research-quality data in the ILD population. Our objectives were to (1) identify ILD patients and extract clinical data from an EHR system and (2) assess the performance of ILD data capture. METHODS/STUDY POPULATION: Case validated algorithms were implemented to identify patients from the University of California San Francisco EHR and extract key ILD clinical information including, demographic variables, process measures and patient outcomes. Key clinical information were defined based on consensus statements and ILD clinical trials. A subset of ILD patients, had variables recorded in both the EHR and a separate ILD longitudinal research database. The completeness of EHR data capture and level of agreement were compared between three data collection methods: (1) data manually and systematically collected for an ILD research database (gold standard), (2) data automatically extracted from structured fields in the EHR, and (3) data extracted from unstructured data sources. RESULTS/ANTICIPATED RESULTS: We identified 5857 ILD patients in the EHR, of which 2100 patients had data available in the both the EHR and research database. Baseline demographic variables, co-morbidities, use of diagnostic testing, pharmacotherapy were accurately extracted from structured fields. Outcome measures, including lung physiology, radiographic patterns, pathology results, and health related quality of life (HRQoL) were unevenly extracted from structured fields alone. With the exception of HRQoL, these measures were accurately captured in unstructured EHR sources. Notably, certain metrics were better defined in the EHR, including health care resource utilization metrics, acute exacerbations, medication side effects, supplemental oxygen use and specialty care referrals (rheumatology, lung transplant, palliative care, etc). DISCUSSION/SIGNIFICANCE OF IMPACT: A large real-world ILD cohort can be algorithmically extracted from the EHR along with key clinical variables with accuracy comparable to protocol-driven research databases. Rigorous assessment of the types of disease-specific variables that are present in EHR-derived data will inform future interventions to improve the fidelity, accessibility and use of the EHR in clinical research.

4471

Interactions of the Infant Nasopharyngeal Microbiota and Subjects' Clinical Traits in Development of Viral Upper Respiratory Tract Infections and Acute Otitis Media

Kamil Khanipov[1], George Golovko[1], Anna Nia[1], Lorraine Evangelista[1], and Yuriy Fofanov[1]
[1]University of Texas Medical Branch

OBJECTIVES/GOALS: Identify the interactions between nasopharyngeal bacterial pathogens, commensals, and patient clinical characteristics in relation to the development of viral upper respiratory tract infections (URI) and acute otitis media (AOM) in infants. METHODS/STUDY POPULATION: The subjects were part of a prospective, longitudinal study (2008–2014) of infants to evaluate the prevalence and risks for the development of URI and AOM. Healthy infants (n = 362) were enrolled from near birth and followed to the first episode of AOM up to 12 months of age. Nasopharyngeal specimens and clinical traits were collected at monthly intervals between 1-6 months, month 9, and during viral URI episodes. Subjects were closely followed for AOM development. 16S rRNA sequencing was performed on the nasopharyngeal swabs to identify their bacterial composition. Multidimensional (2, 3, and 4 dimensional) co-presence, co-exclusion, and one-way relation patterns were identified between the microbiome compositions, health status, and other collected clinical traits. RESULTS/ANTICIPATED RESULTS: We analyzed 971 specimens collected monthly and during URI and AOM episodes from 139 infants. Of the 139 enrolled subjects, 96% had 2 or more healthy samples, 77% contracted URI/AOM during the study period, and 60% had at least 1 healthy sample before URI/AOM onset. Otopathogens (Moraxella, Haemophilus, and Streptococcus), Staphylococcus, and Pseudomonas were the most common pathogenic genera. Corynebacterium, Dolosigranulum, and Acinetobacter were 3 most abundant commensal bacterial genera. Samples from infants with AOM in the first year had a significantly higher relative abundance of Haemophilus, Enterobacter, and Yersinia, and lower relative abundance of Corynebacterium, and Pseudomonas compared to samples from infants who did not develop AOM. DISCUSSION/SIGNIFICANCE OF IMPACT: Identification of complex multidimensional interaction patterns within microbial communities and environmental factors is vital to understanding disease onset risk and prevention. Prophylactic microbiome and environmental factor modulation between enterotypes could be used to reduce URI/AOM onset in infants.

4088

Longitudinal Assessment of Metabolic Syndrome as a Modifiable Risk factor of World Trade Center Particulate Matter Exposure Associated Lung Disease

Sophia Kwon[1], Myeonggyun Lee[1], Theresa Schwartz[1], Rachel Zeig-Owens[1], David Prezant[1], Mengling Liu[1], and Anna Nolan[1]
[1]NYU School of Medicine

OBJECTIVES/GOALS: Metabolic syndrome (MetSyn) is a risk for World Trade Center-Lung Injury (WTC-LI; defined as developing FEV^1<lower limit of normal [LLN]). Metabolic health is a modifiable disease risk factor. We propose to characterize how time-dependent covariates of MetSyn are longitudinally associated with WTC-LI. METHODS/STUDY POPULATION: WTC-particulate exposed firefighters, consented, with pre-9/11 FEV_1 LLN (N = 5,746). Data assessed from last pre-9/11 till August 1, 2017. Longitudinal MetSyn characteristics were assessed using 3 models: ***i.*** A linear mixed effect model to assess the effect size of longitudinal MetSyn and its components on longitudinal FEV_1% predicted as an outcome; ***ii.*** a time-dependent Cox regression to assess the associations of MetSyn to time of onset of WTC-LI; iii. a novel, partially linear single index regression model with repeatedly measured MetSyn to assess their joint effects and delineate their relative contribution on the longitudinal lung function in the WTC-FDNY cohort. RESULTS/ANTICIPATED RESULTS: In **Model I**, BMI 30 kg/m^2 had the largest effect size compared to ever-smoking, with −2.524 (95% CI: −2.708,−2.340) compared to −1.681(−2.325,−1.038) respectively. Having MetSyn, defined as 3/5 risk factors, had an effect size of −2.319(−2.526,−2.112). In **Model II**, hazards of triglycerides 150mg/dL were highest at 1.497(1.336, 1.677), followed by BMI 30 kg/m^2 at 1.406(1.256, 1.575), and HDL<40mg/dL 1.355(1.176-1.561), compared to ever-smoking (1.201, p = 0.002). Having high exposure to PM by being present in the morning of 9/11 was a significant covariate only in Model II investigating HDL<40mg/dL or triglycerides 150mg/dL. **Model III** The proposed methods will be applied to our cohort study. DISCUSSION/SIGNIFICANCE OF IMPACT: MetSyn is both a predictor and concurrent marker of WTC-LI. The single index model can not only reduce dimensionality of the covariates, but also provides efficient estimates of the joint MetSyn effects, allowing linear or nonlinear effects. Future studies

https://doi.org/10.1017/cts.2020.179 Published online by Cambridge University Press