



Chronic obstructive pulmonary disease comorbidities network

Miguel J. Divo¹, Ciro Casanova², Jose M. Marin³, Victor M. Pinto-Plata¹, Juan P. de-Torres⁴, Javier J. Zulueta⁴, Carlos Cabrera⁵, Jorge Zagaceta⁴, Pablo Sanchez-Salcedo⁴, Juan Berto⁴, Rebeca Baz Davila², Ana B. Alcaide⁴, Claudia Cote⁶ and Bartolome R. Celli¹, The BODE Collaborative Group

Affiliations: ¹Pulmonary and Critical Care Division, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. ²Pulmonary Department, Hospital Universitario La Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain. ³Respiratory Service, Hospital Universitario Miguel Servet, Zaragoza, Spain. ⁴Department of Pulmonology, University Clinic of Navarra, Pamplona, Spain. ⁵Pulmonary Department, Hospital Universitario de Gran Canaria Dr Negrin, Las Palmas de Gran Canarias, Spain. ⁶Division of Pulmonary and Critical Care Medicine, The Bay Pines Veterans Affairs Healthcare System, University of South Florida, Tampa, FL, USA.

Correspondence: Miguel J. Divo, Pulmonary and Critical Division, Brigham and Women's Hospital, 75 Francis St., Boston, MA 02115, USA. E-mail: mdivo@partners.org

ABSTRACT Multimorbidity frequently affects the ageing population and their co-existence may not occur at random. Understanding their interactions and that with clinical variables could be important for disease screening and management.

In a cohort of 1969 chronic obstructive pulmonary disease (COPD) patients and 316 non-COPD controls, we applied a network-based analysis to explore the associations between multiple comorbidities. Clinical characteristics (age, degree of obstruction, walking, dyspnoea, body mass index) and 79 comorbidities were identified and their interrelationships quantified. Using network visualisation software, we represented each clinical variable and comorbidity as a node with linkages representing statistically significant associations.

The resulting COPD comorbidity network had 428, 357 or 265 linkages depending on the statistical threshold used ($p \leq 0.01$, $p \leq 0.001$ or $p \leq 0.0001$). There were more nodes and links in COPD compared with controls after adjusting for age, sex and number of subjects. In COPD, a subset of nodes had a larger number of linkages representing hubs. Four sub-networks or modules were identified using an inter-linkage affinity algorithm and their display provided meaningful interactions not discernible by univariate analysis.

COPD patients are affected by larger number of multiple interlinked morbidities which clustering pattern may suggest common pathobiological processes or be utilised for screening and/or therapeutic interventions.



@ERSpublications

COPD patients are affected by interlinked comorbidities forming structured networks

<http://ow.ly/MT4XT>

Received: Sept 19 2014 | Accepted after revision: March 25 2015

Supplementary information: This article has supplementary material available from erj.ersjournals.com

Support statement: M. Divo was supported by the European Respiratory Society Maurizio Vignola Award.

Conflict of interest: None declared.

Copyright ©ERS 2015

Introduction

Increased life expectancy has resulted in an ageing population suffering from one or more chronic, non-communicable diseases [1]. 60% of individuals aged ≥ 65 years have multiple medical conditions referred to as multimorbidity [2], and the presence of these conditions is thought to be a combination of individual susceptibility to each condition, cumulative noxious exposures [3] and the effect of ageing. The extent to which these factors interact likely explains some of the inter-subject variability in disease burden but this process is poorly understood. Applying the classic epidemiological approach in the study of noncommunicable diseases has its limitations because the success of this model depends on a short interval from exposure to disease manifestation and a small number of exposures, vectors or interactions. A more integrated multidimensional approach is needed if we want to further understand the complex relationships of chronic multimorbidity.

Chronic obstructive pulmonary disease (COPD) is one of the most prevalent chronic diseases worldwide [4]. It is a complex multicomponent disease better characterised when using, in addition to the degree of airway obstruction, extra-pulmonary clinical variables like body mass index (BMI), exercise capacity, dyspnoea, smoking status, exacerbations and age captured in the validated multidimensional indices like BODE (BMI, airflow obstruction, dyspnoea and exercise capacity) [5], ADO (age, dyspnoea and airflow obstruction) [6] and DOSE (dyspnoea, airflow obstruction, smoking, exacerbation) [7]. COPD patients suffer from a high proportion of comorbid conditions [8–11] and up to two-thirds of individuals suffering from COPD die of non-pulmonary causes [12, 13]. Comorbidities also influence important patient-centred outcomes, such as health status [14, 15], response to therapeutic intervention [16], and frequent hospital readmissions [17, 18]. The majority of studies of COPD comorbidities report the relationship between a single or small number of comorbidities [19–22]. The inclusion of multiple comorbidities provides a broader view of the complex interaction amongst them that may reveal previously unappreciated associations.

In a previous report, our group described the presence of 79 different comorbidities in the BODE COPD cohort. While prevalence of these conditions ranged from 0.1 to 52%, only 12 comorbidities conferred an increased risk of death when co-occurring with COPD [13]. We propose, however, that the true clinical significance of the other 67 comorbidities may not be adequately assessed by standard regression analysis. In addition, prior reports suggest that comorbidities cluster differently among clinical, demographic and anthropometric characteristics (included in the BODE, ADO and DOSE indices) therefore exploring the interactions between comorbidities and those variables could provide a unique opportunity to better understand COPD complexity beyond a single disease perspective [22–25].

Network science has emerged as a field focussed on the understanding of complex systems by mapping the interconnectivity of such data as objects, persons, proteins, mobile phones or diseases [26–29]. Networks graphics are composed of individual components called nodes and a grid of interconnecting edges representing some type of association between nodes. These interconnected nodes can then be readily visualised revealing the structural basis of the system as either hierarchical, random or scale free networks. The structure of the network will help to identify highly connected individual nodes and specific communities of nodes called modules with great potential applicability in medicine [27]. We hypothesised that such a network-based approach would provide a better understanding of the interaction among comorbidities, clinical and demographic variables in patients with COPD and compared the results with a non-COPD control group.

