HHS Public Access

Author manuscript

J Occup Environ Med. Author manuscript; available in PMC 2025 February 01.

Published in final edited form as:

J Occup Environ Med. 2024 February 01; 66(2): 179–184. doi:10.1097/JOM.0000000000003023.

Cluster analysis of World Trade Center (WTC) related lower airway diseases (LAD)

Rafael E. de la Hoz[1], **Yunho Jeon**[1], **John T. Doucette**[2], **Anthony P. Reeves**[3], **Raúl San José Estépar**[4], **Juan C. Celedón**[5]

[1]Divisions of Occupational and Environmental Medicine, New York, NY, USA

[2]Biostatistics, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[3]School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA

[4]Department of Radiology, Brigham and Women's Hospital, Boston, MA, USA

[5]Division of Pediatric Pulmonary Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

Abstract

Introduction: Cluster analysis can classify without *a priori* assumptions the heterogeneous chronic lower airway diseases (LAD) found in former workers at the World Trade Center (WTC) disaster site.

Methods: We selected the first available chest CT scan with quantitative CT (QCT) measurements on 311 former WTC workers with complete clinical, and spirometric data from their closest surveillance visit. We performed a non-hierarchical iterative algorithm K-prototype cluster analysis, using gap measure.

Results: A 5-cluster solution was most satisfactory. Cluster 5 had the healthiest individuals. In cluster 4 smoking was most prevalent and intense but there was scant evidence of respiratory disease. Cluster 3 had symptomatic subjects with reduced forced vital capacity impairment (low FVC). Clusters 1 and 2 had less dyspneic subjects, but more functional and QCT evidence of chronic obstructive pulmonary disease (COPD) in cluster 1, or low FVC in cluster 2. Clusters 1 and 4 had the highest proportion of rapid FEV_1 decliners.

Conclusion: Cluster analysis confirms low FVC and COPD/pre-COPD as distinctive chronic LAD phenotypes on long term surveillance of the WTC workers.

Keywords

Occupational lung disease; smoke inhalation injury; Chronic obstructive pulmonary disease; World Trade Center Attack, 2001; longitudinal changes in lung function; Spirometry

Corresponding author: Rafael E. de la Hoz, MD, MPH, MSc, Division of Occupational and Environmental Medicine, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, WTC HP CCE Box 1059, New York, NY 10029, United States of America, Telephone: (212) 241-7996, Facsimile: (212)-241-5516, Rafael.delaHoz@mssm.edu.

Authors' contributions:

RED, YJ, and JCC designed and oversaw the study and selected analytical strategies. APR and RSJ performed all quantitative CT measurements. YJ and JTD performed all statistical analyses. All authors contributed to writing, reviewed and revised the drafts, and approved the final manuscript.

Introduction

Among workers engaged in rescue and recovery efforts, occupational exposures to dust and fumes at the World Trade Center (WTC) disaster site have been associated with lower respiratory symptoms (1), worse lung function (2) and accelerated longitudinal lung function decline trajectories (3, 4), asthma (5) and other chronic airway diseases (6–8), and both qualitative (9) and quantitative (4, 8, 10–12) abnormalities on computed tomography of the chest. Common co-existing morbidities such as obesity are linked to greater severity of WTC-related respiratory symptoms, airway disorders, and morbidity (13–19).

The heterogeneity of the WTC-related lower airway diseases (LAD) poses a challenge for research studies and clinical care. Although stringent case definitions have been used to address this issue in clinical research studies (6), cluster analysis can be particularly helpful in classifying heterogeneous WTC-related chronic LAD (20) into groups without any *a priori* assumptions. To achieve this goal, cluster analysis examines multiple clinical, physiologic and imaging characteristics, including novel quantitative chest CT (QCT) features.

Identifying categories or clusters of WTC-related chronic LAD could help us better study and understand their pathogenesis, which should ultimately lead to improved diagnosis and management. We sought to, first, identify clusters of WTC-related chronic LAD and then examine their association with longitudinal lung function trajectories (4, 10) as clinical outcomes, in a cohort of WTC workers with QCT imaging data.

Methods

Subject recruitment and study procedures

All study subjects were members of the Mount Sinai WTC General Responders' Cohort (GRC) (21) who had chest CT images and quantitative chest CT data in the WTC Chest CT Imaging Archive (9) and who also had complete clinical and spirometric data from the surveillance visit closest to the CT scan. The procedures conducted during the surveillance visits (22) and the chest CT scanning procedures (4, 9) have been previously described in detail, and we selected the first available chest CT scan with QCT measurements. As in our previous studies, all QCT measurements were performed with the Chest Imaging Platform (https://chestimagingplatform.org) or the SIMBA Image Management and Analysis System (http://www.via.cornell.edu/visionx/simba/) by their respective developers (RSJ and APR) (4, 8, 10–12).

We selected spirometries that met our quality selection requirement, namely forced exhalatory time>6 seconds and overall spirometry (reproducibility) grades A, B, or (if at least five trials were available) C (2, 8, 23). As previously reported (23), our spirometry program quality assurance included both daily inspiratory and expiratory volume calibration checks and weekly expiratory linearity checks.

