001

Quantifying Lead-Time Bias in Risk-Factor Studies of Cancer through Simulation. *Rick J. Jansen, Bruce H. Alexander, Kristin E. Anderson, Timothy R. Church

Purpose Lead-time is inherent in early detection and creates bias in observational studies of screening efficacy, but its potential to bias effect estimates in risk-factor studies is not always recognized. We describe a form of this bias that conventional analyses cannot address and develop a model to quantify it. Methods: Surveillance Epidemiology and End Results (SEER) data form the basis for estimates of age-specific preclinical incidence and log-normal distributions describe the preclinical duration distribution. Simulations assume a joint null hypothesis of no effect of either the risk factor or screening on the preclinical incidence of cancer, and then quantify the bias as the risk-factor odds ratio (OR) from this null study. This bias can be used as a factor to adjust observed OR in the actual study. Results: Results showed that for this particular study design, as average preclinical duration increased, the bias in the total-physical-activity OR monotonically increased from 1% to 22% above the null, but the smoking OR monotonically decreased from 1% above the null to 5% below the null. Conclusion: The finding of nontrivial bias in fixed risk-factor effect estimates demonstrates the importance of quantitatively evaluating it in susceptible studies.

002

Synthetic likelihood approach for quantifying multi-scale epidemic processes from large and complex data sets. *C. M. Barbu, K. Sethuraman, J. Manne, J. E. Quintanilla Calderón, M. Z. Levy

Adequate planning of epidemic control strategies often relies on correct estimation of the frequency of agent dispersal at multiple scales, including rare long range events. Stochastic models are well suited to handle such complex processes but tend to present intractable likelihoods, making parameter estimation difficult. A recent solution, the synthetic likelihood approach, aims to represent data through summary statistics that follow well-known multivariate distributions to compute a usable likelihood. Building on the original proof of concept on stochastic time-series, we expanded the idea to spatio-temporal processes. We further propose that the synthetic likelihood approach can be used to infer the relative frequency of dispersal at different scales from spatial patterns present in cross-sectional data. In a Bayesian framework, this information can be combined with repeated observations at a smaller scale to infer the frequency of medium and large scale dispersal events. We explored the feasibility of this approach by quantifying the frequency of dispersal of Triatoma infestans, the main vector of Chagas disease south of the equator, at multiple scales. We jointly analyzed infestation data from 400 households observed twice two years apart and cross-sectional infestation data from 25,000 additional households collected during vector control operations in Arequipa, Peru. We found that our estimates of dispersal frequencies are compatible with genetic diversity patterns observed using microsatellite markers. This experimental validation together with rigorous statistical coverage checks demonstrates the validity of the approach to characterize multi-scale epidemic processes.

003

Broken Windows by Way of Microsoft Windows: Ecometrics and Validity of a Neighborhood Physical Disorder Measure Constructed by Virtual Street Audit. *Stephen J. Mooney, Michael D. M. Bader,
Gina S. Lovasi, Kathryn M. Neckerman, Andrew G. Rundle, Julien O. Teitler

BACKGROUND: Physical characteristics of neighborhoods, including physical disorder (or 'broken windows'), have been difficult to measure objectively and reliably in multiple locales, limiting the use of such measures in neighborhood health research. Google Street View provides a novel source of imagery for virtual street audits to assess neighborhood physical characteristics efficiently. We evaluated the ecometric properties of a neighborhood physical disorder measure constructed from virtual street audit data. METHODS: Eleven trained auditors assessed nine previously validated items developed to capture physical disorder (e.g. litter, graffiti, abandoned buildings, etc.) on 1,826 block faces using Google Street View imagery from four US cities (San Jose, California; Detroit, Michigan; New York, New York; and Philadelphia, Pennsylvania). We constructed a 2-parameter Item Response Theory scale to estimate latent levels of disorder on each audited block face and used a geostatistical method known as kriging to interpolate a physical disorder level with confidence estimates for any point in each city. RESULTS: Kappa scores for audit items ranged from 0.34 to 0.80, indicating fair to substantial agreement. Internal consistency reliability of the resulting scale was 0.93. The final interpolated measure of disorder demonstrated convergent validity with US Census data: it was positively correlated with unemployment and housing vacancy and negatively correlated with owner-occupied housing. DISCUSSION: These results illustrate a novel method to combine observed street audits items to construct a valid, reliable, and spatially flexible measure of neighborhood physical disorder.