Methods

The cohort

The BODE registry is an observational prospective multicenter cohort of COPD patients attending outpatient pulmonary clinics at one of the six BODE study centres (Tampa, FL and Boston, MA (USA) and Pamplona, Tenerife, Las Palmas de Gran Canarias and Zaragoza (Spain)). In brief, COPD was defined by history of smoking and spirometric measures of lung function which followed American Thoracic Society/European Respiratory Society standards [30]. We compared our COPD cohort with non-COPD controls recruited from the same centres and same period. Non-COPD controls were smokers and non-smoker individuals with no evidence of airway obstruction on spirometry who volunteered to participate in our cohort. All participants were in clinically stable condition and receiving standard therapy when appropriate. Participants were excluded if they had an illness that was likely to result in death within a year, or inability to take the lung-function and 6-min-walk tests or any condition that could unacceptably increase the subject's risk of performing any of the testing. Between November 1997 and June 2013, a total of 1969 COPD patients and 316 non-COPD controls were enrolled and data collected at the initial visit was used for this analysis. The ethics committee at each of the participating centres approved the study and all patients signed informed consent prior to enrolment.

Comorbidities ascertainment

All comorbidities were systematically recorded by means of patient's interview including those diseases listed in the Charlson Comorbidity Index [31]. Comorbidities not reported by the patient but documented in their

medical records were also included. The presence of each comorbidity was then confirmed by the investigators at each site through a detailed review of available medical records and medication or therapy specific to any disease.

Clinical variables

Demographics, anthropometrics (height and weight), pulmonary function testing, 6-min walk test distance and Modified Medical Research Council dyspnoea score were collected at the initial visit. We chose the following 10 relevant clinical variables included in the multidimensional indices of COPD [5–7] to explore the associations with the 79 comorbidities in the COPD cohort. The selected cut-offs are based on previous published studies: 1) Age categories: a) Participants 55 years old or younger, b) participant between the ages of 56 and 64 years, c) and those 65 years or older [32]; 2) BMI: a) underweight BMI $\leq 21 \text{ kg}\cdot\text{m}^{-2}$ or b) obese BMI $\geq 30 \text{ kg}\cdot\text{m}^{-2}$ [23, 33]; 3) symptoms: MMRC score ≥ 2 points [34]; 4) obstruction: a) FEV₁ $\leq 50\%$ predicated, b) FEV₁ 60–80% of predicted (mild obstruction) [35]; 5) exercise capacity as those with a 6MWD ≤ 350 m [36]; and 6) those who are current smokers [7].

Analysis

We identified a total of 79 distinct comorbidities and compared, using Fisher's exact test, their cumulative number and prevalence between the COPD and no-COPD subjects correcting for sex and age. We determined the correlation between all comorbidities and the 10 clinical variables by calculating the Pearson's correlation (Φ) for binary variables as described in the study by HIDALGO *et al.* [26]. To correct for family-wise error rate due to multiple comparisons, a correlation was considered significant if the p-value was ≤ 0.01 . Additional analyses were performed with more stringent thresholds at $p \leq 0.001$ and $p \leq 0.0001$.

The network graph was constructed with the 79 comorbidities and for the identification of network modules we also included the 10 clinical variables. Each comorbidity and clinical variable is represented in the graph by a specific node with two attributes: the diameter of the node is proportional to the prevalence of the comorbidity or clinical variable and the colour represents the organ system to which the comorbidity belongs (cardiovascular, pulmonary, gastrointestinal, *etc.*). Links or edges between nodes represent statistically significant associations. The edges' thickness represents the strengths of their association (Pearson's Φ). We used the Gephi Graph visualisation and Manipulation software V-0.8.2 beta [37] to create the network graphics. We calculated the degree (number of edges) of each node and plotted their distribution in order to classify the network as either random (most nodes have similar number of edges that follow a Poisson's distribution) or scale free network (few nodes are highly connected and the distribution follow a one-sided heavy tail) [27]. We determined if there was a relationship between disease prevalence and the number of links (degree) and identified those nodes representing hubs. Finally we search for the existence of modules represented by highly interlinked topological clusters in the network using the computational algorithm proposed by BLONDEL *et al.* [38] included in the Gephi statistical module. This community detection algorithm is based on a network property called modularity. Modularity is the fraction of the edges that fall within the given groups of nodes minus the expected such fraction if edges were distributed at random. The value of the modularity lies in the range: -1 to 1 . If positive then the number of edges within groups exceeds the number expected on the basis of chance. Further the algorithm used in our analysis factor the weight of the links (in our case strength of association between comorbidities or Φ) [39]. All analyses were performed using SAS JMP Pro software, version 11.0 (SAS Institute).

Results

The baseline characteristics of the 1969 patients and 316 non-COPD controls are presented in table 1. The COPD cohort is predominantly represented by ever-smokers males, with a mean age of 67 ± 9 years and with moderate to severe obstruction at baseline while the controls have significantly higher number of current smokers (48%), higher proportion of females and are younger. To account for the statistical difference in mean age, all comparisons amongst COPD and non-COPD controls were stratified using the following age brackets: 55 years old or younger, age between 56 and 64 years and those 65 years or older (table 1). In addition, we randomly selected 311 COPD subject matching the non-COPD cohort by age, sex and smoking status and performed the same analysis (table E1 in the supplementary material).

The median number of comorbidities per individual differed significantly between the 2 groups and the difference remains significant when the comparison was adjusted by sex, age and matched numbers (table 1 and table E1 in the supplementary material).

The comparison of comorbidities prevalence between COPD and non-COPD individuals stratified by age categories and sex are presented in figure E1 and E2 in the online supplement. We observed an increase in the number of different comorbidities from 65 to 73 and 79 diseases with increasing age bracket and the prevalence tended to be higher in the COPD cohort compared to controls. The ranking for the top 20 most frequent comorbidities also differs by age group as shown in table E2.

TABLE 1 Subject characteristics

Demographic	COPD	Controls	p-value
Subjects n	1969	316	
Age	67±9	58±11	<0.0001
Age categories			
Age ≤55 years	236 (12)	132 (43)	<0.0001
Age >55 and <65	528 (27)	84 (26)	NS
Age ≥65 years	1205 (61)	100 (31)	<0.0001
Male	1734 (89)	194 (56)	
Current smoker	603 (33)	164 (48)	<0.0001
Pack-years	75±45	33±22	<0.0001
Spirometry			
FEV1 % predicted	52±21	100±15	<0.0001
GOLD I	221 (11)		
GOLD II	757 (38)		
GOLD III	692 (35)		
GOLD IV	299 (15)		
BODE index	3 (1–5)		
BODE 0–2	795 (42)		
BODE 3–4	528 (28)		
BODE 5–6	319 (17)		
BODE 7–10	261 (14)		
BMI	27.2±5.4	27.8±4.9	0.115
≤21 kg·m ⁻²	268 (14)	29 (9)	NS
>21–<25 kg·m ⁻²	355 (18)	59 (19)	NS
≥25–<30 kg·m ⁻²	759 (39)	130 (42)	NS
≥30–<35 kg·m ⁻²	409 (21)	71 (22)	NS
≥35 kg·m ⁻²	175 (9)	27 (9)	NS
Comorbidities per patient[#]			
Comorbidities (whole)	4 (2–7; 0–20)	2 (0–3; 0–14)	<0.0001
Comorbidities by age			
≤55 years	4 (1–6; 0–14)	1 (0–2; 0–7)	<0.0001
>55–<65 years	4 (2–6; 0–19)	2 (1–4; 0–8)	<0.0001
≥65 years	5 (3–7; 0–20)	2 (3–4; 0–14)	<0.0001
Comorbidities by sex			
Male	4 (2–7; 0–20)	2 (1–3; 0–14)	<0.0001
Female	3 (1–5; 0–14)	1 (0–3; 0–8)	<0.0001