For the cluster analysis, we examined twenty one characteristics, including those in the following groups: (1) Self-reported demographic information such as age on 11-

September-2001, sex, and ethnicity/race (grouped into Latino of any race, non-Latino White, and all other races); (2) Smoking status (never, former, and current) and intensity of smoking (in pack-years), as previously defined (2, 24); (3) self-reported indicators of occupational exposures at the WTC site, such as early arrival (within 48 hours) at the site (25) and cumulative exposure duration (2, 6). From the periodic surveillance visit closest to the chest CT scan we included: (4) Clinically significant dyspnea (defined as a modified Medical Research Council [mMRC] score ≥2 [“I walk slower than people of my same age on the level because of breathlessness, or have to stop for breath when walking at my own pace on the level”] (26)); (5) Spirometric pattern, classified as normal, low forced vital capacity (FVC<lower limit of normal, LLN, with normal ratio of first-second forced expiratory volume to FVC, FEV_1/FVC), or obstructive (FEV_1/FVC<LLN) (2, 27). We also considered (6) Bronchodilator response at any visit (BDRany), most often the baseline, defined as an increase in either FEV_1 or FVC of at least 12% and 200 ml in response to inhaled albuterol administration (10)); (7) body mass index (BMI) in kg/m^2 measured at the baseline visit; (8) Weight gain (as indicated by longitudinal measured BMI trajectory (BMIslope) in kg/m^2/year (10)). Lastly, from the quantitative chest CT data: (9) Total lung volume (TLV_{CT}), with calculated predicted values (28) and assessed by z-scores, (10) lung attenuation volume percent, either high (from −600 to −250 Hounsfield units, HAV%, with >10% as cut point) and low (less than −950 Hounsfield units, LAV%, with >2.5% as cut point) (4, 29)), indirect distal airway measurements such as (11) air trapping at −856 Hounsfield units ($AT_{EXP}856$) (8, 29), and (12) mean expiratory:inspiratory lung density ratio (MLD_{EI}) (8), and proximal airway metrics such as (13) wall area percent, WAP) directly measured on the 3rd generation of the right upper lobe (8, 10, 29), and (14) Pi10, the average wall thickness of a hypothetical airway with a 10-mm luminal perimeter on CT (29, 30).

As in previous studies of longitudinal lung function trajectories among participants in the Mount Sinai WTC GRC cohort (4, 10), we conducted a secondary analysis of the FEV_1 trajectories for the resulting clusters. As in previous work (4), we estimated the longitudinal FEV_1slopes for each subject with at least 3 good quality (as defined above) spirometries performed between July 2002 and December 2018, and used them to define three trajectories as follows: rapid FEV_1 decline (“rapid decliner”) by an FEV_1slope<−66.5 ml/year (i.e., less than the group mean-0.5 SD), FEV_1 gain (“gainer”) by an FEV_1slope>0 ml/year, and intermediate FEV_1 decline (“intermediate decliner”) by an FEV_1slope between 0 and −66.5 ml/year (i.e., between 0 and the group mean-0.5 SD).

The Mount Sinai Program for the Protection of Human Subjects approved this study (HS 12–00925). We adhered to the STROBE guidelines for cross-sectional studies (see supplementary digital content).

Statistical analysis

Given our mix of continuous and categorical variables, we chose a non-hierarchical iterative method, the K-prototype clustering algorithm. This algorithm integrates the K-means, for continuous variables, and K-modes, for categorical variables, based on partitioning a set of subjects into homogeneous clusters (31–33). The K-prototype algorithm is available in the KCLUS procedure in SAS Viya (SAS Institute, Cary, NC), a cloud-enabled, in-memory

analytics engine that provides quick, accurate and reliable analytical insights. All continuous variable measurements were standardized using z-scores. The optimal number of clusters was determined by the first peak of the gap statistic (34) on the K-prototype clustering analysis. The gap measure is obtained by subtracting the logarithm of the within-cluster sum of squares error from the logarithm of its expectation for clustering solutions over a range of possible number of clusters. Differences between clusters were compared using Pearson's chi-squared test or analysis of variance (ANOVA), as appropriate. For the secondary analysis comparing FEV_1 trajectory groups (rapid and intermediate decline, and gain) among the resulting clusters, we used Pearson's chi-squared test. For the comparison of the subjects included vs. those excluded from the study, we utilized standardized differences (StD) (35). For dichotomous variables, average differences between proportions expressed in standard deviation units were calculated. For skewed distribution continuous variables (e.g., smoking intensity and WTC exposure duration), we used a rank-based method to calculate StD. For categorical variables with k levels (e.g., ethnicity/race and smoking status), we used a multivariate Mahalanobis distance method to generalize the standardized difference metric to handle a multinomial sample. StD values of 0.2–0.5 are generally considered small. For all other statistical tests, a p-value<0.05 was considered significant. All analyses were conducted in SAS Viya 3.5 and SAS version 9.4 (SAS Institute, Cary, NC). For graphical purposes, the results were displayed as a heat map, using Microsoft Office Excel version 365, and conditional formatting with color scaling.

Results

From the WTC Chest CT Imaging Archive cohort (n=1,630), 311 subjects had all quantitative chest CT measurements, as well as complete clinical and spirometric data from the surveillance visit closest to the CT scan. None of the standardized differences (StD) in the comparison of variables between the subjects included in this study, and those with incomplete or missing data and thus excluded (n=1319, see Table OS1) suggested an important difference in age, baseline BMI, smoking status, early WTC exposure arrival and cumulative duration, prevalence of dyspnea or evidence of bronchodilator response. Included subjects' smoking intensity was slightly higher than among excluded subjects (7.8 SD 11.6 vs. 6.3 SD 12.8 pack-years, StD 0.22).

Table 1 summarizes the main characteristics of the 311 study participants included in the study and the results of the cluster analysis. Chest CT scans were performed 6.65 (standard deviation or SD 1.87) years and nearest spirometries 6.43 (SD 2.12) years after 9/11/2001. Consistent with the characteristics of the WTC General Responders Cohort (2), the participants' mean age (SD) was 42.9 (8.5) years on September 11, 2001, most were male (~84%) and nearly half (49.8%) arrived at the WTC disaster site within 48 hours (early).