004

Beta-diversity of upper digestive tract microbiome is associated with obesity S.W. Lin, N.D. Freedman, J. Shi, S.M. Dawsey, C.C. Abnet

Recent studies implicate the gut microbiome in the etiology of obesity, but its contribution remains uncertain due to conflicting results. Previous studies focused on the fecal microbiome, while few studies have examined the role of microbial diversity at other body sites. To explore the association between obesity and the microbial diversity of the upper digestive tract, we studied in cross-section a cohort of 659 healthy Chinese adults (ages 34 to 67 years) with body mass index (BMI) range of 15 to 36. We characterized the upper gastrointestinal tract microbiome using the HOMIM DNA microarray, which tests for ~300 bacterial species. We characterized alpha-diversity using taxa counts and created beta-diversity metrics using an unweighted unifrac distance matrix of pairwise comparisons of the whole cohort. We performed unsupervised clustering in this matrix and found support for three clusters. BMI was used as a continuous variable on the log scale in regression models adjusted for age, sex, tobacco, alcohol, and antibiotics. We found that alpha-diversity (numbers of genera, families, etc.) was not associated with BMI. However, beta-diversity, assessed as cluster membership (first compared with third cluster, p=0.01; second compared with third cluster, p=0.0002), was associated with BMI. The first (median BMI 22.4) and second (median BMI 22.6) clusters had lower median BMIs within the optimum Asian BMI range than the median BMI (23.6) in the third cluster, which is in the range of increased risk for obesity-related diseases in Asians. Our results demonstrate for the first time that beta-diversity in the upper digestive tract microbiome is associated with BMI. Future longitudinal, prospective studies should address whether the upper digestive tract microbiome plays a causal role for obesity.

005

Validation of a Method to Reconstruct Historical Smoking Prevalence Rates.*Usama Bilal, Esteve Fernandez, Paula Beltran, Ana Navas-Acien, Francisco Bolumar, Manuel Franco

Objective: Our objective was to assess the validity of a method to reconstruct historical smoking prevalence rates. Methods: We reconstructed smoking prevalence rates for each calendar year from 1940 to 2007 for men and women in Spain using data on ages of smoking initiation and cessation available in the Spanish National Health Surveys of 2003-2004 and 2006-2007. We adjusted the reconstructed prevalence for the excess mortality of smokers. To assess the validity of the reconstruction we computed the differences between the reconstructed (from 1987 to 2001) and the contemporary observed smoking prevalence measured in the Spanish National Health Surveys of 1987, 1993, 1995, 1997 and 2001. We also compared reconstructed smoking prevalence trends from 1940 to 1986 with 35-year lagged lung cancer mortality rates in Spain as a proxy for the real prevalence trends. We conducted sensitivity analyses to check the robustness of our assumptions: changing excess mortality ratios for the mortality correction, using either survey instead of two for the reconstruction and changing the lag of lung cancer mortality rates in the trend validation. Results: Reconstructed smoking prevalence rates compared to contemporary measured showed small differences in men (between -2.1% and 2.1%) and an overestimation in women (between +2.0% and +5.7%). Reconstructed smoking prevalence trends were significantly correlated with lagged lung cancer mortality trends ($p=0.004$ for men, $p<0.0001$ for women). Sensitivity analyses showed no major changes in our reconstruction or validation. Discussion: The reconstruction of smoking prevalence rates through this methodology offers a feasible tool for countries with no previous smoking surveys to understand their historical tobacco epidemic trends.

006

Application of latent growth curve modeling to 5 years of longitudinal data; a randomized community based trial for primary prevention of cardiovascular disease. Shervin Assari, Lakshmi Muthukumar, Jennifer Dykstra, Krista Bobo, Carla Schnieder, Peggy Manchester, Arthur Riba

