Word count main body: 3,367 (Abstract 233)

‘HIGH RISK’ CLINICAL AND INFLAMMATORY CLUSTERS IN COPD OF CHINESE DESCENT

Pei Yee Tiew, MD[1,2], Fanny Wai San Ko, MD[3], Jayanth Kumar Narayana, BS-MS[1,4], Mau Ern Poh, MD[5], Huiying Xu, MD[6], Han Yee Neo, MD[6], Li-Cher Loh, MD[7], Choo-Khoo Ong, MD[7], Micheál Mac Aogáin, PhD[1], Jessica Han Ying Tan, MD[8], Nabilah Husna Kamaruddin[5], Gerald Jiong Hui Sim[9], Therese S. Lapperre, MD, PhD[2,10], Mariko Siyue Koh, MD[2], David Shu Cheong Hui, MD[3], John Arputhan Abisheganaden, MD[6], Augustine Tee, MD[9], Krasimira Tsaneva-Atanasova, PhD[11,12], Sanjay H. Chotirmall, MD, PhD[1#].

[1]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore
[2]Department of Respiratory and Critical Care Medicine, Singapore General Hospital, Singapore
[3]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong
[4]Indian Institute of Science Education and Research, Pune, India
[5]Department of Medicine, University of Malaya, Kuala Lumpur, Malaysia
[6]Department of Respiratory and Critical Care Medicine, Tan Tock Seng Hospital, Singapore
[7]Department of Medicine, RCSI-UCD Malaysia Campus, Georgetown, Penang, Malaysia
[8]Department of General Medicine, Sengkang General Hospital, Singapore
[9]Department of Respiratory and Critical Care Medicine, Changi General Hospital, Singapore
[10]Department of Respiratory Medicine, Bispebjerg University Hospital, Copenhagen, Denmark
[11]Living Systems Institute and Department of Mathematics, College of Engineering, Mathematics and Physical Sciences, University of Exeter
[12]PSRC Centre for Predictive Modelling in Healthcare, University of Exeter

Corresponding author: Sanjay H. Chotirmall, Lee Kong Chian School of Medicine, Nanyang Technological University, 11 Mandalay Road, Singapore 308232. Email: schotirmall@ntu.edu.sg

Running head:

COPD in Chinese patients exhibits clinical-inflammatory clustering where 'cardiovascular' and 'ex-tuberculosis' groups illustrate highest mortality. Risk stratification of Chinese patients with COPD is necessary for targeted intervention.

Conflict of interest: All authors have no conflicts of interest to declare.

Funding: This research is supported by the Singapore Ministry of Health's National Medical Research Council under its Research Training Fellowship (NMRC/Fellowship/0049/2017) (P.Y.T) and a Clinician-Scientist Individual Research Grant (MOH-000141) (S.H.C); the Singapore Ministry of Education under its Singapore Ministry of Education Academic Research Fund Tier 1 (2016-T1-001-050) (S.H.C); the NTU Integrated Medical, Biological and Environmental Life Sciences (NIMBELS), Nanyang Technological University, Singapore [NIM/03/2018] (S.H.C) and the Ageing Research Institute for Society and Education (ARISE), Nanyang Technological University, Singapore [ARISE/2017/6] (S.H.C). KTA gratefully acknowledges the financial support of the EPSRC via grant EP/N014391/1.

Parts of this article have been presented in the form of an abstract at the European Respiratory Society Annual Congress 2019, Madrid, Spain.

Abbreviation List

BMI (body mass index), COPD (chronic obstructive pulmonary disease), CVS (cardiovascular), FEV_1 (forced expiratory volume in the first second), GOLD (global initiative for chronic obstructive lung disease), Hazard ratio (HR), IQR (interquartile range), LCHR (low-comorbidity high risk), LCLR (low comorbidity low risk), PDGF (platelet derived growth factor), Regularised Discriminant Analysis (RDA), SD (Standard deviation), TB (tuberculosis), TNF (tumor necrosis factor), VEGF (vascular endothelial growth factor).

Abstract

Introduction: COPD is a heterogeneous disease demonstrating inter-individual variation. A high COPD prevalence in Chinese populations is described but little is known about disease clusters and prognostic outcomes in the Chinese population across South-East Asia. We aim to determine if clusters of Chinese patients with COPD exist and their association with systemic inflammation and clinical outcomes. **Methods:** Chinese patients with stable COPD were prospectively recruited into two cohorts (derivation and validation) from six hospitals across three South-East Asian countries (Singapore, Malaysia and Hong Kong; n=1,480). Each patient was followed over two-years. Clinical data (including co-morbidities) were employed in unsupervised hierarchical clustering (followed by validation) to determine the existence of patient clusters and their prognostic outcome. Accompanying systemic cytokine assessments were performed in a subset (n=336) of COPD patients to determine if inflammatory patterns and associated networks characterised the derived clusters. **Results:** Five patient clusters were identified including (1) Ex-tuberculosis (2) Diabetic (3) Low co-morbidity: low-risk (4) Low co-morbidity: high-risk and (5) cardiovascular. The 'cardiovascular' and 'ex-tuberculosis' clusters demonstrate highest mortality (independent of GOLD assessment) and illustrate diverse cytokine patterns with complex inflammatory networks. **Conclusions:** We describe novel 'clusters' of Chinese COPD patients, two of which represent 'high-risk' clusters. The 'cardiovascular' and 'ex-tuberculosis' patient clusters exhibit high mortality, significant inflammation and complex cytokine networks. Clinical and inflammatory risk stratification of Chinese patients with COPD should be considered for targeted intervention to improve disease outcomes.

Keywords: COPD, Chinese, Cardiovascular, Tuberculosis, Mortality

Introduction

Chronic obstructive pulmonary disease (COPD) in the Asian sub-continent is under-estimated owing to poor disease awareness and significantly delayed diagnosis [1-2]. COPD burden in Asia is expected to further increase, driven by ageing populations, increased tobacco consumption and rapid urbanization [3-5].

While an inimitable group of COPD-associated risk factors exists in Asia, a diverse range of geographic environments, climates, cultural practices, healthcare policies and resource availability influences diagnosis and management [6]. Taken together, this contributes to clinically different COPD phenotypes in the region, for instance, higher males, non-smokers and less symptoms compared to western cohorts [7-8]. Prior work focused on racial differences includes few Asian-based patients and is predominantly conducted outside Asia [9-11].

More recently, work on ethnic Chinese populations reveals the prevalence and scale of COPD [12]. The China Pulmonary Health study evaluated a nationally representative sample of adults across mainland China. Of the 57,779 individuals studied, COPD prevalence was 8.6%, accounting for almost 100 million individuals, a clear public health priority [12]. A second study involving 67,752 Chinese adults confirms these findings and estimates COPD burden at 13.6% [1]. Because of large COPD numbers, it is plausible that disease 'sub-groups' with differing prognostic outcomes exist necessitating individualized intervention. Significant numbers of ethnically Chinese individuals reside in countries across the Asian sub-continent including Malaysia, Singapore and Hong Kong and little published data has assessed COPD in these populations. The limited data does however suggest high mortality in Chinese COPD patients [13].

In view of high COPD burden and mortality in Chinese patients, a need to better understand potential disease clusters (or sub-groups) exists to allow improved risk stratification and

targeted intervention. Here, in a multi-centre study across three countries in South-East Asia, we evaluate a large group of Chinese patients with COPD and describe clusters with prognostic and inflammatory relevance.

Methods

COPD patient recruitment: Patients of Chinese ethnicity (defined as an individual where both parents were of Chinese lineage) aged ≥40 with stable COPD as their predominant diagnosis were prospectively recruited over a seven-year period from January 2012 – December 2018 when attending respiratory outpatient clinics at six tertiary hospital sites across three countries: (1) Singapore (3 sites: Singapore General Hospital, Changi General Hospital and Tan Tock Seng Hospital); (2) Malaysia (2 sites: RCSI-UCD Malaysia Campus and University Malaya Hospital) and (3) Hong Kong (1 site: Prince of Wales Hospital). COPD was defined according to the global initiative for chronic obstructive lung disease (GOLD) criteria [14]. Bronchiectasis was excluded by chest radiography in the absence of tram tracking, ring opacities and tubular structures[15-16]. Disease stability was defined as the absence of exacerbation over the preceding six weeks prior to study recruitment and all patients were receiving COPD therapy (including smoking cessation counseling, inhaler assessment, COPD action plans, inhalers as long acting β-agonists, long acting muscarinic antagonists, inhaled corticosteroids and/or short acting bronchodilators in addition to vaccination as appropriate) based on GOLD guidelines[14].

Derivation and Validation COPD cohorts: A complete cohort of n=1,480 COPD patients of Chinese ethnicity were recruited into the study in two separate arms: a derivation and validation cohort. A total of n=911 patients made up the derivation cohort and were recruited from four different sites over the study period from 2012: Singapore General Hospital and Changi

General Hospital (Singapore), RCSI-UCD Malaysia Campus (Malaysia) and Prince of Wales Hospital (Hong Kong). An independent and unrelated validation cohort of n=569 patients were recruited from five different sites from 2013 onward: Singapore General Hospital, Tan Tock Seng Hospital, Changi General Hospital (Singapore), University Malaya Hospital (Malaysia) and Prince of Wales Hospital (Hong Kong). Clinical data was obtained for all subjects and the institutional review board of all participating hospitals approved the study.

Clustering analysis: Clinical variables and comorbidities were pre-processed with non-metric multidimensional scaling followed by hierarchical clustering. A trained Regularised Discriminant Analysis model was used to assign cluster membership for the validation cohort. Each derived cluster was defined based on the predominant (or lack of) clinical features, comorbidities and outcome (mortality).

Full details on ethical approvals, patient recruitment, clinical data collation, specimen collection and processing, inflammatory assessments and statistics are provided in the e-Appendix 1-3.

Results

Independent unrelated derivation (n=911) and validation (n=569) cohorts of Chinese patients with COPD were recruited. Demographic profiles between cohorts were comparable and similar proportions of patients (from each country) formed the final cohorts (e-Table 1). The majority of patients (in both cohorts) were male, and, while more current smokers and ex-

smokers were identified in the derivation and validation cohorts respectively ($p<0.001$), no significant difference in overall smoking pack years between cohorts was observed (e-Table 1). Greater proportions of prior pulmonary tuberculosis (18.2% versus 8.4%; $p<0.001$), osteoporosis (3.7% versus 1.4%; $p<0.01$) and asthma (7.1% versus 1.9%; $p<0.001$) were identified in the derivation while more malignancy (8.0% versus 17.2%; $p<0.001$) detected in the validation cohort (e-Figure 1a). Prior pulmonary tuberculosis was highest in the Malaysian derivation and lowest in the Hong Kong validation cohort ($p<0.001$) reflective of contrasting TB prevalence (e-Figure 1b). Coronary artery disease was highest in both cohorts from Singapore while peptic ulcer disease was greatest in the Malaysian derivation and lowest in their validation cohort. Airway microbiology demonstrates more *S. pneumoniae* and *H. influenzae* in the derivation cohort (e-Figure 1c).

Unsupervised hierarchical clustering of the derivation (n=911) cohort revealed five clusters of Chinese COPD each defined by their predominant (or lack of) clinical features: (1) prior pulmonary tuberculosis (Ex-TB; n=156); (2) co-existing diabetes (diabetic; n=109); (3) low co-morbidity: low risk (LCLR; n=192); (4) low-comorbidity: high risk (LCHR; n=339) and (5) co-existing cardiovascular disease (CVS; n=115) (Figure 1 and e-Table 2). While two clusters demonstrate low co-morbidity, the LCHR and CVS clusters closely resembled one another, separated only by dendrogram branching (Figure 1). Prognostic outcome between clusters varied based on two-year all cause and respiratory-related mortality (Figures 2a-c). Two-year mortality was highest in the CVS (42.6%), Ex-TB (34.0%) and LCHR (25.7%) clusters (log rank test, $p<0.05$) (Figure 2a), which remained significant after adjustment for age, sex, BMI, FEV_1 and smoking pack year exposure (Figure 2b). Hazard ratio for death in each cluster (compared to LCLR) was as follows: CVS (HR: 2.94; 95% CI 1.51-5.72; $p<0.01$), Ex-TB (HR: 2.10; 95% CI 1.16-3.80; $p<0.05$), LCHR (HR: 2.01; 95% CI 1.18-3.43; $p=0.01$) and diabetic

(HR: 1.67; 95% CI 0.82-3.40; p=ns). When only respiratory-related causes of death are considered, a similar pattern (to that with all-cause mortality) is observed (Figure 2c). Smoking status had no influence on mortality (e-Figure 2a) or exacerbation severity (e-Figure 2b) within each cluster however did differ between the low co-morbidity clusters with predominantly current smokers in the LCHR and ex-smokers in the LCLR clusters (e-Table 2).

Having identified five clinical clusters demonstrating different prognostic outcomes, we next validated these findings in an independently recruited validation cohort (n=569). All patients in the validation cohort were assigned to a cluster using Regularised Discriminant Analysis (RDA) with a high Leave One Out Cross Validation accuracy (97.8%) illustrating robustness of both our model and the previously derived clusters (e-Table 3 and e-Figure 3). The validation cohort was assigned as follows: Ex-TB (n=102; 17.9%), diabetic (n=72; 12.7%), LCLR (n=88; 15.5%), LCHR (n=193; 33.9%) and CVS (n=114; 20.0%). The mean RDA assigned probability of a patient from the validation cohort belonging to each of the derived clusters is as follows: Ex-TB (mean 93% ± SD 13%), diabetic (mean 96% ± SD 8%), CVS (mean 99%± SD 6%), LCHR (mean 88%± SD 17%), and LCLR (mean 69%± SD 13%).

Overall proportion of patients in each cluster was comparable to the derivation cohort despite a lower overall prevalence of co-morbidities (including TB) (e-Tables 2 and e-Table 4). Baseline co-morbidities (Figure 3) and two-year all-cause mortality followed the derivation cohort with poorest survival observed in the CVS (43.0%) and Ex-TB (26.0%) clusters (log rank test, p=0.001) (Figure 4a) which remained significant after adjustment for age, sex, BMI, FEV_1 and smoking pack year exposure (Figure 4b). Hazard ratio for death in each cluster (compared to LCLR) was as follows: CVS (HR: 3.08; 95% CI 1.74-5.44; p<0.0001), Ex-TB (HR: 2.01; 95% CI 1.07-3.80; p<0.05), LCHR (HR: 1.59; 95% CI 0.89-2.85; p=ns) and diabetic

(HR: 1.73; 95% CI 0.86-3.45; p=ns). Where respiratory-related causes of death are considered, the CVS, Ex-TB and LCHR clusters again illustrate the highest risk (Figure 4c). Smoking status, as in the derivation cohort, did not influence mortality (e-Figure 2c) or exacerbation severity (e-Figure 2d) in any cluster although the predominance of current smokers in the LCHR and ex-smokers in the LCLR clusters was reproduced (e-Table 4). Finally, to further verify the accuracy obtained in clustering the validation cohort using RDA Leave One Out Cross Validation, we generated a decision tree to classify Chinese individuals with COPD into one of our proposed five clusters (e-figure 4). This alternate classification methodology produced accuracy results of 72.4%, comparable to the RDA classification accuracies reported above.

Following an unbiased semi-quantitative cytokine-array screen (evaluating 120 cytokines; data not shown), six cytokines (TNF-R1, TNF-R2, VEGF, PDGF-AA, PDGF-BB, PDGF-AB) were selected for confirmatory validation between clusters and compared to a group of non-COPD (healthy) controls (n=24) (e-Table 5). Independent of cluster, elevated TNF-R2 significantly associates with symptoms and severe exacerbations (e-Figure 5). Individual cytokines relating to each cluster were as follows: TNF-R2 and PDGF-AA in the CVS cluster; VEGF in the ex-TB cluster and PDGF-AB and –BB in the LCHR cluster (Figure 5a; e-Table 6). Given observed differences for individual cytokines between clusters, we next assessed how cytokines 'interact' within an inflammatory network and their respective complexity (Figure 5b). The CVS cluster (highest mortality) exhibits the most complex network with the highest number of positive cytokines and cytokine interactions. In line with the observed mortality in the derivation and validation cohorts, the ex-TB cluster followed by the LCHR, diabetic and LCLR respectively demonstrate gradients of decreasing cytokine network complexity (Figure 5b).

We next assessed our 'clusters' in comparison to GOLD ABCD group and conventional GOLD staging. All patients were assigned to their respective GOLD grouping: A (17.0%; n=252), B (29.1%; n=430), C (12.9%; n=191) and D (41.0%; n=607) at study enrolment. All patients were also classified by conventional GOLD grade (FEV_1 criteria) as follows: I (6.4%; n=95), II (34.5%; n=510), III (43.9%; n=650) and IV (15.2%; n=225) [14]. Mortality at two-year follow up was assessed by cluster membership within each GOLD group (Figure 6) and GOLD grade (Figure 7). Across all GOLD groups and grades, the CVS and ex-TB clusters demonstrate highest mortality (Figure 6a, Figure 7). On univariate analysis, stratifying patients into GOLD groups, CVS and ex-TB clusters demonstrate highest mortality in GOLD A, C, and D, while CVS and LCHR in GOLD B (Figure 6a). CVS and ex-TB clusters demonstrate significantly higher mortality in multivariate logistic regression after adjustment for age, sex, BMI, FEV_1, smoking pack year exposure, and GOLD group compared to LCLR cluster (Figure 6b). Adjusted odds ratio (ORs) for mortality in the CVS and ex-TB clusters (compared to LCLR group) were 2.98 (95% CI 1.88-4.73; $p<0.001$) and 1.852 (95% CI 1.17-2.92; $p<0.01$) respectively (Figure 6b). Similar trends were seen with adjusted hazard ratios (e-figure 6). FEV_1 was not significant in independently predicting mortality in our logistic regression model. When the clusters were stratified by GOLD grade, significantly poorer two-year survival was observed in the CVS and ex-TB clusters irrespective of underlying (conventional) GOLD grade (Figure 7; $p<0.05$). Taken together, these data suggest that the CVS and ex-TB clusters perform poorly despite classification as 'low risk' by conventional GOLD group and staging necessitating early identification in Chinese populations.

Discussion

In this multi-center study across South-East Asia, we evaluated 1,480 Chinese patients with COPD and describe five validated 'clusters' with prognostic relevance. The two 'highest-risk'

clusters were CVS and ex-TB which demonstrate high mortality risk. Our 'cluster' classification demonstrates differences in mortality outcome and associates with inflammatory signatures and cytokine network complexity.