The cluster analysis showed that two, three, and five clusters generated the highest gap statistic (1.21 to 1.26), but that a five-cluster solution seemed most satisfactory. Except for HAV%, all variables were significantly different among the five clusters (see Table 1), which are displayed graphically in the heat map of Figure 1. Clusters 3 and 5 had a lower proportion of male participants (~68% to ~72%) than that in other clusters (ranging from

~86% to ~95%). In all clusters, the mean baseline BMI was ≥27.7 kg/m^2, with cluster 3 having the highest mean value (31.3 kg/m^2). On the other hand, cluster 3 was the only one with a slight weight loss instead of weight gain over time. Regarding smoking status, former and current smoking were most frequent in clusters 1 (72.4%) and 4 (84.9%) and least so in clusters 2 (30.5%) and 5 (28.4%). Consistent with such smoking status, clusters 1 (15.8 pack-years) and 4 (14.1 pack-years) had the highest smoking intensity, while clusters 2 (3.6 pack-years) and 5 (2.0 pack-years) had the lowest smoking intensity. While early arrival to the WTC disaster site was most commonly reported among subjects in clusters 2 (96.3%) and 3 (72.4%) (2), subjects in cluster 3, 4, and 5 had longer cumulative WTC occupational exposure than those in clusters 1 and 2. With regard to spirometric patterns, airflow obstruction was most common in cluster 1 (65.5% of subjects), and a low FVC pattern prevailed in clusters 2 and 3 (respectively 54.9% and 75% of subjects). Conversely, a normal spirometry largely predominated in subjects in cluster 4 and 5 (93% and 87.8%, respectively), with relatively infrequent BDR (11.6 and 8%, respectively).

Based on the observed clinical, spirometric, and imaging characteristics, the five clusters could be defined as follows (Figure 1):

1. Cluster 1 – Pauci-symptomatic with high prevalence of airflow obstruction: slightly older age and including predominantly (~69%) non-Hispanic white subjects; frequent former and current smoking; relatively common late arrival at WTC site; low frequency of clinically significant dyspnea; lowest mean FEV_1%predicted with both airflow obstruction and bronchodilator responsiveness most commonly observed; high TLV_{CT}, WAP, AT856, MLD_{EI}, and Pi10; and lowest frequency of high HAV%. Most of these subjects were diagnosed clinically with COPD (6) or pre-COPD (36).
2. Cluster 2 – Pauci-symptomatic with intermediate prevalence of a low FVC pattern: predominantly never smoking, male and non-Latino subjects, most of whom arrived early at the WTC site; low frequency of dyspnea; slightly reduced mean FEV_1%predicted; and lowest HAV% without other positive QCT markers.
3. Cluster 3 – Symptomatic with high prevalence of a low FVC pattern: highest proportion of women and Latinas(os); highest baseline BMI but no weight gain; dyspnea in all subjects with predominance of a low FVC pattern, slightly reduced mean FEV_1%predicted and relatively common bronchodilator responsiveness (second only that of subjects in cluster 1); frequent early arrival and longer stay at the WTC disaster site. Most subjects in this group were diagnosed clinically as having either chronic nonspecific bronchitis or asthma (6, 7).
4. Cluster 4 – Non-susceptible smokers: slightly older and predominantly male non-Latino white subjects; highest proportion of former and current smokers; uncommon early arrival but high cumulative exposure to the WTC site; intermediate proportion of dyspnea with highest mean FEV_1%predicted and predominantly normal spirometry pattern with rarely observed bronchodilator responsiveness; high TLV_{CT} and highest Pi10.

5. Cluster 5 – Relatively healthy subjects: slightly younger and predominantly never smoking Latino subjects with highest average weight gain; very infrequent early arrival but highest cumulative exposure to the WTC disaster site; dyspnea in a quarter of subjects; high mean FEV_1%predicted with very frequent normal spirometry pattern and very rare bronchodilator responsiveness; and average or fairly normal QCT markers.

Of note, 39.2% of the Non-Latino White subjects were Polish workers, almost all of them ever-smokers[37] and, not surprisingly, mostly in clusters 1 (31%, 9 of 29) or cluster 4 (54.8%, or 46 of 84).

Lastly, we analyzed FEV_1 longitudinal trajectory groups as a clinical outcome for the five clusters (Table 2). For this longitudinal analysis, we had a mean of 5.0 (SD 1.7) good quality spirometries in each of 256 (82.3%) of our subjects. In this analysis, the highest proportion of subjects with rapid decline in FEV_1 was in cluster 1, while the highest proportion of subjects with intermediate decline in FEV_1 was in clusters 2 and 3. Although cluster 1 had the highest proportion of subjects with gain in FEV_1, this was based on a low count. Of note, subjects in cluster 4, who had one of the highest mean FEV_1%predicted and predominance of normal spirometry, but also the highest Pi10, had a proportion of subjects with rapid decline in FEV_1 only second to that of subjects in cluster 1.

Discussion

Cluster analysis confirms a low FVC pattern (clusters 2 and 3) and chronic airflow obstruction (suggestive or consistent with COPD, cluster 1) (8, 2, 48) as distinctive chronic LAD phenotypes in WTC responders, while also supporting the presence of two distinct subgroups with unique clinical characteristics and longitudinal trajectories, seemingly non-susceptible smokers in cluster 4 and relatively healthy subjects in cluster 5. Our analysis also illustrated the adverse cross-sectional and longitudinal lung function effects of tobacco smoking, obesity and longitudinal weight gain trends.

