Using Applied Regression and Biostatistics to Assess Outcomes of Community-Based Health Programs

OLAITAN KEMI ATOBATELE[1], OPEOLUWA OLUWANIFEMI AJAYI[2], AKONASU QUDUS HUNGBO[3], CHRISTIANA ADEYEMI[4]
[1]*Johns Hopkins University*
[2]*Amazing Grace Adult Home, Akure, Ondo State, Nigeria*
[3]*Saskatchewan Health Authority, Saskatoon, SK, Canada*
[4]*University of the Cordilleras. Baguio, Philippines*

Abstract- ***Assessing the outcomes of community-based health programs requires rigorous analytical methods that can account for complex, real-world variables influencing health interventions at the population level. Applied regression analysis and biostatistics offer robust tools for evaluating the effectiveness, efficiency, and equity of these programs by enabling precise estimation of intervention impacts while adjusting for confounding factors. Regression models, including linear, logistic, and multilevel regressions, facilitate the quantification of relationships between program interventions and health outcomes, such as service utilization, morbidity reduction, and behavioral change. These models allow researchers to control for socio-demographic, environmental, and programmatic variables, providing clearer causal inferences in non-randomized, observational study designs typical of community interventions. Biostatistical methods further enhance the evaluation process by guiding sampling strategies, determining appropriate statistical power, and addressing issues related to missing data, measurement errors, and intra-cluster correlations. Advanced techniques, such as propensity score matching and difference-in-differences (DiD) analysis, strengthen causal inference by mitigating selection bias and isolating program effects from external influences. The integration of time-series and geospatial regression models also allows for the dynamic assessment of intervention outcomes across temporal and spatial dimensions. Incorporating applied regression and biostatistics into community health program evaluations not only improves the accuracy and validity of findings but also supports data-driven decision-making by program managers and policymakers. These methodologies provide actionable insights that inform program refinement, resource allocation, and scaling strategies. As global health systems increasingly emphasize evidence-based interventions, enhancing the capacity of public health practitioners to apply these analytical approaches is essential for achieving more effective and equitable health outcomes in diverse community settings.***

Indexed Terms- ***Applied regression, Biostatistics, Assess outcomes, Community-based, Health programs***

I. INTRODUCTION

Community-based health programs are critical pillars of public health strategies, especially in low- and middle-income countries (LMICs) where access to formal healthcare infrastructure is limited (Ubani-Ukoma *et al*., 2018; Oni *et al*., 2019). These programs, which often focus on health promotion, disease prevention, and service delivery at the grassroots level, aim to improve health outcomes by engaging directly with populations in their local environments. Assessing the outcomes of such interventions is essential to determine their effectiveness, inform policy decisions, allocate resources efficiently, and ensure accountability to funders and communities (Head, 2016; Harris *et al*., 2017). Outcome assessments provide evidence on whether interventions are achieving their intended health impacts, uncover areas needing improvement, and guide the scaling of successful models to broader populations (Brownson *et al*., 2018; Hallingberg *et al*., 2018).

However, evaluating community-based health programs poses unique methodological challenges. Unlike clinical trials that are often conducted in

controlled environments, community interventions are implemented in real-world settings characterized by complex social, cultural, and environmental factors (Sherman *et al*., 2016; Shoveller *et al*., 2016). These programs frequently involve multi-component interventions—such as health education, behavior change communication, service delivery, and community mobilization—whose effects are interdependent and difficult to isolate. Furthermore, the absence of randomized control groups in most community health programs complicates causal attribution of observed outcomes to specific interventions (Chatterji, 2016; Craig *et al*., 2017). Selection bias, confounding variables, and contextual heterogeneity across communities are common challenges that can distort evaluation findings if not properly addressed through rigorous analytical techniques (Rooney *et al*., 2016; Witte and López-Torres, 2017).

Another significant challenge is the reliance on routine program data, which may vary in completeness, accuracy, and standardization across intervention sites (Hemkens *et al*., 2016; Wagenaar *et al*., 2016). Measurement errors, missing data, and variations in data collection practices further complicate the analytical process. Evaluators must navigate these data limitations while accounting for hierarchical data structures (e.g., individuals nested within households or communities) and the potential for intra-cluster correlations that can bias standard statistical estimates. These complexities necessitate robust analytical frameworks that can adjust for confounders, control for biases, and generate valid, interpretable estimates of program impact (Dorie *et al*., 2016; Kuang *et al*., 2017).

Applied regression analysis and biostatistical techniques offer powerful solutions to these challenges, providing evaluators with a structured methodology to assess the relationships between program interventions and health outcomes while accounting for confounding and contextual variability. Regression models—such as linear, logistic, Poisson, and multilevel regressions—allow for the quantification of associations between intervention exposure and outcome measures, adjusting for demographic, socio-economic, and environmental covariates (Austin *et al*., 2018; Kontopantelis, 2018). These models can accommodate both continuous and categorical outcome variables, making them highly adaptable to diverse evaluation contexts.

Moreover, biostatistical methods extend the analytical toolkit beyond basic regression to include techniques specifically designed for causal inference in observational settings. Methods such as propensity score matching (PSM), difference-in-differences (DiD) analysis, and instrumental variable (IV) regression are instrumental in mitigating selection biases and establishing credible counterfactual scenarios. These approaches enhance the internal validity of evaluations by approximating the conditions of randomized controlled trials (RCTs) within non-randomized program contexts.

Advanced analytical strategies, including time-series regression for longitudinal data, geospatial regression for spatially distributed interventions, and survival analysis for time-to-event outcomes, further enable evaluators to capture the dynamic and multi-dimensional impacts of community health programs. Importantly, biostatistics provides the framework for sample size estimation, statistical power calculations, and the management of complex data structures, ensuring that evaluation findings are statistically robust and generalizable (Hazra and Gogtay, 2016; Blaise *et al*., 2016).

The application of regression and biostatistical methods also facilitates the generation of actionable insights for program managers and policymakers. By providing adjusted estimates of program impact, identifying population subgroups that benefit most (or least) from interventions, and revealing contextual factors that influence program effectiveness, these analytical approaches support evidence-based decision-making and strategic resource allocation. Furthermore, visualizing regression outputs through dashboards and data visualization tools enhances the interpretability of complex analytical findings, bridging the gap between data analysts and program implementers.

Given the critical role of rigorous evaluation in informing the design, scaling, and refinement of community-based health programs, it is imperative to institutionalize the application of applied regression and biostatistics within program evaluation

frameworks. However, this requires concerted efforts in capacity building, data system strengthening, and methodological standardization across implementing organizations and health systems.

The purpose of this, is to explore how applied regression methods and biostatistical techniques can be effectively utilized to assess the outcomes of community-based health programs. This will provide an overview of key regression models suited for program evaluation, discuss biostatistical methods for enhancing causal inference in non-randomized settings, and highlight advanced analytical approaches for dealing with complex data scenarios. Additionally, it will address practical considerations related to data quality, sample design, and analytical capacity building. Through illustrative examples and best practices, this aims to equip public health practitioners, program managers, and evaluators with a methodological roadmap for conducting robust, data-driven evaluations that inform policy and practice (Brownson *et al*., 2018; Huynh *et al*., 2018).

In an era where evidence-based programming is paramount, the integration of applied regression and biostatistics into community health program evaluations is not a technical luxury but a strategic necessity. These methodologies offer the analytical rigor needed to navigate real-world complexities, ensure accountability, and ultimately enhance the impact and sustainability of health interventions at the community level.

II. METHODOLOGY

A systematic literature review was conducted to identify relevant studies, methodological papers, and empirical research on the application of regression analysis and biostatistical methods in evaluating community-based health programs. A comprehensive search strategy was employed across electronic databases including PubMed, Scopus, Web of Science, and Google Scholar. The search covered publications from January 2000 to July 2025, ensuring the inclusion of both foundational methodologies and contemporary innovations. Keywords used in combination with Boolean operators included “community-based health programs,” “program evaluation,” “applied regression,” “biostatistics,” “causal inference,” “propensity score matching,” “difference-in-differences,” “multilevel modeling,” and “public health interventions.”