Data are presented as mean±sd, n (%) or median (interquartile range), unless otherwise stated. #: number of comorbidities per patient presented as median (interquartile range; range). COPD: chronic obstructive pulmonary disease; FEV1: forced expiratory volume in 1 s; GOLD: Global Initiative on Obstructive Lung Disease; BODE: body mass index, airflow obstruction, dyspnoea and exercise capacity; BMI: body mass index; NS: not significant.

In the COPD cohort, female patients had fewer comorbidities than males (median=4, IQR 2–6 compared to 5, IQR 3–8, $p=0.0001$). In addition to the sex-predominant comorbidities (*i.e.* breast cancer), female patients had a significantly higher proportion of osteoporosis, hypothyroidism and venous insufficiency. The detailed comparison of comorbidities prevalence by sex is presented in table E3 of the online supplement and their Networks are illustrated in figure E3.

Comorbidities network

The comorbidities network for COPD and non-COPD controls are presented in figure 1. The COPD Comorbidities Network comprises 79 nodes representing each comorbidity and a total of 428 links (degree) representing those correlations with a p -value of ≤ 0.01 , while the non-COPD controls network have 56 nodes and 149 links. If we compare the control group ($n=316$) with a randomly selected COPD subset matched by number, age, sex and current smoking status, the resulting network has 69 nodes and 214 edges (figure E4 of the online supplement). Comorbidities are not exclusive to COPD patients, however the prevalence, diversity of diseases and density of associations (links) are higher in COPD compared to controls.

If we use a p -value cut-off of ≤ 0.001 the number of links decreases to 357 and to 265 if the significance threshold is lowered to a $p \leq 0.0001$ (figure E5a). The correlation coefficients values ranged from 0.06 to 0.42 (median=0.09, IQR=0.07–0.12) and for negative correlations from -0.05 to -0.16 (median= -0.08 , IQR -0.07 to -0.12) (figure E5b online supplement).