Based on established diagnostic criteria, we previously noted the heterogeneity of the LAD among WTC responders (6), while most other studies have been based on symptom surveys (1, 3, 38), self-reported diagnoses (5, 39), or single lung function measurements (40, 41). In contrast, cluster analysis can integrate multiple variables, does not rely on *a priori* diagnostic assumptions and has been used to identify distinct phenotypes in heterogeneous diseases such as asthma (42–44), COPD (45), and obstructive sleep apnea (46, 47). To our knowledge, cluster analysis has not been previously conducted in the WTC cohorts, or in other inhalational disaster situations.

The functional and QCT characteristics of COPD predictably identified individuals that fit that disease phenotype in our population, as we demonstrated in a previous studies (8). COPD in these workers has been mild to moderate (48), with close to a third of affected subjects being never smokers (48). Our data in these workers with COPD thus far appear to suggest a predominance of the airway- over an emphysema-predominant subtype, as only 4% and 9.6%, respectively, had QCT evidence of emphysema, as indicated by low

attenuation volume percent (LAV%) of ≥5% and ≥2.5%, respectively (8, 10), and half of the cases having functional features of asthma COPD overlap (48).

Low FVC (equivalent to a preserved ratio impaired spirometry, PRiSM) (49) has long been described as the most frequent abnormal spirometric pattern in the WTC occupational cohorts (1, 2, 6, 50). This airway disease pattern is likely heterogeneous (49), and 86.2% (75/87) of our subjects with it were in clusters 2 and 3. Unlike other cohorts (e.g., COPDGene®), where subtypes of low FVC impairment have been recently identified (51), the WTC cohort was unselected for cigarette smoking. Our low FVC clusters 2 and 3 showed the strongest association (2) with early arrival at the WTC disaster site, when environmental exposures were likely highest (25). Low FVC subjects in cluster 3 were more dyspneic, obese, and also showed slightly more restriction (based on TLV_{CT}), and QCT evidence of proximal airway inflammation (WAP), the latter most closely matching (as in our previous study (8)) that of COPD subjects. In contrast, cluster 2 subjects had the highest proportion of never smokers (similar to the relatively healthy subjects of cluster 5) and were less dyspneic. Importantly perhaps for our long-term surveillance, clusters 2 and 3 had lower proportions of subjects with rapid decline in lung function.

Our overall findings confirm some previously identified functional and imaging traits associated with accelerated lung function decline in the WTC cohort, such as BDR, WAP (10), and LAV% (4). Consistent with our previous study (37), Latinos (predominantly laborers) tended to be slightly younger, never smoking, and late arrivals at the disaster site, and have less functional and imaging evidence of LAD (cluster 5). Of note, cluster 4 had a high proportion of subjects with rapid lung function decline (who may thus be at high risk for COPD) and Pi10 as the only significantly associated QCT marker. Given those results with cluster 4, Pi10, as well as other airway imaging metrics, deserve further investigation in the WTC cohort, as studies have shown its association with persistent chronic bronchitis (52), and possibly with more adverse but possibly preventable long-term functional outcomes. Cluster 4 also showed weight gain, and our previous studies had suggested the association of obesity (12) and weight gain (10) with quantitative CT markers of airway inflammation, and accelerated longitudinal expiratory flow decline, respectively. While cluster 4 had the highest proportion of current and former smokers, our studies have shown the remarkable success of smoking cessation interventions in this cohort (2, 8).

Our analysis benefited from an abundance of cross-sectional clinical, functional and respiratory structural data, and longitudinal functional trajectories to assess the potential adverse clinical impact of the identified clusters on respiratory health, in a population not selected for smoking as has been the case with other respiratory cohorts. Limitations on the other hand include the duration of the longitudinal follow up, which may have limited the detection of more severe lung diseases, such as COPD and interstitial lung diseases. Surveillance is, however, ongoing, and should provide additional information in the future. Sample size may have been limiting for only one of 21 variables examined, but the number of cases of low FVC and subjects with apparently normal spirometries allowed cluster analysis to suggest underlying pathophysiological features and indicators to investigate further as the surveillance continues. We used a sample of complete data cases

but our comparison of included vs. excluded subjects suggests that our study sample is quite representative of the WTC Chest Imaging Archive cohort.

In conclusion, we identified five clusters of chronic lower respiratory disease in WTC responders, which have basic functional and QCT characteristics and clinical relevance for the treatment and longitudinal follow up of this population. Although much has been published about the management of disease types such as COPD and asthma, investigations like this shed light on the characteristics and evolution of other clusters such as our clusters 2, 3, and 4. Additional investigations are warranted of those subtypes, for further characterization of their clinical features, identification of additional QCT markers of early lung injury (and, possibly, repair) and longitudinal lung function trajectory.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to thank all participants in this study, and the staff of the Mount Sinai WTC Health Program Clinical Center of Excellence, and the WTC General Responders Cohort Data Center. Jill K. Gregory, MFA, CMI, contributed the illustration presented in Figure 1. The contents of this article are the sole responsibility of the authors and do not necessarily represent the official views of the CDC/NIOSH. Interim results of this work were presented as an abstract at the 2021 International Congress of the European Respiratory Society (Eur Respir J 2021;58 (Suppl 65):PA3350, doi: 10.1183/13993003.congress-2021.PA3350). We adhered to the STROBE guidelines for cross-sectional studies (see supplementary digital content).

Funding:

This work was supported by cooperative agreement No. U01 OH010401 (specific aim 1, RED, PI; YJ, JTD, RSJ, and JCC co-Is, APR co-I in 2012-2016), and contract 200-2017-93325 (WTC General Responders Cohort Data Center) from the Centers for Disease Control and Prevention/National Institute for Occupational Safety and Health (CDC/NIOSH).