The initial search yielded a total of 1,456 articles. After removing duplicates, 1,072 unique articles were screened by title and abstract for relevance to the topic. Studies were included if they described or applied regression models and biostatistical techniques specifically in the context of evaluating community health interventions or population-level programs. Exclusion criteria comprised articles focusing solely on clinical trials, purely descriptive epidemiological studies without inferential analysis, and papers addressing hospital-based interventions without community linkage. Based on these criteria, 216 articles were selected for full-text review.

Full-text assessment led to the inclusion of 82 studies that met the predefined eligibility criteria. These studies provided empirical applications of regression analyses (linear, logistic, Poisson, multilevel models) and biostatistical methods (propensity score matching, difference-in-differences analysis, instrumental variable regression) in community health program evaluations. Additionally, papers detailing methodological frameworks, analytical best practices, and challenges specific to non-randomized program evaluations were included.

Data extraction focused on study context, type of regression or biostatistical method used, programmatic domain (e.g., maternal health, infectious disease control, health promotion), and key findings related to evaluation outcomes. The synthesis process involved qualitative analysis of methodological trends, strengths and limitations of various analytical approaches, and practical implications for evaluators and policymakers.

2.1 Overview of Applied Regression in Health Program Evaluation

Evaluating the outcomes of community-based health programs necessitates robust analytical methods capable of navigating complex, real-world data environments. Regression analysis stands at the core of such evaluations, offering a versatile statistical framework for quantifying relationships between health interventions and their outcomes while controlling for confounding variables. Applied

regression models enable evaluators to dissect how interventions influence health indicators, assess program effectiveness, and uncover factors that modify intervention impacts across diverse populations (Waters *et al*., 2017; Xie *et al*., 2018). This provides an overview of key regression models used in health program evaluation, highlighting their applications, strengths, and considerations in the context of community-based interventions.

Linear regression is one of the most fundamental and widely used statistical models in health program evaluation, particularly when the outcome variable is continuous (Hauber *et al*., 2016; Dobson and Barnett, 2018). Examples of continuous outcomes in community health programs include blood pressure levels, hemoglobin counts, knowledge scores from health education interventions, or the average number of antenatal visits. The linear regression model estimates the magnitude and direction of the association between independent variables (e.g., intervention exposure, demographic factors) and the outcome of interest.

In program evaluations, linear regression facilitates the assessment of mean differences in outcomes between intervention and control groups, while adjusting for covariates such as age, gender, socio-economic status, and other potential confounders. The model's coefficients provide interpretable estimates of the expected change in the outcome for a one-unit change in the predictor variable, holding other variables constant. Linear regression assumes that relationships between variables are linear, errors are normally distributed, and variance is homogenous across levels of predictors. Evaluators must verify these assumptions to ensure the validity of the estimates.

Many health program outcomes are binary in nature, such as whether a child is fully immunized, a mother attended four or more antenatal care visits, or an individual adopted a recommended health behavior. Logistic regression is the preferred model for analyzing such dichotomous outcomes, as it estimates the probability of an event occurring based on a set of explanatory variables. Unlike linear regression, which can produce predicted values outside the 0-1 range, logistic regression models the log-odds of the outcome, ensuring that predicted probabilities are bounded between 0 and 1.

In program evaluation contexts, logistic regression allows for estimation of adjusted odds ratios (AORs), which quantify the likelihood of a positive outcome among those exposed to an intervention compared to those unexposed, controlling for confounding factors (Singh *et al*., 2018; Limpuangthip *et al*., 2018). For instance, a logistic regression model could assess how participation in a community health education program influences the odds of exclusive breastfeeding, adjusting for maternal education, household income, and access to health services.

The logistic model's flexibility extends to multinomial and ordinal logistic regressions when outcomes have more than two categories or are ordered in nature, expanding its applicability in evaluating complex program outcomes such as stages of behavior change (Hoffmann, 2016; Ke *et al*., 2016; Osman *et al*., 2016).

Community-based health programs often generate count data outcomes, such as the number of clinic visits, health education sessions attended, or disease episodes within a specified period. Poisson regression is commonly used to model such count data, under the assumption that the mean and variance of the outcome variable are equal. However, in real-world settings, count data frequently exhibit overdispersion—where the variance exceeds the mean—making the Poisson model's assumptions inappropriate.

Negative Binomial regression provides a solution for overdispersed count data by incorporating an additional parameter to account for unobserved heterogeneity, thus offering more reliable estimates. Both Poisson and Negative Binomial models estimate incidence rate ratios (IRRs), which indicate the expected change in the count outcome per unit change in the predictor variable. For example, evaluators can use these models to assess whether community health outreach programs are associated with increased utilization rates of antenatal care services, while adjusting for population demographics and service availability.

Zero-inflated models, such as Zero-Inflated Poisson (ZIP) or Zero-Inflated Negative Binomial (ZINB), are valuable extensions for scenarios where the data

contains an excess of zero counts, a common occurrence in health program data where non-utilization is prevalent.

Community-based interventions often involve hierarchical data structures, such as individuals nested within households, households within communities, or health facilities within districts. Traditional regression models that ignore these clustering effects can lead to underestimated standard errors, inflated Type I error rates, and misleading inferences. Multilevel (or hierarchical) regression models explicitly account for the nested nature of data, allowing for simultaneous estimation of individual-level and cluster-level effects.

Multilevel models partition the variance into within-cluster and between-cluster components, providing insights into how much of the outcome variability is attributable to community-level factors versus individual-level characteristics. These models are particularly useful in evaluating programs where contextual factors—such as community norms, health facility characteristics, or geographic attributes—are expected to influence program outcomes.

For example, a multilevel logistic regression model could evaluate the impact of a maternal health intervention on facility delivery rates, accounting for individual-level variables (e.g., maternal education) and community-level variables (e.g., availability of skilled birth attendants, community health committee activity). By properly adjusting for clustering, multilevel models produce more accurate standard errors and enable the assessment of contextual moderators of intervention effectiveness.

One of the primary strengths of regression models in program evaluation is their ability to adjust for confounding variables—factors that are associated with both the exposure and the outcome and may bias the estimated program effect if not controlled. Through the inclusion of relevant covariates in the regression model, evaluators can isolate the independent effect of the intervention from these potential confounders.

Effect modification (or interaction) occurs when the impact of an intervention differs across levels of another variable, such as gender, age group, or socio-economic status. Incorporating interaction terms in regression models enables evaluators to explore whether program effectiveness varies across subpopulations. For example, a significant interaction between intervention exposure and household wealth status in a logistic regression model may reveal that program uptake is more effective among wealthier households, highlighting equity considerations that inform program refinement.

Proper identification and inclusion of confounders and potential effect modifiers are critical steps in model specification. Evaluators must rely on theoretical frameworks, prior empirical evidence, and program logic models to guide variable selection and avoid issues of model overfitting or omitted variable bias.

Applied regression models are indispensable tools in the evaluation of community-based health programs, offering a flexible and rigorous analytical framework for quantifying intervention impacts across a wide range of outcome types and data structures. From linear models for continuous outcomes to logistic, Poisson, Negative Binomial, and multilevel models for binary, count, and clustered data, regression techniques enable evaluators to adjust for confounders, assess effect modifications, and generate robust, interpretable findings. When applied appropriately, regression analysis strengthens the validity and credibility of program evaluations, providing evidence that supports data-driven decision-making and the optimization of health interventions in real-world community settings (Horn *et al*., 2016; Darlington and Hayes, 2016).

2.2 Biostatistical Techniques for Causal Inference

In the evaluation of community-based health programs, establishing causal relationships between interventions and observed outcomes is a central objective. However, real-world program implementations rarely allow for randomized controlled trials (RCTs) due to ethical, logistical, or financial constraints (Monti *et al*., 2018; Handley *et al*., 2018). Instead, evaluators must rely on observational data, where the lack of randomization introduces potential biases that can compromise the validity of causal inferences. Biostatistical techniques provide robust methodologies to mitigate these biases and approximate the causal rigor of experimental designs. Among these, Propensity Score Matching

(PSM), Difference-in-Differences (DiD) analysis, Instrumental Variable (IV) methods, and imputation techniques for handling missing data are pivotal tools in applied program evaluation.

