HANDLING COMPLEXITY VIA STATISTICAL METHODS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Evidence S. Matangi

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2019

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Prof. George P. McCabe
Department of Statistics
Prof. Nilupa Gunaratna
Department of Nutrition Science
Prof. Alexander Gluhovsky
Department of Statistics
Prof. Hao Zhang, Head
Department of Statistics

Approved by:

Prof. Jun Xie
Head of the School Graduate Program

I dedicate this thesis to my sweetheart, wife, friend and confidant Jessey,
and wonderful three (TAM) Arlene, Ardele, and Anele
for their invaluable love and support.

ACKNOWLEDGMENTS

Firstly, I would like to express my sincere gratitude to my co-advisors Prof. George P. McCabe and Prof. Nilupa Gunaratna for the continuous support of my Ph.D study, for their patience, motivation, persistence and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. I could not have imagined having better advisors and mentors for my Ph.D study.

Besides my co-advisors, I would like to thank the rest of my thesis committee: Prof. Alexander Gluhovsky, and Prof. Hao Zhang, for their insightful comments and encouragement, but also for their incise questions which helped me widen my research concepts.

My sincere thanks also goes to the Food Agriculture and Natural Resources Policy Analysis Network (FANRPAN), especially their ATONU team who provided me an opportunity to work with their data for my thesis. Without their precious support it would not be possible to conduct part of this research.

Hats off!!! To my Statistics department cohort for your unwavering support and understanding of me and my family and we journeyed through the academic terrain as Boilermakers. Forever Boilermakers, Go too far!!!

I am also grateful to my sponsors, Fulbright who provided me with such an amazing opportunity to study in USA, Statistical Consulting Services, who not only sponsored but also equipped me for statistical collaboration and consulting work, and the Gunaratna lab, Purdue summer grant, and the department of statistics. The sponsorship meant a lot to my family, and nation Zimbabwe.

Last but not the least, I would like to thank my family: my wife Jessey, and amazing kiddos, Arlene, Ardele and Anele, Maphango family, Jon Smith, Chi Alpha Christian family, and River City Church for supporting me spiritually throughout my Boiler life journey.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

ABBREVIATIONS

AR	Auto-regressive
ATONU	Agriculture to Nutrition
CI	Confidence interval
DGM	Data generating mechanism
FANRPAN	Food Agriculture Natural Resources Policy Analysis Network
LMIC	Low and middle income countries
NEAR	Newer exponential auto-regressive
NGO	Non governmental organization
PAR	product auto-regressive
RCT	Randomized controlled trial
WASH	Water and sanitation hygiene
WRA	Women of reproductive age

ABSTRACT

Matangi, Evidence S. Ph.D., Purdue University, December 2019. Handling Complexity via Statistical Methods. Major Professors: McCabe G.P. Professor, Nilupa S. Gunaratna, Assistant Professor.

Phenomena investigated from complex systems are characteristically dynamic, multi-dimensional, and nonlinear. Their traits can be captured through data generating mechanisms (DGM) that explain the interactions among the systems' components. Measurement is fundamental to advance science, and complexity requires deviation from linear thinking to handle it. Simplifying the measurement of complex and heterogeneity of data in statistical methodology can compromise their accuracy. In particular, conventional statistical methods make assumptions on the DGM that are rarely met in real world, which can make inference inaccurate. We posit that causal inference for complex systems phenomena requires at least the incorporation of subject-matter knowledge and use of dynamic metrics in statistical methods to improve on its accuracy.

This thesis consists of two separate topics on handling data and data generating mechanisms complexities: the evaluation of bundled nutrition interventions and modeling atmospheric data.

Firstly, when a public health problem requires multiple ways to address its contributing factors, bundling of the approaches can be cost-effective. Scaling up bundled interventions geographically requires a hierarchical structure in implementation, with central coordination and supervision of multiple sites and staff delivering a bundled intervention. The experimental design to evaluate such an intervention becomes complex to accommodate the multiple intervention components and hierarchical implementation structure. The components of a bundled intervention may impact targeted outcomes additively or synergistically. However, noncompliance and protocol devia-

tion can impede this potential impact, and introduce data complexities. We identify several statistical considerations and recommendations for the implementation and evaluation of bundled interventions.

The simple aggregate metrics used in clustering randomized controlled trials do not utilize all available information, and findings are prone to the ecological fallacy problem, in which inference at the aggregate level may not hold at the disaggregate level. Further, implementation heterogeneity impedes statistical power and consequently the accuracy of the inference from conventional comparison with a control arm. The intention-to-treat analysis can be inadequate for bundled interventions. We developed novel process-driven, disaggregated participation metrics to examine the mechanisms of impact of the Agriculture to Nutrition (ATONU) bundled intervention (ClinicalTrials.gov Identifier: NCT03152227). Logistic and beta-logistic hierarchical models were used to characterize these metrics, and generalized mixed models were employed to identify determinants of the study outcome, dietary diversity for women of reproductive age. Mediation analysis was applied to explore the underlying determinants by which the intervention affects the outcome through the process metrics. The determinants of greater participation should be the targets to improve implementation of future bundled interventions.

