Data Science @ Westat

Complex Analytics

Westat leads the field in providing complex analytical services using cutting-edge techniques rooted in principles and best-practices of data science. We work to empower data through advanced analytics, machine learning, and artificial intelligence.

Illustrative Projects

By using techniques such as supervised and unsupervised machine learning (ML), predictive modelling, text mining and topic modeling, we offer innovations from data creation and harmonization, to automation and process efficiency improvements, to timely and credible policy insights. The following examples highlight techniques and tools that Westat brings to address our clients' questions.

Creating and harmonizing data to support policy and research-based analyses

- **The Recipient Epidemiology and Donor Evaluation-IV Pediatric (REDS-IV-P) Research Program.** In a project for the National Heart, Lung, and Blood Institute (NHLBI) designed to respond to potential threats to the safety of the blood supply and to address emerging research needs in transfusion medicine, Westat built a comprehensive Research Data Warehouse (RDW) comprised of electronic health record patient data from 22 hospitals, blood donor and blood product data from 4 blood centers, and omics data on donor samples using a customized genotype array. Data are sent to Westat using a secure data pipeline created by Westat, are stored on the Amazon Web Services (AWS) cloud, and harmonized using the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). Westat analyzes large data files to address key research questions in blood banking and transfusion medicine and inform blood policy decisions.

Key Benefits of Westat's Approach

- ▶ Data creation and harmonization
- ▶ Process and operational efficiency
- ▶ Analytics-informed assessment and evaluation
- ▶ Rapid hypothesis testing
- ▶ Credible evidence

Providing cost-effective and timely alternatives to manual methods of analysis

- **Drug Abuse Warning Network (DAWN).** For the Substance Abuse and Mental Health Services Administration (SAMHSA), Westat collects data in 50 hospitals nationwide to identify new and emerging drugs and use patterns and provide an early warning system for drug-related events. To ensure rigorous data quality and keep costs low, Westat developed machine learning models to review and route emergency department (ED) records to expert reviewers who must decide whether a drug caused or contributed to a person's ED visit. The advanced analytics models we developed assign a probability score indicating whether the ED visit is likely to be in scope for DAWN and the likely category of the visit. These models are retrained periodically to increase their efficiency and maintain high quality.

- **NSF Data Analytics and Evaluation Support Services Analytics for Decision Support.** Westat led a 'proof-of-concept' project for the National Science Foundation (NSF) to determine if natural language processing (NLP) algorithms can be trained to reduce the time it takes human coders to accurately produce meaningful results from numerous proposals. We determined that an NLP algorithmic process could replace 85-90% of the labor time and costs for producing analytic findings, with an accuracy of 88-94%, regardless of the complexity of the concepts being studied.

Increasing efficiency of assessment and evaluation through rapid hypothesis testing

- **The Efficacy of From Here to There (FH2T): A Dynamic Technology for Improving Algebraic Understanding.** Westat researchers conducted a methodological study for the Department of Education that compares two approaches for processing and analyzing qualitative data collected from a series of focus groups. Specifically, we compared the coding findings from traditional qualitative coding and topic modeling, and unsupervised machine learning methods. Participants in this study included approximately 3,000 middle school students drawn from approximately 150 mathematics classrooms in 15 public schools in San Antonio, Texas.

- **NSF Data Analytics and Evaluation Support Services Analytics for Decision Support.** Another avenue of investigation on this project for NSF was to assess if using complex analytics on existing program data can be used to provide faster answers before and/or in conjunction with funding deeper evaluations. We found that it is technically feasible to apply machine learning to structured NSF data to inform assessment and evaluation on an ongoing basis, for example to compare characteristics of funded and non-funded proposals; as well as to characteristics, strategies, and outcomes for NSF programs that emphasize broadening participation.

Improving policy and outcomes with credible evidence

- In the **COVID Vision Project** for Centers for Disease Control and Prevention (CDC), Westat uses machine learning-based methods and complex analytics (e.g., gradient boosted regression trees) to help CDC learn more about vaccine-induced immunity. These timely analyses on COVID-19 vaccine effectiveness inform public health models that determine key policies (such as 3rd dose/booster recommendations), guidance, and resource allocations. Westat staff have developed their own R packages to analyze the data, and manage a shared Git repository for central code distribution and reproducibility.

- **How Much Can Evidence from National Studies Improve Local Policy Decisions that Affect Youth?** A key decision faced by policymakers is whether to adopt an intervention. Multi-site randomized controlled trials (RCTs) provide unbiased impact estimates for sites in the study sample. However, their ability to produce accurate predictions of impact for individual sites outside the study is largely unknown, which raises questions about their value to local policymakers. In a project funded by the William T. Grant Foundation, Johns Hopkins University and Westat researchers use advanced analytic techniques, including cross-validation and Bayesian techniques to assess whether evidence from multi-site program impact evaluations can be used to inform policy decisions about youth programs in localities that did not participate in those studies.

- **The Efficacy of From Here to There (FH2T): A Dynamic Technology for Improving Algebraic Understanding.** In this Department of Education project, Westat researchers leveraged text analytics (e.g., sentiment analysis) in addition to standard statistical approaches to provide deeper insights and examine the efficacy of FH2T, an interactive touch-based application that allows students to dynamically interact with numbers and expressions to improve children's procedural fluency, flexibility, and conceptual understanding of algebraic ideas.

CONTACT

Sunitha Mathew
Vice President and Associate Director
SunithaMathew@westat.com

Andreea Balan-Cohen
Vice President and Associate Director
AndreeaBalan-Cohen@westat.com