [2, 4–6, 32–34] and the "systemic COPD group" [2] described in Western populations. Although obesity may be associated with increased systemic inflammation in individuals with asthma [35], the systemic inflammation may also be associated with the comorbidities.

There was no evidence of a similar obese/comorbid group in the China cohort, which may be in part due to the lower prevalence of obesity in this population. A BMI threshold of 28.0 $kg \cdot m^{-2}$ is recommended to define obesity in China [36], as against 30.0 $kg \cdot m^{-2}$ in Western populations [37]. Using these thresholds, 42 out of 157 (26.8%) of the China cohort were defined as obese, as against 77 out of 188 (41.0%) of the New Zealand cohort. The prevalence of obesity is increasing in China [38] and obesity-related phenotypes may therefore become evident in the future. Disease phenotypes are likely to change in prevalence over time as the population characteristics alter in response to environmental exposures.

The two mild/intermittent phenotypes are interesting in that they had normal lung function yet similar requirements for antibiotic and oral steroid courses. These features may reflect groups who are prone to exacerbations, presumably infective. These two mild/intermittent groups were differentiated by their age of symptom onset and underlying atopy, a feature which is consistent with findings of other reports, both in Asia [12] and the USA and Europe [3–5, 39, 40]. They also differed in the proportion of individuals with a doctor's diagnosis of asthma (Cluster D 41%, Cluster E 84%).

The main strength of this study is the use of standardised data collected from populations in different countries. This permitted a combined cluster analysis and comparison of separate cluster analyses between populations. The population sampling approach utilised in both countries maximised the chance that the participants tested were representative of the communities studied. We cannot exclude the possibility of a degree of response or selection bias as we do not know the characteristics of those who did not respond and only a proportion of those eligible for participation were recruited.

Plethysmography data was not available for the China cohort. Any change to the variables of a cluster analysis has the potential to alter the results; however, the fact that similar phenotypes were seen despite a change in variables from the NZRHS [13] may be evidence that these are real disease types rather than artefacts in the data.

There were some differences between the populations which may be relevant. Participants in China were less likely to complain of wheeze and breathlessness, which may be due to differences in both disease prevalence and interpretation of the screening questions. The New Zealand screening questionnaire was translated into Mandarin and then back-translated to ensure consistency; however, it is possible that some

of the difference in eligibility rates is due to differences in how the questions were perceived. Consistent with this, participants from China had more severe airflow obstruction, suggesting that individuals with minimal disease were less likely to report wheeze and breathlessness than in New Zealand. The greater cigarette and biomass smoke exposures reported in China may explain their higher prevalence of COPD-related phenotypes compared with the asthma-related phenotypes predominant in the New Zealand cohort.

In summary, combined cluster analysis using data from participants in China and New Zealand provides evidence to support the validity of distinct asthma/COPD phenotypes across different populations. This study also highlights possible differences in phenotype prevalence between populations and the existence of an obese/comorbid phenotype in a Caucasian Western population.

## Acknowledgements

The NZ–China Study Group: Steering Committee: James Fingleton (clinical coordinating investigator), Mark Weatherall (study biostatistician), Kewu Huang (principal investigator China site 2), Richard Beasley (principal investigator New Zealand), Chen Wang (principal investigator China site 1), Stefan Ivanov (AstraZeneca), Piet Bruijnzeel (AstraZeneca). Wellington, New Zealand: Mathew Williams, Philippa Shirtcliffe (designated safety reviewer), Thomas Charles, Darren Bowles, Rianne Strik, Natalie Dooney, Tanya Baker, Mitesh Patel, Mark Holliday, Maureen Stretch, Alison Pritchard, Denise Fabian, Claire Munro, Alexander Hosking, Alex Brinded (Wellington, MRINZ); Gordon Purdie (Otago University, Wellington). Beijing, China: Yong Lu, Xiaohong Chang, Zhang Jun, Qingyuan Li, Baomei Wu (Beijing Chao-Yang Hospital); Zhang Hongsheng, Chai Di, Li Xiaomeng, Lv Lin, Ma Yali, Yang Yimeng, Wang Yuxia, Tan Xiaoming, Yang Jingjing (Beijing Hospital).