Cardiovascular disease and diabetes are well-recognized COPD co-morbidities and therefore identification of these 'clusters' was foreseen in view of existing evidence [17-22]. In the Asian sub-continent however, rapid urbanization with improved socio-economic status has resulted in higher risks of cardiovascular consequences in COPD, with poorer prognosis, a finding consistent in our study [23-24]. The highest mortality risk in our CVS 'cluster' warrants attention as prior Asian data suggests the under-treatment of COPD with co-existing cardiovascular disease [25]. Several studies report increased mortality with concomitant diabetes in COPD [20, 26]. Interestingly, however, in our work, the presence of diabetes illustrated better prognosis compared to the 'high-risk' CVS and ex-TB 'clusters' with some reports suggesting ethnic differences in diabetes-related mortality in South Asians and in particular Chinese [27]. Two low co-morbidity clusters were identified that differed in mortality outcomes. The LCHR cluster contained high proportions of current smokers in contrast to the LCLR cluster with predominance of ex-smokers illustrating the benefits of smoking cessation to the natural course of COPD and its outcomes [28-29].

Unlike some of the identified 'clusters', the 'high-risk' ex-TB cluster is novel and likely unique to Asians and other regions where TB is endemic. Patients in this 'cluster' completed treatment with clinical and microbiological resolution, however, recognized long-term sequelae of TB such as the high-risk of subsequent pulmonary obstruction persist [30-31]. Post-TB related airways disease is commonly recognized in Asia however precise mechanisms remain unclear [31-33]. Structural lung damage as a consequence of TB with pulmonary cavitation, bronchiectasis and

endobronchial disease associates with increased risks of airflow obstruction, and, even with minimal change on chest radiography, risks of obstruction persist suggesting alternate mechanisms contributing to COPD development [34-35]. One possibility is that pulmonary TB 'primes' lung host defenses dysregulating responses to inhaled toxins, pathogens and cigarette smoke and increasing susceptibility to chronic lung disease despite infection clearance [35-36]. Once chronic lung disease such as COPD has developed post-TB, host genetics and dysregulated immunity may play further roles in determining disease trajectory, progression and outcome illustrating why many patients with post-TB related chronic lung disease demonstrate aggressive clinical phenotypes with poorer outcome [35]. A prior history of TB associates with morbidity and mortality in COPD, findings consistent with our work [37]. Our clinical and inflammatory analyses further demonstrate complex networks and interactions in the 'ex-TB' cluster highlighting the possibility of ongoing systemic inflammation post-infection (some of which may be attributed to higher exacerbation rates) contributing to their poorer outcome. Patient ethnicity influences TB-induced inflammation and therefore is of relevance in Chinese COPD[38-39].

Low-grade systemic inflammation is reported in COPD, however only a defined group of patients exhibit a persistent systemic inflammatory state with high mortality [40]. Additionally, COPD can demonstrate Th-2 inflammatory responses with coexisting airway diseases such as asthma in asthma-COPD overlap syndrome (ACOS)[41]. Systemic inflammation in COPD is complex and involves interactions between multiple cytokines and their pathways [40]. Using network analyses, we illustrate cytokine interactions occurring within each cluster. Interestingly, network complexity correlates with cluster mortality. Systemic inflammation is established in cardiovascular disease and when co-existing with COPD likely contributes to even greater levels in this cluster, where tumor necrosis factor (TNF) is identified. TNF

promotes tissue inflammation, injury and reactive oxygen species, which in turn drives poorer prognostic outcomes. Vascular endothelial growth factor (VEGF), an angiogenesis promoter is described to have paradoxical roles in COPD: decreased in emphysematous and increased in bronchitis phenotypes. Interestingly, we detected significant VEGF levels in our 'ex-TB' cluster. An established body of evidence indicates that VEGF plays an essential role in TB pathogenesis enhancing granulomatous inflammation and its associated angiogenesis [42-43]. VEGF increases in active TB, correlating with disease severity and while its levels are thought to improve post-treatment, its precise role in 'ex-TB' COPD is unclear [44-45]. The role of platelets and their functional consequence in COPD is of interest [46-47]. Platelet derived growth factor (PDGF)-related cytokines associate with the CVS and LCHR clusters and, have roles in cell signaling, lung development and cardiopulmonary disease [48]. As a potent stimulant of smooth muscle proliferation, PDGF associates with small airway remodeling [49-50] and, while its role in COPD is uncertain, our findings suggest it to be an important systemic marker (perhaps associated with active smoking) for consideration in future studies in Chinese COPD.

Our derived clusters demonstrate differences in mortality outcome despite conventional stratification by GOLD approaches used for COPD grouping and staging. This is considering our 'highest-risk' clusters: 'CVS' and 'ex-TB' which demonstrate high mortality and inflammatory complexity unrelated to underlying GOLD group or grade. Importantly, this indicates that some Chinese COPD patients within each GOLD group and grade (including that defined as low risk) do poorly and require attention. Furthermore, our defined clusters did not differ based on clinical COPD features alone (symptoms and lung function), with the exception of 'ex-TB' group with consistently low BMIs. A comprehensive tool incorporating cardiovascular and tuberculosis assessment may be of value to improve risk stratification in Chinese COPD populations.

Here, we describe five clusters of Chinese COPD, at least three of which (cardiovascular, diabetic/metabolic and low-comorbidity) are broadly consistent with comparable studies in non-Chinese COPD populations [17-21]. This work, unlike prior studies however did not identify any one cluster enriched for anxiety which overall demonstrates low prevalence in our dataset, potentially explained by employed diagnostic criteria. Similarly, we did not identify an underweight cluster, however, did observe lower BMIs in the 'ex-TB' group, a novel cluster with high mortality and relevant in any country with high TB prevalence. Understanding pathophysiological mechanisms in this cluster is an important avenue for future studies.

Our work demonstrates clear strengths: it is the largest multi-center study to cluster Chinese COPD patients using strict diagnostic criteria including longitudinal follow up and assessment of separate validation and derivation cohorts [51]. We employed robust statistical models to derive and validate clusters, adjusted for potential confounders and demonstrate high predictive accuracies. Despite strengths, our study does have limitations including being restricted to tertiary hospitals making data less generalizable to community based patients with milder disease. More than 90% of our COPD cohort was male, and while this is largely representative of the COPD population across Asia [7-8], it restricts applicability of our findings to female patients. It remains unclear and not addressed by this work whether Chinese females smoke less or have different susceptibilities to developing COPD. While all patients had chest radiography demonstrating no evidence of co-existing bronchiectasis, gold standard imaging of chest computed tomography (CT) was not available in all patients hence potential particularly TB-related sequelae may not have been identified. Furthermore, the precise timing of a tuberculosis diagnosis was not available in all patients; hence we cannot fully differentiate post-tuberculosis fixed airflow limitation from COPD in appropriate patients. While we

present a simplified CART (decision tree) model to classify patients into our described five clusters, its accuracy was significantly lower than that from our RDA model hence further research is clearly necessary before simplified models for patient classification can be routinely implemented into clinical practice. For inflammatory work, we selected a restricted final panel of six cytokines based on strict semi-quantitative screening criteria (at least 2.5-fold differences between clusters). If we decreased fold difference cut-offs, we may have identified more cytokines of relevance to our clusters. All clinical data used for derivation of clusters was collated at time of recruitment and therefore stability of the clusters or patient membership within them over time was not examined and is an interesting area for follow up. Validation of our clusters in Chinese COPD patients outside Asia (including BODE assessment) and among other ethnic populations (e.g. Malays, Indians) within Asia should be pursued. Despite demonstrating the robustness of our clustering algorithm, clustering in COPD does have its limitations and has been previously shown to result in marked heterogeneity[52]. When comparing our cluster mortality with classic COPD mortality prediction models such as the ADO index[53], we found no significant differences. Future work should assess the additive value of our derived clusters in Chinese COPD using classic COPD mortality prediction models including BODE.

We describe, for the first time, validated clusters of Chinese COPD patients including two 'high-risk' patient groups. TB and COPD while associated, are independent key public health concerns demonstrating an unmet need by our work. Overall, our findings improve risk stratification in Chinese COPD patients identifying those at 'highest risk' requiring close monitoring and appropriate intervention.

Acknowledgments

The authors would like to acknowledge The Academic Respiratory Initiative for Pulmonary Health (TARIPH) for collaboration support.

Author contributions: PYT and SHC take full responsibility for the content of the manuscript, including the data analysis. PYT: Study design, patient recruitment and performance of experimental work, data collection, interpretation and analysis including writing of the final manuscript. JKN, KTA, MMA: data interpretation and statistical analysis. FWSK, HX, MEP, NHK, DSCH, AT: patient recruitment, clinical data and specimen collection. HYN, LCL, CKO, JHYT, GJHS, TSL, MSK, JAA: patient recruitment and clinical data collection. SHC: Study design and conception of experiments, obtained study funding, interpretation of results, data analysis and writing of the final manuscript. All authors reviewed and approved the final draft of the manuscript.

Financial/non-financial disclosures: None declared.

Additional information: The e-Appendixes, e-Figures and e-Tables are provided in the online supplement.

References

1. Fang L, Gao P, Bao H, Tang X, Wang B, Feng Y, et al. Chronic obstructive pulmonary disease in China: a nationwide prevalence study. Lancet Respir Med. 2018;6(6):421-30.

2. Lim S, Lam DC, Muttalif AR, Yunus F, Wongtim S, Lan le TT, et al. Impact of chronic obstructive pulmonary disease (COPD) in the Asia-Pacific region: the EPIC Asia population-based survey. Asia Pac Fam Med. 2015;14(1):4.

3. Teramoto S, Yamamoto H, Yamaguchi Y, Matsuse T, Ouchi Y. Global burden of COPD in Japan and Asia. Lancet. 2003;362(9397):1764-5.

4. Oh YM, Bhome AB, Boonsawat W, Gunasekera KD, Madegedara D, Idolor L, et al. Characteristics of stable chronic obstructive pulmonary disease patients in the pulmonology clinics of seven Asian cities. Int J Chron Obstruct Pulmon Dis. 2013;8:31-9.

5. Guo C, Zhang Z, Lau AKH, Lin CQ, Chuang YC, Chan J, et al. Effect of long-term exposure to fine particulate matter on lung function decline and risk of chronic obstructive pulmonary disease in Taiwan: a longitudinal, cohort study. Lancet Planet Health. 2018;2(3):e114-e25.

6. Tan WC, Ng TP. COPD in Asia: where East meets West. Chest. 2008;133(2):517-27.

7. Martin A, Badrick E, Mathur R, Hull S. Effect of ethnicity on the prevalence, severity, and management of COPD in general practice. Br J Gen Pract. 2012;62(595):e76-81.

8. Wedzicha JA, Zhong N, Ichinose M, Humphries M, Fogel R, Thach C, et al. Indacaterol/glycopyrronium versus salmeterol/fluticasone in Asian patients with COPD at a high risk of exacerbations: results from the FLAME study. Int J Chron Obstruct Pulmon Dis. 2017;12:339-49.

9. Gilkes A, Ashworth M, Schofield P, Harries TH, Durbaba S, Weston C, et al. Does COPD risk vary by ethnicity? A retrospective cross-sectional study. Int J Chron Obstruct Pulmon Dis. 2016;11:739-46.

10. Tran HN, Siu S, Iribarren C, Udaltsova N, Klatsky AL. Ethnicity and risk of hospitalization for asthma and chronic obstructive pulmonary disease. Ann Epidemiol. 2011;21(8):615-22.

11. Davis J, Tam E, Taira D. Disparate Rates of Utilization and Progression to Combined Heart Failure and Chronic Obstructive Pulmonary Disease among Asians and Pacific Islanders in Hawai'i. Hawaii J Med Public Health. 2016;75(8):228-34.

12. Wang C, Xu J, Yang L, Xu Y, Zhang X, Bai C, et al. Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study. Lancet. 2018;391(10131):1706-17.

13. Lopez AD, Shibuya K, Rao C, Mathers CD, Hansell AL, Held LS, et al. Chronic obstructive pulmonary disease: current burden and future projections. Eur Respir J. 2006;27(2):397-412.

14. GOLD. Global Strategy for the Diagnosis, Management and Prevention of COPD, Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2018. (http://goldcopd.org/); *2018*.

15. Cantin L, Bankier AA, Eisenberg RL. Bronchiectasis. AJR Am J Roentgenol. 2009;193(3):W158-71.

16. Milliron B, Henry TS, Veeraraghavan S, Little BP. Bronchiectasis: Mechanisms and Imaging Clues of Associated Common and Uncommon Diseases. Radiographics. 2015;35(4):1011-30.

17. Vanfleteren LE, Spruit MA, Groenen M, Gaffron S, van Empel VP, Bruijnzeel PL, 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(7):728-35.

18. Chubachi S, Sato M, Kameyama N, Tsutsumi A, Sasaki M, Tateno H, et al. Identification of five clusters of comorbidities in a longitudinal Japanese chronic obstructive pulmonary disease cohort. Respir Med. 2016;117:272-9.

19. Garcia-Aymerich J, Gomez FP, Benet M, Farrero E, Basagana X, Gayete A, et al. Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes. Thorax. 2011;66(5):430-7.

505 20. Burgel PR, Paillasseur JL, Janssens W, Piquet J, Ter Riet G, Garcia-Aymerich J, et al.
506 A simple algorithm for the identification of clinical COPD phenotypes. Eur Respir J.
507 2017;50(5).
508 21. Raherison C, Ouaalaya EH, Bernady A, Casteigt J, Nocent-Eijnani C, Falque L, et al.
509 Comorbidities and COPD severity in a clinic-based cohort. BMC Pulm Med. 2018;18(1):117.
510 22. Cavailles A, Brinchault-Rabin G, Dixmier A, Goupil F, Gut-Gobert C, Marchand-
511 Adam S, et al. Comorbidities of COPD. Eur Respir Rev. 2013;22(130):454-75.
512 23. Rabe KF, Hurst JR, Suissa S. Cardiovascular disease and COPD: dangerous liaisons?
513 Eur Respir Rev. 2018;27(149).
514 24. Miller J, Edwards LD, Agusti A, Bakke P, Calverley PM, Celli B, et al. Comorbidity,
515 systemic inflammation and outcomes in the ECLIPSE cohort. Respir Med.
516 2013;107(9):1376-84.
517 25. Kubota Y, Tay WT, Asai K, Murai K, Nakajima I, Hagiwara N, et al. Chronic
518 obstructive pulmonary disease and beta-blocker treatment in Asian patients with heart failure.
519 ESC Heart Fail. 2018;5(2):297-305.
520 26. Mannino DM, Thorn D, Swensen A, Holguin F. Prevalence and outcomes of diabetes,
521 hypertension and cardiovascular disease in COPD. Eur Respir J. 2008;32(4):962-9.
522 27. Khan NA, Wang H, Anand S, Jin Y, Campbell NR, Pilote L, et al. Ethnicity and sex
523 affect diabetes incidence and outcomes. Diabetes Care. 2011;34(1):96-101.
524 28. Godtfredsen NS, Lam TH, Hansel TT, Leon ME, Gray N, Dresler C, et al. COPD-
525 related morbidity and mortality after smoking cessation: status of the evidence. Eur Respir J.
526 2008;32(4):844-53.
527 29. Anthonisen NR, Skeans MA, Wise RA, Manfreda J, Kanner RE, Connett JE, et al.
528 The effects of a smoking cessation intervention on 14.5-year mortality: a randomized clinical
529 trial. Ann Intern Med. 2005;142(4):233-9.

30. Menezes AM, Hallal PC, Perez-Padilla R, Jardim JR, Muino A, Lopez MV, et al. Tuberculosis and airflow obstruction: evidence from the PLATINO study in Latin America. Eur Respir J. 2007;30(6):1180-5.

31. Leung JM, Tiew PY, Mac Aogain M, Budden KF, Yong VF, Thomas SS, et al. The role of acute and chronic respiratory colonization and infections in the pathogenesis of COPD. Respirology. 2017;22(4):634-50.

32. Chotirmall SH, Gellatly SL, Budden KF, Mac Aogain M, Shukla SD, Wood DL, et al. Microbiomes in respiratory health and disease: An Asia-Pacific perspective. Respirology. 2017;22(2):240-50.

33. Budden KF, Shukla SD, Rehman SF, Bowerman KL, Keely S, Hugenholtz P, et al. Functional effects of the microbiota in chronic respiratory disease. Lancet Respir Med. 2019.

34. Jung KH, Kim SJ, Shin C, Kim JH. The considerable, often neglected, impact of pulmonary tuberculosis on the prevalence of COPD. Am J Respir Crit Care Med. 2008;178(4):431; author reply 2-3.

35. Ravimohan S, Kornfeld H, Weissman D, Bisson GP. Tuberculosis and lung damage: from epidemiology to pathophysiology. Eur Respir Rev. 2018;27(147).

36. O'Leary SM, Coleman MM, Chew WM, Morrow C, McLaughlin AM, Gleeson LE, et al. Cigarette smoking impairs human pulmonary immunity to Mycobacterium tuberculosis. Am J Respir Crit Care Med. 2014;190(12):1430-6.

37. Park HJ, Byun MK, Kim HJ, Ahn CM, Kim DK, Kim YI, et al. History of pulmonary tuberculosis affects the severity and clinical outcomes of COPD. Respirology. 2018;23(1):100-6.

38. Coussens AK, Wilkinson RJ, Nikolayevskyy V, Elkington PT, Hanifa Y, Islam K, et al. Ethnic variation in inflammatory profile in tuberculosis. PLoS Pathog. 2013;9(7):e1003468.