Secondly, observed atmospheric records are often prohibitively short with only one record typically available for study. Classical nonlinear time series models applied to explain the nonlinear DGM exhibit some statistical properties of the phenomena being investigated, but have nothing to do with their physical properties. The data's complex dependent structure invalidates inference from classical time series models involving strong statistical assumptions rarely met in real atmospheric and climate data. The subsampling method may yield valid statistical inference. Atmospheric records, however, are typically too short to satisfy asymptotic conditions for the method's validity, which necessitates enhancements of subsampling with the use of approximating models (those sharing statistical properties with the series under study).

Gyrostat models (G-models) are physically sound low-order models generated from the governing equations for atmospheric dynamics thus retaining some of their fundamental statistical and physical properties. We have demonstrated statistic that using G-models as approximating models in place of traditional time series models results in more precise subsampling confidence intervals with improved coverage probabilities. Future works will explore other types of G-models as approximating models for inference on atmospheric data. We will adopt this idea for inference on phenomena for AstroStatistics and pharmacokinetics.

1. INTRODUCTION

1.1 Chapter Overview

This chapter introduces the research problem and outlines the background and rationale for the present study. It subsequently describes the research questions and provides a chapter by chapter overview of the thesis.

1.2 Introduction and Background

Complexity is an attribute of a system under investigation, and not necessarily a trait of the mechanism through which an investigation is conducted [1]. It is defined through the dynamical interactions of the processes underlying or generated through a system. It is distinguished on the metaphor used to define the system under investigation as either a machine or an organism. The former advocates for linear thinking that is associated with simplicity, predictability, and that knowledge of the whole machine can be learnt from what is gathered from its parts. The latter view of a system as an organism accommodates for the interconnection of its parts, nonlinearity, and unpredictability in its dynamics. The basis of most statistical methods has been the machine-view of systems with assumptions that are being observed to be rarely met in real world systems. On the other hand, complexity in scientific research questions an be attributed to technological advancements and the emergence of new scientific research fields contributing to the complexity in their associated data. To make accurate inference for complex systems it is important to consider how measurement is conducted, and how are the statistical models generated and under what assumptions for the data generating mechanisms. In order for Statistics to contribute to the scientific goals and challenges exemplified by the 2030 Agenda for Sustainable Development and the global warming, there is a need to account for the context-dependent public health interventions, and the contributions of the underlying dynamics on atmospheric phenomena.

Addressing estimation and reliable inference problems is integral for the application of Statistics to other fields of study. The objective is to improve on solving real world problems through the contributions of statistical methods. The central mandate of statistical inference is the separation of signal from noise in data [2], as we seek to relate data with hypotheses. We endeavor to ensure that statistical significance complement subject-matter significance, and gain traction in their appeal to subject-matter audience. The complexity of the questions that scientists are seeking to solve, and the varying dynamics in the generation of their data, point to the need for advanced statistical inference methods [3]. Under such situations, inferential problems can be handled through considerations on how Statistics handle measurement and contextual factors, and the statistical assumptions postulated on the underlying data generating mechanisms (DGM) for the systems or organisms under study.

Systems and organisms consist of multiple and often interconnected components [4, 5], and they are characteristically dynamic, unpredictable, and multidimensional. They typically generate complex and heterogeneous data, whose reality can be lost in the simplification by statistical models. Climate, societies, and ecology are complex systems, and assuming that they work like machines leads to misleading estimation and inference [5]. Complexity theory moves from complex to simple, based on the interchange amongst a systems' components [6]. In order to understand phenomena in complex systems, complexity theory requires that we comprehend how things are connected, configured and constrained by systematic perturbations. The nature of causality in complex systems is non-linear (*small change can have big effects*) which introduces disproportionality in causal statements between machines and systems [5]. Emerging scientific fields, such as implementation science, translation science, complexity science, and systems science are a reservoir of theory that seek to handle such challenges. Investigations of complexity challenges traditional scientific approaches that uphold linear causal statements [7].