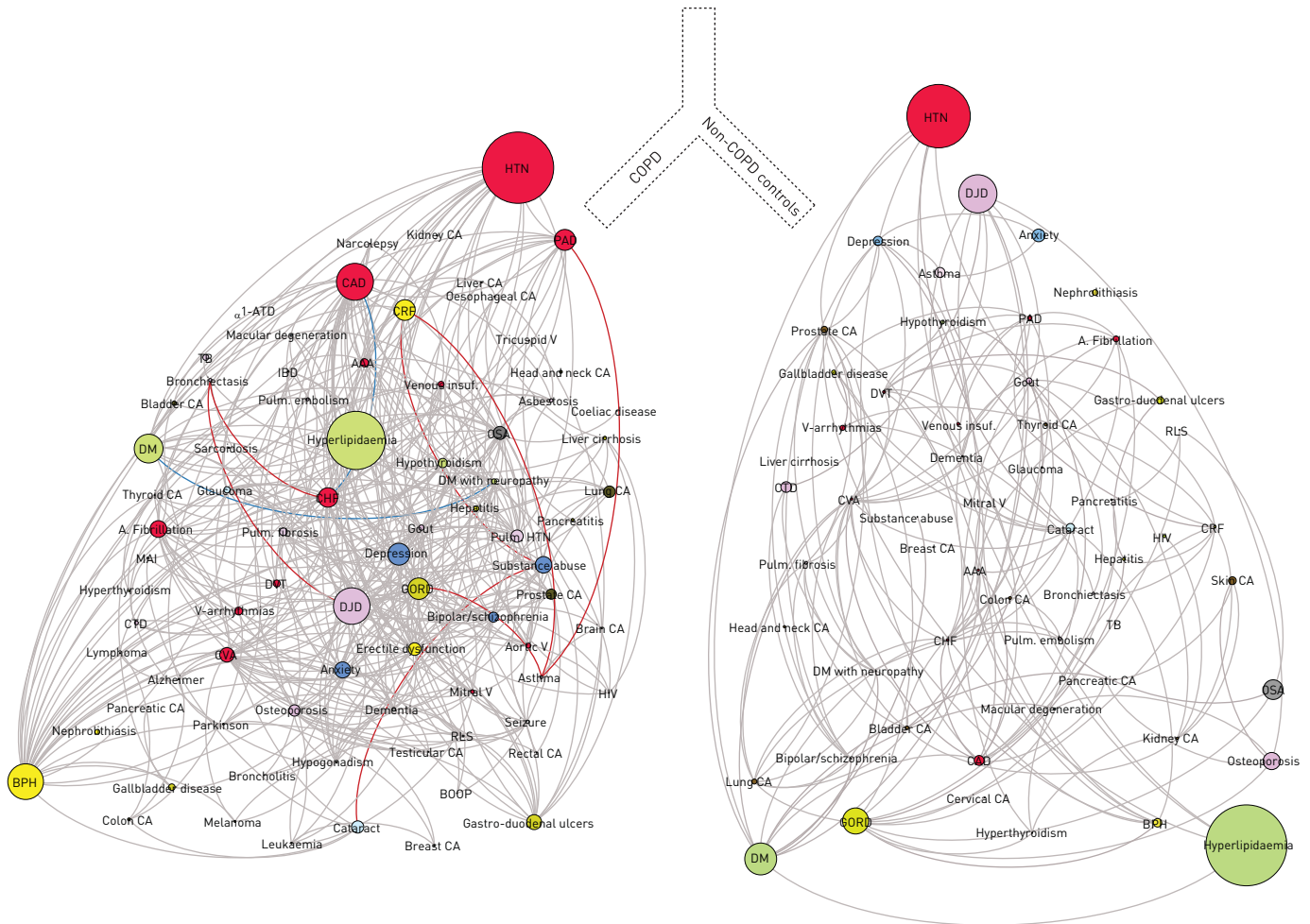


FIGURE 1 The chronic obstructive pulmonary disease (COPD) comorbidities network (left-hand side) is compared with the non-COPD controls (right-hand side) in a lung-shaped layout. The size of the nodes are proportional to the prevalence of the diseases, the links represent a statistically significant correlation ($p \leq 0.01$) between the two diseases and the colour of the bubble, the organ system to which the disease belongs. Clinical variables were excluded from the graph, since we are comparing comorbidity prevalence, linkage and density in the form of a network and some clinical variables are not pertinent to non-COPD controls. A. fibrillation: atrial fibrillation; α_1 -ATD: α_1 -antitrypsin disease; AAA: abdominal aortic aneurysm; BOOP: bronchiolitis obliterans organising pneumonia; BPH: benign prostatic hypertrophy; CA: cancer; CAD: coronary artery disease; CHF: congestive heart failure; CRF: chronic renal failure; CTD: connective tissue disease; CVA: cerebrovascular accident; DJD: degenerative joint disease; DM: diabetes mellitus; DVT: deep venous thrombosis; GORD: gastro-oesophageal reflux disease; HTN: hypertension; IBD: inflammatory bowel disease; MAI: *Mycobacterium avium-intracellulare*; OSA: obstructive sleep apnoea; PAD: peripheral artery disease; Pulm.: pulmonary; RLS: restless legs syndrome; TB: tuberculosis; V: valve; V-arrhythmias: ventricular arrhythmias; Venous insuf.: venous insufficiency.

The median number of links per node was 8 (IQR 3–17) and the distribution (links per node or degree) for the COPD Comorbidity Network is shown in figure E6 (online supplement). The histogram better fit an exponential distribution suggesting features of scale free networks, where one-third ($n=23$) of the comorbidities possess two-thirds of all the edges (figure E6). 23 comorbidities have 56, 69 or 72% of the links depending of the p-values cut-off used for significant association ($p \leq 0.01$, $p \leq 0.001$ and $p \leq 0.0001$ respectively). With more stringent p-values, less prevalent and connected nodes are lost with little changes seen in the highly connected nodes also referred as hubs (figure 2).

Network based clustering: comorbidities modules

The network (diseases and clinical characteristics) contains distinct clusters or modules of highly interlinked nodes and the visual representation reveal a central theme. Modules were detected in both COPD and non-COPD controls and are displayed for comparison in figure 3. We choose to enumerate each module rather than provide a name to avoid taxonomic bias. For the COPD cohort, Module 1A is comprised of 17 nodes connected by 81 edges around the theme of older COPD individuals, cardiovascular comorbidities and clinical characteristics known to confer worse prognosis in patients with COPD [5]. Note that in Module 1B (non-COPD controls), cardiovascular disease also clusters together,

