Highlights

- Non-genetic influences of cystic fibrosis lung disease are not well characterized
- Clustering yielded social-environmental lung phenotypes that corresponded to rapid decline that occurred in early adolescence, mid adolescence, or later in early adulthood
- Timing of peak decline varied by social-environmental adversity
- Social and environmental health disparities impacted middle decliners the most
- Severity of rapid decline differed by regional community
Article type: Research Paper

Social-Environmental Phenotypes of Rapid Cystic Fibrosis Lung Disease Progression in Adolescents and Young Adults Living in the United States

Anushka K. Palipana, PhD,1,2 Andrew Vancil, MS,2 Emrah Gecili, PhD,2,3 Erika Rasnick, MS,2 Daniel Ehrlich,1,2 Teresa Pestian, BS,2 Eleni-Rosalina Andrinopoulou, PhD4,5 Pedro M. Afonso, MS,4,5 Ruth H. Keogh, DPhil,6 Yizhao Ni, PhD,7 Judith W. Dexheimer, PhD,3,8 John P. Clancy, MD,9 Patrick Ryan, PhD,2,3 Cole Brokamp, PhD,2,3 Rhonda D. Szczesniak, PhD2,3,10,*

Affiliations:
1Duke University, Durham, NC, USA
2Division of Biostatistics & Epidemiology, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH, USA
3Department of Pediatrics, University of Cincinnati, Cincinnati, OH, USA
4Department of Biostatistics, Erasmus Medical Center, Rotterdam, The Netherlands
5Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands
6London School of Hygiene and Tropical Medicine, London, UK
7Kaiser Permanente, Denver, CO, USA
8Division of Biomedical Informatics, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH, USA
9Cystic Fibrosis Foundation, Bethesda, MD
10Division of Pulmonary Medicine, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH, USA

*Correspondence: 3333 Burnet Ave (MLC 5041), Cincinnati Children’s Hospital Medical Center, Cincinnati, OH, USA, 45229. Email: Rhonda.Szczesniak@cchmc.org
ABSTRACT

Background: Cystic fibrosis (CF) is a genetic disease but is greatly impacted by non-genetic (social/environmental and stochastic) influences. Some people with CF experience rapid decline, a precipitous drop in lung function relative to patient- and/or center-level norms. Those who experience rapid decline in early adulthood, compared to adolescence, typically exhibit less severe clinical disease but greater loss of lung function. The extent to which timing and degree of rapid decline are informed by social and environmental determinants of health (geomarkers) is unknown.

Methods: A longitudinal cohort study was performed (24,228 patients, aged 6-21 years) using the U.S. CF Foundation Patient Registry. Geomarkers at the ZIP Code Tabulation Area level measured air pollution/respiratory hazards, greenspace, crime, and socioeconomic deprivation. A composite score quantifying social-environmental adversity was created and used in covariate-adjusted functional principal component analysis, which was applied to cluster longitudinal lung function trajectories.

Results: Social-environmental phenotyping yielded three primary phenotypes that corresponded to early, middle, and late timing of peak decline in lung function over age. Geographic differences were related to distinct cultural and socioeconomic regions. Extent of peak decline, estimated as forced expiratory volume in 1 second of % predicted/year, ranged from 2.8 to 4.1 % predicted/year depending on social-environmental adversity. Middle decliners with increased social-environmental adversity experienced rapid decline 14.2 months earlier than their counterparts with lower social-environmental adversity, while timing was similar within other phenotypes. Early and middle decliners experienced mortality peaks during early adolescence and adulthood, respectively.

Conclusion: While early decliners had the most severe CF lung disease, middle and late decliners lost more lung function. Higher social-environmental adversity associated with increased risk of rapid decline and mortality during young adulthood among middle decliners. This sub-phenotype may benefit from enhanced lung-function monitoring and personalized secondary environmental health interventions to mitigate chemical and non-chemical stressors.

Key words: chemical stressors; cluster analysis; community deprivation; environmental health epidemiology; lung function; medical monitoring

Abstract word count: 298; Manuscript word count: 4,744

Abbreviations: CF: cystic fibrosis; CFFPR: Cystic Fibrosis Foundation Patient Registry; CFRD: cystic fibrosis related diabetes; CI: confidence interval; FEV₁: forced expiratory volume in 1 s; FPC₁: first functional principal component; FPCA: functional principal components analysis; MRSA: Methicillin-resistant *Staphylococcus aureus*; OR: odds ratio; Pa: *Pseudomonas aeruginosa*; PC: principal component; PCA: principal components analysis; PEx: pulmonary exacerbation; Q1: first quartile; Q3: third quartile; SD: standard deviation; ZCTA: ZIP Code tabulated area
1. Introduction

Cystic fibrosis (CF) is a rare genetic disease, but nearly 50% of the variation in lung function is attributable to non-genetic (social, environmental and stochastic) factors (Collaco, Blackman et al. 2010). The crucial role of social and environmental exposures in the development and exacerbation of lung diseases such as asthma and chronic obstructive pulmonary disease (COPD) is widely acknowledged (Guarnieri and Balmes 2014, Hansel, McCormack et al. 2016, Louisias, Ramadan et al. 2019). Air pollutants can harm lung tissue directly upon exposure and indirectly by producing reactive oxygen species and causing systemic inflammation. Despite the well-established link between air pollution and respiratory disorders, the impact of social and environmental factors on people living with CF remains largely unexplored. Furthermore, the effects of social and environmental exposures on CF lung function decline have primarily been studied one exposure domain at a time. Examples of two separate but frequently studied exposures are air pollution (Blayac, Coll et al. 2022) and Medicaid insurance use (Schechter, Shelton et al. 2001). While the former has been linked primarily to pulmonary exacerbation onset (Goss, Newsom et al. 2004, Goeminne, Kicinski et al. 2013), which is largely driven by acute drops in lung function, the latter has often corresponded to prolonged drops in lung function (Schechter and Margolis 1998). Recent tobacco smoke exposure studies using pediatric data from the U.S. Cystic Fibrosis Foundation Patient Registry (CFFPR) found that first- or second-hand smoking cessation associated with improved lung function over time (Oates, Baker et al. 2021), and, for individuals treated with tezacaftor/ivacaftor, there was dampened benefit from modulator therapy shown by continued divergence in lung function trajectories of exposed and unexposed groups (Baker, Harris et al. 2021).