The reasoning behind most statistical model building is data-driven, which may fail to incorporate subject-matter expertise thereby limiting inference. The role of higher order statistical moments for climate and atmosphere phenomena emanates from the acknowledgement that their data are non-normal [8]. Heterogeneity in the generation of atmosphere data that is attributed to the underlying dynamics is a typical cause of skewness. Modern statistical methods such as bootstrap, and subsampling have taken a lead in making inference on complex data based on the empirical distribution function of the observed data. These methods are alternatives to statistical inference which often hinge on the assumption of parametric model underlying the observed data or where parametric inference requires complicated formulas for the calculation of standard errors. We envisage that there is a need for subject matter knowledge (*data-centric approach*) to be employed in the approximating distribution to ensure retention of data attributes from the complex DGM, getting away from the often rigid assumption on the DGM, for comprehensive and contextual relevant inference to be obtained.

The need to address the emerging and underlying determinants of public health problems has led to the development of bundled interventions, as an implementation innovation. These are multi-faceted interventions whose components work simultaneously to promote positive outcomes. The multi-pronged dimensions of public health challenges such as nutrition as exemplified by the double burden of malnutrition on obesity and under-nutrition, necessitate complex intertwining of strategies and approaches to handle them. Bundled interventions have been labeled "high-impact investments" but are often offset by the low quality of implementation in low-resource settings [9]. Culturally acceptable health promoting programs (*behavioral change communication*) together with the harnessing of agriculture can help alleviate malnutrition [10].

There is a need for interventions to illuminate the processes and mechanisms leading to the outcome, thereby providing useful information for their adoption for different populations and context [11]. The potential heterogeneity in populations of low and middle income countries' (LMICs) communities can put a strain on the reliability and relevance of bundled interventions inference, since the expected inter-cluster differences may be huge leading to possible confounding associations. The cultural/gender norms within the wider low-resource communities such as gendered relationships/patriarchy can deter the implementation of counter-cultural components of bundled interventions such as women empowerment, and may have a domino effect on the whole intervention participation, and consequently adoption. Non-consideration of masculine issues in development initiatives can challenge women's participation in patriarchal societies [12], which can gloss over the distinction between implementation effectiveness and intervention effectiveness resulting in non-adoption of potentially effective practices to curb public health issues.

In complex interventions, contextual dynamics impact on data quality and quality; and the hierarchical structure is a potential source of variation and bias which can influence the decisions on effectiveness evaluations. Implementation effectiveness precludes intervention effectiveness, and is immensely influenced by context. Adjusting for clustering and covariates offer a great advantage in the evaluation of complex interventions. The interactions between hierarchy and intervention components can contribute to the process dynamics in the implementation of bundled interventions which can be a helpful source for the explanation of the variation in the outcome of interest. The ability to capture the traits of the process-driven metrics allows for the understanding of the interplay of context, delivery and reception of interventions. These will serve to inform implementation quality and attribution of change in outcome of interest to the intervention, objectively. Such metrics can facilitate actionable courses to be undertaken for implementation improvement, which helps make the causal pathways become more clear.

Under a hierarchy structure and contextual dynamics, observational data are prone to the effects of immeasurable confounding variables, limiting the relevance of inference made. Process data can capture some of the confounding effects through metrics that are tied to the process dynamics, which are often not easily capture through conventional data collection methods. The role of technology in data collection allows for the capture of such intricate and yet vital data, as exemplified by the open data kit (ODK). This is a useful tool especially for resource-constrained environments that ensures privacy, and high participation rates, and also helps curb the prevalent challenge of social desirability bias. Statistical considerations on the implementation complexities can improve the understanding of the process dynamics of interventions to ensure sound recommendations on practice based on research findings. This allows for the adoption, sustainability, and scaling of interventions within the contexts of their study.

1.3 The Rationale for the Study

Complexity in systems cannot be explained objectively through linear thinking, when it is evident that such systems are inherently nonlinear. Creative approaches to the statistical inference are required to handle data arising from complex systems. Accommodating for this reality in our investigations aids our quest to address estimation and inference problems in statistical applications. Such adjustments puts traditional and conventional metrics, statistical methods and data generating mechanisms (DGM) assumptions on the spotlight, and calls for data-centric approaches that combine expertise knowledge and data for objective inference. The data revolution and the emergence of new scientific fields allows for more avenues for statistical applications requiring that we be confident of our tools on their relevance to such challenges. The endeavor to lead with Statistics entails that there is a need for statisticians to be pro-active and not necessarily reactive to the myriad of issues at the centre of scientific exploits. Developments in statistical sciences should strive to meet and address the needs of the ever-exploding world of science.