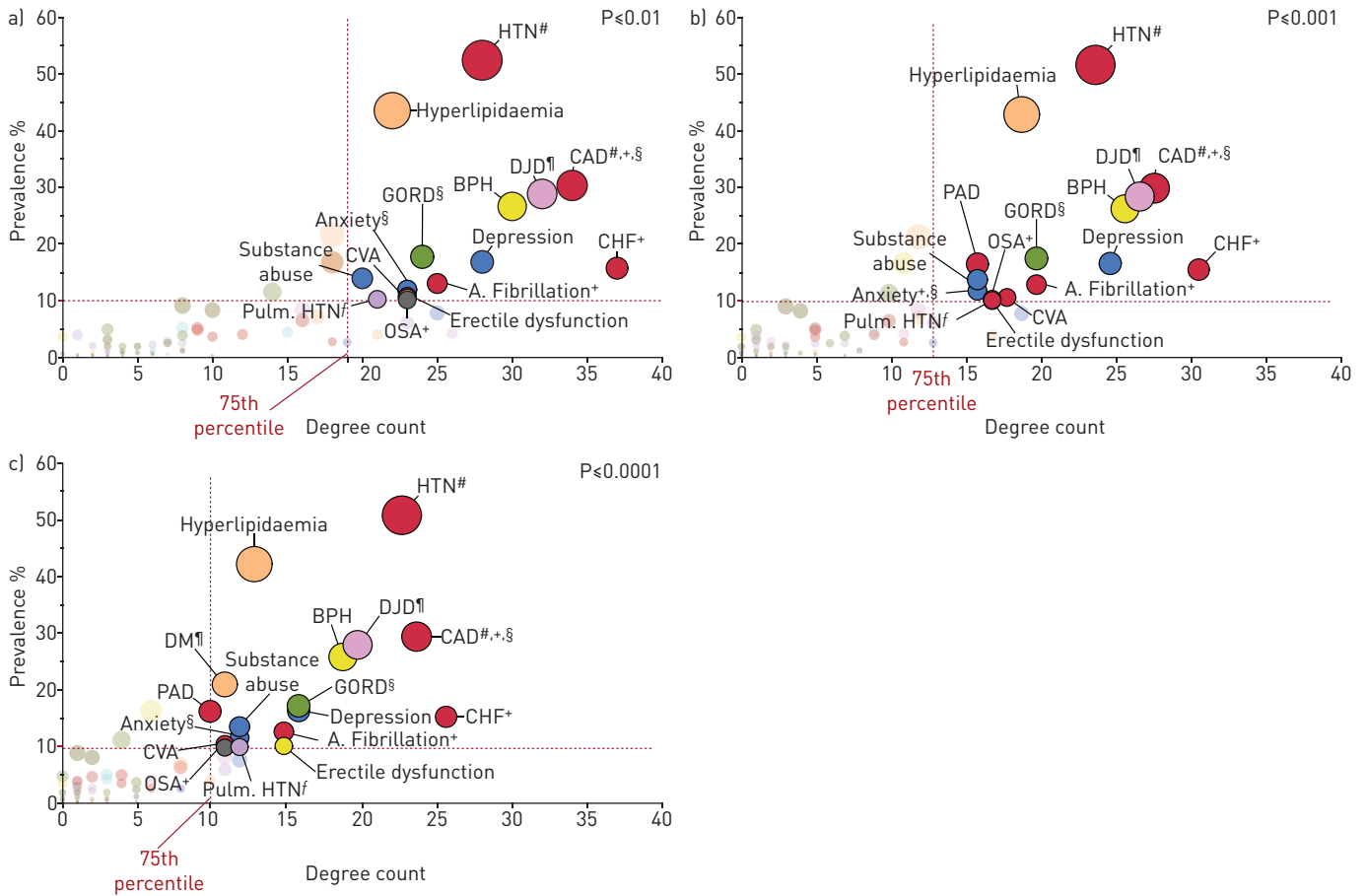


FIGURE 2 Relation between comorbidity prevalence and number of links (degree): Identification of the “hubs”. The graphs demonstrate the relationship between comorbidity prevalence (represented in the y-axis and the size of the bubbles) and degree of connection. 23 comorbidities have 56, 69 or 72% of the links depending on the p-value cut-off used for significant association: a) $p < 0.01$, b) $p < 0.001$ and c) $p < 0.0001$. With more stringent p-values, less prevalent and connected nodes are lost with little changes seen in the highly connected nodes also called hubs. The dotted lines provide the visual cut-off at 10% prevalence (x-axis) and the 75th percentiles of the degree distribution (y-axis). Those bubbles representing comorbidities outside this quadrant are faded to enhance the visualisation of the hubs. The bubble are coloured according to the organ system to which the comorbidity belongs. #: comorbidities described to be associated to lower response to pulmonary rehabilitation [16]; [†]: Comorbidities described to impact in quality of life [14]; [‡]: comorbidities described to be associated with increased mortality [40]; [§]: comorbidities described to be associated with increased acute exacerbations [17, 41]; [∫]: comorbidities described to be associated with increased risk for hospitalisations [42]. A. fibrillation: atrial fibrillation; BPH: benign prostatic hypertrophy; CAD: coronary artery disease; CHF: congestive heart failure; CVA: cerebro-vascular accident; DJD: degenerative joint disease; DM: diabetes mellitus; GORD: gastro-oesophageal reflux disease; HTN: hypertension; OSA: obstructive sleep apnoea; PAD: peripheral artery disease.

however the prevalence and number of links are smaller. Module 2A is composed of 23 nodes connected by 60 links centred around the theme of younger, currently smoking COPD patients with behavioural risk factors and psychiatric conditions manifesting diseases resulting from such risks like hepatitis, liver cirrhosis, pancreatitis and HIV. Similarly in the non-COPD controls, we observe higher prevalence of current smokers in the younger individuals (Module 1B) and the association with anxiety and depression. Interestingly in the control group, this module includes another respiratory disease (asthma). Module 3A is composed of 17 nodes and 42 edges in which COPD individuals with less obstruction and higher BMI whose comorbidities cluster around the components of the metabolic syndrome. In contrast, those components of the metabolic syndrome are present in the non-COPD controls but in the older age group bracket (Module 3B). Module 4A has 11 nodes and 25 links centralised around the theme of gastro-oesophageal reflux disease, gastric and duodenal ulcers, osteoporosis and degenerative joint disease also seen in the non-COPD controls in Module 3B.

Discussion

This study using a novel data visualisation tool to study the integrated relationship among 79 different comorbidities and 10 important clinical variables affecting patients with COPD and the comparison with a cohort of non-COPD controls had three main findings. First, comorbidities are not exclusive of patients with COPD however the prevalence and number of simultaneous comorbidities are higher in COPD.