COI disclosures:

RED received honoraria from the European Respiratory Society. APR received royalty payment from Cornell Center for Technology Licensing for patients licensed to General Electric (GE), and has stock options in and is a member of the Scientific Advisory Board of HeartLung Technologies. RSJ has a contract to serve as Image Core for studies sponsored by Lung Biotechnology and Insmed, respectively, has a sponsored research agreement with Boehringer Ingelheim, receives consulting fees from Leuko Labs and Icahn School of Medicine at Mount Sinai, has three patients pending in the space of lung cancer risk assessment using machine learning technology, and is co-founder and stock holder of Quantitative Imaging Solutions. JCC has received research materials (inhaled corticosteroids) from Merck, in order to provide medications free of cost to participants in an NIH-funded study, unrelated to the current work.

References

1. Prezant DJ, Weiden M, Banauch GI, et al. Cough and bronchial responsiveness in firefighters at the World Trade Center site. N Engl J Med 2002;347:806–815. [PubMed: 12226151]
2. de la Hoz RE, Shapiro M, Nolan A, Celedón JC, Szeinuk J, Lucchini RG. Association of low FVC spirometric pattern with WTC occupational exposures. Respir Med 2020;170:106058.
3. Zeig-Owens R, Singh A, Aldrich TK, et al. Blood leukocyte concentrations, FEV_1 decline, and airflow limitation - a 15-year longitudinal study of World Trade Center-exposed firefighters. Ann Am Thor Soc 2018;15:173–183.

4. Liu X, Reeves AP, Antoniak K, et al. Association of quantitative CT lung density measurements with divergent FEV_1 trajectories in WTC workers. Clin Respir J 2021;15:613–621. [PubMed: 33244876]
5. Wheeler K, McKelvey W, Thorpe L, et al. Asthma diagnosed after September 11, 2001 among rescue and recovery workers: findings from the World Trade Center Health Registry. Environ Health Perspect 2007;115:1584–1590. [PubMed: 18007989]
6. de la Hoz RE, Shohet MR, Chasan R, et al. Occupational toxicant inhalation injury: the World Trade Center (WTC) experience. Int Arch Occup Environ Health 2008;81:479–485. [PubMed: 17786467]
7. de la Hoz RE. Occupational asthma and lower airway disease in former World Trade Center workers and volunteers. Curr Allergy Asthma Rep 2010;10:287–294. [PubMed: 20424998]
8. Weber J, Reeves AP, Doucette JT, et al. Quantitative CT evidence of airway inflammation in World Trade Center workers and volunteers with low FVC spirometric pattern. Lung 2020;198:555–563. [PubMed: 32239319]
9. de la Hoz RE, Weber J, Xu D, et al. Chest CT scan findings in World Trade Center workers. Arch Environ Occup Health 2019;74:263–270. [PubMed: 29543564]
10. de la Hoz RE, Liu X, Doucette JT, et al. Increased airway wall thickness is associated with adverse longitudinal first-second forced expiratory volume trajectories of former World Trade Center workers. Lung 2018;196:481–489. [PubMed: 29797069]
11. de la Hoz RE, Jeon Y, Reeves AP, et al. Increased pulmonary artery diameter is associated with reduced FEV_1 in former World Trade Center workers. Clin Respir J 2019;13:614–623. [PubMed: 31347281]
12. de la Hoz RE, Liu X, Celedón JC, et al. Association of obesity with quantitative chest CT measured airway wall thickness in WTC workers with lower airway disease. Lung 2019;197:517–522. [PubMed: 31254057]
13. de la Hoz RE, Shohet MR, Bienenfeld LA, Afilaka AA, Levin SM, Herbert R. Vocal cord dysfunction in former World Trade Center (WTC) rescue and recovery workers. Am J Ind Med 2008;51:161–165. [PubMed: 18213642]
14. de la Hoz RE, Christie J, Teamer J, et al. Reflux symptoms and disorders and pulmonary disease in former World Trade Center rescue and recovery workers and volunteers. J Occup Environ Med 2008;50:1351–1354. [PubMed: 19092489]
15. de la Hoz RE, Aurora RN, Landsbergis P, Bienenfeld LA, Afilaka AA, Herbert R. Snoring and obstructive sleep apnea among former World Trade Center rescue workers and volunteers. J Occup Environ Med 2010;52:29–32. [PubMed: 20042888]
16. de la Hoz RE, Shohet MR, Cohen JM. Occupational rhinosinusitis and upper airway disease: the World Trade Center experience. Curr Allergy Asthma Rep 2010;10:77–83. [PubMed: 20425498]
17. Napier CO, Mbadugha O, Bienenfeld LA, et al. Obesity and weight gain among former World Trade Center workers and volunteers. Arch Environ Occup Health 2017;72:106–110. [PubMed: 27268046]
18. de la Hoz RE, Jeon Y, Miller GE, Wisnivesky JP, Celedón JC. Post-traumatic stress disorder, bronchodilator response, and incident asthma in World Trade Center rescue and recovery workers. Am J Respir Crit Care Med 2016;194:1383–1391. [PubMed: 27548615]
19. Friedman SM, Alper H, de la Hoz RE, Osahan S, Farfel M, Cone JE. Change in asthma is associated with change in PTSD in World Trade Center Health registrants, 2011 to 2016. Int J Environ Res Public Health 2022;19:7795. [PubMed: 35805453]
20. Wardlaw AJ, Silverman M, Siva R, Pavord ID, Green R. Multi-dimensional phenotyping: towards a new taxonomy for airway disease. Clin Exp Allergy 2005;35:1254–1262. [PubMed: 16238783]
21. Woskie SR, Kim H, Freund A, et al. World Trade Center disaster: assessment of responder occupations, work locations, and job tasks. Am J Ind Med 2011;54:681–695. [PubMed: 23236634]
22. Wisnivesky JP, Teitelbaum S, Todd AC, et al. Persistence of multiple illnesses in September 11 rescue workers. Lancet 2011;378:888–897. [PubMed: 21890053]
23. Enright PL, Skloot GS, Cox-Ganser JM, Udasin IG, Herbert R. Quality of spirometry performed by 13,599 participants in the World Trade Center Worker and Volunteer Medical Screening Program. Respir Care 2010;55:303–309. [PubMed: 20196879]