A Cochrane systematic review published in 2011 reported that behavioral modifications might not be effective for primary prevention of cardiovascular disease. This report has resulted in considerable uncertainty regarding public health investments on life style modifications for primary prevention of cardiovascular disease. Healthy Heart Women is a five year randomized clinical trial (RCT) that tested the efficacy of a tailored multiple risk factor intervention (MRFI) for primary prevention of cardiovascular disease among a community sample of at risk women. Participants were women who did not have any cardiovascular disease, but were at high risk for cardiovascular disease. In this study 400 women were randomized into intervention (n=200, 84% Caucasian) or control (n=200, 77% Caucasian) groups. Intervention was composed of a multiple risk factor intervention that delivered tailored coaching, motivational interviewing and health communication. Primary outcomes included blood pressure, diabetes, diet, smoking, BMI, exercise, lipids, blood glucose, ESR, and CRP. Psychosocial outcomes such as anxiety, depression, quality of life, risk communication, and perceived change were also measured as possible mediators of change. Outcomes were measured at baseline, 6th month, 12th month, and then annually for the next 4 years. Non-linear modeling of data using latent growth curve modeling showed some effects for the intervention. Intervention resulted in significant improvements in diet, systolic blood pressure, diastolic blood pressure, risk communication, and general mental health. For many of these outcomes, modeling linear change failed to show an effect of intervention. Based on our findings, latent growth curve modeling might be a useful tool for analysis of data of long term clinical trials. Testing only linear changes may fail to reveal efficacy of some interventions.

007

An application of finite mixture models to longitudinal analyses of place-health dynamics*Peter Lekkas, Catherine Paquet, Mark Daniel

Within modern epidemiology the life course approach features prominently. A feature of the life course approach is that it seeks to contextualize exposures hierarchically and in reference to spatial and temporal factors unique to a population. Also embodied within this framework is recognition of the dynamic nature of the contexts within which people are embedded. Extant statistical approaches afford the capacity to generate insights into these changing landscapes so as to advance understanding of how 'context gets under the skin' to affect population-level health. Latent transition analysis, a longitudinal statistical model that explores change in a latent variable over time, will be presented as applied to small-area, suburban, residential contexts. In so doing this presentation aims to promulgate the utility of latent transition analysis as a modelling approach able to illuminate the 'life course of place' in analyses that seek to consider the affiliated health of people 'in place'. A special case of finite mixture models, latent transition analysis encapsulates two modelling approaches: latent class analysis and autoregressive modelling. Latent transition analysis can also be viewed as a structural equation model wherein latent class analysis forms the measurement model, identifying unique 'classes' of residential contexts at each point in time, whilst the autoregressive, structural, component substantiates discrete transitions among levels of the latent variable over time. Through an applied example, the latent transition analysis model will be presented in a pedagogical manner. We will demonstrate the process of model specification, identification and interpretation. In addition focus will be directed to the fundamental assumption underlying latent transition models, local independence, with specific reference to applications involving spatial data.

008

Positive and Negative Affect Following Marijuana Use: A Case-crossover Study Embedded within an Ecological Momentary Assessment Design. Craig S. Ross, MBA*, Daniel Brooks, DSc, Ann Aschengrau, ScD, Lydia A. Shrier, MD, MPH, Michael B. Siegel, MD, MPH, Janice Weinberg, ScD

Ecological Momentary Assessment (EMA) studies make use of portable devices such as smart phones to capture context-specific health behaviors and information about transient exposures in real-world settings. In this study, we sought to improve the ability to make causal inferences in such study designs by embedding a case-crossover structure within the EMA study. We propose to use a study of momentary affect and marijuana use to assess the association between marijuana use and subsequent momentary affect. Study participants were trained to use a portable data collection device to respond to signals generated randomly 4 to 6 times during their waking hours for a period of 2 weeks. A total of 3,594 reports were generated. At each report participants completed an assessment of the momentary affect and most recent marijuana use. The proposed case-crossover design treats the EMA reports as samples from each participants' base experience time to identify periods when affect is abnormally high or low relative to each person's average. These periods of unusually high or low affect are considered the outcomes of interest. We next sample from the participant's time preceding the outcome to identify time under the influence of marijuana. The discordant exposed and unexposed time periods are analyzed relative to the outcome to estimate the association. To develop unbiased estimates of the association, we will use three methods to analyze the data. First, we will examine first-occurrences only for each participant, as is typical in a case-crossover design. Next we used Within Subject Pairwise Resampling (WSPR), a bootstrap-type method to sample from the series of all outcomes for each person. Finally, we will use Weighted Estimating Equations (WEE) – a method based on marginal weights of the number of outcome events for each person.