555 39. Pareek M, Evans J, Innes J, Smith G, Hingley-Wilson S, Lougheed KE, et al.
556 Ethnicity and mycobacterial lineage as determinants of tuberculosis disease phenotype.
557 Thorax. 2013;68(3):221-9.
558 40. Agusti A, Edwards LD, Rennard SI, MacNee W, Tal-Singer R, Miller BE, et al.
559 Persistent systemic inflammation is associated with poor clinical outcomes in COPD: a novel
560 phenotype. PLoS One. 2012;7(5):e37483.
561 41. Poh TY, Mac Aogain M, Chan AK, Yii AC, Yong VF, Tiew PY, et al. Understanding
562 COPD-overlap syndromes. Expert Rev Respir Med. 2017;11(4):285-98.
563 42. Harding JS, Herbath M, Chen Y, Rayasam A, Ritter A, Csoka B, et al. VEGF-A from
564 Granuloma Macrophages Regulates Granulomatous Inflammation by a Non-angiogenic
565 Pathway during Mycobacterial Infection. Cell Rep. 2019;27(7):2119-31 e6.
566 43. Datta M, Via LE, Kamoun WS, Liu C, Chen W, Seano G, et al. Anti-vascular
567 endothelial growth factor treatment normalizes tuberculosis granuloma vasculature and
568 improves small molecule delivery. Proc Natl Acad Sci U S A. 2015;112(6):1827-32.
569 44. Matsuyama W, Hashiguchi T, Matsumuro K, Iwami F, Hirotsu Y, Kawabata M, et al.
570 Increased serum level of vascular endothelial growth factor in pulmonary tuberculosis. Am J
571 Respir Crit Care Med. 2000;162(3 Pt 1):1120-2.
572 45. Kumar NP, Banurekha VV, Nair D, Babu S. Circulating Angiogenic Factors as
573 Biomarkers of Disease Severity and Bacterial Burden in Pulmonary Tuberculosis. PLoS One.
574 2016;11(1):e0146318.
575 46. Harrison MT, Short P, Williamson PA, Singanayagam A, Chalmers JD, Schembri S.
576 Thrombocytosis is associated with increased short and long term mortality after exacerbation
577 of chronic obstructive pulmonary disease: a role for antiplatelet therapy? Thorax.
578 2014;69(7):609-15.

47. Fawzy A, Putcha N, Paulin LM, Aaron CP, Labaki WW, Han MK, et al. Association of thrombocytosis with COPD morbidity: the SPIROMICS and COPDGene cohorts. Respir Res. 2018;19(1):20.

48. Noskovicova N, Petrek M, Eickelberg O, Heinzelmann K. Platelet-derived growth factor signaling in the lung. From lung development and disease to clinical studies. Am J Respir Cell Mol Biol. 2015;52(3):263-84.

49. Raines EW. PDGF and cardiovascular disease. Cytokine Growth Factor Rev. 2004;15(4):237-54.

50. Churg A, Tai H, Coulthard T, Wang R, Wright JL. Cigarette smoke drives small airway remodeling by induction of growth factors in the airway wall. Am J Respir Crit Care Med. 2006;174(12):1327-34.

51. Bhatt SP, Balte PP, Schwartz JE, Cassano PA, Couper D, Jacobs DR, Jr, et al. Discriminative Accuracy of FEV1:FVC Thresholds for COPD-Related Hospitalization and MortalityDiscriminative Accuracy of FEV1:FVC Thresholds for COPD-Related Hospitalization and MortalityDiscriminative Accuracy of FEV1:FVC Thresholds for COPD-Related Hospitalization and Mortality. JAMA. 2019;321(24):2438-47.

52. Castaldi PJ, Benet M, Petersen H, Rafaels N, Finigan J, Paoletti M, et al. Do COPD subtypes really exist? COPD heterogeneity and clustering in 10 independent cohorts. Thorax. 2017;72(11):998-1006.

53. Puhan MA, Hansel NN, Sobradillo P, Enright P, Lange P, Hickson D, et al. Large-scale international validation of the ADO index in subjects with COPD: an individual subject data analysis of 10 cohorts. BMJ Open. 2012;2(6).

Figure legends

Figure 1: Unsupervised clustering (of the derivation cohort) reveals five clinically relevant patient clusters of ethnic-Chinese patients with chronic obstructive pulmonary disease (COPD) demonstrating variable prognostic outcome. Dendrogram illustrating the five derived COPD clusters using non-metric multidimensional scaling followed by hierarchical clustering. Different clusters are represented by colours: Ex-tuberculosis (green), diabetic (blue), low co-morbidity: low-risk (violet red), low co-morbidity: high-risk (yellow) and cardiovascular (pink). Ex-TB: ex-tuberculosis, LCLR: low co-morbidity: low-risk. LCHR: low co-morbidity: high-risk, CVS: cardiovascular

628 **Figure 2:** Kaplan-Meier curve (a) demonstrating survival differences between clusters for two-
629 year all-cause mortality: worst prognosis in the cardiovascular and ex-tuberculosis (TB)
630 clusters which (b) remains significant after adjustment for age, sex, BMI, FEV_1 and smoking
631 pack year exposure illustrated as cox regression survival curves. (c) Cumulative incidence
632 curves for each of the five derived COPD clusters for respiratory causes of death: greatest
633 incidence of respiratory cause of death observed in the cardiovascular, ex-tuberculosis and low
634 co-morbidity: high-risk clusters. Different clusters are represented by colours: Ex-tuberculosis
635 (green), diabetic (blue), low co-morbidity: low-risk (violet red), low co-morbidity: high-risk
636 (yellow) and cardiovascular (pink). Ex-TB: ex-tuberculosis, LCLR: low co-morbidity: low-
637 risk. LCHR: low co-morbidity: high-risk, CVS: cardiovascular
638
639 **Figure 3:** Validation cohort (V; n=569) of patients with Chronic Obstructive Pulmonary
640 Disease (COPD) of Chinese ethnicity illustrates comparable co-morbidity profiles by cluster
641 when compared to the Derivation (D; n=911) cohort. Bubble size corresponds to the percentage
642 of patients demonstrating each comorbidity within their respective cohort and bubble colour
643 represents cluster membership: Ex-tuberculosis (green), diabetic (blue), low co-morbidity:
644 low-risk (violet red), low co-morbidity: high-risk (yellow) and cardiovascular (pink). PUD:
645 peptic ulcer disease, pTB: history of prior pulmonary tuberculosis, PAD: peripheral arterial
646 disease, Other Ca: all other malignancies excluding lung, esophageal, pancreatic or breast
647 carcinoma, DM: diabetes mellitus, CVA: cerebrovascular disease, CKD: chronic kidney
648 disease, CHF: congestive heart failure, CAD: coronary artery disease, Ca: lung, esophageal,
649 pancreatic or breast carcinoma, AF: atrial fibrillation.
650
651 **Figure 4:** Survival outcomes between the identified clusters in the validation cohort for two-
652 year all-cause mortality (as demonstrated independently in the Derivation (D; n=911) cohort;

see Figure 2): (a) Kaplan-Meier curves demonstrating worst prognosis in the cardiovascular and ex-tuberculosis clusters which (b) remains significant after adjustment for age, sex, BMI, FEV_1 and smoking pack year exposure illustrated as cox regression survival curves. (c) Comparable cumulative incidence curves for each of the five derived COPD clusters (as demonstrated independently in the Derivation (D; n=911) cohort; see Figure 2) for respiratory causes of death: greatest incidence of respiratory cause of death observed in the cardiovascular, ex-tuberculosis and low co-morbidity: high-risk clusters. Different clusters are represented by color: Ex-tuberculosis (green), diabetic (blue), low co-morbidity: low-risk (violet red), low co-morbidity: high-risk (yellow) and cardiovascular (pink). Ex-TB: ex-tuberculosis, LCLR: low co-morbidity: low-risk, LCHR: low co-morbidity: high-risk, CVS: cardiovascular

Figure 5: The five derived COPD clusters of Chinese ethnicity illustrate inflammatory signatures that associate with all cause and respiratory-related mortality. (a) Radar plot illustrating variation in systemic cytokine profile between each of the five derived COPD clusters. The median normalized score for each selected cytokine (in the final assessed panel) is plotted on the radar chart for comparison between clusters. Each dot represents the median normalized value of the cytokine and the colour indicates the specific cluster: Ex-tuberculosis (green), diabetic (blue), low co-morbidity: low-risk (violet red), low co-morbidity: high-risk (yellow) and cardiovascular (pink). TNF-R1: tumor necrosis factor receptor 1, TNF-R2: tumor necrosis factor receptor 2, VEGF: vascular endothelial growth factor, PDGF: platelet derived growth factor. *$p \leq 0.05$, #$p < 0.1$. (b) Network plots demonstrating inflammatory grids detected within each cluster. Increased cytokine interaction and a greater number of positive cytokines are detectable in the cardiovascular, ex-tuberculosis and low co-morbidity: high-risk clusters compared to the diabetic and low co-morbidity: low-risk clusters (indicated by 'greater than' (>) symbols). Each circle (node) represents a cytokine from the final assessed panel: circle size

corresponds to percentage of patients within that cluster demonstrating a positive value. Lines connecting two nodes indicate positive detection of both cytokines and line thickness illustrates the proportion of patients with a positive result for both cytokines. Node color corresponds to the respective clinical cluster: Ex-tuberculosis (green), diabetic (blue), low co-morbidity: low-risk (violet red), low co-morbidity: high-risk (yellow) and cardiovascular (pink).

Figure 6: Cardiovascular and ex-tuberculosis clusters illustrate highest two-year mortality (a) Tree map illustrating cluster related mortality within each GOLD group [A, B, C and D]. Rectangles represent the proportion of deceased patients within each group, presented as percentages. Rectangle colour indicates cluster membership: Ex-tuberculosis (green), diabetic (blue), low co-morbidity: low-risk (violet red), low co-morbidity: high-risk (yellow) and cardiovascular (pink). (b) The mortality differences remain significant after adjustment for age, sex, BMI, smoking pack year exposure, lung function (by FEV_1) and GOLD group illustrated by forest plot using multivariate logistic regression. The dot represents the odds ratio with colour indicating significance levels: red ($p < 0.05$), grey ($p > 0.05$; ns). Error bar indicates the 95% confidence interval (CI). FEV_1: forced expiratory volume in the 1_{st} second. Ex-TB: ex-tuberculosis, LCLR: low co-morbidity: low-risk, LCHR: low co-morbidity: high-risk, CVS: cardiovascular.

Figure 7: Cardiovascular and ex-tuberculosis clusters demonstrate poorest two-year survival irrespective of underlying COPD grade. Kaplan-Meir curves illustrating two-year mortality of each cluster by conventional COPD staging (defined by FEV_1). *$p \leq 0.05$, **$p=0.001$ by log-rank test. Different clusters are represented by colours: Ex-tuberculosis (green), diabetic (blue), low co-morbidity: low-risk (violet red), low co-morbidity: high-risk (yellow) and

cardiovascular (pink). Ex-TB: ex-tuberculosis, LCLR: low co-morbidity: low-risk. LCHR: low co-morbidity: high-risk, CVS: cardiovascular

Figure 1

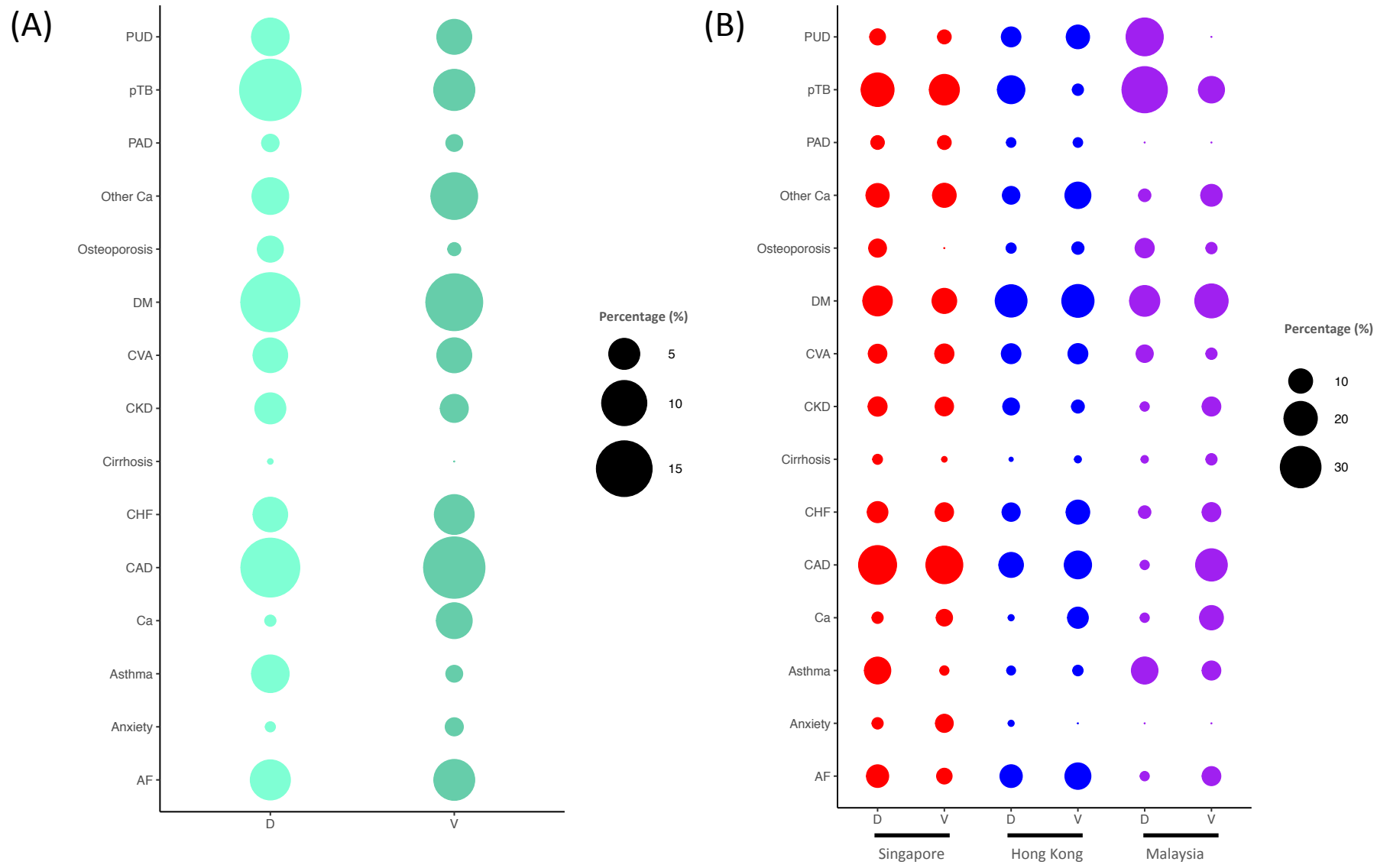
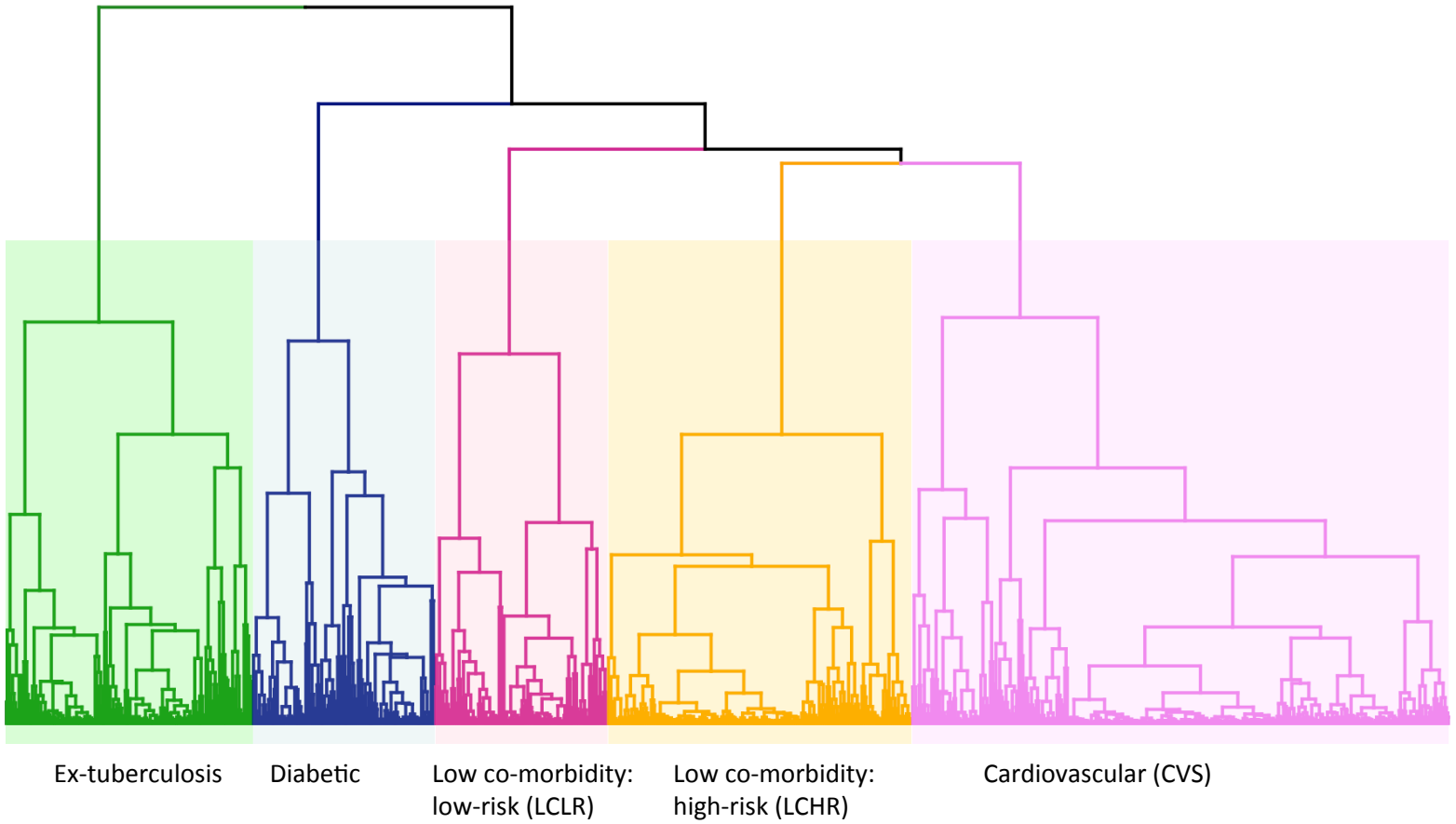
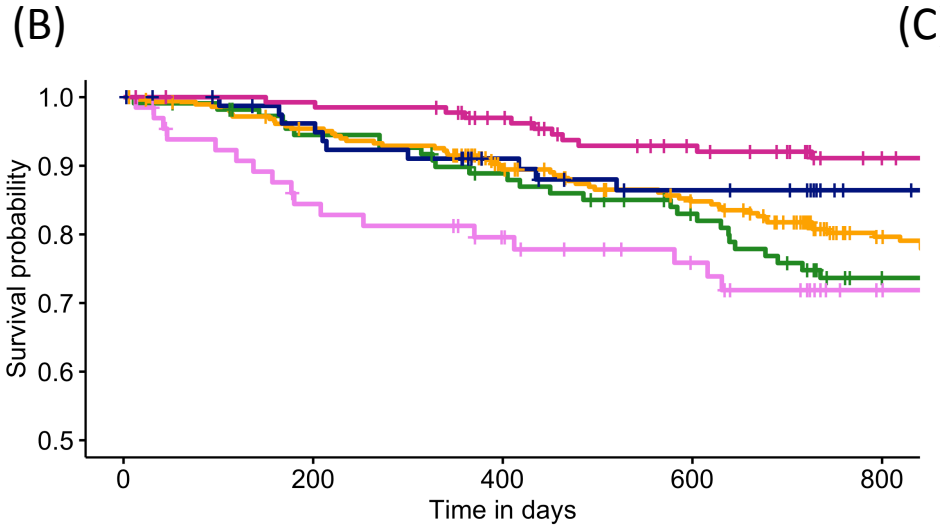
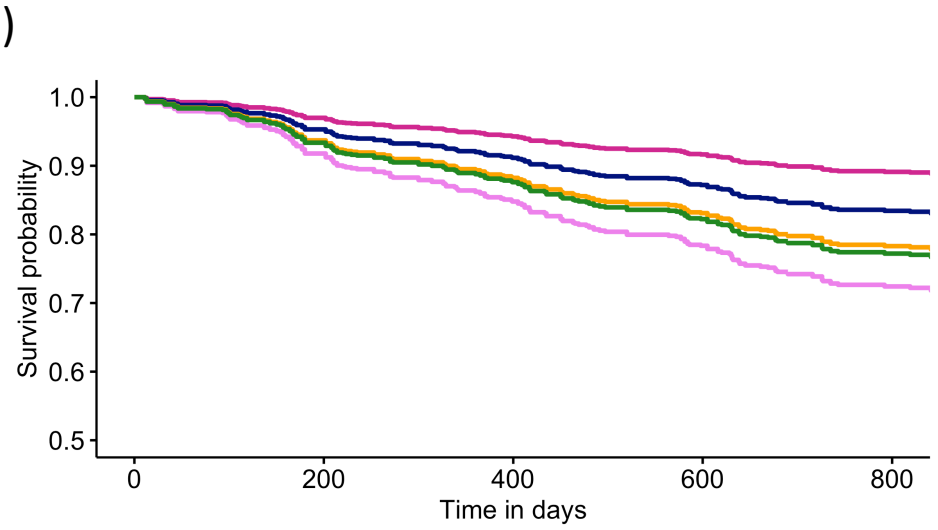