We seek to clarify the role and importance of Statistics methods and subject-matter theory in the evaluation and analysis of nonlinear systems whose underlying dynamics contribute to both the complexity and variability in data. Statistical significance should contribute to substantive or subject-matter significance for meeting the actual needs of the users. Measurement variation at cluster (*aggregate*) and individual (*dis-aggregate*) levels pose a difficult in causal statements for cluster randomized trials of complex interventions. This coupled with the fact that components of bundled interventions are often key facilitators to the expected positive changes, their combined effort makes it no mean endeavor ascertaining the causal pathway in a bundled intervention. We assert that causality can be attributed to the dynamics introduced from each of the levels of administration of the intervention leading to the outcome of interest. Practical and statistical considerations should be embedded in the implementation and evaluation design of bundled interventions, especially under resource-constrained environments.

The focus on first and second moments have ensured that statistical models assume on higher moments to validate inference on the former, which can be a source of missing information for the science being investigated as such higher moments could be containing the crucial information for their understanding. Given that atmosphere data is non-normal, inference on higher order moments, starting with skewness will present useful information on endeavors to understand them, The empirical data-driven distributions approximating the underlying DGM for the original data for subsampling method estimation and inference are simple and exhibit some of the statistical properties of the data. They however, have nothing to do with the subject matter properties of the original data which impedes on the relevance of inference that is obtained from them.

1.4 Contributions of the Study

This study will contribute to the current literature in the following ways. The proliferation and acknowledged relevance of bundled interventions in handling public health problems requires a statistical address on their implementation and evaluation design. This is particularly so for low resource settings where their postulated iterative and integrated design is flouted due to complexities attributed to the bundles' interactions with context within the hierarchy structure of their implementation. We highlighted the statistical issues that point to the implementation quality for bundled interventions and their consequence on effectiveness assessment and offered recommendations for handling them. Unlike traditional study designs that answer to specified problems singly, bundled interventions answer to a host of problems, which creates complexities in streamlining the implementation dynamics to adequately assess their effectiveness on the particular problems being investigated. The interplay amongst the intervention components contribute to their additive, synergistic, and antagonistic effects on the outcome of interest.These effects should be acknowledged in the theory of change to ensure the attribution of the change in outcome to the intervention, which is pivotal for their adoption, and sustainability. We developed and applied process-driven participation metrics that capture the implementation dynamics that are missed by the traditional simple and aggregate metric for intervention evaluations. We proposed a different set of statistical methodology for variation decomposition and identification of the determinants of the participation levels for bundled interventions. Different strategies and decisions were recommended for addressing the variation structures for the participation metrics to enhance the mechanism of impact for bundled interventions. Further assessment was conducted on how the proposed process-driven metrics enhanced the link between the intervention and the outcomes while accounting for the effect of contextual factors on them and the outcome.

The assumptions and necessary conditions for each problem assessment should be handled both uniquely and objectively within the confines of both the evaluation and implementation design with recognition of contextual influence. An application of these statistical consideration in the analysis of a bundled intervention will serve to highlight the importance of process data in handling them for low resource settings and giving credence to the process-outcome links envisaged. The hierarchical structure of bundled interventions is mainly for the purpose of applying an intervention on a wide spectrum of area and population settings. It can also emanate from the multi-disciplinary of the research team members and the multi-sectoral nature of the intervention components, including nutrition, agriculture, water and sanitation hygiene (WASH), that often work simultaneously. The hierarchical influence on the process dynamics, in particular on process variation attribution and how it relates to the variations in the outcomes of interest allows for process improvement through addressing how these impact implementation quality.

A data-centric approach to atmosphere data handling enhances the foray of statistical analysis and modeling in the geosciences. We seek to show how time series models derived from the governing equations of the underlying dynamics of the atmosphere can be used in statistical inference on atmospheric data. We seek to widen the applicability of subsampling methods in handling data with a dependent structure through a relaxation on the assumption on their underlying data generating mechanism (DGM). This is essential in ensuring the reliability of the inference made as they retain both the physics and statistical properties of the original data. The flexibility of such models to incorporate more mechanisms akin to the explanation of the underlying dynamics, offers a leeway for their further expansion to ensure that the DGM captures the reality of the original data.

The possibility of adopting such models opens a door for statistical modeling of data in domains where mathematical modeling has mostly been used, which include but not limited to pharmacokinetics, disease modeling, and the linking of astrostatistics data to its underlying theory.

1.5 Research Questions

This research seeks to address the following research questions emanating from two studies undertaken concurrently on bundled nutrition intervention and atmospheric data handling.

(i) What are the statistical issues that need to be taken into consideration for the successful implementation and evaluation of bundled interventions?

(ii) Does controlling for clustering together with process-driven participation metrics improve causality statements for bundled interventions?

(iii) Can data-centric approximating models for the underlying atmospheric dynamics facilitate reliable inference on atmospheric data?

1.6 Methodology and Main Findings

The use of process data which captures dis-aggregate data, and reveals the sources of variation in its hierarchy structure, in the linear mixed modeling of bundled interventions data allows for process improvement. This highlights the areas that need to be improved on for implementation quality, and ascertain the effectiveness assessment of such interventions on addressing the problems consortium under investigation.