Selection bias arises when individuals who receive an intervention differ systematically from those who do not, leading to confounding in the estimation of program effects. Propensity Score Matching (PSM) is a statistical technique designed to address this issue by balancing observed covariates between treatment and control groups in observational studies. The propensity score is the probability of an individual receiving the intervention, given their observed characteristics. By matching individuals with similar propensity scores from the treatment and control groups, PSM creates a quasi-experimental setting that mimics the balance achieved through randomization.

The implementation of PSM involves estimating the propensity score using logistic regression (or other suitable models) with relevant covariates such as age, gender, socio-economic status, and other factors influencing program participation. Various matching algorithms—such as nearest-neighbor, caliper, or kernel matching—are then applied to pair treated and untreated individuals with similar scores. This process ensures that comparisons of outcomes between groups are not confounded by systematic differences in baseline characteristics.

For instance, in evaluating the impact of a community health worker (CHW) intervention on antenatal care visits, PSM can adjust for differences in maternal education, household income, and geographic access to health facilities between intervention recipients and non-recipients. By achieving covariate balance, PSM strengthens the validity of causal inferences drawn from observational data.

Despite its strengths, PSM has limitations. It only adjusts for observed covariates, leaving unmeasured confounders unaddressed. Sensitivity analyses are therefore recommended to assess the robustness of findings to potential hidden biases.

Difference-in-Differences (DiD) analysis is a quasi-experimental approach that leverages temporal variation to estimate causal impacts. DiD compares changes in outcomes over time between a treatment group exposed to the intervention and a control group that is not. The key assumption is that, in the absence of the intervention, both groups would have experienced parallel trends in the outcome of interest. Any divergence in these trends post-intervention can thus be attributed to the program's effect.

The DiD estimator is calculated as the difference in the average outcome change in the treatment group minus the difference in the average outcome change in the control group over the same period. This method effectively controls for time-invariant unobserved heterogeneity, as well as external factors that affect both groups equally over time.

For example, DiD can be used to evaluate the impact of a health promotion campaign on immunization rates by comparing changes in vaccination uptake before and after the campaign between intervention districts and comparable non-intervention districts. By controlling for common time effects, DiD isolates the program's unique contribution to observed outcome changes.

However, the validity of DiD hinges on the "parallel trends" assumption, which may not always hold. Pre-intervention trend analysis and placebo tests are critical to validate this assumption. Additionally, DiD is limited in addressing time-varying confounders that differentially affect the treatment and control groups during the study period (Bonander, 2018; Arora and Wolf, 2018).

When unobserved confounding is a significant concern, Instrumental Variable (IV) methods offer a powerful strategy for obtaining unbiased causal estimates. An IV is a variable that is correlated with the treatment assignment but affects the outcome only through its influence on treatment, not through any direct pathway. IV methods are particularly useful when treatment selection is endogenous—meaning that factors influencing program participation are correlated with the outcome in ways that cannot be fully observed or measured.

A classic example in health program evaluation is the use of geographic proximity to health facilities as an instrument for healthcare utilization. For instance, in assessing the impact of facility-based deliveries on maternal health outcomes, distance to the nearest

facility can serve as an IV, assuming it influences the likelihood of facility delivery but is unrelated to unobserved determinants of maternal health.

Two-Stage Least Squares (2SLS) regression is the standard technique used in IV analysis. In the first stage, the treatment variable is regressed on the instrument(s) to obtain predicted treatment values. In the second stage, the outcome is regressed on these predicted values, yielding an estimate of the causal effect.

While IV methods can provide unbiased estimates under strong assumptions, identifying valid instruments is often challenging. The instrument must satisfy the relevance condition (strongly correlated with treatment) and the exclusion restriction (no direct effect on the outcome). Weak instruments can lead to biased estimates, making rigorous testing of instrument validity essential.

Missing data is a pervasive challenge in program evaluations, especially in community-based settings where data collection may be hampered by logistical difficulties, non-response, or inconsistent reporting. Ignoring missing data or using simplistic approaches like complete-case analysis can lead to biased estimates and reduced statistical power. Biostatistical imputation techniques offer principled solutions to this problem, ensuring that the integrity of the analysis is maintained.

Multiple Imputation (MI) is a widely used method that involves creating several imputed datasets by replacing missing values with plausible estimates drawn from a predictive distribution based on observed data. Each dataset is analyzed separately, and results are pooled to account for the uncertainty associated with imputation. MI assumes that data are Missing at Random (MAR), meaning the probability of missingness is related to observed variables but not to unobserved factors.

Alternatively, Maximum Likelihood Estimation (MLE) methods, such as Expectation-Maximization (EM) algorithms, can be employed when data are MAR. For data that are Missing Not at Random (MNAR), more sophisticated models, like selection models or pattern-mixture models, may be necessary, though these require strong assumptions and sensitivity analyses to validate.

Imputation methods not only preserve sample size and statistical power but also minimize bias that arises from systematic differences between cases with complete and incomplete data. Proper imputation models must include all variables related to the probability of missingness and the outcome to produce valid inferences.

Biostatistical techniques for causal inference are indispensable in addressing the methodological challenges inherent in evaluating community-based health programs using observational data. Propensity Score Matching (PSM) offers a robust method for reducing selection bias by balancing observed covariates between treatment and control groups. Difference-in-Differences (DiD) analysis leverages temporal variation to isolate program effects while controlling for common time trends. Instrumental Variable (IV) methods address unobserved confounding by exploiting external sources of variation that influence treatment assignment but not the outcome directly. Additionally, imputation techniques ensure that missing data do not compromise the validity and power of evaluations.

When applied appropriately, these techniques enhance the credibility of causal claims, providing evaluators, program managers, and policymakers with reliable evidence to inform strategic decision-making and optimize health interventions. As the complexity of community-based health programs increases, the rigorous application of these biostatistical methods will be essential for generating high-quality, actionable insights that drive effective public health responses (Cunanan *et al*., 2016; Keller *et al*., 2018).

2.3 Advanced Analytical Approaches

The evaluation of community-based health programs increasingly demands analytical techniques capable of capturing complex, dynamic patterns in health outcomes. Traditional cross-sectional analyses, while useful, often fail to account for temporal and spatial variations or the timing of critical events, limiting the depth and precision of program insights. Advanced analytical approaches such as time-series regression, geospatial regression, and survival analysis offer

powerful tools to overcome these limitations as shown in figure 1(Schabenberger and Gotway, 2017; Casals *et al*., 2018). By integrating these methods into program evaluation frameworks, evaluators can generate more nuanced, actionable evidence that reflects the real-world dynamics of health interventions.

Time-series regression models are designed to analyze data collected over time, making them particularly valuable for monitoring the longitudinal effects of health interventions. Unlike cross-sectional analyses that provide a static snapshot, time-series models capture trends, seasonality, and the temporal dynamics of program impacts. This approach is crucial in evaluating interventions where outcomes are expected to evolve gradually, such as behavioral change campaigns, disease control initiatives, or service utilization programs.

Autoregressive Integrated Moving Average (ARIMA) models are a cornerstone of time-series analysis, enabling evaluators to model outcome trajectories, adjust for autocorrelation, and forecast future trends. ARIMA models can be extended to incorporate exogenous variables (ARIMAX) to assess the effect of specific interventions on health outcomes over time. For instance, an ARIMAX model could evaluate how the introduction of a community-based malaria prevention program influences monthly malaria incidence rates while adjusting for climatic variables such as rainfall or temperature.

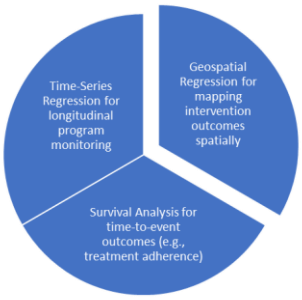

Figure 1: Advanced Analytical Approaches

Interrupted Time-Series (ITS) analysis is another powerful variant, used to assess the impact of an intervention by examining changes in outcome trends before and after program implementation. ITS is particularly useful in quasi-experimental designs where randomization is not feasible, allowing for robust causal inference by controlling for pre-existing trends.