(A)
PUD
pTB
PAD
Other Ca
Osteoporosis
DM
CVA
CKD
Cirrhosis
CHF
CAD
Ca
Asthma
Anxiety
AF
D
V
Percentage (%)
5
10
15
(B)
PUD
pTB
PAD
Other Ca
Osteoporosis
DM
CVA
CKD
Cirrhosis
CHF
CAD
Ca
Asthma
Anxiety
AF
D
V
D
V
D
V
Singapore
Hong Kong
Malaysia
Percentage (%)
10
20
30

Figure 2

(A)


Ex-tuberculosis
Diabetic
Low co-morbidity: low-risk (LCLR)
Low co-morbidity: high-risk (LCHR)
Cardiovascular (CVS)


(B)
Survival probability
1.0
0.9
0.8
0.7
0.6
0.5
0
200
400
600
800
Time in days


(C)
Survival probability
1.0
0.9
0.8
0.7
0.6
0.5
0
200
400
600
800
Time in days

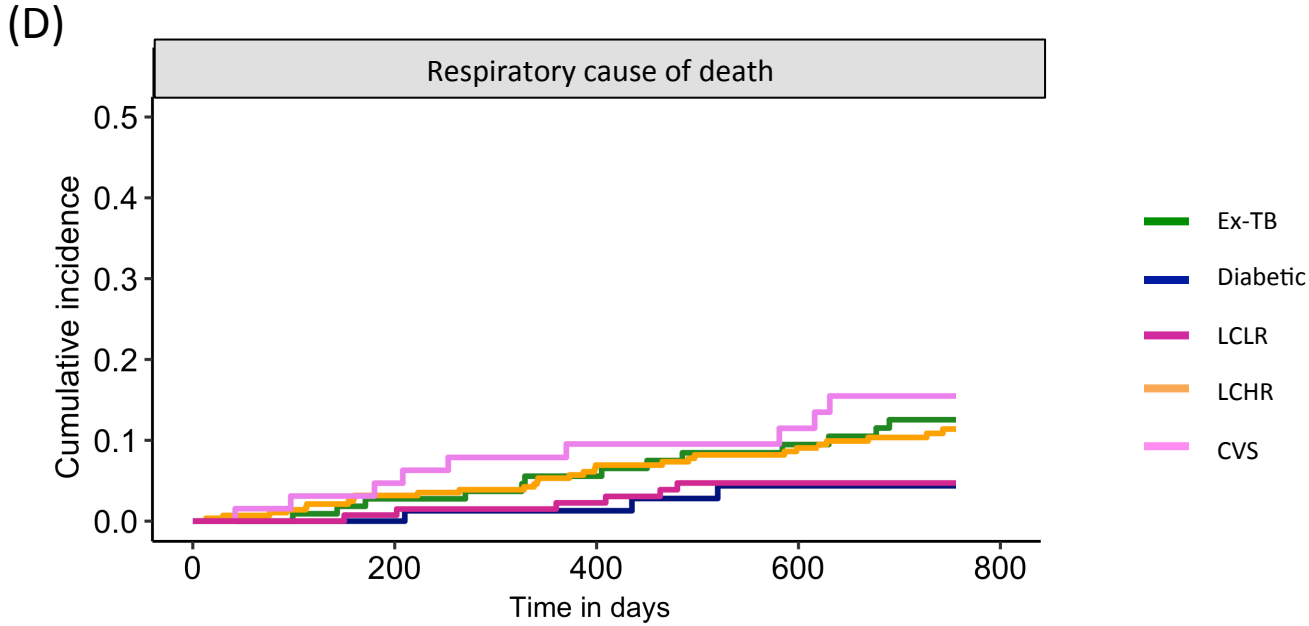

(D)
Respiratory cause of death
Cumulative incidence
0.5
0.4
0.3
0.2
0.1
0.0
0
200
400
600
800
Time in days
Ex-TB
Diabetic
LCLR
LCHR
CVS

Figure 3

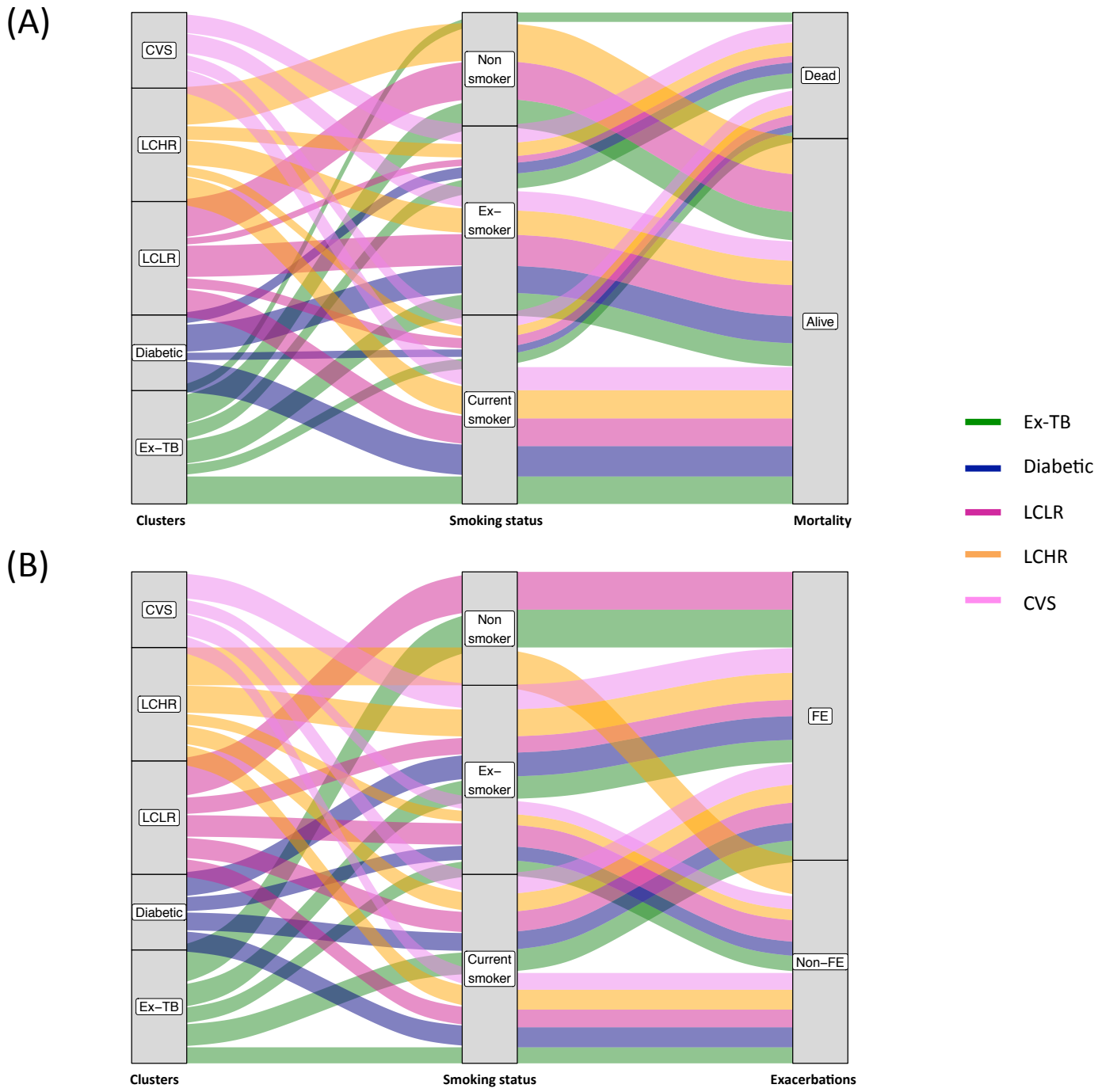
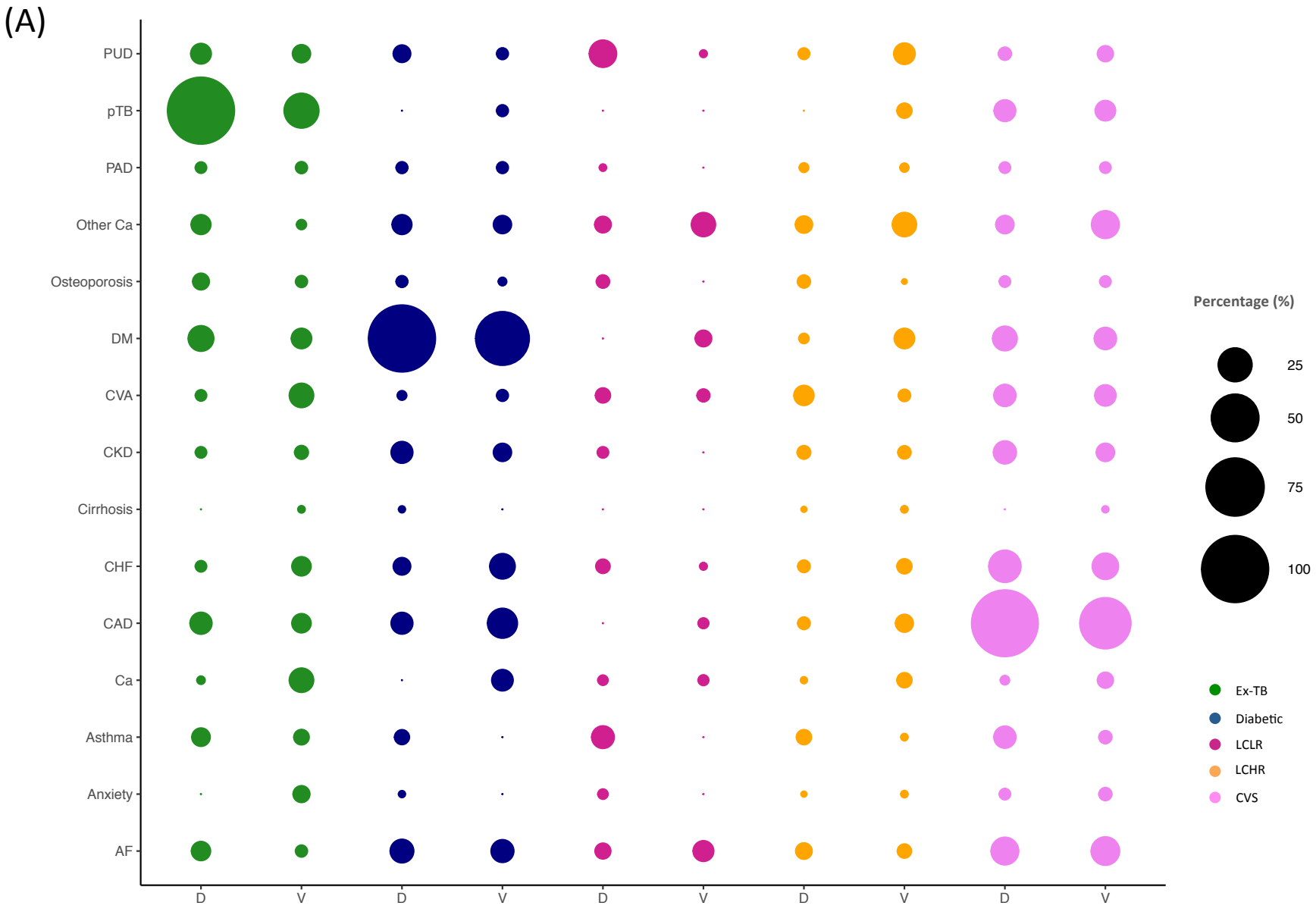
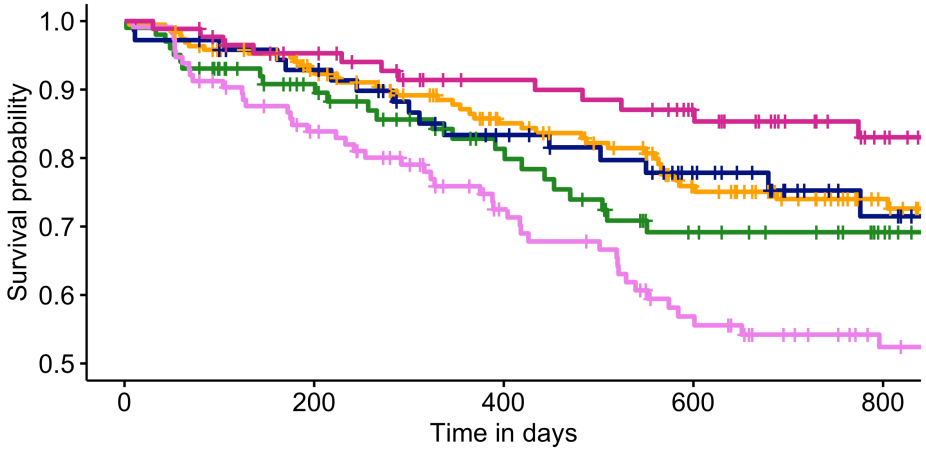

(A)
CVS
LCHR
LCLR
Diabetic
Ex–TB
Non smoker
Ex–smoker
Current smoker
Dead
Alive
Clusters
Smoking status
Mortality
(B)
CVS
LCHR
LCLR
Diabetic
Ex–TB
Non smoker
Ex–smoker
Current smoker
FE
Non–FE
Clusters
Smoking status
Exacerbations
Ex-TB
Diabetic
LCLR
LCHR
CVS

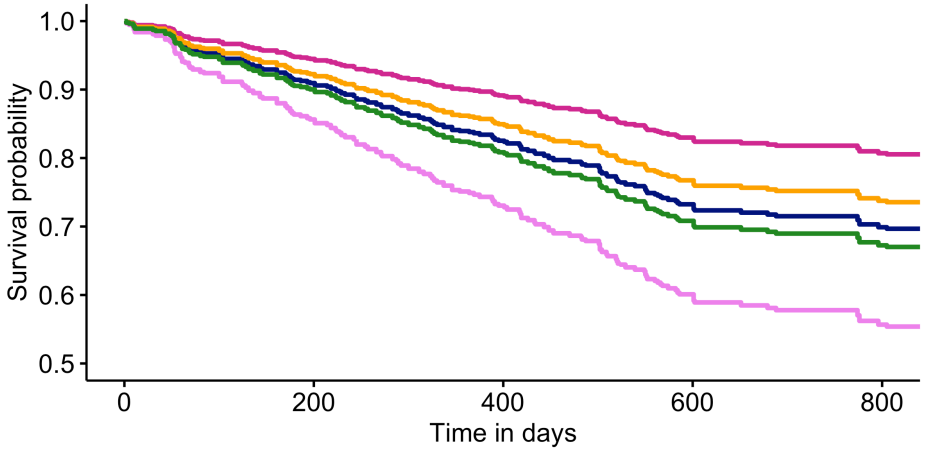
Figure 4


(A)
PUD
pTB
PAD
Other Ca
Osteoporosis
DM
CVA
CKD
Cirrhosis
CHF
CAD
Ca
Asthma
Anxiety
AF
D
V
D
V
D
V
D
V
D
V
Percentage (%)
25
50
75
100
Ex-TB
Diabetic
LCLR
LCHR
CVS

(B)


Survival probability
1.0
0.9
0.8
0.7
0.6
0.5
0
200
400
600
800
Time in days

(C)


Survival probability
1.0
0.9
0.8
0.7
0.6
0.5
0
200
400
600
800
Time in days

(D)