The implementation of the Agriculture to Nutrition (ATONU) nutrition sensitive agriculture bundled intervention in Ethiopia and Tanzania was characteristically heterogeneous. This had an impact on the intervention's implementation quality and effective assessment, and vital statistical considerations have to be adjusted for to handle these aspects. Process-driven participation metrics for ATONU intervention on the dietary diversity index for women of reproductive age (WRA) in Ethiopia showed that significant variation in them was attributed to both intrahousehold and inter-household variation within the unit of randomization. In conventional clustering randomized controlled trial (cRCT) studies such information is not revealed as metrics are often aggregated at cluster level for the assessment of population level change. Both ecological fallacy and aggregation bias (*loss of detail due to aggregation*) can be attributed to the challenges that so often surrounds the adoption of misaligned effective interventions that fail to be translated to practice and policy for public health issues.

Statistical procedures seek to reach a decision on postulated hypotheses, and to do so they rely on the assumptions of the statistical models. In our aim to make inference on atmosphere data, we employed G-models for subsampling confidence interval construction, and obtained narrower intervals. These are physically sound models that are derived from the underlying governing equations for atmospheric dynamics [13]. AR(1) models have been frequently used to model climate data because of their ability to handle correlated time series [14]. G-models' advantage over AR(1)-based nonlinear models is in their ability to capture both the physics and the statistical properties of the atmospheric data. The accuracy of such confidence intervals hinge on the determination of the subsample size, otherwise considered as the block size b, which helps in ensuring that the actual coverage is in sync with the target coverage for appropriate interpretation of the findings. The block sizes we obtained was comparable to those used in previous works done for subsampling confidence intervals for atmosphere data. The subsampling confidence intervals obtained with G-models as approximations of the underlying dynamics were narrower than all previously computed ones.

1.7 Structure of Thesis

Here is an overview of the chapters in this thesis; chapter one focuses on the introduction, rationale, motivation, the research problems being investigated, and the major findings made. Chapter two focuses on highlighting the statistical considerations in the implementation and evaluation design for bundled interventions and possible solutions to address them. Chapter three offers an application of process-driven metrics in the evaluation of a bundled intervention, showcasing some solutions on handling statistical considerations on heterogeneity in implementation. Chapter four gives an overview on investigating inferential relevance based on Monte Carlo (MC) simulations for atmospheric data through time series models generated from their underlying dynamics. The study seeks to utilize these models for subsampling confidence interval for parameters of non-normal atmospheric data, as they allow for the incorporation of the physics defining the data. Lastly, chapter five offers conclusions drawn from the studies, recommendations, and future research suggestions.

2. STATISTICAL CONSIDERATIONS FOR HIERARCHICALLY IMPLEMENTED BUNDLED INTERVENTIONS

2.1 Abstract

Although the randomized controlled trial (RCT) is considered the gold standard for assessing interventions, many nutrition studies use experimental designs with more complex structures. We examine one class of such designs, hierarchically-implemented bundled nutrition interventions, with particular focus on the unique statistical issues associated with these studies. Hierarchically-implemented studies involve several levels, such as the individual, the household, the village, and the region, that must be carefully taken into account in the planning and execution of the study. A bundled intervention includes a collection of interventions, with separate but often complementary objectives, that can be implemented at different levels of the hierarchy. Statistical considerations for bundling and hierarchical implementation are described, and recommendations are proposed which include the development of process-driven participation metrics, power and sample size optimization, context and spillover measurement, and the use of analytical methods that take into account both clustering and covariates.

2.2 Introduction and background

Nutrition interventions often address problems with connected underlying causes such as the double burden of malnutrition. They require a sound evidence base for adoption for the at-risk-populations. Such interventions need to be implemented in the context of sound Theory of Change (ToC); which are often complex and consist of multiple pathways to the target nutrition outcomes. Communal public health issues are often multidimensional and cannot be addressed through single interventions. The bundling of interventions is an innovative design, which is defined as multiple interventions combined to address public health problems. The bundled components can be instructional sessions, reminder messages, and activities. They can contribute additively or synergistically to the target nutritional outcomes. The effectiveness for bundled interventions hinges on the accounting for the complexity involved in implementing their components.

When bundled interventions are scaled to target geographically dispersed populations, their implementation becomes hierarchical structured. The dissemination of their components requires consolidated support systems through hierarchical structures to achieve the desired public health impact [15]. Decision-making, mobilization initiatives, and interpersonal communication during the implementation process can influence participation dynamics.

