Published in final edited form as:

Epidemiology. 2018 January ; 29(1): 76–77. doi:10.1097/EDE.0000000000000768.

New Designs for New Epidemiology

Timothy L. Lash^a and Enrique F. Schisterman^b

^aDepartment of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA

^bEpidemiology Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD.

Resources to support epidemiologic research are almost always constrained, so efficient study designs that sample relatively informative individuals are crucial. In the descriptive setting, surveillance studies may be population-wide when the number of data elements is small and the ascertainment costs are low. Otherwise, they must rely on sampling strategies to select an informative subgroup, with rereighting commonly used to ensure generalizability to the broader population. In the analytic setting, cohort designs and case–control studies nested in cohorts follow a parallel progression.

Longitudinal and clustered outcome studies extend univariate studies by allowing researchers not only to examine exposure–outcome relationships between distinct individuals, but also within individuals (e.g., over time) or clusters (e.g., families). Even though such study designs are ubiquitous in public health and medical research, until recently, efficient sampling strategies for longitudinal and correlated data have not been well characterized. To encourage further development, the Epidemiology Branch of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development convened a group of experts in study design to push the boundaries in this important field. A large portion of the group’s work is presented in three articles published in this issue of *EPIDEMIOLOGY*.

In the article titled “Extending the case–control design to longitudinal data: stratified sampling based on repeated binary outcomes” by Schildcrout et al, the authors explain how to enhance case-control designs when longitudinal binary outcome data are already collected as part of a primary cohort study, but new and expensive exposure data must be retrospectively collected. By stratifying subjects into those who never, sometimes, and always experienced the event of interest during longitudinal follow-up, and then sampling individuals for whom the costly secondary exposure will be measured, one can substantially improve the study’s cost efficiency. The sampling strategies enable efficient estimation of the effects of time-varying and time-fixed covariates.

Correspondence to: Timothy L. Lash Department of Epidemiology Rollins School of Public Health Emory University 1518 Clifton Rd. NE, CNR 3 Atlanta, GA 30322 404.712.1270 tlash@emory.edu.

Disclosure: The authors report no conflicts of interest.

In a second article titled, “Outcome-related, auxiliary variable sampling designs for longitudinal binary data,” Schildcrout et al aimed to gain statistical efficiency by prospectively over-sampling relatively informative subjects. In this case, the case-control or sampling variable is measured at the screening visit, cases are sampled with high probability, controls are sampled with low probability, and individuals are then followed over time. Because, the case-control variable is closely related to the longitudinal binary outcome variable, the sample is enriched with events, so efficiency gains can be realized. The authors then describe sequential offsetted regression analysis that is a relatively novel analytic approach for valid estimation in this setting.

Finally, in the article titled “On the analysis of case-control studies in cluster-correlated data settings” by Haneuse and Rivera, the authors considered outcome dependent sampling in a case-control setting with cluster correlated data. This type of data is common where the cost for collecting individual level data is impossible and only aggregate data are available. With constrained research resources, this situation will become more common, especially in low- and middle income countries. The authors propose case-control sampling within clusters and use inverse-probability-weighted generalized estimating equations with a robust sandwich estimator to account for the specialized sampling scheme. They ultimately demonstrate the small sampling characteristics as well as the potential trade-offs associated with standard case-control sampling or when case-control sampling is performed within clusters.

In all three articles, the authors have developed novel sampling strategies to make the most efficient use of research resources. With the accumulation of longitudinal cohort data, many with linked archived biospecimens, strategies to selectively sample the most informative subjects is of ever-growing importance. Bioassays such as epigenetics methylation arrays that are prone to vary over time are often expensive, and biospecimens are a finite resource, so judicious use of these resources is imperative. The cost of collecting bioanalytical data provides just one compelling example for when sampling strategies are important to assure cost-efficient research. Nested case-control designs are the paradigm for efficient sampling, but repeated measurements of the outcome, informative variables beyond the outcome, and clustered data, add complexities to the conventional design. Use of more nuanced sampling strategies, such as the three described in this issue, to improve the information density of sampled participants promises to optimize the utility of scarce research resources. We encourage all epidemiologists to familiarize themselves with these new designs to pave the way for new epidemiology.

REFERENCES

1. JS Schildcrout, EF Schisterman, ND Mercaldo, et al. Extending the case-control design to longitudinal data: stratified sampling based on repeated binary outcomes. *Epidemiology*. 2017;29:67–75
2. JS Schildcrout, EF Schisterman, MC Aldrich, et al. Outcome-related, auxiliary variable sampling designs for longitudinal binary data. *Epidemiology*. 2017;29:58–66
3. Haneuse S, Rivera C. On the analysis of case-control studies in cluster-correlated data settings. *Epidemiology*. 2017;29:xxx