The strength of time-series regression lies in its ability to capture the dynamic nature of program impacts, identify lagged effects, and differentiate between immediate and sustained intervention outcomes. However, time-series models require high-frequency data of sufficient duration, posing data availability challenges in resource-limited settings.

Geospatial regression techniques are essential for evaluating the spatial heterogeneity of program impacts. Community-based health interventions are often implemented across diverse geographic contexts where environmental, socio-economic, and infrastructural factors can influence program effectiveness. Traditional regression models that ignore spatial dependencies risk producing biased estimates due to spatial autocorrelation—where observations located near each other are more likely to exhibit similar characteristics.

Spatial regression models, such as Spatial Lag Models (SLM) and Spatial Error Models (SEM), explicitly account for spatial dependencies by incorporating spatial weight matrices that define the relationship between observations based on geographic proximity or connectivity. SLMs assess how outcomes in one area are influenced by outcomes in neighboring areas, capturing spillover effects that are common in public health interventions. SEMs, on the other hand, adjust for spatial autocorrelation in model residuals, improving estimation accuracy.

Geographically Weighted Regression (GWR) is a more flexible spatial analytical technique that allows regression coefficients to vary across space, revealing local variations in the relationship between predictors and outcomes. GWR is particularly valuable in identifying geographic "hotspots" or "coldspots" of program effectiveness, guiding targeted interventions. For example, GWR can be used to assess how the impact of a maternal health program on facility delivery rates varies across districts with different levels of healthcare accessibility or socio-economic status.

Integrating Geographic Information Systems (GIS) with spatial regression enables evaluators to visualize program impacts spatially, facilitating intuitive communication of findings to stakeholders and enhancing programmatic decision-making. However, spatial regression analyses require high-quality geocoded data, and model selection must carefully consider the spatial scale and contextual factors influencing program outcomes.

Survival analysis is a statistical approach designed to examine the timing of events, making it particularly suitable for evaluating outcomes where the time until an event occurs is of interest. In community health program evaluations, survival analysis is commonly applied to outcomes such as time to treatment adherence, time to recovery, duration of program participation, or time to dropout from health services (Mukherjee *et al*., 2016; Nachega *et al*., 2016).

The Kaplan-Meier estimator is a non-parametric method used to estimate survival functions, providing a visual representation of the probability of event-free survival over time. While useful for descriptive analyses, Kaplan-Meier curves do not adjust for covariates, necessitating the use of more advanced models for multivariable analysis.

The Cox Proportional Hazards Model (Cox PH) is the most widely used survival regression model in health program evaluation. It estimates the hazard ratio associated with covariates, quantifying their impact on the likelihood of experiencing the event of interest at any given time. The semi-parametric nature of the Cox model allows for flexible baseline hazard functions while adjusting for multiple covariates.

For example, in evaluating a tuberculosis treatment adherence program, a Cox PH model can assess how factors such as patient education, socioeconomic status, and distance to health facilities influence the time to treatment completion. Stratified Cox models and time-varying covariate models further extend the analytical capacity to handle non-proportional hazards or changing risk factors over time.

Parametric survival models, such as the Weibull, Exponential, and Log-normal models, offer alternative approaches when the shape of the hazard function is of analytical interest. These models provide more efficient estimates when the distributional assumptions are appropriate, enabling the prediction of survival probabilities at specific time points.

Survival analysis not only accommodates censored data (where the event has not yet occurred for all subjects) but also allows for the estimation of median survival times, hazard functions, and survival probabilities, providing a comprehensive understanding of time-to-event dynamics in program outcomes.

Advanced analytical approaches such as time-series regression, geospatial regression, and survival analysis significantly enhance the rigor and depth of community-based health program evaluations. Time-series models capture temporal dynamics, allowing evaluators to monitor intervention impacts over time and forecast future trends. Geospatial regression techniques reveal spatial heterogeneity and contextual influences on program effectiveness, guiding geographically targeted strategies. Survival analysis provides robust tools for examining time-to-event outcomes, offering insights into the duration and timing of program impacts.

The integration of these advanced methods into evaluation frameworks ensures that complex patterns in health data are accurately modeled and interpreted. However, their effective application requires high-quality longitudinal and spatial data, analytical expertise, and careful consideration of underlying assumptions. By leveraging these advanced analytical tools, evaluators can generate more nuanced, actionable insights that drive evidence-based decision-making and optimize the design, implementation, and scaling of health interventions across diverse community settings.

2.4 Data Considerations in Community Program Evaluation

The effectiveness of community-based health program evaluations is heavily dependent on the quality and appropriateness of the data collected and analyzed. Unlike controlled clinical trials, community interventions operate in dynamic and often unpredictable environments, where data challenges such as small sample sizes, clustered data structures, and measurement errors can compromise the validity

and reliability of findings. Rigorous attention to data considerations—including sample size calculations, management of intra-cluster correlations, and strategies to address data quality and measurement errors—is essential to ensure robust and credible evaluations (Flight *et al*., 2016; Lachman *et al*., 2017). This discusses these critical aspects and outlines methodological strategies for their mitigation.

Accurate sample size estimation is a fundamental component of study design in community program evaluations. Sample size determines the statistical power of a study—the probability of detecting a true program effect if it exists. Underpowered studies may fail to identify meaningful impacts (Type II error), while overpowered studies may waste resources without added analytical benefit. In the context of community interventions, sample size calculations must account for expected effect sizes, variability in outcomes, and the study design, including any clustering of data.

Effect size estimates are often derived from prior studies, pilot data, or program logic models. Larger effect sizes require smaller sample sizes to detect, while smaller expected effects necessitate larger samples to achieve adequate power. In addition, outcome variability—measured by the standard deviation for continuous outcomes or baseline proportions for binary outcomes—must be accurately estimated to avoid under- or over-estimating sample requirements.

The choice of significance level (typically set at 0.05) and desired power (commonly 80% or 90%) further informs sample size needs. In community evaluations where interventions are delivered at the group level (e.g., village, health facility catchment area), the study design effect—driven by intra-cluster correlations—must be incorporated into sample size formulas to avoid underestimating the required sample size.

Sample size calculations should also consider potential attrition, non-response rates, and missing data, which can reduce the effective sample size and consequently diminish the study's statistical power. Including a buffer for these contingencies ensures the final dataset retains sufficient power for hypothesis testing.

Community-based interventions often involve clustered data structures, where individuals are nested within larger units such as households, villages, or health facility catchment areas. Individuals within the same cluster are likely to share similar characteristics or be influenced by common contextual factors, leading to intra-cluster correlation (ICC). Ignoring ICC can result in underestimated standard errors, inflated Type I error rates, and misleading conclusions about program effectiveness.

To address clustering effects, evaluators must first quantify the extent of ICC in their data. The ICC represents the proportion of total variance in an outcome attributable to between-cluster variability. Higher ICC values indicate greater homogeneity within clusters, necessitating larger sample sizes to achieve adequate statistical power. In designing cluster-randomized trials or observational evaluations with cluster-level interventions, the design effect—calculated as $1 + (\text{average cluster size} - 1) \times \text{ICC}$—adjusts the effective sample size to account for clustering.

Analytically, multilevel (hierarchical) regression models are the preferred method for handling clustered data, as they partition variance into within-cluster and between-cluster components, ensuring accurate estimation of standard errors. Alternatively, generalized estimating equations (GEE) with robust standard errors provide another approach for accounting for intra-cluster correlations in population-averaged models.

When designing community evaluations, balancing the number of clusters versus the number of individuals per cluster is crucial (Levorato *et al*., 2017; Xiao *et al*., 2017). Increasing the number of clusters generally has a more significant impact on statistical power than merely increasing cluster sizes, given that adding more clusters reduces between-cluster variability. Evaluators must therefore optimize sampling strategies to maximize power while considering resource constraints.

High-quality data are the bedrock of credible program evaluations. However, community-based health programs often contend with data quality challenges arising from incomplete reporting, inaccurate measurements, inconsistent data collection

procedures, and contextual constraints such as low literacy or limited infrastructure. Measurement error—whether systematic (bias) or random—can significantly distort estimates of program effects, undermine causal inferences, and erode stakeholder confidence in evaluation findings.