The analysis of complex social interventions as single entities without comprehensive integration of the components is challenging [11]. There is a need to address the possible consequences of interactions among bundled components and with the hierarchy levels. A complex systems approach to such interventions, viewed as events in systems, emphasizes the role of context [16]. The careful examination of the implementation process can help to assess how the target effects are attained [15]. The ToC should provide a framework for describing the pathway on how and why a desired change can be achieved through the intervention. The "implementation gap" is the challenge for translating research evidence into routine practice. The dynamics of the five domains of the Implementation Science in Nutrition (ISN) framework [17] are crucial for addressing this "implementation gap". The five domains are:

(i) The object of implementation.

(ii) Implementation organizations and staff.

(iii) Enabling environment.

(iv) Participants.

(v) Implementation process.

Bundling and hierarchical structure can enhance the effectiveness of bundled interventions at the individual level through engagement with the different components. These innovations also present statistical challenges for the intervention's evaluation that requires a critical analysis of the whole implementation process for appropriate conclusions to be drawn [18].

The bundling of interventions has been shown to be an efficient technique [19] which has been applied for public health as care bundles, community-based, and nutrition-sensitive agriculture interventions. They have been effective for acute health problems in high resource settings [20]. The personalized nutrition care bundle that was created by the American Society of Parenteral and Enteral Nutrition (ASPEN) in conjunction with the Society of Critical Care Medicine (SCCM), sought to optimize patients' nutrition statuses during acute care admissions [21]. It consisted of the following six components:

(i) Malnutrition assessment.

(ii) Initiation and maintenance of enteral feeding.

(iii) Reduction of aspiration.

(iv) Implementation of enteral feeding protocols.

(v) Avoidance of gastric residual volumes use for tolerating enteral nutrition.

(vi) Non-initiation for early parenteral nutrition when enteral feeding.

Its effectiveness depended on patients' demographics and the involvement of diversified professional staff handling varying components of the bundle. The additive, and synergistic effects of the components need to be acknowledged for the bundled intervention to be viewed as a single entity [22], for aggregate beneficial effects on the outcome [23].

As interventions are scaled, their hierarchy structures promote planning for easy and efficient use by implementers thereby facilitating intervention effectiveness [24]. The Realigning Agriculture for Improved Nutrition (RAIN) was a hierarchically implemented bundled intervention focusing on child nutrition in rural Zambia. RAIN's structure involved a primary level (infants at baseline and their parents), a secondary level (women's groups), and a tertiary level (implementing organizations) [25]. Strong implementation emphasis and effective monitoring were significant for RAIN's effectiveness. Understanding the change process within hierarchy helps in the identification of factors that promote the development and implementation of interventions [26].

The agriculture to nutrition (ATONU) bundled intervention was implemented in Ethiopia and Tanzania to improve the nutrition status for subsistence farmers through behavioral change communication [27]. It consisted of the following five thematic components:

(a) Family nutrition.

(b) Dietary diversity.

(c) Maternal infant and young child feeding (IYCF).

(d) Women empowerment.

(e) Home gardening.

Figure 2.1 shows the hierarchical structure designed for ATONU implementation.