The CF lung disease process is typically characterized by acute and prolonged drops in lung function. Attenuated decreases in lung function relative to patient- and/or center-level norms, clinically termed rapid decline, typically manifest during adolescence and early adulthood (Konstan, Morgan et al. 2007, Vandenbranden, McMullen et al. 2012). Early rapid
decline tends to happen in individuals with below-average lung function, infections, poor/deteriorating nutrition, increased antibiotic use, and more frequent hospitalizations. Meanwhile, regardless of clinical phenotype, lung function trajectories exhibit nonlinear, heterogeneous patterns over this age range (Vandenbranden, McMullen et al. 2012, Szczesniak, McPhail et al. 2013, Harun, Wainwright et al. 2016). A classification study of F508del homozygotes indicated that lung function (initial level and rate of decline) coupled with survival percentiles at 20 years of age accurately distinguished mild and severe pulmonary phenotypes from the Gene Modifier Study (Schluchter, Konstan et al. 2006). An effort to phenotype rapid decline was made using the aforementioned CFFPR data source while accounting for both age-related nonlinearity and heterogeneity in lung function trajectories suggested that the magnitude of peak lung function decline is similar; however, timing of peak decline differed according to early, middle, and late ages during adolescence/early adulthood: 12.9, 16.3, and 18.5 years, respectively (Szczesniak, Li et al. 2017). In addition to the above risk factors of rapid decline, the early phenotype was associated with Medicaid insurance use. A single-center study including young adults with CF developed an index score of rapid decline using quantiles of lung function trajectories, showing that nutrition- and sex-related effects were more influential in higher quantiles corresponding to worsening disease severity (Denaro 2020).

Past approaches have illuminated subsets of social and environmental exposures that are correlates of rapid decline, primarily studied independent from phenotyping. As a result, social-environmental phenotypes of rapid decline have not been characterized. In this study, we hypothesized that a comprehensive evaluation of social and environmental exposures (geomarkers), including the formation of an adversity index, would yield distinct age-related phenotypes of rapid decline. Additionally, we hypothesized that more severe social-environmental phenotype scores would associate with worsening patient characteristics that are routinely surveilled as part of CF care, including genotype, lung infections, smoking status, and comorbidities.
2. Methods

2.1 Study population and health data

Inclusion criteria: We performed a longitudinal cohort study of patients aged 6–21 years followed in the CFFPR (over the timeframe 1997–2017) approved by the local Institutional Review Board (Protocol ID: 2018-4839). The CFFPR has been used to track demographic and clinical characteristics of people living with CF in the U.S. for many decades with a recently estimated coverage of 77% allowing for generalizability across the population (Cromwell, Ostrenga et al. 2023). The primary outcome was forced expiratory volume in 1 second (FEV₁) of % predicted from Global Lung Function Initiative reference equations (Quanjer, Stanojevic et al. 2012). Age (in years) was the time variable. Individuals with less than 7 quarterly FEV₁ observations over the study timeframe or who never reported ZIP Code were excluded. Post-lung transplantation data was censored. Baseline was individually defined as first observed FEV₁ during the study period once the patient was aged ≥ 6 years.

 Routinely collected health measures: Non-time-varying CFFPR variables were genotype defined by F508del alleles (homozygous, heterozygous, or neither/unknown), race (White or non-White), ethnicity (Hispanic or non-Hispanic), sex (male or female), age at CF diagnosis, any recorded use of pancreatic enzymes (yes/no). Time-varying variables were private insurance use (yes or no); lung infections (each coded as yes or no): *Pseudomonas aeruginosa* (Pa) and Methicillin-resistant *Staphylococcus aureus* (MRSA); CF-related diabetes (CFRD) status; modulator use: ivacaftor or lumacaftor/ivacaftor; reporting either first- or secondhand tobacco smoke exposure (obtained from annual self-report). Pulmonary exacerbation (PEx) frequency was defined by counting the number of events with intravenous antibiotic use that required hospitalization within the prior year.

2.2 Geomarker assessments
Social and environmental exposures (geomarkers) were linked to ZIP Codes in the CFFPR by using ZIP Code tabulation areas (ZCTAs, a census-derived geography for 5-digit residential ZIP Codes). Geomarkers (respective data sources) included traffic proximity, ozone and PM$_{2.5}$ concentrations, diesel particulate matter, and respiratory hazard index (Environmental Protection Agency Environmental Justice Screen Index (Agency 2015-2019)), landcover (% of greenspace, impervious, and tree canopy areas derived from the National Land Cover Database (Jin 2019)), total crime (Applied Geographic Solutions (Solutions 2018)), neighborhood material deprivation index (Brokamp, Beck et al. 2019) (assesses extent of poverty, vacant housing, assisted income, education level, median income and health insurance coverage for a given neighborhood), lengths and densities of primary and secondary roads (derived from geospatial data in the Topologically Integrated Geographic Encoding and Referencing system (US Department of Commerce 2018)). Rationale for including each geomarker was based on components of a postulated CF disease-outcome model (Schechter 2011) and their roles as potential correlates of rapid CF lung disease progression from more recent literature review (Szczesniak, Rice et al. 2020) (e-Table 1). We imputed missing data on demographic/clinical variables and ZIP Codes based on the nature of collection for each CFFPR variable (e-Table 2). Backward-imputing was employed for geomarkers and ZIP Codes that were not reported until after baseline.