Mitigating data quality issues begins with the development of standardized data collection tools and protocols. Training data collectors thoroughly on questionnaire administration, measurement techniques, and ethical considerations is essential to ensure consistency and accuracy. In settings where self-reported data is used, employing validated instruments and triangulating with objective measures (e.g., facility records, biometrics) enhances data reliability.

Routine data audits, back-checks, and supervision during data collection help identify and correct errors early in the process. The use of digital data collection platforms—such as mobile data collection apps—reduces transcription errors, enforces skip logic, and allows for real-time data validation. Automated checks for out-of-range values, missing data, and internal inconsistencies provide immediate feedback to enumerators, improving data completeness and accuracy.

Addressing measurement error analytically involves various techniques. For continuous variables prone to measurement inaccuracies (e.g., weight, blood pressure), implementing repeated measurements and averaging them can reduce random error. In cases of misclassification in categorical variables, sensitivity analyses can be conducted to assess the robustness of findings under varying assumptions about classification accuracy.

Missing data is a prevalent challenge in community evaluations. Rather than excluding incomplete cases (complete-case analysis), which can introduce bias if data are not missing completely at random, multiple imputation or maximum likelihood methods offer statistically sound alternatives that preserve sample size and reduce bias.

Data harmonization is another critical aspect, particularly in multi-site evaluations where variations in data collection practices or definitions across sites can compromise comparability. Establishing clear data dictionaries, variable coding schemes, and centralized data management systems ensures uniformity in data handling and facilitates pooled analyses.

Finally, stakeholder engagement in the data management process fosters a culture of data quality. Involving community members, program staff, and policymakers in the development of data tools and feedback loops encourages ownership, accountability, and responsiveness to data-driven insights.

Robust data considerations are indispensable for credible and impactful evaluations of community-based health programs. Accurate sample size calculations that account for effect sizes, variability, and clustering ensure that studies are adequately powered to detect true program effects. Managing intra-cluster correlations through design and analytical strategies preserves the validity of statistical inferences in clustered data structures. Addressing data quality and measurement error through standardized protocols, training, real-time validation, and appropriate imputation techniques safeguards the integrity of evaluation findings.

As data-driven decision-making becomes central to global health initiatives, investments in data system strengthening, analytical capacity building, and stakeholder engagement in data processes will be pivotal. By prioritizing these data considerations, evaluators can generate high-quality evidence that not only informs programmatic improvements but also builds trust and accountability among communities, funders, and policymakers alike (Haby *et al*., 2016; Li *et al*., 2017).

2.5 Practical Applications

The use of regression analysis and biostatistical techniques has become a cornerstone of modern public health program evaluation (Hayes and Moulton, 2017; Panaretos *et al*., 2018). These methods offer robust frameworks for understanding the impact of interventions, adjusting for confounding factors, and presenting actionable insights to program managers and policymakers. While the theoretical underpinnings of these techniques are well-established, their practical application in real-world

settings demonstrates their true value. This presents illustrative examples of how logistic regression, Difference-in-Differences (DiD) analysis, and data visualization of regression outputs are applied in evaluating community-based health programs to inform strategic decision-making.

Vaccination programs in rural communities often face challenges such as low health literacy, access barriers, and cultural resistance, which impact program uptake. Evaluating these programs requires analytical methods that can identify determinants of vaccination while controlling for potential confounders. Logistic regression is particularly suited for this task when the outcome variable is binary—such as whether a child is fully vaccinated (yes/no).

In a case study evaluating a rural immunization program in Sub-Saharan Africa, researchers collected data from 1,200 households, focusing on whether children under five had received the full schedule of recommended vaccines. Key predictor variables included maternal education, household income, distance to the nearest health facility, exposure to community health worker (CHW) visits, and participation in health education sessions.

Using logistic regression, the evaluators estimated adjusted odds ratios (AORs) for each predictor. The analysis revealed that children whose mothers had completed secondary education were 2.5 times more likely to be fully vaccinated compared to those whose mothers had no formal education (AOR = 2.5, 95% CI: 1.8–3.6). Similarly, households within 5 km of a health facility had significantly higher odds of complete vaccination (AOR = 1.9, 95% CI: 1.4–2.7). Participation in CHW-led health education sessions emerged as a strong programmatic factor, with an AOR of 3.2 (95% CI: 2.2–4.5), underscoring the intervention's effectiveness in promoting immunization.

By adjusting for socio-demographic confounders, logistic regression provided a clear picture of which factors most influenced vaccination uptake. These findings allowed program managers to prioritize scaling CHW visits and develop targeted messaging for lower-education households, thus refining intervention strategies for greater impact.

Difference-in-Differences (DiD) analysis is a powerful quasi-experimental approach used to assess program impacts where randomized controlled trials are not feasible. An illustrative example involves evaluating a maternal health program aimed at increasing facility-based deliveries in rural districts of South Asia.

The program, implemented in 10 intervention districts, included demand-side incentives (e.g., conditional cash transfers), transportation vouchers, and supply-side enhancements (e.g., training midwives, improving facility infrastructure). To evaluate its impact, researchers used data from 10 comparable control districts where no intervention was applied.

Baseline data on facility-based deliveries were collected six months prior to program implementation, and follow-up data were gathered one year post-intervention. The DiD model compared the change in facility delivery rates over time between the intervention and control districts, controlling for time-invariant differences.

Results indicated that facility delivery rates in intervention districts increased from 45% to 68%, while control districts experienced a modest increase from 47% to 50% due to general health system improvements. The DiD estimate of program impact was a 20 percentage point increase in facility deliveries attributable to the intervention (DiD coefficient = 0.20, $p < 0.01$).

The DiD approach effectively controlled for external factors that affected both groups, isolating the intervention’s unique contribution to the observed outcome. Furthermore, subgroup analyses revealed larger gains among women in the lowest income quintile, highlighting the program’s success in addressing socio-economic disparities.

These findings informed policymakers about the program’s cost-effectiveness and equity impact, providing empirical justification for scaling the intervention to additional districts.

Communicating complex statistical findings to program managers, donors, and policymakers is a critical but often overlooked component of program evaluation. While regression coefficients and odds

ratios provide analytical rigor, visual representations of these outputs enhance comprehension and facilitate data-driven decision-making among non-technical stakeholders.

Effective visualization techniques include forest plots, predictive probability graphs, marginal effects plots, and interactive dashboards (Stadler *et al*., 2016; He *et al*., 2018). For example, in the vaccination uptake case study, a forest plot depicting adjusted odds ratios for various determinants (with confidence intervals) offers a clear, at-a-glance summary of which factors are most influential. Color coding significant predictors further aids quick interpretation.

Predictive probability graphs transform abstract logistic regression coefficients into tangible probabilities, showing, for instance, the likelihood of full immunization across different levels of maternal education or distance to health facilities. These visualizations help program managers understand practical implications, such as how much improvement in coverage can be expected from specific intervention strategies.

For DiD analyses, line graphs depicting pre- and post-intervention trends for treatment and control groups clearly illustrate the intervention's effect over time. Highlighting the divergence in trends post-implementation reinforces the program's impact in a visually intuitive manner.

Interactive dashboards, developed using platforms like Tableau or Power BI, enable real-time exploration of regression outputs across various dimensions, such as geographic areas, demographic subgroups, and intervention components. These tools empower program managers to engage with the data actively, simulate “what-if” scenarios, and tailor strategies based on localized insights.

Visualization also plays a vital role in policy advocacy. Simplified infographics summarizing key regression findings can be presented to funding agencies or used in public communication campaigns to demonstrate program effectiveness and mobilize further support.

The practical application of regression and biostatistical techniques in program evaluation bridges the gap between data analysis and actionable program insights. Logistic regression enables evaluators to dissect the determinants of binary outcomes, as illustrated in assessing vaccination uptake in rural areas. Difference-in-Differences analysis provides a rigorous framework for attributing program impacts in non-randomized settings, exemplified by maternal health interventions. Furthermore, effective visualization of regression outputs ensures that complex analytical findings are translated into intuitive, decision-relevant formats for program managers and policymakers.