24. Ferris BG. Epidemiology standardization project (American Thoracic Society). Am Rev Respir Dis 1978;118:1–120.
25. Thurston G, Maciejczyk P, Lall R, Hwang J, Chen LC. Identification and characterization of World Trade Center disaster fine particulate matter air pollution at a site in Lower Manhattan following September 11. Epidemiology (Cambridge, Mass) 2003;14:S87–S88.
26. Williams N. The MRC breathlessness scale. Occup Med (Oxford, England). 2017;67:496–497.
27. Pellegrino R, Viegi G, Brusasco V, et al. Interpretative strategies for lung function tests. Eur Respir J 2005;26:948–968. [PubMed: 16264058]
28. Stocks J, Quanjer PH. Reference values for residual volume, functional residual capacity and total lung capacity. ATS Workshop on Lung Volume Measurements. Official Statement of The European Respiratory Society. Eur Respir J 1995;8:492–506. [PubMed: 7789503]
29. Coxson HO. Lung parenchyma density and airwall thickness in airway diseases. Breathe (Sheffield, England). 2012;9:36–45.
30. Ross JC, Castaldi PJ, Cho MH, et al. Longitudinal modeling of lung function trajectories in smokers with and without chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2018;198:1033–1042. [PubMed: 29671603]
31. Huang Z. Clustering large data sets with mixed numeric and categorical values. Proceedings of the First Pacific Asia Knowledge Discovery and Data Mining Conference. Singapore: World Scientific; 1997:21–34.
32. Huang Z. A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining. Proceedings of the SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery. New York: ACM Press; 1997:1–8.
33. Huang Z. Extensions to the k-means algorithm for clustering large datasets with categorical values. Data Mining and Knowledge Discovery 1998;2:283–304.
34. Tibshirani R, Walther G, Hastie T. Estimating the Number of Clusters in a Dataset via the Gap Statistic. Journal of the Royal Statistical Society, Series B 2001;63:411–423.
35. Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. Communications in Statistics - Simulation and Computation 2009;38:1228–1234.
36. Han MK, Agustí A, Celli BR, et al. From GOLD 0 to pre-COPD. Am J Respir Crit Care Med 2020;203:414–423.
37. de la Hoz RE, Hill S, Chasan R, et al. Health care and social issues of immigrant rescue and recovery workers at the World Trade Center site. J Occup Environ Medicine 2008;50:1329–1334. DOI: 10.1097/JOM.0b013e31818ff6fd).
38. Brackbill RM, Hadler JL, DiGrande L, et al. Asthma and posttraumatic stress symptoms 5 to 6 years following exposure to the World Trade Center terrorist attack. JAMA 2009;302:502–516. [PubMed: 19654385]
39. Haghighi A, Cone JE, Li J, de la Hoz RE. Asthma-COPD overlap in World Trade Center Health Registry enrollees, 2015–2016. J Asthma 2021;58:1415–1423. [PubMed: 32930623]
40. Banauch GI, Brantly M, Izbicki G, et al. Accelerated spirometric decline in New York City firefighters with α−1-antitrypsin deficiency. Chest 2010;138:1116–1124. [PubMed: 20634282]
41. Aldrich TK, Vossbrinck M, Zeig-Owens R, et al. Lung function trajectories in World Trade Center-exposed New York City firefighters over 13 years: the roles of smoking and smoking cessation. Chest 2016;149:1419–1427. [PubMed: 26836912]
42. Haldar P, Pavord ID, Shaw DE, et al. Cluster analysis and clinical asthma phenotypes. Am J Respir Crit Care Med 2008;178:218–224. [PubMed: 18480428]
43. Weatherall M, Travers J, Shirtcliffe PM, et al. Distinct clinical phenotypes of airways disease defined by cluster analysis. Eur Respir J. 2009;34:812–818. [PubMed: 19357143]
44. Moore WC, Meyers DA, Wenzel SE, et al. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. Am J Respir Crit Care Med 2010;181:315–323. [PubMed: 19892860]
45. Young AL, Bragman FJS, Rangelov B, et al. Disease progression modeling in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2020;201:294–302. [PubMed: 31657634]

46. Ye L, Pien GW, Ratcliffe SJ, et al. The different clinical faces of obstructive sleep apnoea: a cluster analysis. Eur Respir J. 2014;44:1600–1607. [PubMed: 25186268]

47. Keenan BT, Kim J, Singh B, et al. Recognizable clinical subtypes of obstructive sleep apnea across international sleep centers: a cluster analysis. Sleep 2018;41:zsx214.