009

Use of Arabin pessaries to prevent preterm delivery among twin pregnancies: Needs for a priori sub-group analyses targeting PreCocious Cervical Ripening. Qing Li*, Claudia Holzman, Nigel Paneth, Mathew Reeves, Zhiying You, Louis Keith

To evaluate whether use of Arabin pessaries is better to prevent preterm delivery (PTD) among sensitive subgroups targeting PreCocious Cervical Ripening (PCCR: multidimensional cervical changes prior to the onset of PTD including softening, shortening, funneling and dilation) compared to targeting short cervices alone. We conducted retrospective subgroup analyses of the nonrandomized concurrent control study of twin pregnancies, which were partially published (Arabin et al., 2003). Transvaginal ultrasonography (TVU) longitudinally assessed cervical length (CL) and funneling (CF) in two positions in 315 twin pregnancies at a Dutch center (1994-2001). Forty mothers placed with Arabin pessaries were matched with 40 mothers without pessaries through existing pairs and nearest neighbor matching by CL within 2 mm at the same gestational week period. A short cervix was defined as CL<38 mm (supine). PTD was defined as delivery before 34 weeks of gestation. Paired t-Test, McNemar's Test and conditional logistic regression were conducted to calculate p values, odds ratios (ORs) and 95% confidence intervals (CIs). The presence of CF occurred in 38% and 15% of pessary and control groups, respectively. Mean delivery was 35+6 vs. 34+3 weeks (p=0.01), and mean intervals between TVU and delivery were 72 vs. 62 days (p=0.01). PTD was 77% less frequent in the pessary group (OR=0.23, 95% CI 0.07-0.81) and 83% less in a subgroup of 34 pairs with short cervices (OR=0.17, 95% CI 0.04-0.75). Among 34 women with the pessary treatment and short cervices, those with CF had 3% less PTD compared to those without CF (1/14 [7.1%] vs 2/20 [10.0%], OR=0.97, 95% CI 0.79-1.19). Mothers of twins with Arabin pessaries had less PTD. Subgroups with short CL and CF appear amenable to pessaries. This exploratory analysis calls attention to quality and a prior biologic justification for subgroup analyses considering PCCR in new pessary trials in a consortium of a prospective meta-analysis.

010

The Interrelationships between Substance Abuse, Mental Health and Problem Behaviors in Illinois 12th Graders: A Latent Class Analysis. *B. Mutyala

To identify the patterns of substance abuse, mental health and problem behaviors in Illinois 12th graders (n=2272), latent class analysis (LCA) was performed on the 2012 Illinois Youth Survey data. The association between latent class membership and demographic variables such as gender, race and SES; and risk factors such as depression, suicidal thoughts, gambling and drunk driving was examined with multinomial logistic regression. A LCA model with three latent classes appears to best fit the data on Illinois 12th graders' use of alcohol, including binge drinking, cigarettes, marijuana, prescription pain killers, and other prescription drugs. The overall membership probabilities of Illinois 12th graders in latent classes one, two and three were 58.28%, 7.47% and 34.25% respectively. Illinois 12th graders in class one had a lower probability and those in class two had a higher probability of substance use, mental health problems and other problem behaviors. Illinois 12th graders in class three had a higher probability of use for alcohol, cigarettes and marijuana, but not for over the counter drugs or prescription drugs. Multinomial logistic regression results, with class one as the reference group, indicate that Illinois 12th graders who gambled, had suicidal thoughts or were involved in drunk driving had significantly higher odds of class two or class three membership. Male students also had significantly higher odds of membership in class two. There appears to be three distinct groups of substance abusing students in relationship to mental health and problem behaviors. These three latent classes should be taken into consideration when planning prevention and treatment programs.

011

Modelling and Forecasting of COPD Admissions in Portuguese Hospitals: a Box–Jenkins Methodology in Medical Research. M. Rodrigues*, P. Santana

In epidemiology, data often arise in the form of time series e.g. notifications of diseases, entries to a hospital, mortality rates etc. are frequently collected at weekly or monthly intervals. In the health sector where hospital admissions (HA) data, resulting from the monitoring of hospital services, can help predict the number of people presenting to hospitals, thereby offering an advantage for the planning of health service delivery requirements.The main objective of this study is to apply autoregressive integrated moving average (ARIMA) models to make predictions on the daily number of patients visiting the hospitals for Chronic Obstructive Pulmonary Disease (COPD) (ICD9 490-499) in the Portuguese Metropolitan Areas (PMAs). The health data considered in this study consist of HA in the Lisbon Metropolitan Area (LMA) and in Porto Metropolitan Area (PMA) for COPD. Data were collected from the information system database of the public hospitals. Six years of data from January 2003 to December 2009 data were used to develop a forecasting model for the following six consecutive months and were processed for validation. The Index of Agreement (ID), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) and average absolute percentage error (AAPE) were calculated between predicted and observed admissions to evaluate the predictability of the model. We found that the ARIMA (2,1,2) model was able to describe and predict the number of HA for COPD in MAL and ARIMA (2,0,1) model has been developed for the PMA. The models evaluation statistics suggest that considerably satisfactory real-time forecasts of HA can be generated using the Box–Jenkins approach. The implementation of a forecasting service for use by general practitioners in the PMAs may help reduce hospital admissions and associated costs.