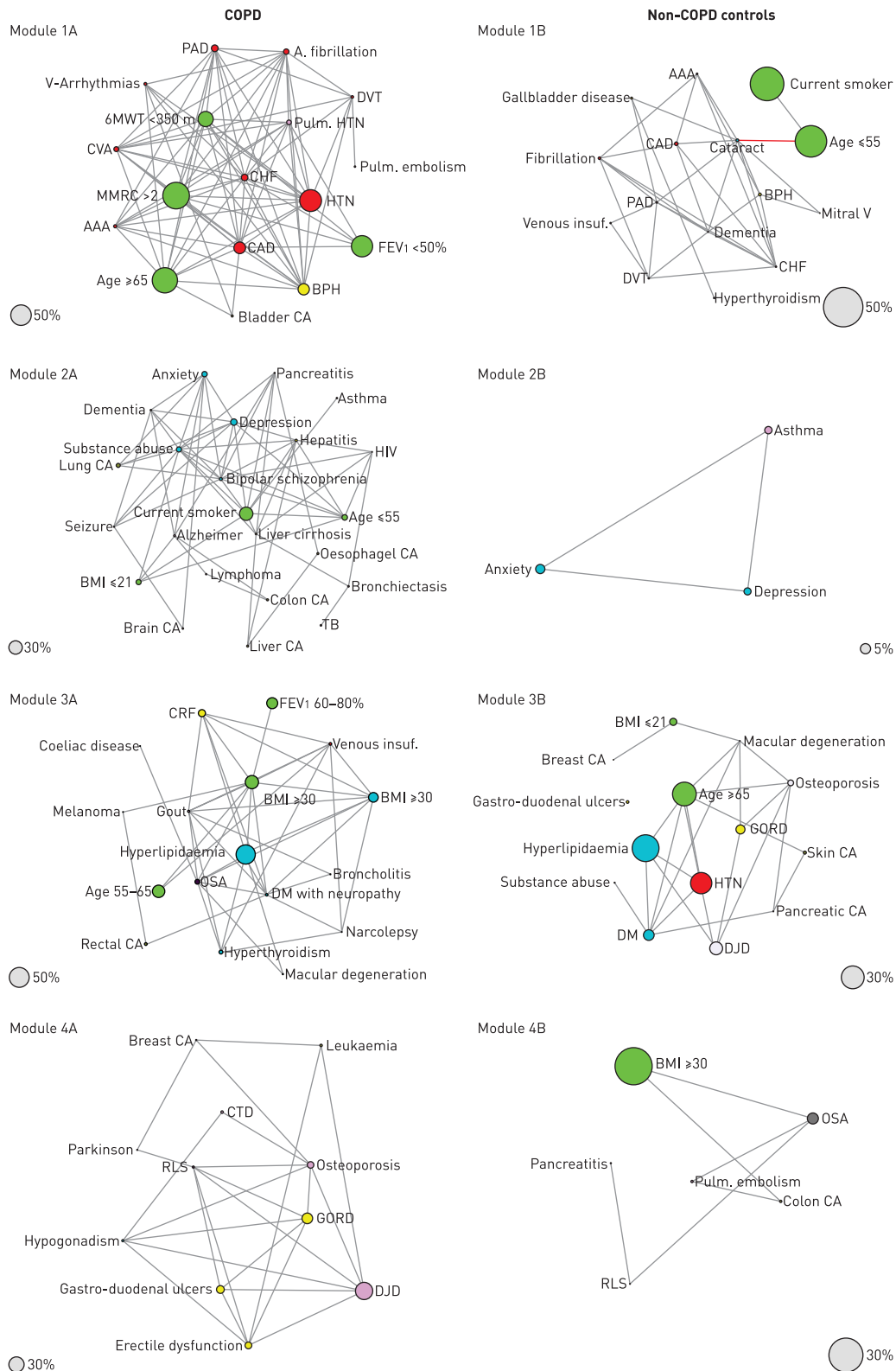


FIGURE 3 Comorbidities modules. Each panel represent a module, the right column contain four chronic obstructive pulmonary disease (COPD) modules, while the left column consists of modules detected in the non-COPD controls. Each panel contains in the right lower corner a reference node with a percentage representing the prevalence scale for each module. Links coloured in grey represent positive correlations, while those coloured in red (see module 1B) represent a negative correlation. We juxtaposed COPD and non-COPD controls modules that have some similarities in order to facilitate comparison. 6MWT: six-min walk test; A. fibrillation: atrial fibrillation; AAA: abdominal aortic aneurism; BMI: body mass index; BPH: benign prostatic hypertrophy; CA: cancer; CAD: coronary artery disease; CHF: congestive heart failure; CRF: chronic renal failure; CVA: cerebro-vascular accident; DVT: deep venous thrombosis; FEV1: forced expiratory volume in 1 s; GORD: gastro-oesophageal reflux disease; HTN: hypertension; MMRC: Modified Medical Research Council dyspnoea scale; OSA: obstructive sleep apnoea; PAD: peripheral artery disease; PAD: peripheral artery disease; Pulm.: pulmonary; RLS: restless legs syndrome; TB: tuberculosis; V: valve; V-arrhythmias: ventricular arrhythmias; Venous insuf.: venous insufficiency.

Second, diseases coexisting are interlinked beyond simple coincidence, a phenomenon difficult to discern if we study comorbidities in isolation. Diseases aggregate modules with meaningful syndromic associations. Third, the COPD Comorbidity Network has characteristics suggesting a scale-free architecture revealing the presence of an influential group of highly connected comorbidities (hubs). The scale-free structure of this network and the stability of hubs are preserved independent of the threshold levels of statistical significance used. Based on this principle, we can propose that those influential comorbidities could be targeted for specific intervention or for screening.

It is known that after adjusting for confounding factors, patients with COPD have a higher prevalence of cardiovascular diseases and lung cancer [8, 10, 11] than patients without COPD. Much less is known about the many other comorbidities that those patients frequently have. In this study, we applied network analysis to explore the associations between the 79 comorbidities detected in the BODE cohort and the 10 clinically relevant variables that characterise COPD multidimensionality and compared the results with a non-COPD control cohort. Using an unbiased approach, we constructed a graphic expression of the complex relationship between prevalence, connections between diseases and clinical variables, the presence of hubs and the clustering of comorbidities. Interestingly, the highly connected comorbidities (figure 2 and table 2) are known to be associated with important patient-related outcomes [13, 14, 17, 19, 20, 42, 43]. If we assume that the COPD Comorbidity Network is similar in behaviour to common networks seen in daily life, like airports composed of important hubs with many incoming and outgoing connecting flights with less influential airports in minor destinations, then disruption of air traffic will be mostly felt when delays affect large hubs [44]. Likewise we could hypothesise that perturbation of the comorbidities network could be achieved by selecting highly connected and treatable comorbidities and measure the impact of specific treatments [27, 45, 46]. This postulation is provocative and untested, however prior evidence indirectly supports this idea. This may be the case of CPAP and its benefit on patients with overlap COPD-OSA, beta receptor-blockers and angiotensin-converting enzyme-inhibitor therapy and their potential impact in reducing mortality in patients with COPD and cardiovascular disease [40, 47, 48]. Our integrative method differs from previously published COPD comorbidities studies by the capacity of visualising a wide range of comorbidities rather than focussing on single or small number of them [14, 17, 20–22, 40, 49–54] and relating them to clinical variables. Those studies have suffered from a reductionist approach and may have

TABLE 2 Top 25th percentile of comorbidities with highest number of links

Comorbidity	Prevalence %	Degrees at p-value cut-off			Association with patient-centred outcome [ref.]
		p<0.01	p<0.001	p<0.0001	
Systemic hypertension	52.4	28	24	23	Response to PR [16]
Hypertlipidaemia	443.5	22	19	13	
Coronary artery disease	30.3	34	28	24	Mortality [13], response to PR [16]
Degenerative joint disease	298.8	32	27	20	QOL [14]
BPH	276.6	30	26	19	
DM	221.6			11	QOL [14]
GORD	187.7	24	20	16	AE [41]
Depression	176.8	28	25	16	AE [17]
Peripheral arterial diseases	176.7		16	10	
Congestive heart failure	165.7	37	31	26	Mortality [13]
Substance abuse	143.9	20	16	12	
Atrial fibrillation/flutter	13	25	20	15	Mortality [13]
Anxiety	121.9	23	16	12	Mortality (females) [13], AE [17]
Cerebrovascular accident	110.7	23	18	11	
Erectile dysfunction	10.4	23	17	15	
Pulmonary hypertension	10.2	21	17	12	Hospitalisation [42]
OSA	10.1	23	17	11	Mortality [40]
Osteoporosis	8.4			11	
Bipolar/schizophrenia	87.8	25	19	12	
Pulmonary fibrosis	6.1	23	13	11	Mortality [13]
Gout	4.2	26			
DM with neuropathy	4	21	17	10	Mortality [13], QOL [14]
Dementia	32.6	19	13	8	

PR: pulmonary rehabilitation; QOL: quality of life; BPH: benign prostatic hypertrophy; DM: diabetes mellitus; GORD: gastro-oesophageal reflux disease; AE: acute exacerbations; OSA: obstructive sleep apnoea.

missed important associations provided by other diseases thought to be “innocent bystanders” that have a more prominent role when analysed in group rather than individually.

This network system analysis expands our current knowledge by showing modules that follow identifiable clinical observations. One such module (figure 3, Module 3A and 3B) shows a cluster of diseases or clinical characteristics related to the metabolic syndrome, interestingly in the non-COPD controls this pattern is seen the older age bracket. Other modules reveal relationship that are less evident but equally logical such as in module 2A (figure 3) with the clustering of behavioural risk factors (substance abuse) and diseases associated with those behaviours like hepatitis, pancreatitis, liver cirrhosis and liver cancer. Moreover, diseases or behavioural risks present in module 2A like substance abuse, depression and anxiety are highly connected (more than 25 links) providing support for the role of behavioural and social ties in a complex diseases such as COPD [55–58]. In addition to its practical application, the visualisation of comorbidities into modules or subnetworks provide clues that are hypothesis generating as they suggest the possibility of shared genetics or pathobiological mechanisms within highly correlated comorbidities [59]. When combining disease modules with clinical characteristics we observed that not all comorbidities affect all COPD individuals similarly as is the case with modules 1A, 2A and 3A (figure 3). Indeed, age or BMI are correlated with particular types of comorbidities, an observation that has implications for clinicians attempting to relate diseases to specific clinical sub-groups [23, 59].