48. de la Hoz RE, Shapiro M, Nolan A, et al. Association of World Trade Center (WTC) occupational exposure intensity with chronic obstructive pulmonary disease (COPD) and asthma COPD overlap (ACO). Lung. 2023;201:325–334. DOI: 10.1007/s00408-023-00636-4 [PubMed: 37468611]

49. Godfrey MS, Jankowich MD. The vital capacity is vital: epidemiology and clinical significance of the restrictive spirometry pattern. Chest 2016;149:238–251. [PubMed: 26356330]

50. de la Hoz RE. Occupational lower airway disease in relation to World Trade Center inhalation exposure. Curr Opin Allergy Clin immunol 2011;11:97–102. [PubMed: 21325944]

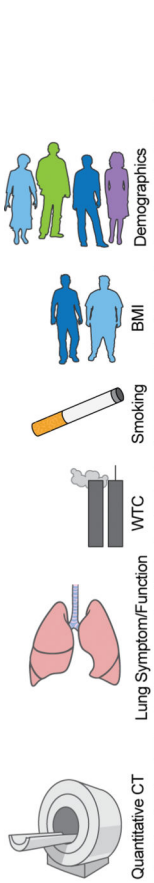
51. Wan ES, Castaldi PJ, Cho MH, et al. Epidemiology, genetics, and subtyping of preserved ratio impaired spirometry (PRISm) in COPDGene. Respir Res 2014;15:89. [PubMed: 25096860]

52. Bhatt SP, Bodduluri S, Kizhakke Puliyakote AS, et al. Structural airway imaging metrics are differentially associated with persistent chronic bronchitis. Thorax 2021;76:343–349. [PubMed: 33408194]

Learning Outcomes

After reading this article readers will be able to:

- Understand the value of cluster analysis in the clinical and pathophysiological characterization of heterogeneous airway disorders.
- Appraise the value of longitudinal surveillance and detailed objective clinical functional, and imaging evaluation in the characterization of adverse respiratory health effects from an exposure.



	CLUSTERS	1: Pauci-symptomatic with high prevalence of airflow obstruction	2: Pauci-symptomatic with intermediate prevalence of a low FVC pattern	3: Symptomatic with high prevalence of a low FVC pattern	4: Non-susceptible smokers	5: Relatively healthy subjects	Total	P value
	N	29	82	40	86	74	311	
Demographics: Age on 9/11/2001	Mean	46.41	42.15	42.18	44.19	41.08	42.86	0.0241
Demographics: Gender	Male	86.2	91.5	67.5	94.2	71.6	83.92	<.0001
Demographics: Ethnicity/race	Latino/any race	20.7	15.9	65.0	0.0	81.1	33.76	<.0001
	Non-Latino/White	69.0	64.6	20.0	89.5	10.8	53.38	
	Non-Latino/Other	10.3	19.5	15.0	10.5	8.1	12.86	
Demographics: Height	Mean (cm)	172.62	175.31	167.31	175.35	164.62	171.50	<.0001
BMI: Baseline BMI	Mean (kg/m^2)	27.71	28.73	31.25	28.83	28.33	28.89	0.0042
BMI: BMIslope	Mean (kg/m^2)	0.13	0.05	-0.02	0.13	0.18	0.10	0.0116
Smoking: Smoking Status	Never	27.6	69.5	50.0	15.1	71.6	48.55	<.0001
	Former	44.8	18.3	32.5	50.0	9.5	29.26	
	Current	27.6	12.2	17.5	34.9	18.9	22.19	
Smoking: Smoking pack-years	Mean	15.75	3.59	7.56	14.06	2.03	7.76	<.0001
WTC: WTC arrival within 48 hr	Yes	37.9	96.3	75.0	30.2	12.2	49.84	<.0001
WTC: WTC exposure duration	Mean	85.03	70.24	97.43	98.43	104.59	91.08	0.0313
Lung Symptom/Function: Dyspnea	Yes	17.2	15.9	100.0	31.4	23.0	32.80	<.0001
Lung Symptom/Function: FEV$_1$% predicted	Mean	70.18	78.35	76.01	92.90	96.68	85.67	<.0001
Lung Symptom/Function: Spirometry pattern	Normal	24.1	29.3	15.0	93.0	87.8	58.52	<.0001
	Low FVC	10.3	54.9	75.0	5.8	5.4	27.97	
	Obstruction	65.5	15.9	10.0	1.2	6.8	13.50	
Lung Symptom/Function: BDRany	Yes	79.3	30.5	42.5	11.6	8.1	26.05	<.0001
Quantitative CT: TLV$_{CT}$ (z-score)	Mean	1.04	0.07	-0.63	0.78	-0.11	0.22	<.0001
Quantitative CT: WAP	Mean	63.97	61.59	63.88	59.18	61.11	61.33	0.0031
Quantitative CT: AT$_{856}$	Mean	22.25	4.70	3.03	6.59	7.37	7.28	<.0001
Quantitative CT: MLD$_{EI}$ratio	Mean	88.12	81.04	81.51	81.89	83.94	82.69	<.0001
Quantitative CT: PI$_{10}$	Mean	2.75	2.72	2.67	2.76	2.69	2.72	0.004
Quantitative CT: LAV>2.5%	Yes	72.4	3.7	0.0	9.3	9.5	12.54	<.0001
Quantitative CT: HAV>10%	Yes	3.5	3.7	12.5	5.8	6.8	6.11	0.3885

Min Max

Figure 1.

Cluster analysis with K-prototype of 311 WTC rescue and recovery workers from the WTC GRC and the WTC Chest CT Imaging Archive. With the exception of QCT measured HAV%, unadjusted analyses demonstrated statistically significant differences for all variables among the 5 clusters.

Table 1.

Characteristics and functional and quantitative chest computed tomography (QCT) findings of the study population (n=311), presented as counts and proportions, or means and standard deviations (SD).

