

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<sup>1</sup>, George Golovko<sup>1</sup>, Anna Nia<sup>1</sup>, Lorraine Evangelista<sup>1</sup>, and Yuriy Fofanov<sup>1</sup>

<sup>1</sup>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<sup>1</sup>, Myeonggyun Lee<sup>1</sup>, Theresa Schwartz<sup>1</sup>, Rachel Zeig-Owens<sup>1</sup>, David Prezant<sup>1</sup>, Mengling Liu<sup>1</sup>, and Anna Nolan<sup>1</sup>

<sup>1</sup>NYU School of Medicine

**OBJECTIVES/GOALS:** Metabolic syndrome (MetSyn) is a risk for World Trade Center-Lung Injury (WTC-LI; defined as developing FEV<sub>1</sub> < 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<sub>1</sub> 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<sub>1</sub>% 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<sup>2</sup> 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<sup>2</sup> 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

will investigate dietary intervention as a potential disease-modifying factor. CONFLICT OF INTEREST DESCRIPTION: NA, nothing to disclose.

4183

### Micro-consults: An effective tool for meeting statistical support needs in an academic medical research center

Sandra Taylor<sup>1</sup>, and Susan L. Stewart<sup>1</sup>

<sup>1</sup>University of California, Davis

**OBJECTIVES/GOALS:** Access to biostatistics expertise is essential for a successful clinical and translational research program. However, demand for statistical support at academic research centers can strain the capacity of biostatistics units. Our objective was to efficiently increase access to statistical expertise. **METHODS/STUDY POPULATION:** In cooperation with the Cancer Center Biostatistics Shared Resource, we replaced an informal 1-hour drop-in consultation program with structured office hours to provide statistical support to clinical and translational researchers at the University of California, Davis Medical Center. We doubled office hours to 2 hours per week and established six 20-minute appointments. Two Ph.D. level statisticians staff office hours. Researchers schedule appointments through Acuity Scheduling, a free on-line resource. Availability of the service is advertised monthly by sending an informational flyer to various university listservs. **RESULTS/ANTICIPATED RESULTS:** Prior to implementing the program in 2014, we averaged 91 office hour consults per year. Subsequently, consultations jumped to 171 in 2014 and have averaged 150 per year since then. Office hours attract students, residents, staff and faculty from a wide range of disciplines including the Schools of Medicine, Nursing, Veterinary Medicine and basic science departments. Project types span the clinical and translational spectrum covering lab, animal, clinical and population-level studies. Most consults related to data analysis and interpretation (57%) followed by sample size calculations/study design (29%) and response to reviewers (4%), with general statistical advice as the remainder. **DISCUSSION/SIGNIFICANCE OF IMPACT:** With 6 micro-consults per week, we can meet with many investigators and triage their statistical support needs. This program has proved very popular and was highly rated in a recent user survey, with several investigators noting that the consults facilitated successful publications and proposals.

4494

### Predictors of Reintubation After Cardiac Surgery

Robert Edward Freundlich<sup>1</sup>, Gen Li<sup>1</sup>, Jonathan P Wanderer<sup>1</sup>, Frederic T Billings<sup>1</sup>, Henry Domenico<sup>1</sup>, Daniel Byrne<sup>1</sup>, and Pratik Pandharipande<sup>1</sup>

<sup>1</sup>Vanderbilt University Medical Center

**OBJECTIVES/GOALS:** We modeled risk of reintubation within 48 hours of cardiac surgery using variables available in the electronic health record (EHR). This model will guide recruitment for a prospective, pragmatic clinical trial entirely embedded within the EHR among those at high risk of reintubation. **METHODS/STUDY POPULATION:** All adult patients admitted to the cardiac intensive care unit following cardiac surgery involving thoracotomy or sternotomy were eligible for inclusion. Data were obtained from

operational and analytical databases integrated into the Epic EHR, as well as institutional and departmental-derived data warehouses, using structured query language. Variables were screened for inclusion in the model based on clinical relevance, availability in the EHR as structured data, and likelihood of timely documentation during routine clinical care, in the hopes of obtaining a maximally-pragmatic model. **RESULTS/ANTICIPATED RESULTS:** A total of 2325 patients met inclusion criteria between November 2, 2017 and November 2, 2019. Of these patients, 68.4% were male. Median age was 63.0. The primary outcome of reintubation occurred in 112/2325 (4.8%) of patients within 48 hours and 177/2325 (7.6%) at any point in the subsequent hospital encounter. Univariate screening and iterative model development revealed numerous strong candidate predictors (ANOVA plot, figure 1), resulting in a model with acceptable calibration (calibration plot, figure 2),  $c = 0.666$ . **DISCUSSION/SIGNIFICANCE OF IMPACT:** Reintubation is common after cardiac surgery. Risk factors are available in the EHR. We are integrating this model into the EHR to support real-time risk estimation and to recruit and randomize high-risk patients into a clinical trial comparing post-extubation high flow nasal cannula with usual care. **CONFLICT OF INTEREST DESCRIPTION:** REF has received grant funding and consulting fees from Medtronic for research on inpatient monitoring.

4099

### Principles of Statistical Education for Translational Scientists in the Age of Rigor, Reproducibility, and Reporting

Emilia Bagiella, PhD<sup>1</sup>, Paul Christos<sup>1</sup>, Mimi Kim, ScD<sup>2</sup>, Shing Lee<sup>1</sup>, Roger Vaughan<sup>3</sup>, and Judy Zhong<sup>4</sup>

<sup>1</sup>Mount Sinai School of Medicine; <sup>2</sup>Albert Einstein College of Medicine; <sup>3</sup>Rockefeller University; <sup>4</sup>New York University General Clinical Research Center

**OBJECTIVES/GOALS:** To describe principles, best practices, and techniques recommended to instill deep understanding of the application and interpretation of statistical techniques and statistical inference among translational scientists and trainees, that best support the concepts of scientific Rigor, Reproducibility and Reporting. **METHODS/STUDY POPULATION:** Each of the six New York City Area Biostatistics, Epidemiology and Research Design (BERD) resources have strong educational programs, novel curricular components, and creative strategies, implemented by award winning educators. To capitalize on shared knowledge, innovation, and resources, the six teams formed the **New York City Area BERD Collaborative (NYC-ABC)** comprised of BERD resources from Mt. Sinai, Cornell, Einstein, Columbia, Rockefeller, and NYU. The collaborative suggests principles, concepts, tools and approaches to support the concepts of scientific Rigor, Reproducibility and Reporting in translational science. **RESULTS/ANTICIPATED RESULTS:** Principles:

- Value of team science approach and including biostatisticians early and often.
- Carefully designing experiments to reduce bias and increase precision.
- Trainees' focus is often on "statistical significance" and the p-value. Consequences of data dredging/p-hacking, and the impact of sample size and other factors on statistical significance.