2.3 Statistical analyses

We summarized all variables at baseline and over follow-up as mean (standard deviation or SD) for continuous variables and n (%) for categorical variables. Analyses were implemented using R and MATLAB (code and packages described in e-Section A). We performed a two-stage approach to obtain social-environmental phenotypes of rapid decline. In the first stage, principal components analysis (PCA) was used to characterize relationships between geomarkers at baseline and acquire a composite geomarkers score for subsequent cluster analysis. For the
second stage, we performed covariate-adjusted functional principal components analysis (FPCA) for longitudinal data (Jiang 2010), using mean quarterly FEV₁% predicted as the outcome variable.

Since the available FPCA method only accommodates one covariate aside from the time variable (which is age in our approach), we included the aforementioned social-environmental adversity index as the covariate. Detailed implementation and results, including selecting the first functional principal component and forming tertiles to represent phenotypes of rapid decline according to age, are provided in e-Section B.

After performing the two-stage analysis, we examined associations between derived phenotypes with respect to individual demographic/clinical characteristics using Wilcoxon rank sum with Bonferroni-adjusted p-values and Chi-square tests. A p-value less than 0.05 was considered statistically significant.

2.4 Sensitivity analyses
We assessed the impact of identifying phenotypes using average rather than baseline values of the geomarkers in PCA (first stage). Second-stage sensitivity analyses evaluated i) potential selection bias from requiring a minimum number of FEV₁ measurements; ii) how FPCA findings may be impacted by variable length of follow-up between patients; iii) impact of modulator initiation; iv) impact of loss-to-follow-up in the CFFPR due to 2003-2006 intake changes, which included introduction of a web-based platform to capture encounter-level data in 2003 followed by detailed medication collection that began in 2006 (Knapp, Fink et al. 2016); v) extent to which imputation of later-observed geomarkers into earlier time periods (largely prior to 2015) affected results.
3.1 Study population and routinely monitored characteristics

There were 24,228 individuals with 664,267 quarterly measurements who met inclusion criteria (e-Figure 1). The analysis cohort primarily included F508del homozygotes, had slightly more males, and many did not have private insurance during the analysis timeframe (Table 1 – Overall cohort). Infections, CFRD prevalence, and enzyme use increased expectedly over follow up. Very few individuals used modulators at baseline, but prevalence increased over follow up. Reported tobacco smoke also increased over time. Most individuals did not undergo lung transplant and remained alive through follow up. Geomarker data was variable across the overall cohort (Table 2). Deprivation index, which ranges from 0 to 1 with higher values being associated with higher levels of material community deprivation, for the overall CF cohort was below the national average, estimated to be 0.37 and 0.35 (computed by weighting each tract-level deprivation index by its population under age 18) for 2015 and 2018, respectively.

The first PC score was retained from scree-plotting (e-Figure 3) and used to collectively measure extent of social-environmental adversity (higher scores implied more negative social and/or environmental exposure levels; details shown in e-Section A). Social-environmental adversity was primarily driven by crime, landcover, traffic-related air pollution, and density and length of primary and secondary roadways, respectively (loadings for rotated PC1, e-Table 6). Although not used to form the social-environmental adversity index, the second principal component mainly linked to deprivation index, ozone concentration, primary roadway length, and secondary roadway density (loadings for rotated PC2, e-Table 6).

3.2 Social-environmental lung phenotypes over age

Functional principal component analysis (FPCA) suggested highly heterogeneous lung function decline according to age and environmental exposure, although all individuals lost some degree of lung function by early adulthood (Figure 1A). Overall, increased social-environmental adversity corresponded to higher lung function at younger ages, but the degree of rapid decline
was more pronounced in adolescence and early adulthood, indicating declines of more than 4% predicted/year for some individuals (Figure 1B). Upon segmenting these unusual results, subjects with less social-environmental adversity also experienced lung function decline, but at a slower rate, compared to counterparts with high social-environmental adversity (e-Figure 4). Specifically, excluding individuals with outlier values for the social-environmental adversity index showed that higher social-environmental adversity associated with more rapid lung function decline over age (e-Figure 5).

We classified individuals into three distinct social-environmental phenotypes of rapid decline over age (early, middle, or late) based on FPC\textsubscript{1} first and third quantiles from FPCA cluster analysis (e-Table 5). Further segmenting early, middle, and late decliners into subgroups according to median social-environmental adversity, there was within-phenotype variability (Figure 2). Early decliners experienced similar timing and extent of peak decline regardless of social-environmental adversity (Figure 2A versus 2D). Middle decliners had similar peak decline, but those with greater extent of social-environmental adversity experienced peak decline an average of 14.4 months or 1.2 years earlier than those with lower social-environmental adversity (Figure 2B versus 2E). Late decliners with lower social-environmental adversity had slightly earlier peak decline (approximately four months earlier), compared to their counterparts with higher social-environmental adversity (Figure 2C versus 2F).

Specific geomarkers primarily drove differences in social-environmental adversity among the early, middle, and late decline phenotypes (Table 2). The earlier declining phenotypes tend to reside in areas with higher community deprivation and crime. The average deprivation index for late decliners was lower than the aforementioned national averages, while average deprivation for early decliners was similar to national averages. Compared to late decliners, early decliners tended to reside in areas with slightly higher green space and imperviousness. Early decliners also tend to live in areas with higher air pollution, measured by the respiratory hazard index, ozone concentration, proximity to traffic, and diesel particulate matter. Middle
decliners tend to live closer to longer/denser secondary roadways compared to late decliners. While early decliners tend to live closer to longer secondary roadways compared to late decliners, they also live closer to denser secondary roadways compared to middle decliners.