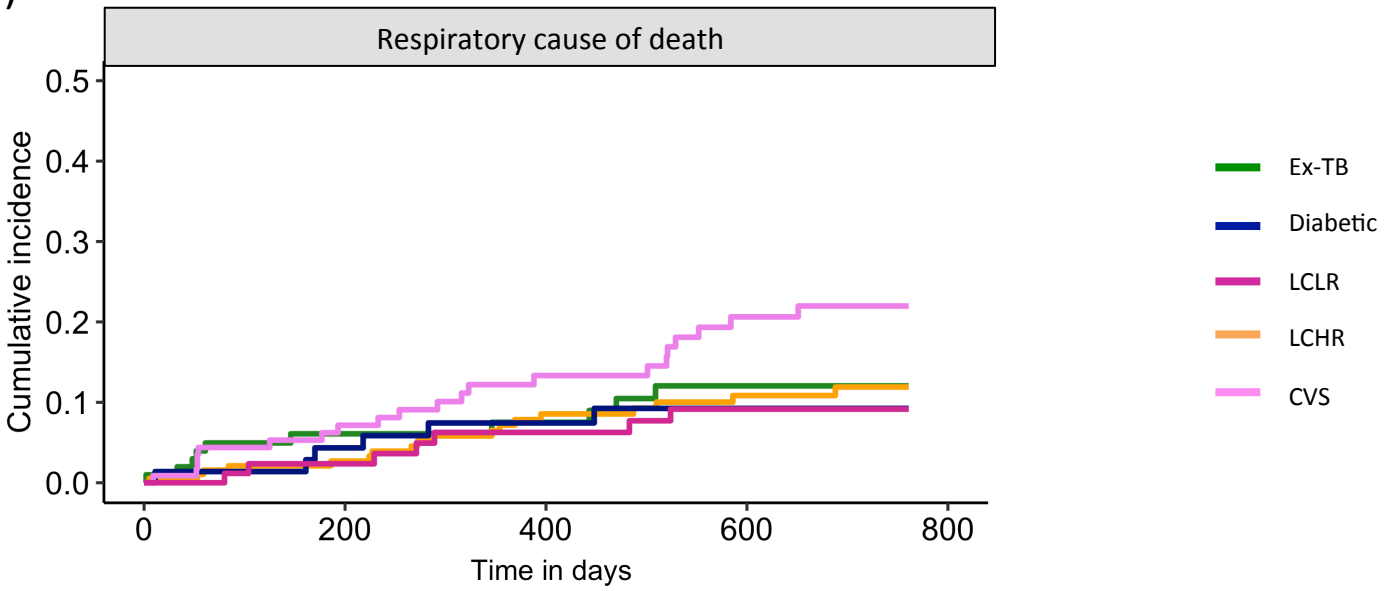
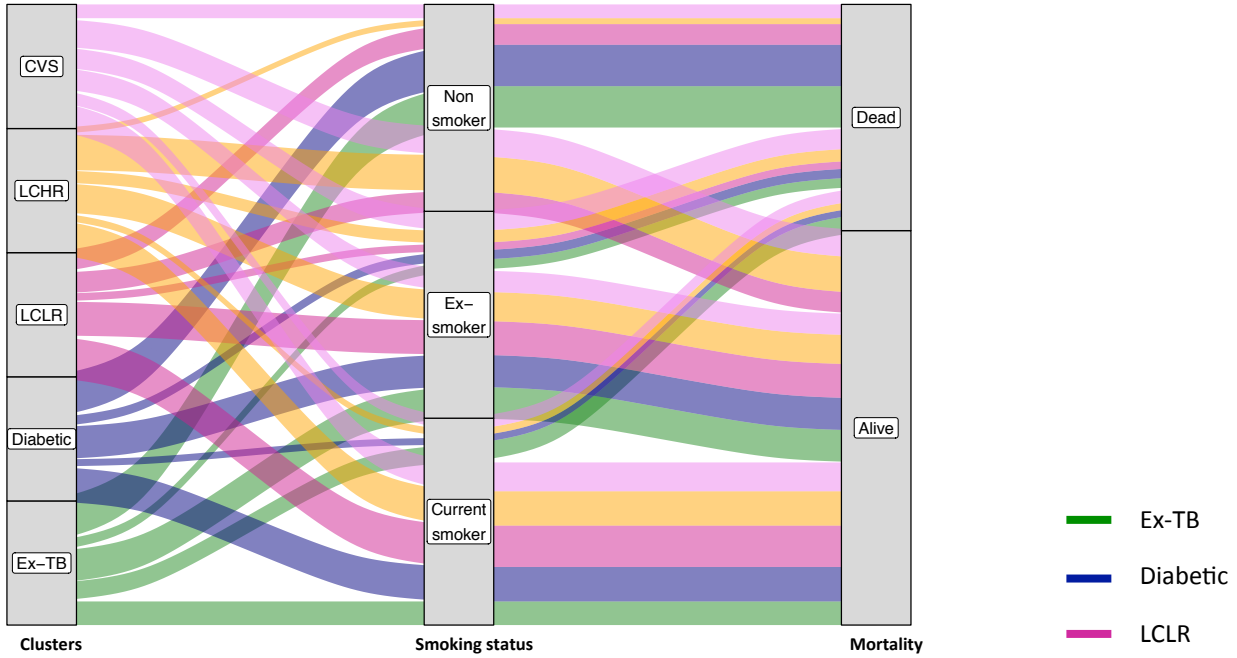
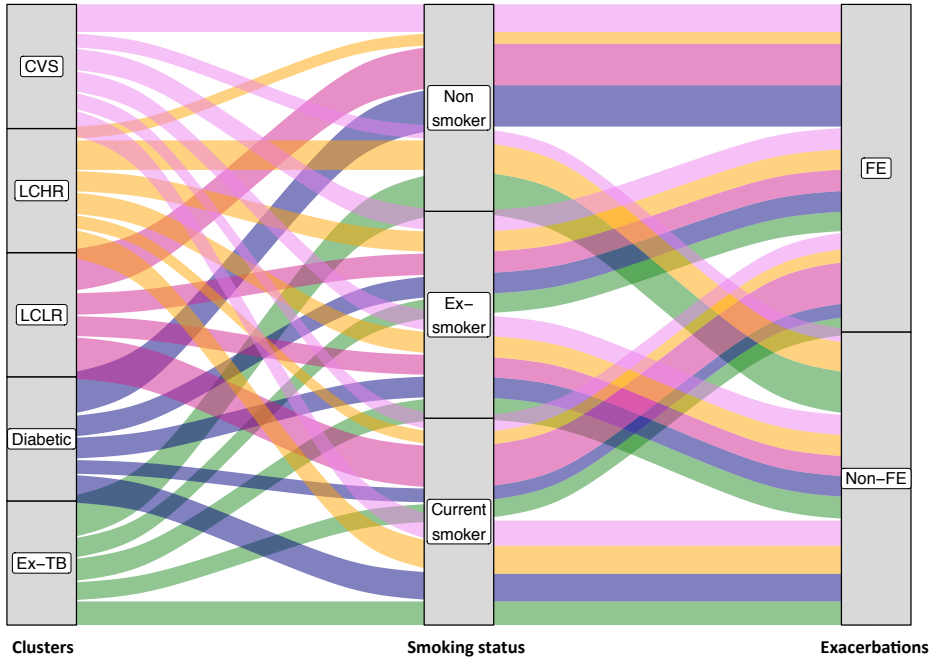

Respiratory cause of death
Cumulative incidence
0.5
0.4
0.3
0.2
0.1
0.0
0
200
400
600
800
Time in days
Ex-TB
Diabetic
LCLR
LCHR
CVS

Figure 5

(A)


CVS
LCHR
LCLR
Diabetic
Ex–TB
Non smoker
Ex–smoker
Current smoker
Dead
Alive
Clusters
Smoking status
Mortality

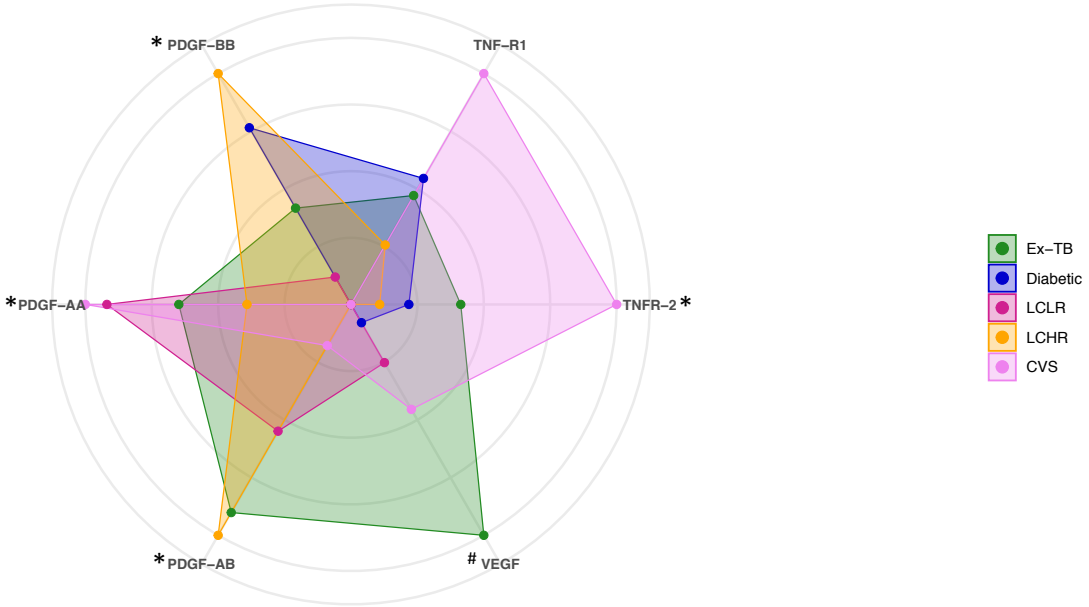
(B)


CVS
LCHR
LCLR
Diabetic
Ex–TB
Non smoker
Ex–smoker
Current smoker
FE
Non–FE
Clusters
Smoking status
Exacerbations

Ex-TB
Diabetic
LCLR
LCHR
CVS

Figure 6

(A)


* PDGF–BB
TNF–R1
*PDGF–AA
TNFR–2 *
*PDGF–AB
VEGF
Ex–TB
Diabetic
LCLR
LCHR
CVS

(B)

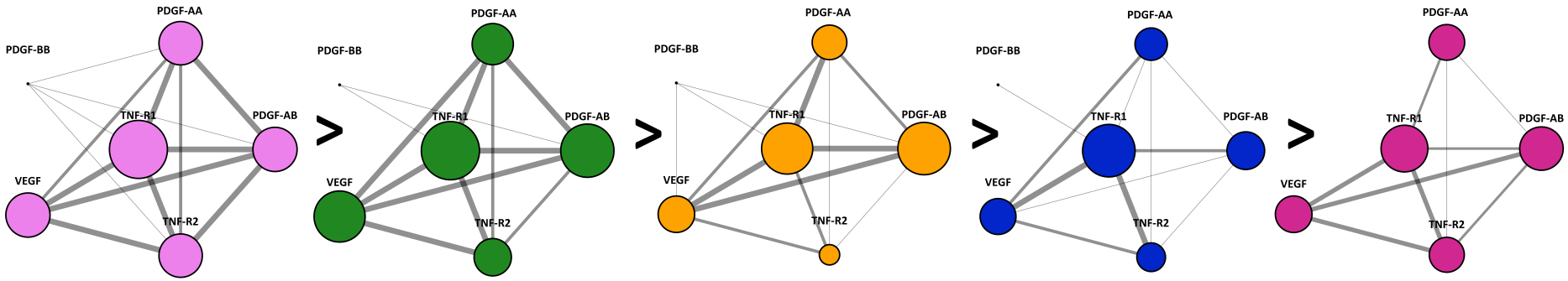
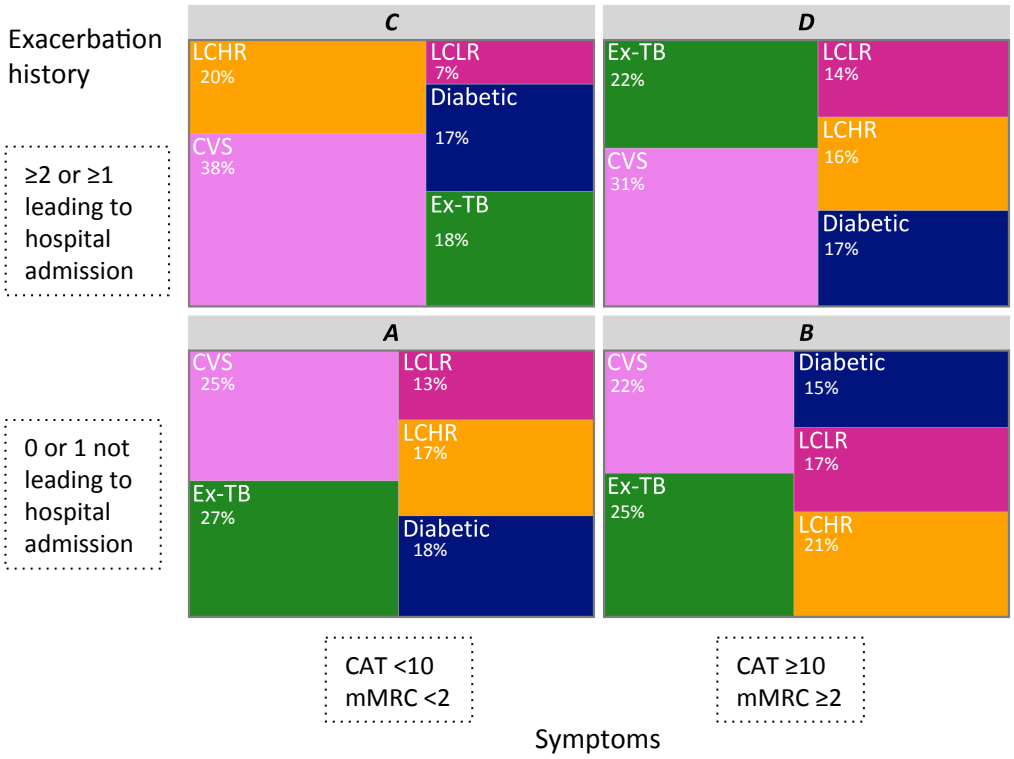
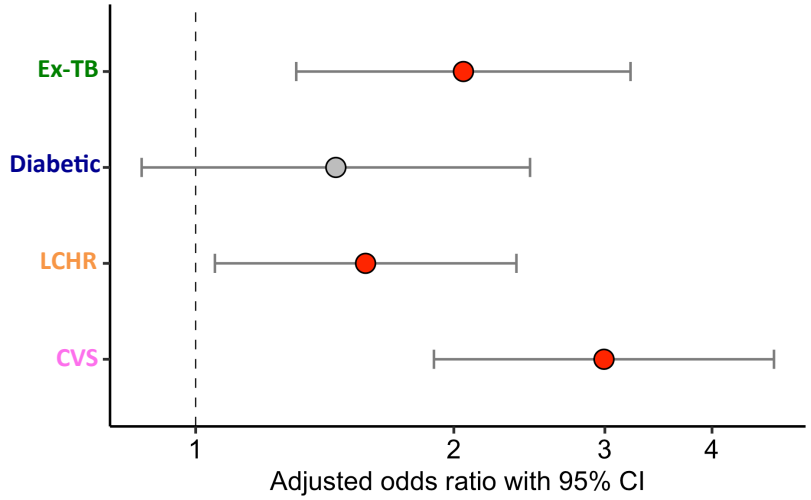

PDGF-AA
PDGF-BB
TNF-R1
PDGF-AB
VEGF
TNF-R2
>
Cardiovascular
>
Ex-TB
>
LCHR
>
Diabetic
>
LCLR

Figure 7

(A)


Exacerbation history
≥2 or ≥1 leading to hospital admission
0 or 1 not leading to hospital admission
C
LCHR 20%
LCLR 7%
Diabetic 17%
CVS 38%
Ex-TB 18%
D
Ex-TB 22%
LCLR 14%
LCHR 16%
CVS 31%
Diabetic 17%
A
CVS 25%
LCLR 13%
LCHR 17%
Ex-TB 27%
Diabetic 18%
B
CVS 22%
Diabetic 15%
LCLR 17%
Ex-TB 25%
LCHR 21%
CAT <10 mMRC <2
CAT ≥10 mMRC ≥2
Symptoms

(B)


Ex-TB
Diabetic
LCHR
CVS
1
2
3
4
Adjusted odds ratio with 95% CI

Online Supplement

‘HIGH RISK’ CLINICAL AND INFLAMMATORY CLUSTERS IN COPD OF CHINESE DESCENT

Pei Yee Tiew, MD[1,2], Fanny Wai San Ko, MD[3], Jayanth Kumar Narayana, BS-MS[1,4], Mau Ern Poh, MD[5], Huiying Xu, MD[6], Han Yee Neo, MD[6], Li-Cher Loh, MD[7], Choo-Khoo Ong, MD[7], Micheál Mac Aogáin, PhD[1], Jessica Han Ying Tan, MD[8], Nabilah Husna Kamaruddin[5], Gerald Jiong Hui Sim[9], Therese S. Lapperre, MD, PhD[2,10], Mariko Siyue Koh, MD[2], David Shu Cheong Hui, MD[3], John Arputhan Abisheganaden, MD[6], Augustine Tee, MD[9], Krasimira Tsaneva-Atanasova, PhD[11,12], Sanjay H. Chotirmall, MD, PhD[1#].

e-Appendix 1: Ethical approval

The institutional ethics review boards of all participating hospitals approved the study as follows: DSRB 2012/01110, CIRB 2018/2186 (2013/184/C), CIRB 2016/2715, CIRB 2017/3010, CIRB 2017/2933 (all mutually recognized by DSRB, Singapore), NNMR-13-313-15138, UMMC 2018725-6524 (Malaysia), CREC 2011.146, CREC 2015.164 and CREC 2018.042 (Hong Kong). Non-COPD (healthy) patient recruitment was approved by the NTU institutional ethics review board under IRB-2017-12-010.

e-Appendix 2: Additional methods

Clinical data collation: Clinical data was obtained for all subjects at recruitment and included demographics, smoking history (including pack year exposure where relevant), number of acute exacerbations requiring hospitalization in the previous year (from time of recruitment), sputum (culture positive) microbiology, COPD assessment test (CAT) scores and, co-morbidities by the COPD specific co-morbidity test (COTE) [1]. Comorbidities recorded through COTE include: anxiety, atrial fibrillation/flutter, coronary artery disease, congestive heart failure, diabetes mellitus, liver cirrhosis, malignancy, gastric or duodenal ulcers (peptic ulcer disease). Additional comorbidities noted by the recruiting clinician at study entry and verified through patient medical records were also documented. These include chronic kidney disease, cerebrovascular disease/stroke, peripheral arterial disease, asthma, prior documented pulmonary tuberculosis, chronic respiratory failure, pulmonary hypertension and osteoporosis. Co-morbidities were assessed in all patients at enrolment and definitions of each comorbidity are provided below. Sputum microbiology was obtained through spontaneous expectorated and representative sputum (<10 squamous epithelial cells and >25 leucocytes per low-power microscopic field [2-3]. Two-year mortality (as documented on the death certification) and number of exacerbations including hospitalizations (for COPD

exacerbations) in the subsequent year following recruitment was prospectively recorded. A COPD exacerbation was defined as a sudden deterioration of respiratory symptoms requiring additional therapy (steroids and/or antibiotics) [4]. Severe exacerbators were defined as the occurrence of one or more exacerbations requiring hospitalization [4]. Respiratory cause of death was defined as death secondary to pneumonia, chronic obstructive pulmonary disease or respiratory failure.