Author contributions: R. Beasley and J. Fingleton conceived the study, and R. Beasley was principal investigator for the study. C. Wang was the senior investigator in China. C. Wang and K. Huang were principal investigators for the two sites in China. J. Fingleton wrote the first draft of the manuscript and was clinical coordinating investigator for the study. C. Wang, K. Huang, J. Fingleton, M. Weatherall, P. Bruijnzeel, R. Beasley and S. Ivanov all contributed to the design of the study as well as analysis and data interpretation. C. Wang, H. Zhang, K. Huang, J. Fingleton, Y. Guo and W. Wang performed data collection with the help of members of the NZ–China Study Group. All authors contributed to the revision of the paper and approved the final version.

## References

1 Wardlaw AJ, Silverman M, Siva R, *et al.* Multi-dimensional phenotyping: towards a new taxonomy for airway disease. *Clin Exp Allergy* 2005; 35: 1254–1262.
2 Garcia-Aymerich J, Gómez FP, Benet M, *et al.* Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes. *Thorax* 2011; 66: 430–437.
3 Haldar P, Pavord ID, Shaw DE, *et al.* Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med* 2008; 178: 218–224.
4 Weatherall M, Travers J, Shirtcliffe PM, *et al.* Distinct clinical phenotypes of airways disease defined by cluster analysis. *Eur Respir J* 2009; 34: 812–818.
5 Moore WC, Meyers DA, Wenzel SE, *et al.* Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med* 2010; 181: 315–323.
6 Sutherland ER, Goleva E, King TS, *et al.* Cluster analysis of obesity and asthma phenotypes. *PLoS One* 2012; 7: 1–7.
7 Boudier A, Curjuric I, Basagaña X, *et al.* Ten-year follow-up of cluster-based asthma phenotypes in adults. A pooled analysis of three cohorts. *Am J Respir Crit Care Med* 2013; 188: 550–560.
8 Jo K-W, Ra S, Chae E, *et al.* Three phenotypes of obstructive lung disease in the elderly. *Int J Tuberc lung Dis* 2010; 14: 1481–1488.
9 Soriano JB. The proportional Venn diagram of obstructive lung disease: two approximations from the United States and the United Kingdom. *Chest* 2003; 124: 474–481.
10 Marsh SE, Travers J, Weatherall M, *et al.* Proportional classifications of COPD phenotypes. *Thorax* 2008; 63: 761–767.
11 Kim L, El-Dahr JM. Skin test reactivity in children with asthma and rhinitis in New Orleans prior to hurricane Katrina. *J Allergy Clin Immunol* 2006; 117: Suppl., S6.
12 Kim TB, Jang AS, Kwon HS, *et al.* Identification of asthma clusters in two independent Korean adult asthma cohorts. *Eur Respir J* 2013; 41: 1308–1314.
13 Fingleton J, Travers J, Williams M, *et al.* Treatment responsiveness of phenotypes of symptomatic airways obstruction in adults. *J Allergy Clin Immunol* 2015; 136: 601–609.
14 Miller MR, Hankinson J, Brusasco V, *et al.* Standardisation of spirometry. *Eur Respir J* 2005; 26: 319–338.
15 Marsh S, Aldington S, Williams M, *et al.* Complete reference ranges for pulmonary function tests from a single New Zealand population. *NZ Med J* 2006; 119: U2281.
16 Zheng J, Zhong N. Normative values of pulmonary function testing in Chinese adults. *Chin Med J (Engl)* 2002; 115: 50–54.
17 Putinati S. ATS/ERS recommendations for standardized procedures for the online and offline measurement of exhaled lower respiratory nitric oxide and nasal nitric oxide, 2005. *Am J Respir Crit Care Med* 2005; 171: 912–930.
18 Janson, C, Anto, J, Burney, P, *et al.* The European Community Respiratory Health Survey: what are the main results so far? *Eur Respir J* 2001; 18: 598–611.
19 Everitt BS, Landau S, Leese M, *et al.* Cluster Analysis. 5th Edn. Chichester, Wiley, 2011.
20 Weatherall M, Shirtcliffe P, Travers J, *et al.* Use of cluster analysis to define COPD phenotypes. *Eur Respir J* 2010; 36: 472–474.