Estimated peak decline and degree of social-environmental adversity were jointly mapped across the contiguous U.S. (Figure 3). While findings were heterogeneous, the most severe declines coupled with greatest social-environmental adversity (purple shaded areas) were concentrated in western and southern regions. These regions also included areas with higher social-environmental adversity but less rapid decline (pink shaded areas). In examining designations from the American Communities Project, areas with the highest social-environmental adversity and peak decline levels corresponded to Hispanic Centers, Working Class Country (rural communities) and Big City (densely populated urban areas). Social-environmental adversity appeared lowest in the Central Plains, but these areas were marked with higher peak declines (green-shaded areas). A subset of the map was not estimable (black shaded areas). The map referred to as Figure 3 has been made available in an interactive R Shiny app. This app allows readers to explore the joint variation of estimated peak decline and the degree of social-environmental adversity on a state-by-state basis. This app can be accessed using the following link: https://medmonitoring.shinyapps.io/Interactive_map3digZip/

3.3 Clinical subgrouping of phenotypes

Preliminary comparisons between early, middle, and late decline phenotypes (Table 1) showed that middle decliners had higher representation of F508del homozygotes and males, compared to other phenotypes. Middle decliners were slightly older with substantially lower baseline lung function. Having no insurance or use of public insurance, lung infections and CFRD diagnosis were more prevalent in early decliners. However, late decliners had the highest prevalence of reported impaired glucose tolerance. Middle decliners had the highest reported use of pancreatic enzymes. Ivacaftor use was most common among late decliners, followed by middle,
then early decliners. Middle decliners had the highest reported use of lumacaftor/ivacaftor, followed by late, then early decliners. Reported tobacco smoke exposure was most prevalent among middle decliners. Early decliners had the lowest PEx frequency but highest rates of hospitalizations prior to baseline and lung transplant and death over follow-up. Kaplan–Meier analysis of phenotype-specific survival probabilities extending beyond early adulthood suggest mortality peaks occurred in early adolescence and again in early adulthood, followed by exceedingly higher rates of lung transplant/death in the early phenotype (e-Figure 2).

3.4 Sensitivity Analyses
The social-environmental adversity index was similar when performing PCA on average geomarker values, compared to using baseline geomarker values (e-Table 6). When comparing the included individuals in the analysis cohort to those who were excluded due to an insufficient number of FEV₁ measurements, we found that the excluded cohort had higher prevalence of F508del homozygotes, and those with no completed genotype were older, on average, and had slightly higher rates of Pa infection and modulator use (e-Table 7). However, distributions of FPC₁ scores were similar (e-Table 8). Variable length of follow-up between patients did not apparently impact FPCA results based on correlation analyses (e-Figure 7). We further examined how rate of follow-up (number of visits/year) varied geographically and found that urban and rural areas had similar median years of follow-up but differing ranges: 5.67 (1.52-18.02) versus 5.56 (1.34-27.13). To consider how loss to follow up may have been impacted by CFFPR-related intake changes during 2003–2006, we summarized social-environmental adversity and degree of rapid decline separately for those individuals entering before versus after 2006. Distributions of scores from FPCA and geographical associations were similar to primary analysis (e-Figure 10), indicating that results were not impacted by data intake changes. Detailed results are provided in e-Section C. We evaluated impact of geomarker imputation by comparing distributions of FPC₁ scores from individuals with observations only from 2015 or
later to the overall cohort and to those who only had observations prior to 2015 (e-Figure 11). Scores were largely similar among these groups; those observed 2015 and onward had a slightly higher median score.

3.5 Converting phenotypes into digestible clinical information

While the results have different implications among the three main social-environmental phenotypes, we found that existing strategies for maintaining lung function in CF under each phenotype could be layered with enhanced social and environmental health interventions (Table 3). Routine or more frequent social needs screening, for example, could be made available for people with CF who live in communities with high deprivation or crime levels. We found that the nature of potential environmental health interventions ranges from personalized, such as installing HEPA filters in the home, to policy-level alerts or actions, including smog notifications during periods in which air pollution spikes are present or anticipated due to extreme climate events. We observed higher within-area variation compared to between areas with respect to social-environmental adversity and risk of rapid decline (e-Figure 12), implying that people with CF and their families who make long-distance moves to different geographic regions (e.g., moving from the Midwest to the Eastern seaboard) may be subject to the same risk of rapid decline if social-environmental adversity is similar between regions. However, relocating within the same city but to a community with less social-environmental adversity may result in lower risk of rapid decline; although such a measure may be unrealistic for many with CF.

Another clinically informative takeaway for individual patients, providers and researchers may be the review of each social and environmental exposure based on how pronounced it is for their given residential area or where they will spend a significant amount of time (e-Figure 13), in light of their individual clinical risk factors (Table 3). From the patient and clinician perspective, this more granular review could identify risk of environmental triggers of PEx, such as seasonal or event-specific spikes in air pollution levels that could be mitigated with masking.
Masking strategies could be developed and implemented, similar to those being undertaken by some people with CF and their families during peak influenza season. For health equity researchers, these data may pinpoint communities suffering outcome disparities and enable clinics to implement enhanced screening of psychosocial needs.