Co-morbidity definitions: Anxiety was diagnosed based on DSM-V criteria [5]. Atrial fibrillation/flutter was defined by the presence of electrocardiogram criteria including absence of P waves, irregular R-R intervals and atrial activity [6]. Congestive heart failure was diagnosed by echocardiography [7]. Coronary artery disease was defined based on functional testing, radiological imaging and/or coronary angiography [8]. Diabetes mellitus was defined as the presence of either fasting blood glucose levels of ≥7.0 mmol/l, blood glucose levels of ≥11.1 mmol/l two-hour post oral glucose tolerance test, random blood glucose level ≥11.1 mmol/l with hyperglycemia symptoms, or HbA1c ≥6.5% [9]. Liver cirrhosis was defined by the presence of clinical, biochemical and radiological features of liver failure with or without liver biopsy [10]. Respective malignancies were diagnosed by radiological imaging and/or histological confirmation. Peptic ulcer disease was diagnosed by endoscopy [11]. Chronic kidney disease was defined as an estimated glomerular filtration rate (eGFR) <60 ml/min, according to the National Kidney Foundation Kidney Disease Outcome Quality Initiative guidelines [12]. Cerebrovascular disease/stroke was defined as a previously documented central nervous system infarction or hemorrhage [13]. Peripheral arterial disease was defined as ankle brachial index of ≤0.9 [14]. Asthma was defined according to the Global Initiative of Asthma (GINA) guidelines based on past history, symptoms and the presence of variable expiratory airflow limitation [15]. Previous pulmonary tuberculosis was defined as a prior history of

783 documented tuberculosis with positive sputum analysis for mycobacteria tuberculosis
784 including positive acid-fast bacilli smear, culture or nucleic acid amplification, radiological
785 features (on chest radiography or computed tomography) and/or prior pharmacological
786 treatment for pulmonary tuberculosis [5, 16-17]. Chronic respiratory failure was defined as either
787 requiring long-term oxygen therapy or non-invasive ventilation with $PCO_2 > 52mmHg$ [4, 18].
788 Pulmonary hypertension was defined as an increased mean pulmonary arterial pressure (PAP)
789 ≥25mmHg at rest assessed by transthoracic echocardiography and/or right heart
790 catheterization [19]. Osteoporosis was defined by radiological imaging [20]. Comorbidities were
791 obtained from patient histories and verified through medical records, including medication
792 lists and therapies received for specific disease states. Where data was unavailable or could
793 not be verified for any one particular co-morbidity, the patient was classified as not
794 demonstrating that respective co-morbidity.
795
796 *Non-COPD (healthy) patient recruitment:* Non-COPD subjects were recruited from
797 community volunteers who participated in an exercise program conducted at Nanyang
798 Technological University, Singapore aged >60 years with a measured $FEV_1/FVC>0.7$ with
799 normal FEV_1 (≥80% predicted).
800
801 *Venous blood sampling and processing:* Venous blood samples were collected from non-
802 COPD (healthy) controls (n=24) and COPD patients (n=336) from Singapore (n=74);
803 Malaysia (n=49) and Hong Kong (n=213). Samples were immediately transferred to a local
804 laboratory and centrifuged at 1300g for 10 minutes at 18°C to isolate plasma which was then
805 maintained at -80°C until inflammatory assessment. All assessments were performed in
806 Singapore and temperature-controlled shipments permitted safe transfer of specimens
807 between sites.

Cytokine (Inflammatory) assessment: To determine a 'cytokine panel' of relevance to differentiate the detected COPD 'clusters', we first screened for the presence of 120 different cytokines using the Raybiotech human cytokine array C2000 kit according to the manufacturer's instructions. A pooled plasma sample from each respective detected cluster (randomly selected) was used for screening which contained n=20 separate patients plasma (pooled) where patients were recruited from all different participating sites. Briefly, diluted plasma (1:8) was incubated with cytokine membrane arrays containing antibodies at 4˚C overnight according to the manufacturer's instructions. Following washing with appropriate buffers and incubation with secondary antibodies, detection was performed using horseradish peroxidase labeled streptavidin. Raw data were analyzed with the Raybio@ analysis software tool and six cytokines (PDGF-AA, PDGF-AB, PDGF-BB, VEGF-A, TNFR1, TNFR2) selected (based on a minimum of at least a 2.5-fold difference from at least one other cluster) for quantitative confirmatory validation assays. For conformation and protein quantification, a customized ProcartaPlex immunoassay panel (Thermo Fisher Scientific) consisting of PDGF-BB, VEGF-A, TNF-R1 and TNF-R2 was used in combination with human PDGF-AB (EHPDGFAB) and PDGF-AA (EHPDGFAA) ELISAs (Thermo Fisher Scientific) and experiments performed as per manufacturer's instructions. Twenty-five microliters of plasma (in duplicate) was used for ProcartaPlex panel and the plates read using Bio-Plex-200 system (Bio-Rad). The concentration of each cytokine was generated with standard curves using Bio-Plex manager software 6.1. The proportion of patients in each detected COPD cluster used for inflammatory assessments was comparable to that observed in both the derivation and validation cohorts as follows: Ex-tuberculosis (17%), diabetic (19%), low co-morbidity: low-risk (10%), low co-morbidity: high-risk (39%) and cardiovascular (15%).

e-Appendix 3: Statistical analysis

Statistical analysis was performed using RStudio (Version 1.1.453, Integrated Development for R. RStudio, Inc., Boston, MA) and Python (Version 2.7). All continuous variables were not normally distributed; hence all continuous data is presented as medians (with interquartile ranges). To compare differences between groups, the Kruskal-Wallis and Dunn test (with Benjamini-Hochberg correction for false discovery rate) were employed for continuous variables and Fisher exact or Chi-squared test performed for categorical data as appropriate. Significance level is defined as p-values <0.05(*); <0.01(**) and <0.001(***).

Clustering analysis: Clinical variables, including numerical and categorical data were used for patient clustering. These include age, body mass index (BMI), lung function, CAT scores, sex, smoking status (and smoking pack year exposure where relevant), co-morbidities and sputum microbiology. For data transformation, a Gower dissimilarity matrix was calculated using the R function *'daisy'* from the "Cluster package". *'Sammon'*, a non-metric multidimensional scaling (NMDS) algorithm was implemented on this Gower dissimilarity matrix using an appropriate value of 'k' as determined by a Screen plot. Embedding all patients into a Euclidean Space of k=8, Ward's minimum-variance unsupervised hierarchical clustering method was applied on this transformed/embedded dataset using an agglomerative approach with the *'hclust'* function of the "Cluster" Package. The optimal number of derivation clusters was determined with R package "Nbclust". Cluster stabilities were investigated via computing the Jaccard similarities index, with bootstrapping over 100 iterations using the R package "fpc". The mean Jaccard similarities index was 0.79

858 suggesting stability of the identified clusters. Assignment of cluster membership of the
859 validation cohort was carried out using a trained Regularised Discriminant Analysis (RDA)
860 model. The RDA model with a uniform prior was trained on the data-transformed derivation
861 cohort with cluster membership as class labels using "klaR" package in R. Model parameters,
862 lambda and gamma were tuned to an optimal value. To account for the randomness, a nested
863 approach was implemented over 100 iterations with random seeds. At each iteration, optimal
864 model parameters were determined numerically by minimizing the estimated
865 misclassification rate, which is estimated by dividing the data into 70% training and 30%
866 testing over 100 bootstrap iterations. In order to find the high-density region (median) of the
867 optimal model parameters, a kernel density estimation (KDE) of these parameters was
868 generated using the "seaborn" library in python. The median optimal lambda and gamma,
869 chosen from the darkest region of KDE density plot corresponded to gamma of 0.012 and
870 lambda of 0.130 (e-Figure 3). The leave one out cross validation (LOOCV) accuracy of the
871 RDA model with optimal parameters was found to be 97.9%. To validate the derived cluster
872 groups, the validation cohort was data-transformed and the trained RDA model, on 70% of
873 the derivation cohort (training dataset) was used to predict the class membership probabilities
874 of the patients from the validation cohort (e-Table 3). Class labels were assigned based on the
875 maximum probability of a patient being in that class.
876
877 Decision tree: CART (Classification and Regression trees), a supervised classification
878 algorithm was implemented using sklearn[21] in python 2.7, to derive a simplistic model to
879 predict cluster membership using baseline clinical data from the derivation cohort. This
880 decision tree was pruned to make the model simple for clinical application. Using the
881 validation cohort, the accuracy of the decision tree was 72%.
882

Survival analysis: Survival analysis was performed using R packages "survival" and "survminer". All-cause mortality was assessed and compared with Kaplan-Meier curves, log-rank test and cox proportional hazards regression model adjusted for age, sex, BMI, forced expiratory volume in the 1st second (FEV_1) and smoking pack year exposure [22]. Respiratory causes of death were further compared with non-respiratory causes with competing risk analysis using the R package "cr17". Sankey plots were used to visualize the differences between smoking status in each cluster with mortality and severe exacerbation, using "ggalluvial" package in R. Multivariate logistic regression adjusted for age, sex, BMI, FEV_1, smoking pack year exposure, and GOLD group was performed using *'glm'* function in R, a Hosmer and Lemeshow test for the model was assessed using "generalhoslem" package in R.

Inflammatory analyses: Cytokine data was corrected for batch variation with the R "MdimNormn" package [23]. The median values were normalized to percentage differences between the clusters and plotted on a radar chart using the "ggiraphExtra" package in R. Network diagrams of the six cytokines for each respective clinical cluster was generated using "tidygraph" and "ggraph" package in R. Nodes represent the percentage of patients with positive cytokines, the edge (line) illustrates cytokine connectivity and edge attributes (line thickness) represents the proportion of patients demonstrating positive cytokine connections. An elevated cytokine concentration (i.e. the 'cut off') was defined as a value above the detected 95th percentile of non-COPD subjects in accordance with prior published work [24].

908

909

910

911

912

913 **References**

914 1. Divo M, Cote C, de Torres JP, Casanova C, Marin JM, Pinto-Plata V, et al.
915 Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease.
916 Am J Respir Crit Care Med. 2012;186(2):155-61.
917 2. Murray PR, Washington JA. Microscopic and baceriologic analysis of expectorated
918 sputum. Mayo Clin Proc. 1975;50(6):339-44.
919 3. Van Scoy RE. Bacterial sputum cultures. A clinician's viewpoint. Mayo Clin Proc.
920 1977;52(1):39-41.
921 4. GOLD. Global Strategy for the Diagnosis, Management and Prevention of COPD,
922 Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2018. (http://goldcopd.org/);
923 *2018*.
924 5. Diagnostic Standards and Classification of Tuberculosis in Adults and Children. This
925 official statement of the American Thoracic Society and the Centers for Disease Control and
926 Prevention was adopted by the ATS Board of Directors, July 1999. This statement was
927 endorsed by the Council of the Infectious Disease Society of America, September 1999. Am
928 J Respir Crit Care Med. 2000;161(4 Pt 1):1376-95.
929 6. Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B, et al. 2016 ESC
930 Guidelines for the management of atrial fibrillation developed in collaboration with EACTS.
931 European journal of cardio-thoracic surgery : official journal of the European Association for
932 Cardio-thoracic Surgery. 2016;50(5):e1-e88.

7. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur J Heart Fail. 2016;18(8):891-975.

8. Task Force M, Montalescot G, Sechtem U, Achenbach S, Andreotti F, Arden C, et al. 2013 ESC guidelines on the management of stable coronary artery disease: the Task Force on the management of stable coronary artery disease of the European Society of Cardiology. Eur Heart J. 2013;34(38):2949-3003.

9. American Diabetes A. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2018. Diabetes Care. 2018;41(Suppl 1):S13-S27.

10. Schuppan D, Afdhal NH. Liver cirrhosis. Lancet. 2008;371(9615):838-51.

11. Committee ASoP, Banerjee S, Cash BD, Dominitz JA, Baron TH, Anderson MA, et al. The role of endoscopy in the management of patients with peptic ulcer disease. Gastrointest Endosc. 2010;71(4):663-8.

12. Kovacic B, Sehl C, Wilker B, Kamler M, Gulbins E, Becker KA. Glucosylceramide Critically Contributes to the Host Defense of Cystic Fibrosis Lungs. Cell Physiol Biochem. 2017;41(3):1208-18.

13. Sacco RL, Kasner SE, Broderick JP, Caplan LR, Connors JJ, Culebras A, et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2013;44(7):2064-89.

14. Gerhard-Herman MD, Gornik HL, Barrett C, Barshes NR, Corriere MA, Drachman DE, et al. 2016 AHA/ACC Guideline on the Management of Patients With Lower Extremity Peripheral Artery Disease: Executive Summary: A Report of the American College of

Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation. 2017;135(12):e686-e725.

15. Global Initiative for Asthma. Global Management and Prevention 2018. http://www.ginasthma.org.

16. Lewinsohn DM, Leonard MK, LoBue PA, Cohn DL, Daley CL, Desmond E, et al. Official American Thoracic Society/Infectious Diseases Society of America/Centers for Disease Control and Prevention Clinical Practice Guidelines: Diagnosis of Tuberculosis in Adults and Children. Clin Infect Dis. 2017;64(2):111-5.

17. Jung KH, Kim SJ, Shin C, Kim JH. The considerable, often neglected, impact of pulmonary tuberculosis on the prevalence of COPD. Am J Respir Crit Care Med. 2008;178(4):431; author reply 2-3.

18. Murphy PB, Rehal S, Arbane G, Bourke S, Calverley PMA, Crook AM, et al. Effect of Home Noninvasive Ventilation With Oxygen Therapy vs Oxygen Therapy Alone on Hospital Readmission or Death After an Acute COPD Exacerbation: A Randomized Clinical Trial. JAMA. 2017;317(21):2177-86.

19. Galie N, Humbert M, Vachiery JL, Gibbs S, Lang I, Torbicki A, et al. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). Eur Heart J. 2016;37(1):67-119.

20. Compston J, Cooper A, Cooper C, Gittoes N, Gregson C, Harvey N, et al. UK clinical guideline for the prevention and treatment of osteoporosis. Arch Osteoporos. 2017;12(1):43.

21. Fabian Pedregosa GV, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake

983 Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot,
984 Édouard Duchesnay. Scikit-learn: Machine Learning in Python. J Mach Learn Res.
985 2011(12):2825-30.
986 22. Miller J, Edwards LD, Agusti A, Bakke P, Calverley PM, Celli B, et al. Comorbidity,
987 systemic inflammation and outcomes in the ECLIPSE cohort. Respir Med.
988 2013;107(9):1376-84.
989 23. Mac Aogain M, Tiew PY, Lim AYH, Low TB, Tan GL, Hassan T, et al. Distinct
990 'Immuno-Allertypes' of Disease and High Frequencies of Sensitisation in Non-Cystic-
991 Fibrosis Bronchiectasis. Am J Respir Crit Care Med. 2018.
992 24. Agusti A, Edwards LD, Rennard SI, MacNee W, Tal-Singer R, Miller BE, et al.
993 Persistent systemic inflammation is associated with poor clinical outcomes in COPD: a novel
994 phenotype. PLoS One. 2012;7(5):e37483.
995

996 **SUPPLEMENTARY TABLES AND LEGENDS**

997

998 **e-Table 1:** Demographic table showing baseline characteristics of the derivation and

999 validation cohorts respectively.

Characteristics	**Overall n=1480**	**Derivation cohort n=911**	**Validation cohort n=569**
Age (years), Median (IQR)	74 (68-79)	74 (68-79)	74 (69-81)
Country, n (%)			
Singapore	630 (42.6)	406 (44.5)	224 (39.4)
Hong Kong	720 (48.6)	425 (46.7)	292 (51.8)
Malaysia	130 (8.8)	80 (8.8)	50 (8.8)
Sex (Male), n(%)	1396 (94.6)	868 (95.3)	531 (93.3)
Smoking status, n(%)			
Current smoker	690 (46.6)	547 (60.0)	143 (25.1)
Ex- smoker	773 (52.2)	362 (39.8)	411 (72.2)
Non-smoker	17 (1.2)	2 (0.2)	15 (2.7)
Smoking pack years, Median (IQR)	50.0 (35.0-65.5)	50.0 (32.0-67.5)	50.0 (40.0-60.0)
BMI (kg/m2), Median (IQR)	21.0 (18.0-23.6)	20.9 (18.0-23.5)	21.3 (18.1-23.9)
FEV_1 (%predicted), Median (IQR)	45.0 (34.0-60.0)	47.3 (35.3-61.8)	42.0 (33.0-58.0)

FEV_1/FVC (%predicted), Median (IQR)	49.9 (40.5-59.0)	50.0 (41.3-60.0)	47.0 (39.9-56.1)
Total number of exacerbations per year, Median (IQR)	1 (0-2)	1 (0-3)	1 (0-2)
Hospitalized exacerbations, Median (IQR)	0 (0-2)	1 (0-2)	0 (0-1)
COTE Index, Median (IQR)	0 (0-2)	0 (0-2)	1 (0-2)

Data are presented as number of patients (n) (with percentage; %) or median (and interquartile range; IQR). BMI: body mass index; FEV_1: forced expiratory volume in the 1st second; FVC: forced vital capacity.

e-Table 2: Baseline characteristics of the five derived chronic obstructive pulmonary disease (COPD) clusters (from the derivation cohort).