The third important finding of this analysis is that the COPD Comorbidity Network has characteristics of a scale-free architecture revealing the presence of an influential group of highly connected comorbidities (hubs) suggesting an influential or intermediary role. The scale-free structure of this network and the stability of hubs are well preserved even after adjusting from less to more stringent p-values. At lower p-values, less prevalent and connected nodes are lost with little changes seen within hubs. In fact, 23 comorbidities shown in table 2 hold 56, 69 or 72% of the links depending of the p-values cut-off used for significant association and interestingly many of those comorbidities have been shown to influence important outcomes as described in previous studies [13, 14, 17, 19, 20, 42, 43].

Our study has several limitations. First, our cohort is predominantly male. Nevertheless, the study included 219 females which is not a trivial number. This was large enough to perform a sex sub-analysis allowing us to describe those comorbidities affecting predominately and for the first time sex specific co-morbidities (table E3 in the online supplementary material). Second, a common argument on comorbidity ascertainment relates to the methodology used to report them. Our method is a combination of history taking, direct questioning and medical records documentation with an attempt to confirm the diagnosis. This combined methodology has a high sensitivity to capture a large number of comorbidities (79), which will be difficult to reproduce in those cohorts that only include a selected number of comorbidities. Alternatively, we may underestimate the true prevalence by not objectively repeating or seeking confirmatory tests for each disease, a task that is prohibitively expensive. Third, this cohort is composed of ever-smokers COPD individuals (at least 10 pack-years of cumulative exposure to cigarette smoking), and we can't extrapolate our results to the important never-smokers COPD group whose main exposure is indoor pollution from biomass fuel burning. Lastly, our cohort was compared with a younger control group; however, we validated our analysis by performing the same comparison with a randomly selected, matched COPD sub-group and by performing age stratified analysis.

In conclusion, using system network analysis, we show that many comorbidities encountered in patients with COPD are interlinked among themselves and with clinical variables in a pattern that provides information about common risk factors or disease mechanism difficult to discern if studied individually. This data analysis and visualisation approach provides a hypothesis for potential co-management and “module” targeting in an attempt to disrupt common patterns towards more advanced or irreversible diseases.

References

- 1 Divo MJ, Martinez CH, Mannino DM. Number 2 in the series “Multimorbidity and the lung”: Ageing and the epidemiology of multimorbidity. *Eur Respir J* 2014; 44: 1055–1068.
- 2 Barnett K, Mercer SW, Norbury M, *et al.* Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* 2012; 380: 37–43.
- 3 Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int J Epidemiol* 2002; 31: 285–293.
- 4 Wang H, Dwyer-Lindgren L, Lofgren KT, *et al.* Age-specific and sex-specific mortality in 187 countries, 1970–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380: 2071–2094.
- 5 Celli BR, Cote CG, Marin JM, *et al.* The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N Engl J Med* 2004; 350: 1005–1012.
- 6 Puhan MA, Garcia-Aymerich J, Frey M, *et al.* Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. *Lancet* 2009; 374: 704–711.
- 7 Jones RC, Donaldson GC, Chavannes NH, *et al.* Derivation and Validation of a Composite Index of Severity in Chronic Obstructive Pulmonary Disease: The DOSE Index. *Am J Respir Crit Care Med* 2009; 180: 1189–1195.