4. Discussion

This longitudinal cohort study of children and young adults living with CF in the U.S. shows that both social and environmental stressors associate with greater degree and earlier timing of rapid lung function decline, and that, to some extent, social-environmental phenotypes of lung function decline are associated with a subset of routinely collected demographic and clinical characteristics. We observed that young people with early decline tend to reside in areas with higher total crime rates and community deprivation; in contrast, their late-declining counterparts typically reside in low-crime, low-deprivation areas, which highlights the significant role of non-chemical, socioeconomic stressors on CF disease. We also identified geographic regions with elevated ambient air pollution (traffic proximity and roadway length/density) and less greenspace in which young people with CF may be at the highest risk of early, rapid lung function decline, but we found that these areas represent heterogeneous environmental conditions spread across the US. Social-environmental adversity in CF was heavily influenced by levels of chemical stressors, such as those from traffic-related air pollution exposure. This finding corroborates a recent, pediatric single-center study performed in the Midwest, which identified elemental carbon attributable to traffic as a strong predictor of rapid CF lung function decline; while not statistically significant, greenspace and community deprivation were also selected in the final prediction model (Gecili, Brokamp et al. 2023). In the current study, community deprivation, which is a non-chemical stressor, in the CF population was estimated to be beneath the national average. To provide some context, given that the population level SD in 2018 was 0.14, the difference of 0.02 units in the deprivation index is roughly 14% of the SD.
The current study corroborates earlier work suggesting that individuals with the highest levels of lung function are late decliners but cumulatively lose more lung function over childhood and adolescence than their counterparts who maintain lower lung function (Vandenbranden, McMullen et al. 2012). From characteristics of late decliners identified in the current study, an outlying percentage maintained high lung function initially in the presence of extreme social-environmental adversity, while the majority of this phenotype had higher social-environmental adversity associated with more rapid decline. Identification of three prominent clusters reflects prior joint longitudinal-survival modeling of the CFFPR that also identified three distinct latent classes of FEV1 progression (Andrinopoulou, Nasserinejad et al. 2020). Mortality among phenotypes diverged at early adolescence for those with the most severe declines, while middle- and late-declining phenotypes had similar survival probabilities. These mortality peaks are consistent with prior literature (Vandenbranden, McMullen et al. 2012) but also highlight the risk to a newly identified middle-decliner subgroup susceptible to early mortality and high social-environmental adversity. It is possible that the influence of social-environmental factors may be less pronounced for individuals who decline early or late compared to their middle declining counterparts. Additionally, the impact of social-environmental adversity on lung function may vary depending on the timing of exposure and age could serve as a confounding variable in the clustering of environmental factors and rapid decline. Tailored secondary prevention strategies (e.g.: avoiding living near highways or high pollution areas; use of HEPA filters in the home (James, Bernstein et al. 2020)), could benefit this sub-phenotype the most. How social-environmental adversity mediates effectiveness of established strategies, (e.g., more frequent monitoring and timely treatment of lung function declines (Schechter 2018)) requires additional research.

Our work substantiates recent findings that chemical stressors (e.g., tobacco smoke exposure) have explanatory value distinct from socioeconomic deprivation (Oates, Baker et al. 2020). In COPD, the relationship between ambient air pollution and lung function appears to be
moderated by sex, household income, and occupation type (Doiron, de Hoogh et al. 2019). Specific chemical stressors have not been widely studied in CF, but causal pathways identified in asthma link air pollution to exacerbation (Pfeffer, Mudway et al. 2021). Clinical manifestations among CF, asthma and COPD are distinctive (e.g.: given the prolonged versus acute nature of exacerbation onset in CF versus asthma, or the nature of airway obstruction in CF versus COPD), suggesting that causal pathways between chemical stressors and disease severity/onset differ.

The current study utilized mean quarterly FEV$_1$% predicted measurements, while other CF epidemiological studies have utilized different aggregates, such as the maximum. Recent advances in linear mixed effects models, which have been instrumental in monitoring and predicting the natural history of CF lung disease, suggest that use of mean or maximum quarterly FEV$_1$% predicted measurements result in similar trajectory estimates (Szczesniak, Andrinopoulou et al. 2023). Functional data analysis techniques such as FPCA can be applied to temporally collected FEV$_1$% predicted measurements to identify longitudinal functional data with similar characteristics and cluster them together. Linear mixed effects models, which estimate associations between the longitudinal outcome variable and a set of explanatory variables, have led to a consensus that rapid decline often occurs between adolescence and early adulthood, although the specific timing varies across studies. In this study, using FPCA on CF and social and environmental exposure data not only helped identify patients with phenotypes at risk of rapid lung function decline but also characterized the timing of rapid lung function decline within each phenotype. This approach can enhance opportunities for more timely and targeted interventions for patients clustered into high-risk phenotypes.

This study has inherent limitations, including those previously described with CFFPR analyses (Schechter 2008) and the large extent of missing data in income and education variables (Foundation 2022), which precluded their use in the current study as a measure of individual-level socioeconomic status. There are differing amounts of overlap among the
geomarker and clinical data sources, which make it difficult to produce a single, contemporaneous collection of data. Due to computational issues with FPCA of data with short-term follow-up, we were unable to restrict follow-up to the more contemporaneous time period in which select geomarkers were observed (2015 and up). To offer some insight into how the findings are impacted by contemporality, we conducted several sensitivity analyses, but acknowledge that the findings are still based on the original FPCA with assumed imputations of earlier data and therefore subject to bias especially over earlier time periods of follow-up. There is also the influence of bigger data sets on precision estimates and statistical significance (Cox, Kartsonaki et al. 2018). Variability in social-environmental adversity and rapid decline was observed in aggregated estimates throughout the US, but ZIP-Code-specific inference was intentionally limited for patient privacy. U.S. Postal Service ZIP Codes are designed to facilitate the distribution of mail, which makes the resulting ZCTAs and geomarker resolution dependent upon residential size (i.e.: cities typically have far more ZIP Codes than rural areas). Opposing PCA loadings on roadway lengths and densities show the impact of the area of ZIP Codes. An example of the granularity impact is provided showing secondary roadway densities and aggregation for a given city and rural areas (e-Figure 14). Covariate-adjusted FPCA only allows for a single covariate, which necessitated a two-stage approach, but the resulting social-environmental adversity index was partially driven by previously identified features, such as air pollution (Goss, Newsom et al. 2004), while examining novel geomarkers. This index could enhance research on CF-specific social and environmental determinants of health by aggregating exposures across a breadth of young people with CF living in the US. Covariate adjustment with geomarkers appeared to improve upon marginal FPCA in which identified phenotypes exhibited stronger differential selection according to baseline age (Szcza{~n}iak, Li et al. 2017). Air temperature and seasonality were not considered, given that these characteristics typically vary over time for many locations. However, geomarkers of traffic-related air pollution, which can be driven by increased temperatures within warmer seasons, were included.
Furthermore, sensitivity analysis suggested that incorporating time-varying exposures may not lead to differing conclusions. As the study focused on a specific timeframe, patients receiving treatment with the highly effective CFTR modulator, elexacaftor-tezacaftor-ivacaftor (ETI), were not included in the analysis. However, it is possible to make assumptions about the impact of ETI treatment on the phenotypes described during the ETI era, such as the potential attenuation of lung function decline; however, effects may be dampened depending on social-environmental influences on extent to which CFTR is modulated. Considering the available literature, it is plausible that the average trajectory levels will rise but with a variable degree of rate of decline (Nichols, Paynter et al. 2021).