Characteristics	**Overall**	**Ex-tuberculosis**	**Diabetic**	**Low co-morbidity: low-risk**	**Low co-morbidity: high-risk**	**Cardio-vascular**	**p-value**
n (%)	911	156 (17)	109 (12)	192 (21)	339 (37)	115 (13)	
Country, n (%)							<0.001
Singapore	406 (44.5)	73 (46.8)	45 (41.3)	107 (55.7)	101 (29.8)	80 (69.6)	
Hong Kong	425 (46.7)	53 (34.0)	54 (49.5)	50 (26.1)	234 (69.0)	34 (29.5)	
Malaysia	80 (8.8)	30 (19.2)	10 (9.2)	35 (18.2)	4 (1.2)	1 (0.9)	
Age (years), Median (IQR)	74 (68-79)	74 (67-79)	72 (67-79)	74.5 (68-78)	73 (67-79)	73 (68-78)	0.861
Sex (Male), n (%)	868 (95.3)	148 (94.9)	103 (94.5)	182 (94.8)	327 (96.6)	108 (93.9)	0.766

Smoking status, n (%)	542 (59.5)	66 (42.3)	66 (60.6)	15 (7.8)	323 (95.3)	72 (62.6)	<0.001
Current smoker	362 (39.8)	86 (55.1)	43 (39.4)	176 (91.7)	14 (4.1)	43 (37.4)	
Ex- smoker	7 (0.8)	4 (2.6)	0 (0.0)	1 (0.5)	2 (0.6)	0 (0.0)	
Non-smoker							
Smoking pack years, Median (IQR)	50.0 (32.0-67.5)	50.0 (33.9-77.3)	50.0 (29.5-80.0)	51.0 (40.0-79.9)	47.0 (35.0-61.0)	50.0 (30.0-60.0)	0.015
Body mass index (kg/m2), Median (IQR)	20.9 (18.0-23.5)	19.5 (17.3-23.2)	21.4 (19.0-25.0)	21 (18.3-23.5)	20.9 (18.2-23.0)	21 (17.8-23.8)	0.049
FEV1 (% predicted), Median (IQR)	47.3 (35.3-61.8)	46.0 (33.0-61.8)	50.0 (35.8-62.5)	50.6 (38.7-69.0)	44.3 (32.9-57.0)	50.0 (40.6-63.0)	<0.001
FEV1/FVC (% predicted), Median (IQR)	50.0 (41.3-60.0)	49.9 (40.0-58.0)	53.0 (42.8-60.0)	51.0 (45.1-60.0)	50.0 (40.0-58.6)	50.0 (42.0-60.0)	0.007
COPD Assessment test (CAT), Median (IQR)	14 (9-20)	13 (8-21)	13 (8-19)	13 (9-19)	15 (8-20)	16 (10-23)	0.271
Total number of exacerbations per	1 (0-3)	1 (0-4)	1 (0-3)	1 (0-3)	1 (0-2)	2 (0-5)	0.005

year, Median (IQR)							
Hospitalized exacerbations, Median (IQR)	1 (0-2)	1 (0-2)	1 (0-2)	0 (0-2)	0 (0-2)	1 (0-3)	0.003
COTE Index, Median (IQR)	1 (0-2)	0 (0-2)	2 (2-3)	0 (0-1)	0 (0-0)	1 (1-3)	<0.001
Sputum culture positive, n (%)	185 (20.3)	30 (19.2)	24 (22.0)	14 (7.3)	100 (29.5)	17 (14.8)	<0.001
H.influenzae, n (%)	81 (8.9)	14 (9)	9 (8.3)	0 (0.0)	56 (16.5)	2 (1.7)	<0.001
P.aeruginosa, n (%)	43 (4.7)	9 (5.8)	3 (2.8)	4 (2.1)	22 (6.5)	5 (4.3)	0.154
S.pneumoniae, n (%)	33 (3.6)	6 (3.8)	5 (4.6)	8 (4.2)	11 (3.2)	3 (2.6)	0.917
M.catarrhalis, n (%)	24 (2.6)	4 (2.6)	3 (2.8)	2 (1.0)	11 (3.2)	4 (3.5)	0.606
K.pneumoniae, n (%)	13 (1.4)	1 (0.6)	6 (5.5)	1 (0.5)	3 (0.9)	2 (1.7)	0.004
Acinetobacter spp., n (%)	10 (1.1)	2 (1.3)	1 (0.9)	0 (0.0)	6 (1.8)	1 (0.9)	0.451
Fungi, n (%)	9 (1.0)	3 (1.9)	1 (0.9)	0 (0.0)	5 (1.5)	0 (0.0)	0.259

Other bacteria, n (%)	22 (2.4)	1 (0.6)	1 (0.9)	0 (0.0)	15 (4.4)	5 (4.3)	0.004

Data are presented as number of patients (n) (with percentage; %) or median (and interquartile range; IQR). FEV_1: forced expiratory volume in the 1st second; FVC: forced vital capacity, COPD: chronic obstructive pulmonary disease; COTE: COPD specific co-morbidity test. *H. influenzae: Haemophilus influenzae, P.aeruginosa: Pseudomonas aeruginosa, S.pneumoniae: Streptococcus pneumoniae, M.catarrhalis: Moraxella catarrhalis, K.pneumoniae: Klebsiella pneumoniae.*

e-Table 3: Confusion matrix showing the predicted and actual cluster identities (expressed as number of patients, n and percentage; %) generated from the derivation cohort (n= 911) using Regularised Discriminant Analysis (RDA). We note that the Leave One Out Cross Validation (LOOCV) accuracy of the RDA model with optimal parameters was 97.9% (confusion matrix not shown), illustrative of a robust model. This model was used to predict cluster identity for each patient in the validation cohort based on the maximum probability of belonging to each cluster.

Predicted / Actual	Ex-tuberculosis, n (%)	Diabetic, n (%)	Low co-morbidity: low-risk (%)	Low co-morbidity: high-risk, n (%)	Cardiovascular, n (%)	Total number of patients	Accuracy

Ex – tuberculosis, n (%)	156 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	156	99.9%
Diabetic, n (%)	0 (0.0%)	108 (99.1%)	0 (0.0%)	0 (0.0%)	1 (0.9%)	109	99.0%
Low co-morbidity: low-risk, n (%)	0 (0.0%)	0 (0.0%)	191 (99.5%)	1 (0.5%)	0 (0.0%)	192	99.7%
Low co-morbidity: high-risk, n (%)	0 (0.0%)	7 (2.0%)	2 (0.6%)	324 (95.6%)	6 (1.8%)	339	98.2%
Cardiovascular, n (%)	1 (0.9%)	1 (0.9%)	0 (0.0%)	0 (0.0%)	113 (98.2%)	115	99.0%

e-Table 4: Baseline characteristics of the five derived chronic obstructive pulmonary disease (COPD) clusters (from the validation cohort).

Characteristics	Overall	Ex-tuberculosis	Diabetic	Low co-morbidity: low-risk	Low co-morbidity: high-risk	Cardio-vascular	p-value
n (%)	569	102 (17.9)	72 (12.7)	88 (15.5)	193 (33.9)	114 (20.0)	
Country, n (%)							0.281
Singapore	224 (39.4)	39 (38.2)	19 (26.4)	35 (39.8)	80 (41.5)	51 (44.7)	
Hong Kong	295 (51.8)	50 (50.0)	45 (62.5)	45 (51.1)	99 (51.3)	56 (49.1)	
Malaysia	50 (8.8)	13 (12.8)	8 (11.1)	8 (9.1)	14 (7.2)	7 (6.2)	
Age (years), Median (IQR)	74 (69-81)	77 (71-81)	74 (70-80)	75 (70-82)	72 (67-79)	74 (68-81)	0.019
Sex (Male), n (%)	531 (93.3)	90 (88.2)	66 (91.7)	82 (93.2)	185 (95.9)	108 (94.7)	0.147
Smoking status, n (%)							<0.001
Current smoker	143 (25.1)	94 (92.1)	12 (16.7)	83 (94.3)	98 (50.8)	78 (68.4)	
Ex- smoker		1 (1.0)		2 (2.3)	7 (3.6)	3 (2.6)	

Non-smoker	411 (72.2)		58 (80.5)				
	15 (2.7)		2 (2.8)				
Smoking pack years, Median (IQR)	50.0 (40.0-60.0)	50.0 (40.0-60.0)	60.0 (50.0-80.0)	50.0 (40.0-80.0)	50.0 (33.6-60.0)	50.0 (40.0-60.0)	0.006
Body mass index (kg/m2), Median (IQR)	21.3 (18.1-23.9)	19.9 (18.0-25.4)	22.7 (19.7-25.9)	21.8 (19.0-24.2)	20.5 (17.2-23.6)	21.9 (18.6-23.6)	0.315
FEV1 (% predicted), Median (IQR)	42.0 (33.0-58.0)	47.4 (32.0-65.0)	40.0 (31.0-50.2)	39.7 (29.0-48.5)	43.0 (34.0-61.6)	43.0 (31.2-52.1)	0.001
FEV1/FVC (% predicted), Median (IQR)	47.0 (39.9-56.1)	49.0 (40.8-61.5)	43.6 (38.2-54.8)	46.0 (36.1-52.6)	47.0 (40.0-58.1)	48.3 (40.7-56.1)	0.024
Total number of exacerbations per year, Median (IQR)	1 (0-2)	1 (0-2)	1 (0-1)	1 (0-2)	0 (0-1)	1 (0-2)	0.369
Hospitalized exacerbations, Median (IQR)	0 (0-1)	0 (0-1)	0 (0-1)	1 (0-2)	0 (0-1)	0 (0-2)	0.314
COTE Index, Median (IQR)	1 (0-2)	0 (0-2)	2 (2-3)	0 (0-2)	0 (0-2)	2 (1-3)	<0.001

Sputum culture positive, n (%)	89 (15.6)	30 (29.4)	8 (11.1)	7 (8.0)	28 (14.5)	16 (14.0)	<0.001
H.influenzae, n (%)	30 (5.3)	6 (5.9)	1 (1.4)	3 (3.4)	14 (7.3)	6 (5.3)	0.377
P.aeruginosa, n (%)	19 (3.3)	9 (8.8)	2 (2.8)	0 (0.0)	4 (2.1)	4 (3.5)	0.012
S.pneumoniae, n (%)	9 (1.6)	3 (2.9)	0 (0.0)	1 (1.1)	1 (0.5)	4 (3.5)	0.135
M.catarrhalis, n (%)	13 (2.3)	4 (3.9)	3 (4.2)	1 (1.1)	4 (2.1)	1 (0.9)	0.423
K.pneumoniae, n (%)	7 (1.2)	4 (3.9)	1 (1.4)	0 (0.0)	0 (0.0)	2 (1.8)	0.035
Acinetobacter spp., n (%)	6 (1.1)	1 (1.0)	0 (0.0)	1 (1.1)	3 (1.6)	1 (0.9)	0.970
Fungi, n (%)	3 (0.5)	3 (3.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.011
Other bacteria, n (%)	9 (1.6)	3 (3.0)	1 (1.4)	0 (0.0)	4 (2.1)	1 (0.9)	0.535

Data are presented as number of patients (n) (with percentage; %) or median (and interquartile range; IQR). FEV_1: forced expiratory volume in the 1_{st} second; FVC: forced vital capacity, COPD: chronic obstructive pulmonary disease; COTE: COPD specific co-morbidity test. *H. influenzae: Haemophilus influenzae, P.aeruginosa: Pseudomonas*

aeruginosa, S.pneumoniae: Streptococcus pneumoniae, M.catarrhalis: Moraxella catarrhalis, K.pneumoniae: Klebsiella pneumoniae.

e-Table 5: Demographic table showing baseline characteristics and inflammatory (cytokine) assessment of the non-diseased and chronic obstructive pulmonary disease (COPD) cohorts.

Characteristics	**Overall n=360**	**Non-diseased (Healthy) n=24**	**COPD n=336**	**p-value**
Age (years), Median (IQR)	73 (67-79)	64 (59-68)	74 (67-79)	<0.0001
Sex (Male), n (%)	332 (92.2)	9 (38.0)	323 (96.1)	<0.001
Current smoker, n (%)	221 (61.4)	3 (12.5)	218 (64.9)	<0.0001
Smoking pack years, Median (IQR)	50 (40-68)	40 (37-47)	50 (40-70)	0.364
Body mass index (kg/m2), Median (IQR)	21.4 (18.1-23.8)	15.1 (13.1-20.35)	21.4 (18.1-23.8)	0.070
FEV_1 (% predicted), Median (IQR)	47.0 (34.1-61.8)	84.0 (83.0-90.5)	47.0 (34.0-61.0)	<0.0001
FEV_1/FVC (% predicted), Median (IQR)	50.0 (41.4-59.3)	78.0 (78.0-80.0)	49.7 (41.2-59.1)	<0.0001
COPD assessment test (CAT), Median (IQR)	14 (9-20)	NA	14 (9-20)	NA
Hospitalized exacerbations, Median (IQR)	0 (0-2)	NA	0 (0-2)	NA
COTE Index, Median (IQR)	1 (0-2)	NA	1 (0-2)	NA

TNF-R1 (pg/ml), Median (IQR)	1090.3 (597.0-1926.9)	278.7 (147.9-342.5)	1131.3 (721.8-2036.4)	<0.001
TNF-R2 (pg/ml), Median (IQR)	41.4 (15.7-88.7)	4.6 (2.5-8.4)	45.5 (18.8-91.5)	<0.001
VEGF (pg/ml), Median (IQR)	310.0 (40.0-625.2)	3.3 (0.0-33.0)	363.3 (90.07-689.7)	<0.001
PDGF-AB (pg/ml), Median (IQR)	3965.2 (2129.4-6496.8)	2017.0 (994.2-3074.4)	4054.4 (2352.5-7009.3)	<0.001
PDGF-AA (pg/ml), Median (IQR)	1.9 x10^5 (6.3 x10^4 - 4.8x10^5)	2.2 x10^5 (8.6 x10^4- 2.8 x10^5)	1.9x10^5 (6.3x10^4- 5.1x10^5)	0.648
PDGF-BB (pg/ml), Median (IQR)	197.5 (70.6-455.3)	131.0 (43.1-239.9)	204.0 (71.9-459.9)	0.149

Data are presented as number of patients (n) (with percentage; %) or median (and interquartile range; IQR). FEV_1: forced expiratory volume in the 1_{st} second; FVC: forced vital capacity, COPD: chronic obstructive pulmonary disease; COTE: COPD specific co-morbidity test; TNF-R1: tumor necrosis factor receptor 1; TNF-R2: tumor necrosis factor receptor 2; VEGF: vascular endothelial growth factor; PDGF: platelet derived growth factor.

e-Table 6: : Demographic table showing baseline characteristics and inflammatory (cytokine) assessment of the five derived chronic obstructive pulmonary disease (COPD) clusters.

Characteristics	Overall	Ex-tuberculosis	Diabetic	Low co-morbidity: low-risk	Low co-morbidity: high-risk	Cardio-vascular	p-value
n, (%)	336	59 (17)	63 (19)	33 (10)	131 (39)	50 (15)	
Centre, n (%)							
Singapore	74 (22)	12 (20)	7 (11)	4 (12)	33 (25)	18 (36)	0.035
Hong Kong	213 (63)	35(60)	47 (75)	23 (70)	84 (64)	24 (48)	
Malaysia	49 (15)	12 (20)	9 (14)	6 (18)	14 (11)	8 (16)	
Age (years), Median (IQR)	74 (67-79)	74 (69-81)	73 (68-79)	75 (71-80)	73 (65-79)	74 (69-79)	0.728
Sex (Male), n (%)	323 (96.1)	56 (94.9)	59 (93.7)	32 (97.0)	128 (97.7)	48 (96.0)	0.633
Smoking status, n (%)	218 (64.9)	28 (47.5)	45 (71.4)	10 (30.3)	104 (79.4)	31 (62.0)	<0.001
Current smoker	118 (35.1)	31 (52.5)	18 (28.6)	23 (69.7)	27 (20.6)	19 (38.0)	
Ex-smoker							
Smoking pack years, Median (IQR)	50 (40-70)	50 (40-62)	50 (30-80)	60 (48-77)	50 (40-63)	49 (32-73)	0.437
Body mass index (kg/m^2), Median (IQR)	21.4 (18.1-23.8)	20.8 (17.7-24.5)	22.6 (19.6-25.6)	19.8 (17.3-21.9)	21.5 (18.0-23.5)	21.7 (17.9-23.6)	0.039

FEV_1 (% predicted), Median (IQR)	47.0 (34.0-61.0)	42.0 (31.3-63.8)	47.1 (35.3-59.0)	47.0 (36.5-57.5)	48.0 (33.1-61.0)	42.6 (36.3-57.8)	0.997
FEV_1/FVC (% predicted), Median (IQR)	49.7 (41.2-59.0)	46.0 (40.0-60.9)	55.0 (44.0-61.5)	48.5 (42.0-57.7)	48.9 (40.7-58.3)	47.4 (41.0-57.5)	0.140
COPD assessment test (CAT), Median (IQR)	14 (9-20)	14 (9-21)	13 (8-18)	13 (11-22)	13 (8-19)	14 (10-22)	0.785
Total number of exacerbations per year, Median (IQR)	1 (0-2)	1 (0-2)	1 (0-3)	0 (0-2)	1 (0-2)	1 (0-3)	0.609
Hospitalized exacerbations, Median (IQR)	0 (0-2)	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-2)	1 (0-3)	0.724
COTE Index, Median (IQR)	1 (0-2)	0 (0-2)	2 (2-3)	0 (0-1)	0 (0-0)	2 (1-4)	<0.001
TNF-R1 (pg/ml), Median (IQR)	1131.3 (721.8-2036.4)	1175.4 (792.7-2243.1)	1203.9 (708.7-2111.2)	993.94 (730.4-2042.5)	1093.01 (632.6-1827.5)	1378.5 (795.4-2138.1)	0.556
TNF-R2 (pg/ml), Median (IQR)	45.5 (18.8-91.5)	53.8 (16.4-109.5)	43.8 (21.9-90.8)	32.7 (16.8-122.8)	38.2 (15.7-70.0)	83.8 (34.3-151.3)	0.010

VEGF (pg/ml), Median (IQR)	363.3 (90.1-689.7)	550.4 (234.9-815.4)	294.4 (14.5-536.6)	342.5 (83.8-791.7)	272.4 (39.7-722.0)	398.8 (97.5-793.2)	0.077
PDGF-AB (pg/ml), Median (IQR)	4054.3 (2352.5-7009.3)	4471.8 (2832.7-8977.9)	3395.1 (1680.5-5465.1)	4109.5 (1314.4-9210.4)	4595.8 (2843.3-7755.8)	3609.0 (1998.5-5461.9)	0.031
PDGF-AA (pg/ml), Median (IQR)	1.9 X10^5 (6.3 X10^4-5.1X10^5)	1.9 X10^5 (7.7 x10^4 – 5.6 X10^5)	1.0 X10^5 (3.7 X10^4- 4.7 X10^5)	2.3 X10^5 (5.1 X10^4-5.0X10^5)	1.6 X10^5 (5.3 X10^4 – 3.9 X10^5)	2.4 X10^5 (1.3 X10^5 -5.7X10^5)	0.037
PDGF-BB (pg/ml), Median (IQR)	204.0 (71.9-459.9)	158.2 (82.3-403.9)	218.4 (89.1-511.1)	106.5 (42.2-817.6)	259.1 (121.2-460.2)	85.9 (21.9-332.6)	0.026

Data are presented as number of patients (n) (with percentage; %) or median (and interquartile range; IQR). FEV_1: forced expiratory volume in the 1_{st} second; FVC: forced vital capacity, COPD: chronic obstructive pulmonary disease; COTE: COPD specific co-morbidity test; TNF-R1: tumor necrosis factor receptor 1; TNF-R2: tumor necrosis factor receptor 2; VEGF: vascular endothelial growth factor; PDGF: platelet derived growth factor.

e-Figure 1: Derivation (D; n=911) and validation cohorts (V; n=569) of patients with Chronic Obstructive Pulmonary Disease (COPD) of Chinese ethnicity illustrate comparable comorbidity and sputum microbiology profiles. Bubble charts illustrating the proportion of patients demonstrating an established indicated comorbidity between (a) the overall cohorts, (b) based on country of origin and (c) sputum microbiology. Bubble size corresponds to the percentage of patients demonstrating each comorbidity and detectable microorganisms by sputum culture within their respective cohort and bubble colour represents the country of patient origin: light green: derivation (overall), dark green: validation (overall), red: Singapore, blue: Hong Kong, purple: Malaysia. PUD: peptic ulcer disease, pTB: history of prior pulmonary tuberculosis, PAD: peripheral arterial disease, Other Ca: all other malignancies excluding lung, esophageal, pancreatic or breast carcinoma, DM: diabetes mellitus, CVA: cerebrovascular disease, CKD: chronic kidney disease, CHF: congestive heart failure, CAD: coronary artery disease, Ca: lung, esophageal, pancreatic or breast carcinoma, AF: atrial fibrillation. H. influenzae: *Haemophilus influenzae,* P.aeruginosa: *Pseudomonas aeruginosa,* S.pneumoniae: *Streptococcus pneumoniae,* M.catarrhalis: *Moraxella catarrhalis*, K.pneumoniae: *Klebsiella pneumoniae.*

e-Figure 2: Two-year mortality and exacerbation frequency of each identified clinical cluster in both derivation and validation cohorts are not influenced by smoking status. Sankey diagram illustrating mortality and exacerbation frequency between identified clinical clusters partitioned by smoking status in derivation (a and b) and validation (c and d) cohorts. Horizontal flow colors indicate the five clinical clusters: Ex-tuberculosis (green), diabetic (blue), low co-morbidity: low-risk (violet red), low co-morbidity: high-risk (yellow) and cardiovascular (pink). Width of flow indicates the proportion of patients within each group.