012

Generalized Additive Models in Environmental Epidemiology: Investigate the Association between Air Pollution and Human Health Effects. M. Rodrigues*, P. Santana

Many epidemiologic studies demonstrated associations between high concentrations of air pollutants, hospital admissions, and emergency room visits (Dominici, 2005; Schwartz, 1994). The aim of this study is to evaluate the association between hospital admissions (HA) and concentrations of pollutants in ambient air for Lisbon Metropolitan Area (LMA) and Porto Metropolitan Area (PMA). The daily counts of admissions, by admit date, were computed for asthma (ICD9 493), pneumonia and influenza (ICD9 480-489), chronic obstructive pulmonary disease (COPD) (ICD9 490-499) and respiratory diseases (ICD9 460-519) for the period January 2003 to December 2010. The authors used daily air pollutants levels for particulate matter (PM10), ozone (O3), sulfur dioxide (SO2), nitrogen dioxide (NO2), and carbon monoxide (CO) in a Poisson regression model with generalized additive models (GAM) to adjust for effects of time trend, season, mean temperature, humidity and wind. Daily concentrations of each air pollutant were added separately into the core model to obtain the respective partial regression coefficients (β) and Relative Risks (RR). Relative risks were calculated for respiratory and cardiovascular disease admissions. Conclusively, this epidemiologic study of air pollution in Portugal suggests that short-term exposure to pollutants is associated with hospital admissions risks. Also, our findings indicate that policies oriented to increase the presence of green, such as the creation of "green belts" around cities, may be effective in mitigating pollution generated within urban zone and therefore in the number of hospitalizations for respiratory diseases.

References:

Dominici, F., McDermott, A., Daniels, M., Zeger, S.L. and Samet, J.M. (2005), "Revised analyses of the national morbility, mortality, and air pollution study: mortality, among residences of 90 cities", J. Toxicol. Environ. Health, 68(13-14), 1071-1092.

Schwartz, J. (1994), "Air pollution and daily mortality: a review and a meta-analysis", Environ. Res., 64, 36-52.

013

Time-Series Models: Particulate Air Pollution and Chronic Obstructive Pulmonary Disease in Portugal. *M. Rodrigues, P. Santana

A number of epidemiologic studies have reported variable associations between particulate matter with ≤ 10 μ/m3 (PM10) and the severity of Chronic Obstructive Pulmonary Disease (COPD) hospitalizations and mortality. The purpose of this study was to examined the association between PM10 concentrations and hospital admissions (HA) for COPD in Portuguese Metropolitan Areas (PMAs). The analytical approach used in the examinations of HA in context of weather was also adapted here. Daily admission counts for COPD for the period 1st January 2003 to 31 st December 2010 were analysed to selected air pollutants in PMA's using Sperman's Rank order correlation. Daily air pollution data from all monitoring stations for PM10 in Lisbon Metropolitan Area (LMA) and Porto Metropolitan Area (PMA) were obtained and was computed. Correlation analysis of HA in relation to PM10 concentrations was performed separately for LMA and LMA. The data were analyzed using generalized additive regression (GAM), adjusted for weather and time trend. The results of this investigation, together with results of previous research (Dominici et al., 2006; Rodrigues et al., 2013), demonstrate the significant effect of pollution on various health indicators within PMA's.

References:

Dominici, F., Peng, R.D., Bell, M.L., Pham, L., McDermott, A., Zeger, S.L. and Samet, J.M. (2006), "Fine particulate matter air pollution and hospital admissions for cardiovascular and respiratory diseases", J. Am. Med. Assoc., 295(10), 1127-1134.

Rodrigues, M., Santana, P. (2013), "Prediction of hospital admissions associated with PM10 and O3 pollution in metropolitan areas in Portugal using ARIMA models", European Journal of Epidemiology (2013) 28:S1–S270, DOI 10.1007/s10654-013-9820-0.