5. Conclusion
Cystic fibrosis lung phenotypes are characterized by social-environmental adversity and a subset of routinely monitored demographic and clinical characteristics. Middle decliners have the greatest differences in timing of peak decline related to social-environmental adversity and mortality. The present study offers a comprehensive profile of geographic exposure risks and how individuals with mild to moderate amounts of lung disease who are subject to adverse social-environmental exposures may maximally benefit from personalized (rather than primary or global) environmental health interventions.

Acknowledgments

**Funding:** Supported by grants from the National Institutes of Health (R01 HL141286 and P30DK117467) and the Cystic Fibrosis Foundation (GECILI20F0 and Naren19R0). The sponsors had no role in study design or manuscript preparation. A pre-submission copy of the manuscript was provided to the Cystic Fibrosis Foundation Patient Registry Committee in accordance with data request/use policies.
Contributions: The authors thank the Cystic Fibrosis Foundation for use of Patient Registry (CFFPR) data to conduct this study and the patients, care providers, and clinic coordinators at U.S. CF centers for their contributions to the CFFPR. Requests to access the CFFPR analysis data utilized for this study may be sent to the Foundation for review/approval:
datarequests@cff.org

References


Table 1. Demographic and clinical characteristics, overall and across social-environmental phenotypes.

<table>
<thead>
<tr>
<th></th>
<th>Overall cohort</th>
<th>Early (n = 6,057)</th>
<th>Middle (n = 12,114)</th>
<th>Late (n = 6,057)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 24,228)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F508del mutation*</td>
<td>F508del homozygous</td>
<td>11,618 (48.0%)</td>
<td>2,887 (47.7%)</td>
<td>6,006 (49.6%)</td>
</tr>
<tr>
<td></td>
<td>F508del heterozygous</td>
<td>8,711 (36.0%)</td>
<td>1,983 (32.7%)</td>
<td>4,363 (36.0%)</td>
</tr>
<tr>
<td></td>
<td>Neitherrunknow</td>
<td>3,899 (16.0%)</td>
<td>1,137 (19.6%)</td>
<td>1,745 (14.4%)</td>
</tr>
<tr>
<td>Male*</td>
<td>12,367 (51.0%)</td>
<td>2,769 (45.7%)</td>
<td>6,325 (52.2%)</td>
<td>3,273 (54.0%)</td>
</tr>
<tr>
<td>Race</td>
<td>White</td>
<td>22,486 (92.8%)</td>
<td>5,589 (93.4%)</td>
<td>11,241 (92.8%)</td>
</tr>
<tr>
<td></td>
<td>Non-White</td>
<td>1,742 (7.2%)</td>
<td>468 (7.7%)</td>
<td>873 (7.2%)</td>
</tr>
<tr>
<td></td>
<td>Hispanic ethnicity*</td>
<td>1,973 (8.1%)</td>
<td>650 (10.7%)</td>
<td>916 (7.6%)</td>
</tr>
<tr>
<td></td>
<td>Diagnosis age_a,b,c</td>
<td>2.0 (3.7)</td>
<td>1.35 (2.99)</td>
<td>2.00 (3.71)</td>
</tr>
<tr>
<td></td>
<td>Baseline age_a,b,c</td>
<td>8.6 (6.0 - 20.4)</td>
<td>9.2 (3.8)</td>
<td>8.5 (3.6)</td>
</tr>
<tr>
<td></td>
<td>Baseline FEV1 (% predicted)_a,b,c</td>
<td>89.2 (21.6)</td>
<td>66.1 (19.6)</td>
<td>89.9 (15.4)</td>
</tr>
<tr>
<td></td>
<td>MRSA</td>
<td>At baseline*</td>
<td>4,495 (18.6%)</td>
<td>1,615 (20.6%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ever during follow-up*</td>
<td>19,449 (80.3%)</td>
<td>5,566 (91.9%)</td>
</tr>
<tr>
<td>Microbiology</td>
<td>Pa</td>
<td>At baseline*</td>
<td>1,247 (5.1%)</td>
<td>366 (6.0%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ever during follow-up*</td>
<td>10,897 (45.0%)</td>
<td>2,997 (49.5%)</td>
</tr>
<tr>
<td></td>
<td>Impaired Glucose Tolerance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>At baseline</td>
<td>67 (0.3%)</td>
<td>22 (0.4%)</td>
<td>33 (0.3%)</td>
</tr>
<tr>
<td></td>
<td>Ever during follow-up*</td>
<td>3,220 (13.3%)</td>
<td>736 (12.2%)</td>
<td>1,596 (13.2%)</td>
</tr>
<tr>
<td></td>
<td>CFRD diagnosis</td>
<td>At baseline_a,b,c</td>
<td>117 (6.5%)</td>
<td>44 (0.7%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ever during follow-up*</td>
<td>5,580 (24.3%)</td>
<td>2,347 (38.7%)</td>
</tr>
<tr>
<td></td>
<td>Pancreatic enzymes</td>
<td>22,955 (94.7%)</td>
<td>5,987 (98.8%)</td>
<td>11,471 (94.7%)</td>
</tr>
<tr>
<td></td>
<td>Ivacaftor*</td>
<td>Ever during follow-up*</td>
<td>1,087 (4.4%)</td>
<td>107 (1.8%)</td>
</tr>
<tr>
<td></td>
<td>Lumacaftor/Ivacaftor*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ever during follow-up*</td>
<td>3,930(16.2%)</td>
<td>697 (11.5%)</td>
<td>2,214 (18.3%)</td>
</tr>
<tr>
<td>PEx frequency in year prior to baseline</td>
<td>None_a,b,c</td>
