Symposium Title: The Promise and Peril of Randomness

Symposium Description (1000 max):

Randomness – and random error – are inescapable elements of the design, analysis and interpretation of epidemiologic studies, but the exact nature of the impact of randomness and its effects on our inference are often elided, or assumed to be captured in relatively simplistic representations of uncertainty. This symposium will delve deeper into the role randomness plays in epidemiology, exploring both threats to the validity of our studies as well as ways to exploit randomness to improve and validate our methods.

Presentation Titles and Allocated Time (90 minutes total)

Introduction: 5 minutes

"Causal & statistical challenges of infectious outcomes" (15 minutes) – Laura Balzer

Correlated outcomes in epidemiology are common. For example, individuals are nested within schools, neighborhoods, communities; and multiple measurements are made on a given individual in longitudinal studies. Infectious outcomes, including both those biologically or socially transmitted, pose additional challenges. Without careful consideration of the underlying causal structure, standard methods for dependent data are at risk of biased estimates and misleading inference, including underestimation of uncertainty. This talk explores the circumstances under which these misleading estimates arise, and suggests some potential methods for addressing correlated outcomes arising from biological or social contagion.

"Bad Doctors or Bad Luck?: Stochastic threats to common quality of care metrics" (15 minutes) – Eric Lofgren

Statistical metrics have entered widespread use in the evaluation of the quality of care delivered by hospitals. One such metric, the standardized ratio, where a hospital's performance on a given indicator (such as the level of antibiotic administration) is divided by an expected level generated from a statistical model, is widely used in the evaluation of both antibiotic stewardship programs and infection control activities. While attention has been given to improving the quality of the denominator of these ratios, it is widely assumed that the numerator is deterministic – that is hospitals doing the right things have desirable scores, and those delivering worse quality care have worse scores. In reality, especially on ward-level metrics, these measures are extremely noisy and vulnerable to random fluctuations. Using a simple model of a series of identical simulated ICUs, this talk explores the large amount of variability in these metrics for hospitals delivering the same quality of care. This talk also shows that these problems grow worse when more data is used because of the use of confidence intervals to identify well and poorly performing hospitals. The consequences for the evaluation of quality improvement programs and the financial penalization of hospitals is discussed.

Do these results seem random to you?: A validation approach for investigating multiple exposures using tree-based methods (15 minutes) – Jeanette Stingone

There is a growing interest in using machine learning approaches to generate hypotheses within high-dimensional epidemiologic data. This includes identifying combinations of multiple exposures associated with an outcome, to then be investigated using more targeted methods. Prior to committing resources to these targeted investigations, there is a need to determine if generated hypotheses are more than chance findings resulting from the large number of computations performed. Traditional validation techniques focus on predictive ability when applied to previously-unseen data; yet this may not be the most relevant metric when the proposed goal of the analysis is identification of combinations of exposures for further investigation. This talk will focus on a proposed method that seeks to leverage randomness as a way to evaluate identified exposure combinations generated by tree-based analytic methods. Just as random permutation of individual exposures is used to generate variable importance factors in a random forest, we propose to randomly permute the outcome variable in order to provide a comparison for results of identified combinations of exposures. We will demonstrate how this approach can provide a random threshold of combination frequency that can then be used to identify the exposure combinations to be targeted by future investigations.

"Accounting for selection bias in the measurement of outbreak intervention efficacy" (15 minutes) – Jon Zelner

A common approach to measuring the impact of interventions in the course of infectious disease outbreaks is to measure the daily evolution of a measure of average infectiousness, such as the basic reproduction number, R0, before and after the beginning of interventions. Change in this value over time is often attributed to the protective impact of interventions. In this talk simple models are used to explore the conditions under which another mechanism is likely to explain these effects: Specifically, when early observations represent anomalous 'superspreading' events that pushed the caseload above a threshold at which the outbreak became visible to public health surveillance. When this is followed by a process of regression towards the mean in which secondary cases have infectiousness closer to a population average or R0 near or below 1, we will see stochastic extinction of the outbreak often occurring around the time when interventions are applied. Using simulation of this process, this talk will quantify the bias in R0 estimates expected from data obtained in this fashion, and characterize the conditions in which this selection bias is most consequential.

"Causal identification and randomness: why 'chance confounding' is a contradiction in terms" (15 minutes) – Daniel Westreich

This talk will discuss causal identification conditions and a theory of how each can be conceived as an amalgam of random as well as systematic components.

Panel Discussion: 10 minutes