By embedding these analytical practices into routine evaluation workflows, community-based health programs can enhance their responsiveness, target interventions more effectively, and demonstrate accountability to stakeholders. As health systems increasingly prioritize data-driven decision-making, the integration of robust analytical methods with clear, audience-tailored communication strategies will be essential to optimizing public health outcomes (Moser *et al*., 2017; Frith *et al*., 2017).

2.6 Capacity Building and Institutionalizing Analytical Rigor

As global health programs evolve to address increasingly complex and diverse challenges, the demand for rigorous, data-driven evaluation practices becomes more urgent. Analytical rigor, particularly through applied biostatistics and regression techniques, is essential to generate credible evidence that informs program design, resource allocation, and policy decisions as shown in figure 2 (Sells *et al*., 2018; Bärnighausen *et al*., 2017). However, achieving this level of analytical sophistication requires deliberate investments in capacity building, the development of standardized evaluation protocols, and the fostering of strategic collaborations between academic institutions, governments, and non-governmental organizations (NGOs). This discusses the critical elements of building institutional capacity for rigorous health program evaluation, emphasizing workforce development, methodological standardization, and multi-sector partnerships.

A skilled workforce is the cornerstone of institutionalizing analytical rigor in health program evaluations. While advanced statistical techniques

such as regression modeling, causal inference methods, and time-series analyses are becoming indispensable in contemporary program evaluation, there remains a significant skills gap among public health practitioners, particularly in low- and middle-income countries (LMICs). Addressing this gap necessitates comprehensive training programs that equip practitioners with both theoretical understanding and practical application of biostatistical methods.

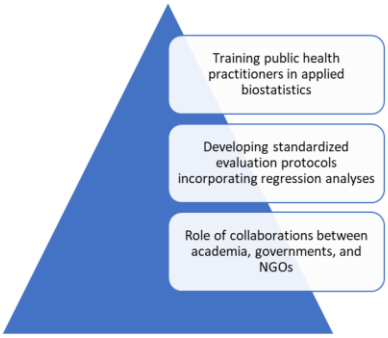

Figure 2: Capacity Building and Institutionalizing Analytical Rigor

Training initiatives should focus on applied biostatistics, ensuring that practitioners can interpret regression outputs, handle clustered and longitudinal data, and perform appropriate model diagnostics. Programs must go beyond traditional classroom instruction and incorporate hands-on, project-based learning where trainees work on real-world datasets relevant to their programmatic contexts. This approach not only reinforces statistical concepts but also cultivates problem-solving skills crucial for addressing data challenges in dynamic field settings.

Moreover, training curricula should be tailored to diverse levels of proficiency. Foundational courses should introduce essential concepts such as hypothesis testing, confidence intervals, and linear/logistic regression, while advanced modules can delve into multilevel modeling, survival analysis, and causal inference techniques. Embedding data visualization and communication skills within these programs ensures that practitioners are not only proficient in analysis but also adept at translating findings into actionable insights for decision-makers.

Professional development pathways—including certifications, continuing education credits, and formal degree programs in biostatistics and public health informatics—can incentivize practitioners to continually enhance their analytical competencies. Additionally, fostering communities of practice through peer-learning networks, online forums, and regional workshops facilitates ongoing knowledge exchange and skill reinforcement.

The absence of standardized evaluation protocols often leads to inconsistent methodologies, fragmented data collection, and analytical practices that compromise the comparability and validity of program evaluations. To institutionalize analytical rigor, it is essential to develop and disseminate standardized evaluation protocols that explicitly incorporate regression analyses and other advanced biostatistical techniques.

Standardized protocols should outline clear guidelines for study design selection (e.g., cross-sectional, longitudinal, quasi-experimental), sample size calculations, handling of clustered data, and appropriate regression models for different outcome types (binary, continuous, count, time-to-event). These protocols must also address strategies for adjusting for confounding, assessing effect modifiers, and performing sensitivity analyses to validate findings.

Developing data dictionaries and metadata standards ensures uniform variable definitions and coding schemes, facilitating data harmonization across programs and sites. Protocols should also include standard operating procedures (SOPs) for data quality assurance, including automated data validation checks, error tracking, and documentation of data cleaning processes (Dirnagl, 2016; Gass *et al*., 2017).

Incorporating practical tools—such as template scripts for regression analyses (using R, Stata, or SPSS) and interactive dashboards for visualizing regression outputs—streamlines analytical workflows and enhances reproducibility. By providing program implementers with readily applicable analytical frameworks, standardized protocols reduce methodological variability and elevate the overall quality of evaluations.

Furthermore, the establishment of institutional review boards (IRBs) or methodological advisory committees within implementing agencies ensures that evaluation

designs and analytical plans undergo rigorous peer-review, fostering accountability and adherence to best practices.

The successful institutionalization of analytical rigor in health program evaluations hinges on robust collaborations between academia, governments, and NGOs. Each of these actors brings unique strengths that, when synergized, can drive substantial improvements in evaluation capacity and practice.

Academic institutions are well-positioned to provide technical expertise, develop training curricula, and advance methodological innovations. Through joint research initiatives, academic partners can co-design evaluation studies with implementers, ensuring that cutting-edge analytical methods are appropriately applied to programmatic contexts. Additionally, academic researchers can mentor government and NGO staff in biostatistics, fostering a pipeline of skilled evaluators embedded within implementing organizations.

Governments play a critical role in setting evaluation standards, integrating analytical capacity building into national health strategies, and allocating resources for institutional strengthening. By mandating the use of standardized evaluation protocols and regression-based analyses in program assessments, governments can ensure methodological consistency across agencies and regions. Moreover, governmental endorsement of data-driven decision-making creates an enabling environment for evidence-based policy formulation and resource allocation.

NGOs, as frontline implementers of community health programs, are essential partners in operationalizing evaluation frameworks. Their proximity to communities and deep programmatic insights ensure that evaluations are contextually relevant and grounded in practical realities. NGOs also facilitate data collection, stakeholder engagement, and the dissemination of evaluation findings to beneficiaries, funders, and policymakers.

Multi-sector collaborations foster knowledge exchange, resource sharing, and joint problem-solving. For instance, university-led capacity-building consortia can develop training-of-trainers (ToT) programs for NGO staff, who in turn disseminate these skills within their networks. Governments can support such initiatives through policy frameworks and funding mechanisms, ensuring scalability and sustainability. Collaborative platforms, such as national evaluation task forces or regional learning hubs, further institutionalize these partnerships by providing structured avenues for continuous dialogue, joint evaluations, and methodological harmonization.

Additionally, global health agencies and donors have a vital role in incentivizing collaborative efforts by embedding capacity-building requirements in funding agreements, supporting open-access analytical resources, and fostering south-south collaborations that leverage regional expertise and experiences.

Institutionalizing analytical rigor in health program evaluation is not merely a technical exercise but a comprehensive capacity-building endeavor that demands strategic investments in workforce development, methodological standardization, and multi-sector collaboration. Training public health practitioners in applied biostatistics equips them with the analytical skills necessary to conduct robust evaluations and generate actionable insights. Developing standardized evaluation protocols ensures methodological consistency, enhances data quality, and streamlines analytical workflows (Yeung *et al*., 2016; Haarbrandt *et al*., 2018). Finally, fostering collaborations between academia, governments, and NGOs leverages complementary strengths, accelerates knowledge translation, and embeds a culture of data-driven decision-making within health systems.

As global health challenges become more complex and resources more constrained, the ability to conduct rigorous, credible, and policy-relevant evaluations will be critical to optimizing program effectiveness and achieving sustainable health outcomes. By institutionalizing analytical rigor, health systems can enhance accountability, foster continuous improvement, and ultimately, deliver more impactful interventions to the communities they serve.

CONCLUSION

Regression and biostatistics play a pivotal role in elevating the rigor and credibility of community-based health program evaluations. These analytical tools enable evaluators to move beyond simple descriptive

statistics, providing robust methods for estimating program impacts, adjusting for confounding variables, and drawing valid causal inferences in complex, real-world settings. Techniques such as logistic regression, Difference-in-Differences (DiD) analysis, and survival modeling offer nuanced insights into intervention effectiveness, while advanced methods like propensity score matching and instrumental variable approaches help mitigate biases inherent in non-randomized designs. Visualization of regression outputs further bridges the gap between complex analyses and actionable program insights, ensuring that data-driven narratives inform policy and programmatic decisions.