<td>21,130 (87.2%)</td>
<td>5,030 (83.0%)</td>
<td>10,551 (87.1%)</td>
</tr>
<tr>
<td></td>
<td>1_a,b,c</td>
<td>1,049 (4.3%)</td>
<td>327 (5.4%)</td>
<td>555 (4.6%)</td>
</tr>
<tr>
<td></td>
<td>2_a,b,c</td>
<td>287 (1.2%)</td>
<td>84 (1.4%)</td>
<td>148 (1.2%)</td>
</tr>
<tr>
<td></td>
<td>3 or more_a,b,c</td>
<td>1,762 (7.3%)</td>
<td>616 (10.2%)</td>
<td>860 (7.1%)</td>
</tr>
<tr>
<td>Hospital visits in the year prior to baseline visit</td>
<td>None_a,b,c</td>
<td>20,115 (83.0%)</td>
<td>4,841 (79.9%)</td>
<td>10,000 (82.5%)</td>
</tr>
<tr>
<td></td>
<td>1_a,b,c</td>
<td>1,147 (4.7%)</td>
<td>305 (5.0%)</td>
<td>605 (5.0%)</td>
</tr>
<tr>
<td></td>
<td>2_a</td>
<td>463 (1.9%)</td>
<td>130 (2.1%)</td>
<td>237 (2.0%)</td>
</tr>
<tr>
<td></td>
<td>3 or more_a,b,c</td>
<td>2,503 (10.3%)</td>
<td>781 (12.9%)</td>
<td>1,272 (10.5%)</td>
</tr>
<tr>
<td>Smoke exposure*</td>
<td>At baseline</td>
<td>1,296 (5.3%)</td>
<td>266 (4.4%)</td>
<td>292 (4.8%)</td>
</tr>
<tr>
<td></td>
<td>Ever during follow-up*</td>
<td>8,130 (33.6%)</td>
<td>1,845 (30.5%)</td>
<td>4,227 (34.9%)</td>
</tr>
<tr>
<td></td>
<td>Lung transplant during follow-up*</td>
<td>1,810 (7.5%)</td>
<td>1,251 (20.7%)</td>
<td>523 (4.3%)</td>
</tr>
<tr>
<td></td>
<td>Alive through follow-up*</td>
<td>20,380 (84.1%)</td>
<td>3,519 (58.1%)</td>
<td>10,982 (90.7%)</td>
</tr>
</tbody>
</table>
Mean (SD) and n (%) are reported for continuous and categorical variables, respectively. P-values from Wilcoxon rank sum test or chi-square test. Statistical significance of comparisons (P-value < 0.05) marked as aEarly versus Middle; bEarly versus Late; cMiddle versus Late for continuous variables and *evidence of overall association between phenotype and categorical variable. ^Indicates insufficient sample size or censoring prohibiting standard statistical comparison. Abbreviations include CFRD = cystic fibrosis-related diabetes; MRSA = methicillin-resistant Staphylococcus aureus; Pa = Pseudomonas aeruginosa; PEx = pulmonary exacerbation. ?Baseline use of modulator therapies was suppressed for patient privacy purposes due to low cell counts (< 5 subjects). ##Types of smoke exposure (firsthand, secondhand and within household) were combined due to low cell counts.
Table 2. Baseline geomarker characteristics, overall and across social-environmental phenotypes.

<table>
<thead>
<tr>
<th></th>
<th>Overall cohort (N = 24,228)</th>
<th>Early (n = 6,057)</th>
<th>Middle (n = 12,114)</th>
<th>Late (n = 6,057)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deprivation index</td>
<td>0.341 (0.0983)</td>
<td>0.359 (0.100)</td>
<td>0.340 (0.0972)</td>
<td>0.324 (0.0950)</td>
</tr>
<tr>
<td>Total crime</td>
<td>87.3 (59.7)</td>
<td>90.4 (57.8)</td>
<td>86.6 (59.2)</td>
<td>85.6 (62.5)</td>
</tr>
<tr>
<td>Landcover</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Green</td>
<td>83.4 (20.9)</td>
<td>83.1 (21.5)</td>
<td>83.6 (20.8)</td>
<td>83.2 (20.3)</td>
</tr>
<tr>
<td>% Impervious</td>
<td>15.9 (18.1)</td>
<td>16.1 (18.6)</td>
<td>15.7 (18.1)</td>
<td>16.2 (17.8)</td>
</tr>
<tr>
<td>% Tree canopy</td>
<td>26.8 (21.6)</td>
<td>27.1 (22.3)</td>
<td>26.7 (21.4)</td>
<td>26.6 (21.4)</td>
</tr>
<tr>
<td>Respiratory hazard index</td>
<td>1.57 (0.811)</td>
<td>1.59 (0.791)</td>
<td>1.56 (0.815)</td>
<td>1.58 (0.823)</td>
</tr>
<tr>
<td>Ozone concentration</td>
<td>46.2 (6.69)</td>
<td>46.4 (6.85)</td>
<td>46.1 (6.62)</td>
<td>45.9 (6.65)</td>
</tr>
<tr>
<td>PM2.5 concentration</td>
<td>9.64 (1.58)</td>
<td>9.68 (1.57)</td>
<td>9.63 (1.58)</td>
<td>9.63 (1.57)</td>
</tr>
<tr>
<td>Traffic proximity</td>
<td>74.4 (128)</td>
<td>75.3 (121)</td>
<td>73.1 (131)</td>
<td>76.1 (128)</td>
</tr>
<tr>
<td>Diesel particulate matter</td>
<td>0.695 (0.615)</td>
<td>0.696 (0.606)</td>
<td>0.689 (0.623)</td>
<td>0.705 (0.608)</td>
</tr>
<tr>
<td>Primary roadways</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>19100 (27600)</td>
<td>19000 (27800)</td>
<td>19400 (28900)</td>
<td>18400 (26300)</td>
</tr>
<tr>
<td>Density</td>
<td>1.24 (2.75)</td>
<td>1.30 (2.87)</td>
<td>1.21 (2.70)</td>
<td>1.23 (2.72)</td>
</tr>
<tr>
<td>Secondary roadways</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>59300 (73200)</td>
<td>63100 (76700)</td>
<td>59900 (74600)</td>
<td>54400 (66400)</td>
</tr>
<tr>
<td>Density</td>
<td>2.54 (4.38)</td>
<td>2.61 (4.45)</td>
<td>2.59 (5.41)</td>
<td>2.46 (4.56)</td>
</tr>
</tbody>
</table>