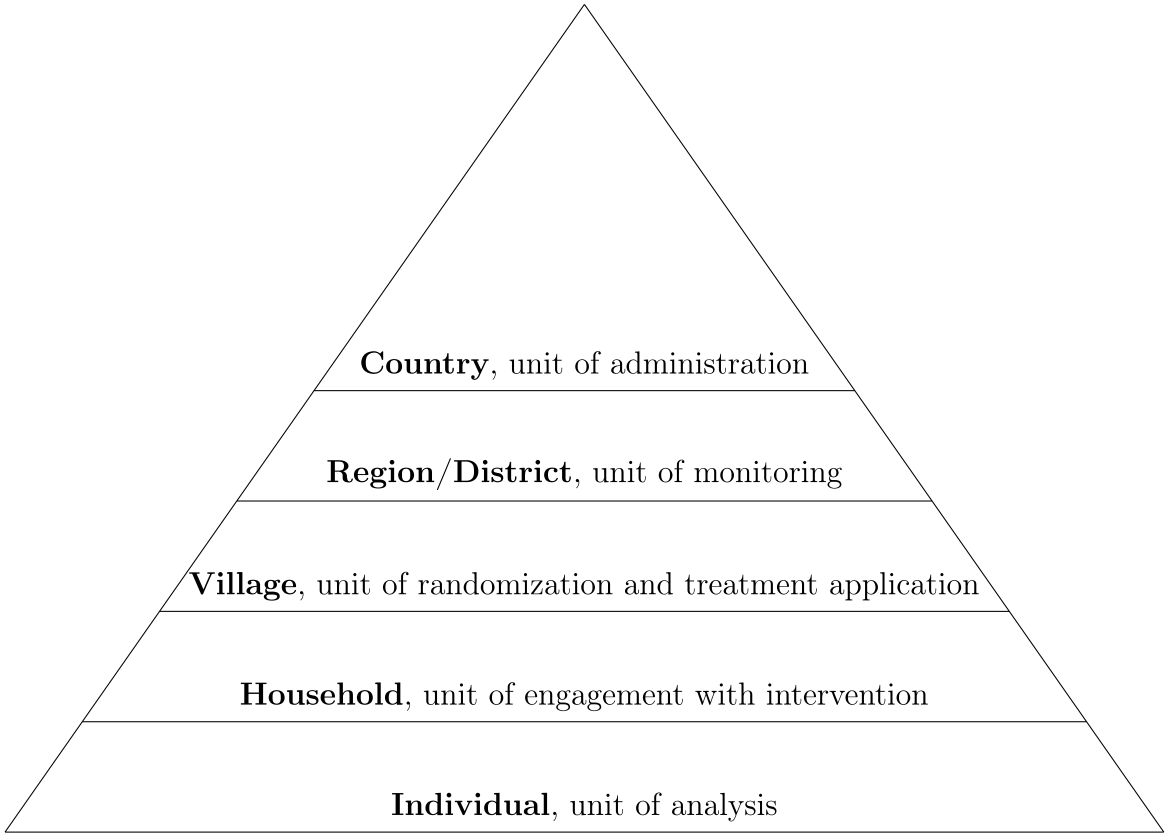

Figure 2.1. Hierarchy structure for ATONU implementation

2.3 Statistical considerations and recommendations

The purpose of this paper is to highlight the statistical considerations for the implementation and process evaluation of bundled nutrition interventions and to make recommendations. These statistical considerations assist in explaining the conduct of bundled nutrition interventions to ensure that precise and unbiased intervention effectiveness are obtained for the possibility of the transition of research evidence to nutrition practice and policy. This underscores the need for implementation effectiveness that can help to ascertain the effectiveness of bundled intervention and identify their success factors.

2.3.1 Bundling innovation

The ToC is a tool for developing and evaluating complex interventions [28], and there is little knowledge about its use for public health interventions [29]. The Medical Research Council (MRC) evaluation guidelines fail to incorporate theory-driven approaches [30]. We posit that the ToC for bundled interventions is complex, involving additive, synergistic, and potentially antagonistic contributions from the components.

The Engaging Fathers for Effective Child nutrition and development in Tanzania (EFFECTS), is a bundled nutrition intervention that seeks to assess the impact of father's involvement on children's nutrition. Its ToC consists of nutrition and parenting pathways that link and explain child nutrition and morbidity outcomes with the intervention's components. They causally connect the messages and activities on water and sanitation hygiene (WASH), infant and young children feeding (IYCF), women empowerment, parenting knowledge and practices, and nutrition knowledge to the target outcomes. The ToC exhibits an integrated and iterative linkage of the components capturing their additive or synergistic effects.

The following are the statistical concerns for bundling, highlighted by the resource constraints to testing the components individually. The intermediate outcomes derived from the components are often not measured yet the ToC suggests additive and synergistic links of the components. The main objective for intervention research is treatment effects, and a measure of how the bundled components evolve to the target outcome needs to be captured. Based on Rubin's motto "no causation without manipulation", the outcome change should be associated with the bundle components' manipulations [31].

The robustness of the causal links needs to consider the impact of implementation quality for bundled interventions. The participants may not get all the bundle components and under such circumstances, the additive and the synergistic effects may not be fully realized. The complex causal structure for bundled interventions may require appropriate process data for their assessment.

We recommend the need to develop ToC based on literature with hypothesized interactions among the bundle components to address the above-mentioned statistical concerns for bundling. Power, effect size, and sample size calculations and justifications should form an integral part of the ToC. Intermediate outcomes need to be measured to explain the implementation dynamics associated with bundling. Process outcomes can potentially offer more evidence than observational or perception measurements. Metrics about the components delivered and received by the participants help to show the extent of interaction with the bundled intervention. They would allow for the contribution of the bundle's components to the effect and variation in the target outcomes.

2.3.2 Heterogeneous implementation

The delivery and reception of bundled interventions can vary in terms of content, capacity, timing, and participants' motivation. Population level risk factors such as poor sanitation, lack of education, infrastructure, illiteracy, and poverty can affect engagement with the bundle components. Given heterogeneity in the target population, implementation may purposely be varied to reach the targeted participants. Implementation heterogeneity can be intentional when adapting to local context for food culture, availability, affordability, and seasonality of diverse foods. It can also be unintentional when there is poor delivery, and variability in the competence of the implementation staff.