To fully harness the potential of regression and biostatistics, it is imperative to integrate these methods into routine monitoring and evaluation (M&E) processes. Standardizing analytical protocols, investing in practitioner training, and fostering a culture of continuous learning are essential steps toward institutionalizing advanced analytics within health systems. Embedding analytical rigor into M&E frameworks ensures that evaluations not only generate credible evidence but also foster adaptive learning loops where data informs real-time program refinement.

The vision for community health interventions must center on data-driven decision-making, where robust analytical evidence guides program design, resource allocation, and policy formulation. Strengthening analytical capacity across governments, NGOs, and local health practitioners will empower stakeholders to navigate the complexities of health interventions with precision and agility. As global health challenges evolve, a commitment to integrating regression-based analytics into program evaluation will be instrumental in achieving impactful, equitable, and sustainable health outcomes across diverse community settings.

REFERENCES

[1] Arora, K. and Wolf, D.A., 2018. Does paid family leave reduce nursing home use? The California experience. *Journal of Policy Analysis and Management*, *37*(1), pp.38-62.

[2] Austin, P.C., Stryhn, H., Leckie, G. and Merlo, J., 2018. Measures of clustering and heterogeneity in multilevel P oisson regression analyses of rates/count data. *Statistics in medicine*, *37*(4), pp.572-589.

[3] Bärnighausen, T., Røttingen, J.A., Rockers, P., Shemilt, I. and Tugwell, P., 2017. Quasi-experimental study designs series—paper 1: introduction: two historical lineages. *Journal of clinical epidemiology*, *89*, pp.4-11.

[4] Blaise, B.J., Correia, G., Tin, A., Young, J.H., Vergnaud, A.C., Lewis, M., Pearce, J.T., Elliott, P., Nicholson, J.K., Holmes, E. and Ebbels, T.M., 2016. Power analysis and sample size determination in metabolic phenotyping. *Analytical chemistry*, *88*(10), pp.5179-5188.

[5] Bonander, C., 2018. Compared with what? Estimating the effects of injury prevention policies using the synthetic control method. *Injury prevention*, *24*(Suppl 1), pp.i60-i66.

[6] Brownson, R.C., Baker, E.A., Deshpande, A.D. and Gillespie, K.N., 2018. *Evidence-based public health*. Oxford university press.

[7] Brownson, R.C., Baker, E.A., Deshpande, A.D. and Gillespie, K.N., 2018. *Evidence-based public health*. Oxford university press.

[8] Casals, J., Garcia-Hiernaux, A., Jerez, M., Sotoca, S. and Trindade, A.A., 2018. *State-space methods for time series analysis: theory, applications and software*. Chapman and Hall/CRC.

[9] Chatterji, M., 2016. Causal inferences on the effectiveness of complex social programs: Navigating assumptions, sources of complexity and evaluation design challenges. *Evaluation and Program Planning*, *59*, pp.128-140.

[10] Craig, P., Katikireddi, S.V., Leyland, A. and Popham, F., 2017. Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. *Annual review of public health*, *38*(1), pp.39-56.

[11] Cunanan, K.M., Carlin, B.P. and Peterson, K.A., 2016. A practical Bayesian stepped wedge design for community-based cluster-randomized clinical trials: The British Columbia Telehealth Trial. *Clinical Trials*, *13*(6), pp.641-650.

[12] Darlington, R.B. and Hayes, A.F., 2016. *Regression analysis and linear models: Concepts, applications, and implementation*. Guilford Publications.

[13] Dirnagl, U., 2016. Quality Control and Standard Operating Procedures. In *Rodent Models of Stroke* (pp. 291-300). New York, NY: Springer New York.

[14] Dobson, A.J. and Barnett, A.G., 2018. *An introduction to generalized linear models*. Chapman and Hall/CRC.

[15] Dorie, V., Harada, M., Carnegie, N.B. and Hill, J., 2016. A flexible, interpretable framework for assessing sensitivity to unmeasured confounding. *Statistics in medicine*, *35*(20), pp.3453-3470.

[16] Flight, L., Allison, A., Dimairo, M., Lee, E., Mandefield, L. and Walters, S.J., 2016. Recommendations for the analysis of individually randomised controlled trials with clustering in one arm–a case of continuous outcomes. *BMC Medical Research Methodology*, *16*(1), p.165.

[17] Frith, N., Catchpole, S., Home, R., Watson, C. and Kessler, H., 2017. A unified data: framework for mapping underground: Project Iceberg. Work package 2, defining the problem space for an integrated data operating system above and below ground.

[18] Gass Jr, J.D., Misra, A., Yadav, M.N.S., Sana, F., Singh, C., Mankar, A., Neal, B.J., Fisher-Bowman, J., Maisonneuve, J., Delaney, M.M. and Kumar, K., 2017. Implementation and results of an integrated data quality assurance protocol in a randomized controlled trial in Uttar Pradesh, India. *Trials*, *18*(1), p.418.

[19] Haarbrandt, B., Schreiweis, B., Rey, S., Sax, U., Scheithauer, S., Rienhoff, O., Knaup-Gregori, P., Bavendiek, U., Dieterich, C., Brors, B. and Kraus, I., 2018. HiGHmed–an open platform approach to enhance care and research across institutional boundaries. *Methods of information in medicine*, *57*(S 01), pp.e66-e81.

[20] Haby, M.M., Chapman, E., Clark, R., Barreto, J., Reveiz, L. and Lavis, J.N., 2016. What are the best methodologies for rapid reviews of the research evidence for evidence-informed decision making in health policy and practice: a rapid review. *Health research policy and systems*, *14*(1), p.83.

[21] Hallingberg, B., Turley, R., Segrott, J., Wight, D., Craig, P., Moore, L., Murphy, S., Robling, M., Simpson, S.A. and Moore, G., 2018. Exploratory studies to decide whether and how to proceed with full-scale evaluations of public health interventions: a systematic review of guidance. *Pilot and feasibility studies*, *4*(1), p.104.

[22] Handley, M.A., Lyles, C.R., McCulloch, C. and Cattamanchi, A., 2018. Selecting and improving quasi-experimental designs in effectiveness and implementation research. *Annual review of public health*, *39*(1), pp.5-25.

[23] Harris, C., Allen, K., Waller, C. and Brooke, V., 2017. Sustainability in health care by allocating resources effectively (SHARE) 3: examining how resource allocation decisions are made, implemented and evaluated in a local healthcare setting. *BMC health services research*, *17*(1), p.340.

[24] Hauber, A.B., González, J.M., Groothuis-Oudshoorn, C.G., Prior, T., Marshall, D.A., Cunningham, C., IJzerman, M.J. and Bridges, J.F., 2016. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research practices task force. *Value in health*, *19*(4), pp.300-315.

[25] Hayes, R.J. and Moulton, L.H., 2017. *Cluster randomised trials*. Chapman and Hall/CRC.

[26] Hazra, A. and Gogtay, N., 2016. Biostatistics series module 5: Determining sample size. *Indian journal of dermatology*, *61*(5), pp.496-504.

[27] He, L., Levine, R.A., Bohonak, A.J., Fan, J. and Stronach, J., 2018. Predictive analytics machinery for STEM student success studies. *Applied Artificial Intelligence*, *32*(4), pp.361-387.

[28] Head, B.W., 2016. Toward more "evidence-informed" policy making?. *Public administration review*, *76*(3), pp.472-484.

[29] Hemkens, L.G., Contopoulos-Ioannidis, D.G. and Ioannidis, J.P., 2016. Routinely collected data and comparative effectiveness evidence: promises and limitations. *Cmaj*, *188*(8), pp.E158-E164.

[30] Hoffmann, J.P., 2016. *Regression models for categorical, count, and related variables: An applied approach*. Univ of California Press.

[31] Horn, A.S. and Lee, G., 2016. The reliability and validity of using regression residuals to measure institutional effectiveness in promoting degree

completion. *Research in Higher Education*, *57*(4), pp.469-496.