21 Postma DS, Boezen HM. Rationale for the Dutch hypothesis: allergy and airway hyperresponsiveness as genetic factors and their interaction with environment in the development of asthma and COPD. *Chest* 2004; 126: 2 Suppl., 96S–104S.
22 Vanfleteren LEGW, Kocks JWH, Stone IS, *et al.* Moving from the Oslerian paradigm to the post-genomic era: are asthma and COPD outdated terms? *Thorax* 2014; 69: 72–79.
23 Postma DS, Weiss ST, van den Berge M, *et al.* Revisiting the Dutch hypothesis. *J Allergy Clin Immunol* 2015; 136: 521–529.
24 Barnes PJ. Therapeutic approaches to asthma–chronic obstructive pulmonary disease overlap syndromes. *J Allergy Clin Immunol* 2015; 136: 531–545.
25 Reddel HK. Treatment of overlapping asthma–chronic obstructive pulmonary disease: can guidelines contribute in an evidence-free zone? *J Allergy Clin Immunol* 2015; 136: 546–552.
26 Agusti A, Bel E, Thomas M, *et al.* Treatable traits: toward precision medicine of chronic airway diseases. *Eur Respir J* 2016; 47: 410–419.
27 Travers J, Marsh S, Williams M, *et al.* External validity of randomised controlled trials in asthma: to whom do the results of the trials apply? *Thorax* 2007; 62: 219–223.
28 Travers J, Marsh S, Caldwell B, *et al.* External validity of randomized controlled trials in COPD. *Respir Med* 2007; 101: 1313–1320.
29 Gelb AF, Nadel JA. Understanding the pathophysiology of the asthma–chronic obstructive pulmonary disease overlap syndrome. *J Allergy Clin Immunol* 2015; 136: 553–555.
30 Han MK, Agusti A, Calverley PM, *et al.* Chronic obstructive pulmonary disease phenotypes: the future of COPD. *Am J Respir Crit Care Med* 2010; 182: 598–604.
31 Bourdin A, Molinari N, Vachier I, *et al.* Prognostic value of cluster analysis of severe asthma phenotypes. *J Allergy Clin Immunol* 2014; 134: 1043–1050.
32 Haldar P, Pavord ID, Shaw DE, *et al.* Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med* 2008; 178: 218–224.
33 Benton AS, Wang Z, Lerner J, *et al.* Overcoming heterogeneity in pediatric asthma: tobacco smoke and asthma characteristics within phenotypic clusters in an African American cohort. *J Asthma* 2010; 47: 728–734.
34 Musk AW, Knuiman M, Hunter M, *et al.* Patterns of airway disease and the clinical diagnosis of asthma in the Busselton population. *Eur Respir J* 2011; 38: 1053–1059.
35 Sutherland TJT, Cowan JO, Young S, *et al.* The association between obesity and asthma: interactions between systemic and airway inflammation. *Am J Respir Crit Care Med* 2008; 178: 469–475.
36 Chen C, Lu FC. The guidelines for prevention and control of overweight and obesity in Chinese adults. *Biomed Environ Sci* 2004; 17: Suppl., 1–36.
37 World Health Organization. Obesity: Preventing and Managing the Global Epidemic. Report of a WHO consultation (WHO Technical Report 894). Geneva, WHO, 2000.
38 Wu Y. Overweight and obesity in China. *BMJ* 2006; 333: 362–363.
39 Fitzpatrick AM, Teague WG, Meyers DA, *et al.* Heterogeneity of severe asthma in childhood: confirmation by cluster analysis of children in the National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. *J Allergy Clin Immunol* 2011; 127: 382–383.
40 Siroux V, Basagan X, Boudier A, *et al.* Identifying adult asthma phenotypes using a clustering approach. *Eur Respir J* 2011; 38: 310–317.