- 8 Cazzola M, Bettoncelli G, Sessa E, *et al.* Prevalence of comorbidities in patients with chronic obstructive pulmonary disease. *Respiration* 2010; 80: 112–119.
- 9 van Manen JG, Bindels PJ, IJzermans CJ, *et al.* Prevalence of comorbidity in patients with a chronic airway obstruction and controls over the age of 40. *J Clin Epidemiol* 2001; 54: 287–293.
- 10 Schnell K, Weiss CO, Lee T, *et al.* The prevalence of clinically-relevant comorbid conditions in patients with physician-diagnosed COPD: a cross-sectional study using data from NHANES 1999–2008. *BMC Pulm Med* 2012; 12: 26.
- 11 Mapel DW, Hurley JS, Frost FJ, *et al.* Health care utilization in chronic obstructive pulmonary disease. A case-control study in a health maintenance organization. *Arch Intern Med* 2000; 160: 2653–2658.
- 12 McGarvey LP, John M, Anderson JA, *et al.* Ascertainment of cause-specific mortality in COPD: operations of the TORCH Clinical Endpoint Committee. *Thorax* 2007; 62: 411–415.
- 13 Divo MJ, Cote C, de Torres JP, *et al.* Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2012; 186: 155–161.
- 14 Putcha N, Puhan MA, Hansel NN, *et al.* Impact of co-morbidities on self-rated health in self-reported COPD: an analysis of NHANES 2001–2008. *COPD* 2013; 10: 324–332.
- 15 van Manen JG, Bindels PJE, Dekker FW, *et al.* The influence of COPD on health-related quality of life independent of the influence of comorbidity. *J Clin Epidemiol* 2003; 56: 1177–1184.
- 16 Crisafulli E, Costi S, Luppi F, *et al.* Role of comorbidities in a cohort of patients with COPD undergoing pulmonary rehabilitation. *Thorax* 2008; 63: 487–492.
- 17 Xu W, Collet JP, Shapiro S, *et al.* Independent effect of depression and anxiety on chronic obstructive pulmonary disease exacerbations and hospitalizations. *Am J Respir Crit Care Med* 2008; 178: 913–920.
- 18 Jun Su, Baker C, Zou K. Risk assessment of readmissions following an initial COPD-related hospitalization. *COPD* 2013; 3: 551–559.
- 19 Patel A, Donaldson GC. The impact of ischemic heart disease on symptoms, health status, and exacerbations in patients with COPD. *Chest* 2012; 141: 851–857.
- 20 Terada K, Muro S, Sato S, *et al.* Impact of gastro-oesophageal reflux disease symptoms on COPD exacerbation. *Thorax* 2008; 63: 951–955.
- 21 Cote C, Zilberberg MD, Mody SH, *et al.* Haemoglobin level and its clinical impact in a cohort of patients with COPD. *Eur Respir J* 2007; 29: 923–929.
- 22 Vanfleteren LEGW, Spruit MA, Groenen M, *et al.* Clusters of comorbidities based on validated objective measurements and systemic inflammation in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2013; 187: 728–735.
- 23 Divo MJ, Cabrera C, Casanova C, *et al.* Comorbidity distribution, clinical expression and survival in COPD patients with different body mass index. *J Copd F* 2014; 1: 229–238.
- 24 Burgel PR, Paillasseur JL, Caillaud D, *et al.* Clinical COPD phenotypes: a novel approach using principal component and cluster analyses. *Eur Respir J* 2010; 36: 531–539.
- 25 Garcia-Aymerich J, Gómez FP, Benet M, *et al.* Identification and prospective validation of clinically relevant (COPD) subtypes. *Thorax* 2011; 66: 430–437.
- 26 Hidalgo CA, Blumm N, Barabási A-L, *et al.* A dynamic network approach for the study of human phenotypes. Meyers LA, editor. *PLoS Comput Biol* 2009; 5: e1000353.
- 27 Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* 2011; 12: 56–68.
- 28 Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. *N Engl J Med* 2008; 358: 2249–2258.
- 29 Bliss CA, Kloumann IM, Harris KD, *et al.* Twitter reciprocal reply networks exhibit assortativity with respect to happiness. *J Comput Sci* 2012; 3: 388–397.
- 30 Celli BR, MacNee W, ATS/ERS Task Force. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J* 2004; 932–946.
- 31 Charlson ME, Pompei P, Ales KL, *et al.* A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; 40: 373–383.
- 32 Sanchez-Salcedo P, Divo MJ, Casanova C, *et al.* Disease progression in young patients with COPD: rethinking the Fletcher and Peto model. *Eur Respir J* 2014; 44: 324–331.
- 33 Schols AM, Slangen J, Volovics L, *et al.* Weight loss is a reversible factor in the prognosis of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998; 157: 1791–1797.
- 34 Jones PW, Adamek L, Nadeau G, *et al.* Comparisons of health status scores with MRC grades in COPD: implications for the GOLD 2011 classification. *Eur Respir J* 2013; 42: 647–654.
- 35 Vestbo J, Hurd SS, Agustí AG, *et al.* Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2013; 187: 347–365.
- 36 Pinto-Plata VM, Cote C, Cabral H, *et al.* The 6-min walk distance: change over time and value as a predictor of survival in severe COPD. *Eur Respir J* 2004; 23: 28–33.
- 37 Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring and Manipulating Networks. International AAAI Conference on Weblogs and Social Media, North America, mar. 2009. www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154 Date last accessed: January 27, 2015.
- 38 Blondel VD, Guillaume J-L, Lambiotte R, *et al.* Fast unfolding of communities in large networks. *J Stat Mech* 2008: P10008.
- 39 Newman MEJ. Modularity and community structure in networks. *Proc Natl Acad Sci U S A* 2006; 103: 8577–8582.
- 40 Marin JM, Soriano JB, Carrizo SJ, *et al.* Outcomes in patients with chronic obstructive pulmonary disease and obstructive sleep apnea: The Overlap Syndrome. *Am J Respir Crit Care Med* 2010; 182: 325–331.
- 41 Hurst JR, Vestbo J, Anzueto A, *et al.* Susceptibility to exacerbation in chronic obstructive pulmonary disease. *N Engl J Med* 2010; 363: 1128–1138.
- 42 Terzano C, Conti V, Di Stefano F, *et al.* Comorbidity, hospitalization, and mortality in COPD: results from a longitudinal study. *Lung* 2010; 188: 321–329.
- 43 Crisafulli E, Gorgone P, Vagaggini B, *et al.* Efficacy of standard rehabilitation in COPD outpatients with comorbidities. *Eur Respir J* 2010; 36: 1042–1048.

- 44 Agusti A, Sobradillo P, Celli B. Addressing the Complexity of Chronic Obstructive Pulmonary Disease: From Phenotypes and Biomarkers to Scale-Free Networks, Systems Biology, and P4 Medicine. *Am J Respir Crit Care Med* 2011; 183: 1129–1137.
- 45 Diez D, Agustí A, Wheelock CE. Network analysis in the investigation of chronic respiratory diseases: from basics to application. *Am J Respir Crit Care Med* 2014; 190: 981–988.
- 46 Grosdidier S, Ferrer A, Faner R, *et al.* Network medicine analysis of COPD multimorbidities. *Respir Res* 2014; 15: 111.
- 47 Etminan M, Jafari S, Carleton B, *et al.* Beta-blocker use and COPD mortality: a systematic review and meta-analysis. *BMC Pulm Med* 2012; 12: 48.
- 48 Mancini GBJ, Etminan M, Bin Z, *et al.* Reduction of morbidity and mortality by statins, angiotensin-converting enzyme inhibitors, and angiotensin receptor blockers in patients with chronic obstructive pulmonary disease. *J Am Coll Cardiol* 2006; 47: 2554–2560.
- 49 Mannino DM, Thorn D, Swensen A, *et al.* Prevalence and outcomes of diabetes, hypertension and cardiovascular disease in COPD. *Eur Respir J* 2008; 32: 962–969.
- 50 Soriano JB, Visick GT, Muellerova H, *et al.* Patterns of comorbidities in newly diagnosed COPD and asthma in primary care. *Chest* 2005; 128: 2099–2107.
- 51 de Torres JP, Marin JM, Casanova C, *et al.* Lung cancer in patients with chronic obstructive pulmonary disease--incidence and predicting factors. *Am J Respir Crit Care Med* 2011; 184: 913–919.
- 52 van Oostrom SH, Picavet HSJ, van Gelder BM, *et al.* Multimorbidity and comorbidity in the Dutch population – data from general practices. *BMC Public Health* 2012; 12: 715.
- 53 van der Molen T. Co-morbidities of COPD in primary care: frequency, relation to COPD, and treatment consequences. *Prim Care Respir J* 2010; 19: 326.
- 54 Frei A, Muggensturm P, Putcha N, *et al.* Five comorbidities reflected the health status in patients with chronic obstructive pulmonary disease: the newly developed COMCOLD index. *J Clin Epidemiol* 2014; 67: 904–911.
- 55 Barabási A-L. Network medicine--from obesity to the “diseasome”. *N Engl J Med* 2007; 357: 404–407.
- 56 Hall SM, Prochaska JJ. Treatment of smokers with co-occurring disorders: emphasis on integration in mental health and addiction treatment settings. *Annu Rev Clin Psychol* 2009; 5: 409–431.
- 57 Prochaska JJ. Smoking and mental illness — breaking the link. *N Engl J Med* 2011; 365: 196–198.
- 58 Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *N Engl J Med* 2007; 357: 370–379.
- 59 Menche J, Sharma A, Cho MH, *et al.* A diVIasive Shuffling Approach (VIStA) for gene expression analysis to identify subtypes in Chronic Obstructive Pulmonary Disease. *BMC Syst Biol* 2014; 8: Suppl. 2, S8.