Mean (SD) are reported. P-values from Wilcoxon rank sum test. Statistical significance of comparisons (P-value < 0.05) marked as aEarly versus Middle; bEarly versus Late; cMiddle versus Late.
Table 3. Clinical translation of social-environmental phenotypes.

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Low social-environmental adversity&lt;sup&gt;a&lt;/sup&gt;</th>
<th>High social-environmental adversity&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
</table>
| Early       | • Proactive indication of clinical risk factors identified in this study, including older age, established lung infections, comorbidities, history of PEx;  
• Hispanic ethnicity (McGarry, Neuhaus et al. 2017);  
• possibly ultra-rare mutations (those with no F508del alleles, modulator ineligible) | • Prone to higher community deprivation and total crime, may consider more frequent psychosocial assessment (Oates and Schechter 2016)  
• Also close proximity to primary roadways, may consider HEPA filters in the home. Needs further study; asthma findings have been mixed (Park, Lee et al. 2021) |
| Middle      | • Routine monitoring of microbiology and PFT per care guidelines (Foundation 2022) | • Most susceptible to respiratory hazards, may consider proper masking during spikes in air pollution levels to avoid triggering PEx (e.g.: in CF (Goeminne, Kicinski et al. 2013) or asthma severity during wildfires (McArdle, Dowling et al. 2023)) |
| Late        | • Avoid untreated declines in individuals with higher lung function (Morgan, Wagener et al. 2013) | • High levels of traffic-related air pollution, may require air quality policies in urban areas (Khreis, Sanchez et al. 2023) |

<sup>a</sup>Includes measures to improve lung function that have been identified in the absence of social and environmental exposure data.  
<sup>b</sup>Potential clinical actions are based on results from this study and other literature; interventional studies are needed for cystic fibrosis.  
Purple arrow denotes that these interventions could be applied amongst all three decline phenotypes depending on personalized factors. Each horizontal arrow and light-to-dark color shading illustrates, for a given phenotype, that the interventions specific to the category of low social-environmental adversity could be layered for deployment with interventions in the corresponding high social-environmental adversity level. Abbreviations: PEx = pulmonary exacerbation; PFT = pulmonary function testing.
Figure 1. Lung function trajectories by age and social-environmental adversity.

Three-dimensional plotting is used to examine how (A) lung function trajectory (FEV₁, % predicted) and (B) rate of change in lung function trajectory (% predicted/year) vary over age (the time variable, measured in years) and extent of social-environmental adversity (higher values imply more negative environmental exposure). Blue areas imply higher lung function or rate of change, while red areas indicate lower lung function or more rapid decline. Results were obtained as fitted curves from covariate-adjusted functional principal components analysis (see Study Design and Methods section). FEV₁ (% pred) indicates forced expiratory volume in 1 second of % predicted. Sub-plots by age are shown as supplemental material (e-Figures 4 and 5).
Figure 2. Peak decline in lung function over age according to social-environmental phenotype.

Average rates of change (y-axis) over age (the time variable, measured in years) are shown for early (A and D), middle (B and E), and late (C and F) environmental phenotypes of rapid decline. Rate of change is shown for each phenotype sub-grouped by degree of social-environmental adversity (lower and higher correspond to blue dashed and red dot-dashed curves, respectively). “X” is used to mark coordinates for the average timing of peak decline and estimated age at which it occurred. Results were obtained by differentiating the fitted curves from covariate-adjusted functional principal components analysis (see Study Design and Methods). FEV₁ (% pred/year) indicates annualized rate of change forced expiratory volume in 1 second of % predicted.
Figure 3. Geographic variation of social-environmental adversity and rapid lung function decline.

Bivariate quantities are shown representing extent of social-environmental adversity and severity of peak decline in lung function in the continental U.S. Higher values of social-environmental adversity imply more negative social and/or environmental exposure levels (vertical arrow pointing upward), and more negative values of severity of peak decline imply greater maximal loss of lung function (horizontal arrow pointing leftward). Dark blue (upper left corner of the grid) represents extremely rapid decline and worst environmental adversity, while gray (lower right corner of the grid) corresponds to at or below average rate of decline (-1.5% predicted/year) with least degree of social-environmental adversity. Results were obtained from two-stage cluster analysis (see Study Design and Methods). Estimates as shown on the map are aggregated to three-digit ZIP Codes and displayed using HIPAA Safe Harbor Guidelines. A three-digit ZIP Code was colored black if it did not contain either (1) a population of at least 20,000 residents or (2) a sufficient number of residents in the analysis cohort to make an estimate.
Declaration of competing interest: Author RDS serves on the Cystic Fibrosis Foundation Patient Registry Committee. The remaining authors have no conflicts of interest to report.