	Pauci-symptomatic with high prevalence of airflow obstruction	**Pauci-symptomatic with intermediate prevalence of a low FVC pattern**	**Symptomatic with high prevalence of a low FVC pattern**	**Non-susceptible smokers**	**Relatively healthy subjects**		
Variables	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Total**	**p**
n	29	82	40	86	74	311	
Age on 9/11/2001, years	46.4 (10.3)	42.2 (8.8)	42.2 (7.4)	44.2 (7.2)	41.1 (9.0)	42.9 (8.5)	0.0241
Male sex	25 (86.2)	75 (91.5)	27 (67.5)	81 (94.2)	53 (71.6)	261 (83.9)	<.0001
Ethnicity/race							
Non-Latino/White	20 (69.0)	53 (64.6)	8 (20.0)	77 (89.5)	8 (10.8)	166 (53.4)	<.0001
Non-Latino/other races	3 (10.3)	16 (19.5)	6 (15)	9 (10.5)	6 (8.1)	40 (12.9)	
Latino/any race	6 (20.7)	13 (15.9)	26 (65.0)	0 (0)	60 (81.1)	105 (33.8)	
Height, cm	172.6 (7.7)	175.3 (9.6)	167.3 (8.5)	175.4 (8.7)	164.6 (9.4)	171.5 (10.1)	<.0001
Baseline BMI, kg/m^2	27.7 (3.7)	28.7 (4.5)	31.3 (5.4)	28.8 (4.0)	28.3 (4.0)	28.9 (4.4)	.0042
BMIslope, kg/m^2/year	0.13 (0.26)	0.05 (0.29)	−0.02 (0.31)	0.13 (0.34)	0.18 (0.37)	0.10 (0.3)	.0116
Smoking status							<.0001
Never Smoker	8 (27.6)	57 (69.5)	20 (50.0)	13 (15.1)	53 (71.6)	151 (48.6)	
Former Smoker	13 (44.8)	15 (18.3)	13 (32.5)	43 (50.0)	7 (9.5)	91 (29.3)	
Current Smoker	8 (27.6)	10 (12.2)	7 (17.5)	30 (34.9)	14 (18.9)	69 (22.2)	
Smoking intensity, pack-years	15.8 (16.4)	3.6 (7.1)	7.6 (11.0)	14.1 (12.6)	2.0 (5.8)	7.8 (11.6)	<.0001
WTC arrival time≤48 hr	11 (37.8)	79 (96.3)	30 (75.0)	26 (30.2)	9 (12.2)	155 (49.8)	<.0001
WTC exposure duration, days	85.0 (68.3)	70.2 (64.9)	97.4 (82.1)	98.4 (75.7)	104.6 (72.5)	91.1 (73.2)	.0313
Dyspnea mMRC≥2	5 (17.2)	13 (15.9)	40 (100.0)	27 (31.4)	17 (23.0)	102 (32.8)	<.0001
Lung function							
FEV_1%predicted	70.2 (17.4)	78.4 (15.0)	76.0 (12.5)	92.9 (11.3)	96.7 (16.5)	85.7 (17.2)	<.0001
Spirometry pattern							<.0001
Normal	7 (24.1)	24 (29.3)	6 (15.0)	80 (93.0)	65 (87.8)	182 (58.5)	
Low FVC	3 (10.3)	45 (54.9)	30 (75.0)	5 (5.8)	4 (5.4)	87 (28.0)	
Obstruction	19 (65.5)	13 (15.9)	4 (10.0)	1 (1.16)	5 (6.8)	42 (13.5)	
BDRany	23 (79.3)	25 (30.5)	17 (42.5)	10 (11.6)	6 (8.1)	81 (26.0)	<.0001

	Pauci-symptomatic with high prevalence of airflow obstruction	Pauci-symptomatic with intermediate prevalence of a low FVC pattern	Symptomatic with high prevalence of a low FVC pattern	Non-susceptible smokers	Relatively healthy subjects		
Variables	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Total**	**p**
QCT metrics							
TLV_{CT}, z-score	1.04 (0.83)	0.07 (0.77)	−0.063 (0.62)	0.78 (0.85)	−0.11 (0.78)	0.22 (0.94)	<.0001
WAP	64.0 (6.6)	61.6 (8.0)	63.9 (7.8)	59.2 (7.4)	61.1 (6.4)	61.3 (7.5)	.0031
AT_{856}	22.3 (17.6)	4.7 (7.0)	3.0 (6.0)	6.6 (7.2)	7.4 (9.0)	7.3 (10.2)	<.0001
MLD_{EI}	88.1 (4.4)	81.0 (5.9)	81.5 (7.0)	81.9 (6.1)	83.9 (5.7)	82.7 (6.3)	<.0001
PI_{10}	2.75 (0.16)	2.72 (0.15)	2.67 (0.12)	2.76 (.015)	2.69 (0.12)	2.72 (0.14)	.004
LAV>2.5%	21 (72.4)	3 (3.7)	0	8 (9.3)	7 (9.5)	39 (12.5)	<.0001
HAV>10%	1 (3.5)	3 (3.7)	5 (12.5)	5 (5.8)	5 (6.8)	19 (6.1)	.3885

Table 2.

FEV_1 trajectory categories (rapid and intermediate decliners, and gainers) for 256 (82.3%) of the 311 subjects in each of the five clusters, who had at least three good quality periodic surveillance spirometries between 2002 and 2018.

	Clusters					
FEV_1 trajectory group	**1**	**2**	**3**	**4**	**5**	**Total**
Rapid decliners (n, %)	9 (39.1)	8 (11.6)	5 (14.7)	21 (30.9)	14 (22.6)	57 (22.3)
Intermediate (n, %)	9 (39.1)	55 (79.7)	25 (73.5)	39 (57.4)	44 (71.0)	172 (67.2)
Gainers (n, %)	5 (21.7)	6 (8.7)	4 (11.8)	8 (11.8)	4 (6.5)	27 (10.5)
Total	23 (9.0)	69 (27.0)	34 (13.3)	68 (26.6)	62 (24.2)	256