[32] Huynh, A.K., Hamilton, A.B., Farmer, M.M., Bean-Mayberry, B., Stirman, S.W., Moin, T. and Finley, E.P., 2018. A pragmatic approach to guide implementation evaluation research: strategy mapping for complex interventions. *Frontiers in public health*, *6*, p.134.

[33] Ke, Y., Fu, B. and Zhang, W., 2016. Semi-varying coefficient multinomial logistic regression for disease progression risk prediction. *Statistics in Medicine*, *35*(26), pp.4764-4778.

[34] Keller, S., Shipp, S., Korkmaz, G., Molfino, E., Goldstein, J., Lancaster, V., Pires, B., Higdon, D., Chen, D. and Schroeder, A., 2018. Harnessing the power of data to support community-based research. *Wiley Interdisciplinary Reviews: Computational Statistics*, *10*(3), p.e1426.

[35] Kontopantelis, E., 2018. A comparison of one-stage vs two-stage individual patient data meta-analysis methods: A simulation study. *Research Synthesis Methods*, *9*(3), pp.417-430.

[36] Kuang, K., Cui, P., Li, B., Jiang, M. and Yang, S., 2017, August. Estimating treatment effect in the wild via differentiated confounder balancing. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 265-274).

[37] Lachman, J.M., Cluver, L., Ward, C.L., Hutchings, J., Mlotshwa, S., Wessels, I. and Gardner, F., 2017. Randomized controlled trial of a parenting program to reduce the risk of child maltreatment in South Africa. *Child abuse & neglect*, *72*, pp.338-351.

[38] Levorato, M., Figueiredo, R., Frota, Y. and Drummond, L., 2017. Evaluating balancing on social networks through the efficient solution of correlation clustering problems. *EURO Journal on Computational Optimization*, *5*(4), pp.467-498.

[39] Li, R., Ruiz, F., Culyer, A.J., Chalkidou, K. and Hofman, K.J., 2017. Evidence-informed capacity building for setting health priorities in low-and middle-income countries: a framework and recommendations for further research. *F1000Research*, *6*, p.231.

[40] Limpuangthip, N., Somkotra, T. and Arksornnukit, M., 2018. Modified retention and stability criteria for complete denture wearers: A risk assessment tool for impaired masticatory ability and oral health-related quality of life. *The Journal of prosthetic dentistry*, *120*(1), pp.43-49.

[41] Monti, S., Grosso, V., Todoerti, M. and Caporali, R., 2018. Randomized controlled trials and real-world data: differences and similarities to untangle literature data. *Rheumatology*, *57*(Supplement_7), pp.vii54-vii58.

[42] Moser, S.C., Coffee, J. and Seville, A., 2017. Rising to the challenge, together. *The Kresge Foundation*.

[43] Mukherjee, J.S., Barry, D., Weatherford, R.D., Desai, I.K. and Farmer, P.E., 2016. Community-based ART programs: sustaining adherence and follow-up. *Current Hiv/Aids Reports*, *13*(6), pp.359-366.

[44] Nachega, J.B., Adetokunboh, O., Uthman, O.A., Knowlton, A.W., Altice, F.L., Schechter, M., Galárraga, O., Geng, E., Peltzer, K., Chang, L.W. and Van Cutsem, G., 2016. Community-based interventions to improve and sustain antiretroviral therapy adherence, retention in HIV care and clinical outcomes in low-and middle-income countries for achieving the UNAIDS 90-90-90 targets. *Current HIV/AIDS Reports*, *13*(5), pp.241-255.

[45] Oni, O., Adeshina, Y.T., Iloeje, K.F. and Olatunji, O.O., ARTIFICIAL INTELLIGENCE MODEL FAIRNESS AUDITOR FOR LOAN SYSTEMS. *Journal ID*, *8993*, p.1162.

[46] Osman, M., Paleti, R., Mishra, S. and Golias, M.M., 2016. Analysis of injury severity of large truck crashes in work zones. *Accident Analysis & Prevention*, *97*, pp.261-273.

[47] Panaretos, D., Koloverou, E., Dimopoulos, A.C., Kouli, G.M., Vamvakari, M., Tzavelas, G., Pitsavos, C. and Panagiotakos, D.B., 2018. A comparison of statistical and machine-learning techniques in evaluating the association between dietary patterns and 10-year cardiometabolic risk (2002–2012): the ATTICA study. *British Journal of Nutrition*, *120*(3), pp.326-334.

[48] Rooney, A.A., Cooper, G.S., Jahnke, G.D., Lam, J., Morgan, R.L., Boyles, A.L., Ratcliffe, J.M., Kraft, A.D., Schünemann, H.J., Schwingl, P. and Walker, T.D., 2016. How credible are the study

results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards. *Environment international*, *92*, pp.617-629.

[49] Schabenberger, O. and Gotway, C.A., 2017. *Statistical methods for spatial data analysis*. Chapman and Hall/CRC.

[50] Sells, S.N., Bassing, S.B., Barker, K.J., Forshee, S.C., Keever, A.C., Goerz, J.W. and Mitchell, M.S., 2018. Increased scientific rigor will improve reliability of research and effectiveness of management. *The Journal of Wildlife Management*, *82*(3), pp.485-494.

[51] Sherman, R.E., Anderson, S.A., Dal Pan, G.J., Gray, G.W., Gross, T., Hunter, N.L., LaVange, L., Marinac-Dabic, D., Marks, P.W., Robb, M.A. and Shuren, J., 2016. Real-world evidence—what is it and what can it tell us. *N Engl J Med*, *375*(23), pp.2293-2297.

[52] Shoveller, J., Viehbeck, S., Di Ruggiero, E., Greyson, D., Thomson, K. and Knight, R., 2016. A critical examination of representations of context within research on population health interventions. *Critical Public Health*, *26*(5), pp.487-500.

[53] Singh, S., Kalb, R.E., de Jong, E.M., Shear, N.H., Lebwohl, M., Langholff, W., Hopkins, L., Srivastava, B. and Armstrong, A.W., 2018. Effect of age of onset of psoriasis on clinical outcomes with systemic treatment in the Psoriasis Longitudinal Assessment and Registry (PSOLAR). *American journal of clinical dermatology*, *19*(6), pp.879-886.

[54] Stadler, J.G., Donlon, K., Siewert, J.D., Franken, T. and Lewis, N.E., 2016. Improving the efficiency and ease of healthcare analysis through use of data visualization dashboards. *Big data*, *4*(2), pp.129-135.

[55] Ubani-Ukoma, U., Osei, Y. and Ilomuanya, M., 2018. Optimisation of clotrimazole emulgel using essential oils of Citrus sinensis and Zingiber officinale as permeation enhancers. *West African Journal of Pharmacy*, *29*(2), pp.19-33.

[56] Wagenaar, B.H., Sherr, K., Fernandes, Q. and Wagenaar, A.C., 2016. Using routine health information systems for well-designed health evaluations in low-and middle-income countries. *Health policy and planning*, *31*(1), pp.129-135.

[57] Waters, E., Gibbs, L., Tadic, M., Ukoumunne, O.C., Magarey, A., Okely, A.D., De Silva, A., Armit, C., Green, J., O'Connor, T. and Johnson, B., 2017. Cluster randomised trial of a school-community child health promotion and obesity prevention intervention: findings from the evaluation of fun 'n healthy in Moreland!. *BMC public health*, *18*(1), p.92.

[58] Witte, K.D. and López-Torres, L., 2017. Efficiency in education: A review of literature and a way forward. *Journal of the operational research society*, *68*(4), pp.339-363.

[59] Xiao, Y., Xue, Y., Nazarian, S. and Bogdan, P., 2017, November. A load balancing inspired optimization framework for exascale multicore systems: A complex networks approach. In *2017 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)* (pp. 217-224). IEEE.

[60] Xie, Q.W., Chan, C.H., Ji, Q. and Chan, C.L., 2018. Psychosocial effects of parent-child book reading interventions: A meta-analysis. *Pediatrics*, *141*(4), p.e20172675.

[61] Yeung, R.S., Albani, S., Feldman, B.M., Mellins, E., Prakken, B. and Wedderburn, L.R., 2016. Enhancing translational research in paediatric rheumatology through standardization. *Nature Reviews Rheumatology*, *12*(11), pp.684-690.