SE: severe exacerbator, Ex-TB: ex-tuberculosis, LCLR: low co-morbidity: low-risk. LCHR: low co-morbidity: high-risk, CVS: cardiovascular

e-Figure 3: Contour plot to determine optimal model parameter selection for Regularised Discriminant Analysis (RDA) model used for determination of validation cluster membership. Kernel density estimation (KDE) is represented in a density plot with marginal distributions of gamma and lambda on the x- and y-axes respectively. Contour colour corresponds to the probability where the optimal gamma and lambda lie (dark blue: highest probability, light blue: lowest probability). The optimal gamma and lambda of the Regularised Discriminant Analysis (RDA) model were chosen from the darkest region of the KDE plot and corresponds to gamma=0.012 and lambda=0.130.

e-Figure 4: Decision tree generated via Classification and Regression trees (CART) for classification of a Chinese COPD patient into one of the derived clusters. Previous pulmonary tuberculosis was defined as a prior history of documented tuberculosis with positive sputum analysis for *mycobacteria tuberculosis* including positive acid-fast bacilli smear, culture or nucleic acid amplification, radiological features (on chest radiography or computed tomography) and/or prior pharmacological treatment for pulmonary tuberculosis. Diabetes mellitus was defined as the presence of either fasting blood glucose levels of ≥7.0 mmol/l, blood glucose levels of ≥11.1 mmol/l two-hour post oral glucose tolerance test, random blood glucose level ≥11.1 mmol/l with hyperglycemia symptoms, or HbA1c ≥6.5%. Coronary artery disease was defined based on functional testing, radiological imaging and/or coronary angiography. Ex-TB: ex-tuberculosis, LCLR: low co-morbidity: low-risk, LCHR: low co-morbidity: high-risk, CVS: cardiovascular.

e-Figure 5: Increased systemic TNF-R2 associates with significantly greater symptoms (by CAT score) and severe exacerbations. Median values (grey line) are illustrated. Dot color indicates patient membership to their respective clinical cluster: Ex-tuberculosis (green), diabetic (blue), low co-morbidity: low-risk (violet red), low co-morbidity: high-risk (yellow) and cardiovascular (pink). CAT: COPD assessment test, SE: severe exacerbator. *$p \leq 0.05$.

e-Figure 6: Significantly higher mortality is detected in the cardiovascular followed by ex-tuberculosis, diabetic and low comorbidity high-risk cluster compared to the low comorbidity low-risk cluster after adjustment for age, sex, BMI, smoking pack year exposure, lung function (by FEV_1) and GOLD group illustrated by forest plot with multivariate hazard regression. The dot represents the hazard ratio with colour indicating significance levels: red ($p<0.05$). Error bar indicates the 95% confidence interval (CI). Ex-TB: ex-tuberculosis, LCLR: low co-morbidity: low-risk, LCHR: low co-morbidity: high-risk, CVS: cardiovascular.

Figure E1

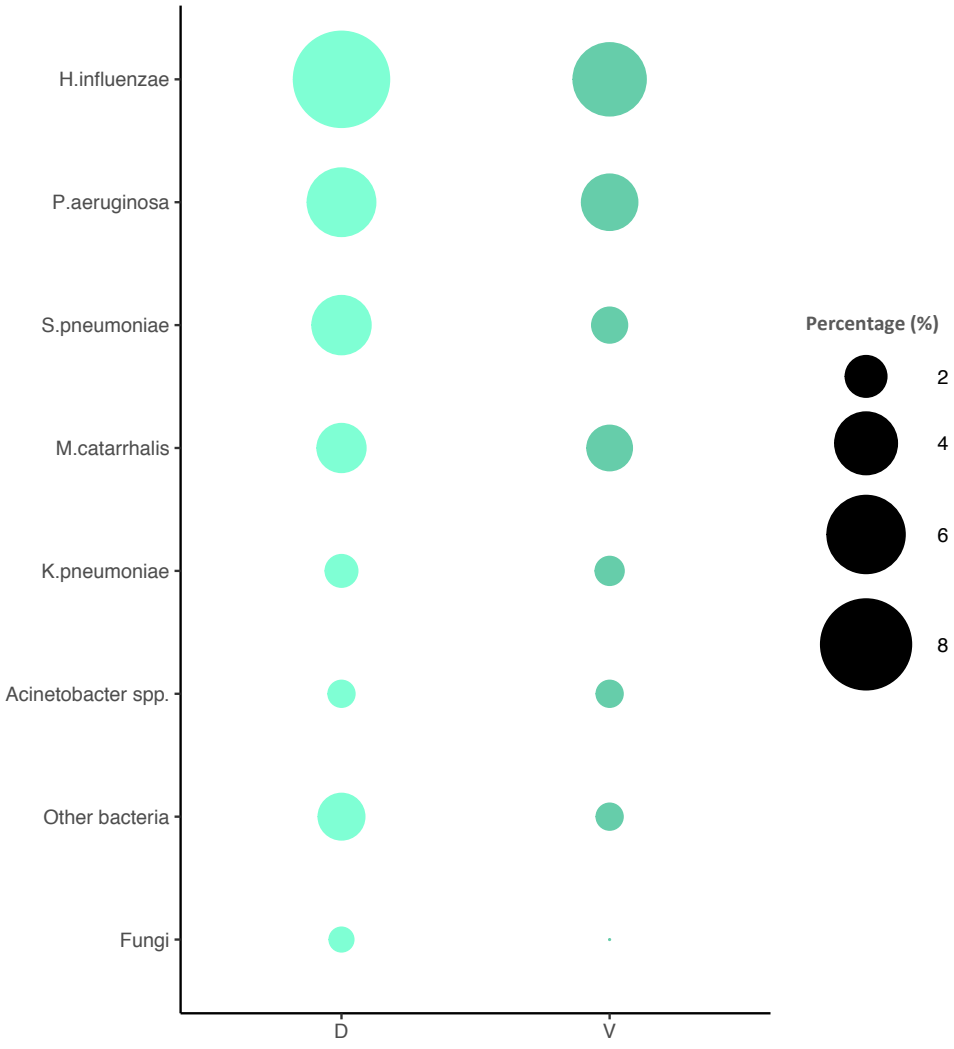

H.influenzae
P.aeruginosa
S.pneumoniae
M.catarrhalis
K.pneumoniae
Acinetobacter spp.
Other bacteria
Fungi
D
V
Percentage (%)
2
4
6
8

Figure E2

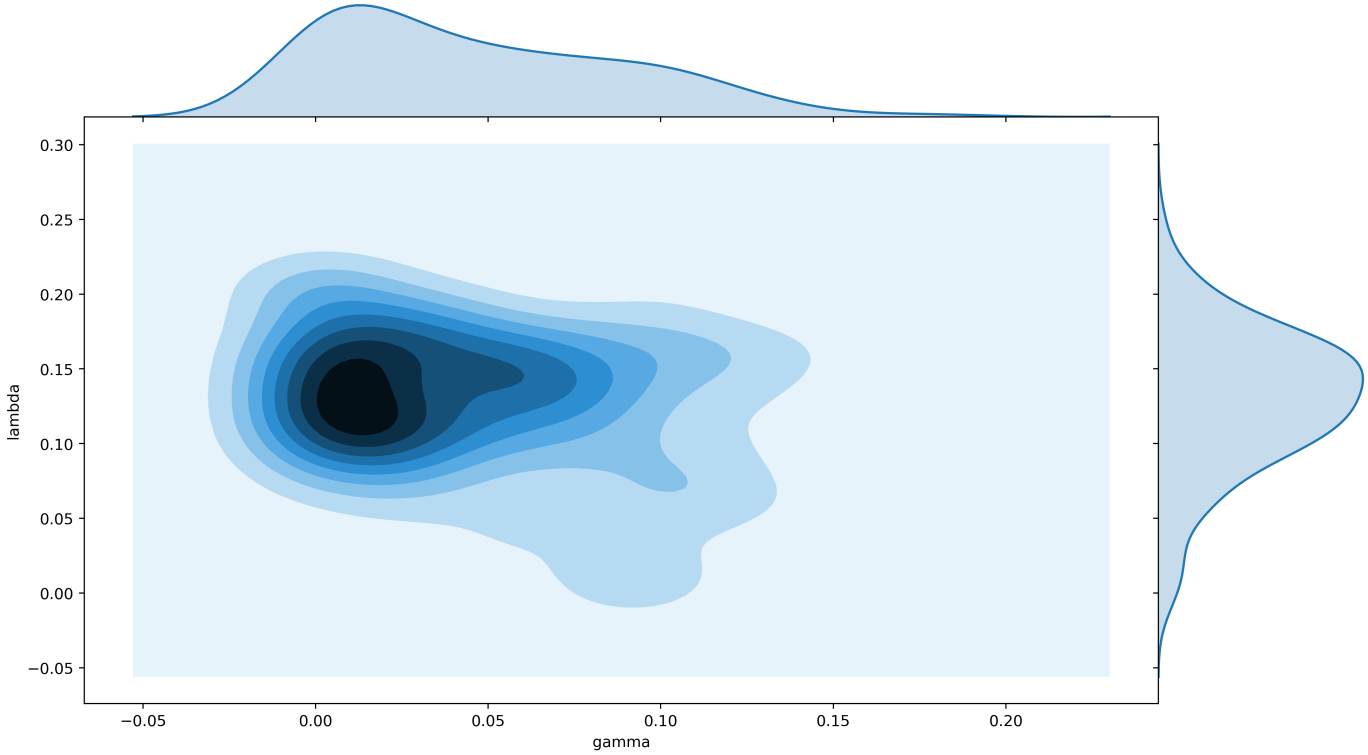

0.30
0.25
0.20
0.15
0.10
0.05
0.00
−0.05
lambda
−0.05
0.00
0.05
0.10
0.15
0.20
gamma

Figure E3

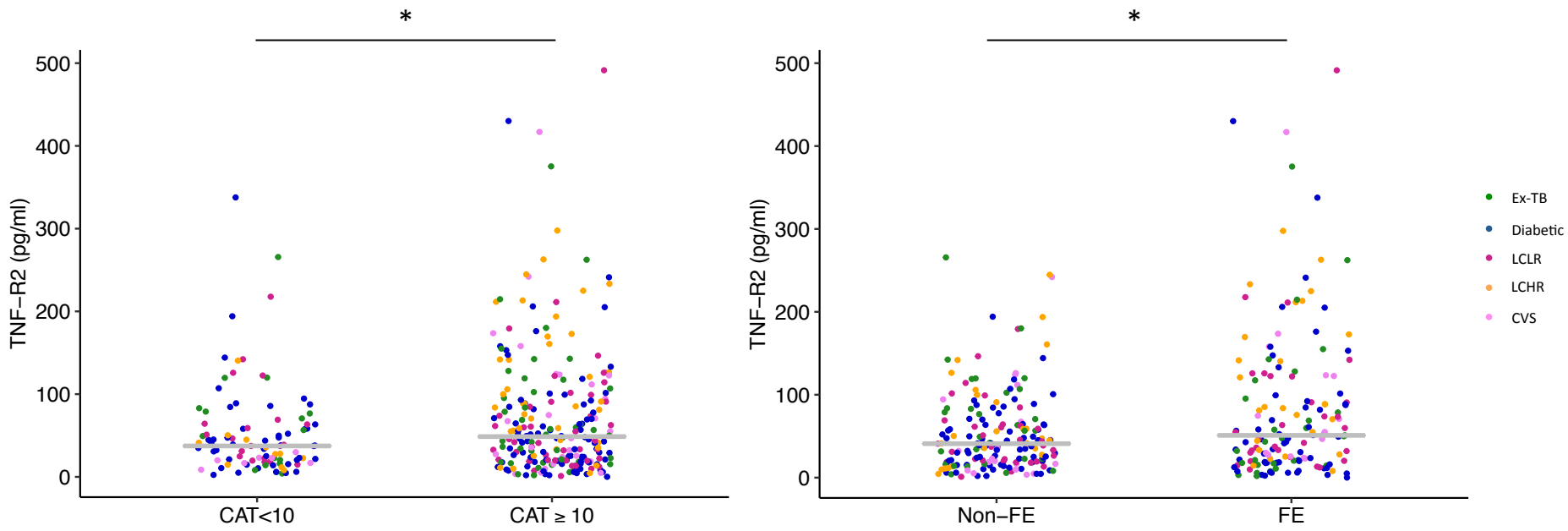
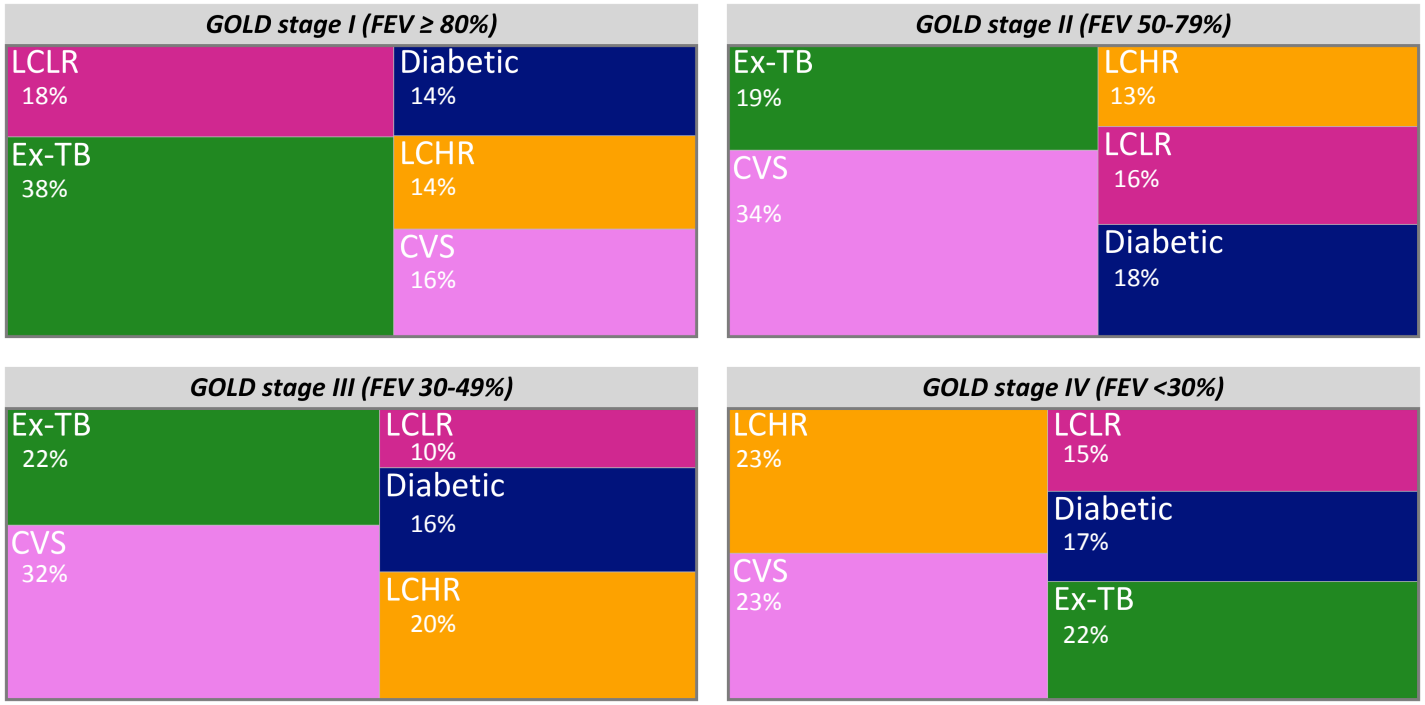

*
500
400
300
200
100
0
TNF–R2 (pg/ml)
CAT<10
CAT ≥ 10
*
Non–FE
FE
Ex-TB
Diabetic
LCLR
LCHR
CVS

Figure E4


GOLD stage I (FEV ≥ 80%)
LCLR 18%
Diabetic 14%
Ex-TB 38%
LCHR 14%
CVS 16%
GOLD stage II (FEV 50-79%)
Ex-TB 19%
LCHR 13%
CVS 34%
LCLR 16%
Diabetic 18%
GOLD stage III (FEV 30-49%)
Ex-TB 22%
LCLR 10%
Diabetic 16%
CVS 32%
LCHR 20%
GOLD stage IV (FEV <30%)
LCHR 23%
LCLR 15%
Diabetic 17%
CVS 23%
Ex-TB 22%