The locally adapted ATONU's implementation was heterogeneous on content delivery, delivery timing, and staff retention due to varying socio-economic, climatic conditions, and staff turnover. These factors may have negatively impacted on the delivery decisions and necessitated implementation heterogeneity. In the evaluation for bundled interventions, unintentional implementation heterogeneity would bias us towards the null hypothesis, while intentional heterogeneity would bias away from the null hypothesis. The non-rejection of the null hypothesis may not entirely be due to failure of the intervention theory, but also implementation challenges. Process metrics need to be considered as they can adequately capture the mechanism of impact through tracking the engagement dynamics. Compliance metrics can be used to capture the retention levels, i.e. the extent of intervention reception. However, small sample size and effects attributed to depressed values on these alternative metrics can exacerbate the low statistical power challenge, and statistical significance may be due to false positive results. Caution must be taken in delivering conclusions for decisions on the adoption of bundled interventions.

Observational assessment of adoption bundle activities or messages may fail to capture the effect of unmeasured confounding variables. These may undermine their contribution in the evaluation of the intervention by biasing towards the null hypothesis. On the contrary, they can heighten the Hawthorne effect in conjunction with the social desirability bias.

Implementation heterogeneity can impact uptake and the effectiveness of the intervention [32]. Adequate sample size and statistical power are needed to improve uptake of bundle components and enhance effectiveness. The delivery and reception of the bundle components maybe heterogeneous which may limit their overall effectiveness [23]. In such scenarios, statistically insignificant conclusions may yield vital trial trends, for which post-hoc power computations are needs to inform future research for sample size considerations to facilitate the detection of significant differences [33].

We recommend the need for comprehensive data collection and development of metrics based on the implementation dynamics, for use in the analysis. A consideration of the intention to treat analysis for evaluating the effects of bundled interventions is a viable alternative to comparison with a control group, or in the presence of treatment heterogeneity [34]. This however, maybe inadequate as spillovers and contamination may be present. Tracking the individuals will capture their compliance to the protocol of the intervention in relation to their randomization assignment. Furthermore, we recommend the establishment of guidelines for monitoring participation levels for the bundle components which allows for composition data analysis.

2.3.3 Hierarchical/vertical implementation

Bundle components may have additive, synergistic or antagonistic effects at all or some of the hierarchical levels [35]. Hierarchy can limit bias and promote the internal validity of the bundled intervention studies through facilitating for adjusting for clustering in evaluations. This reduces the likelihood of spillovers, and can help capture the sources of variation for bundled interventions.

Participation dynamics can effect variation in the target outcomes, which often respond to the implementation framework defined by the hierarchy structure. Ecological fallacy is an inherent misconception on causal inference [36] that is shown in the assumption that what is true for a group holds for the individuals. Group positions can be influenced by stereotypes attributed to research lag typified by female disadvantage on education [37], which may not translate to individual females. The ATONU bundled intervention's implementation varied in terms of ecology, governance, socio-culture characteristics, which mirrored its hierarchical structure.

The main statistical concern for hierarchy structured bundled intervention is the need to adjust for clustering. This helps in capturing and explaining the sources of variation both in the implementation and outcome metrics. The failure to account for clustering can lead to spurious conclusions [38]. We need to have appropriate power/sample size for the experimental design to facilitate for the objective assessment of bundled interventions' effectiveness. The analysis of bundled interventions, calls for the use of multilevel models that adjust for clustering at all necessary levels. Mediation analysis can help identify the facilitators and inhibitors for their effectiveness.

Design effects and variability obtained at the appropriate hierarchical levels can be used to correct for statistical inference as cluster randomization is prone to spillover effects that bias towards the null hypothesis due to social interference [38]. There is need to define, identify, and estimate spillover effects, and control for them in the process evaluation for bundled interventions.

2.3.4 Varying context

Contextual variation in the target population or the physical, social, or institutional environment, becomes visible in the interactions of context with the bundle components in their implementation. Context shows the prevalence or severity of the challenges under investigation [11]. An understanding of intervention adaptation to context can help elaborate the processes leading to the target outcome [39]. Another challenge in understanding effect modification is the population heterogeneity for which personal attributes are the stand-out factors [40]. The key characteristics of individuals tend to vary in the clusters, and may confound on the observational data on the intervention [41]. Internal validity and causal pathways consolidation requires a consideration of the potential socio-economic inhibitors in the intervention clusters [42].

Social drivers of causality in interventions cannot be controlled under different contexts and they also accentuate the variation in intermediate outcomes for the bundle components. The intervention's effects on the outcomes can be heterogeneous and context-specific and dependent on the quality of implementation [43].

Contextual implementation research should endeavor to define the acceptable methodological rigor for sound results under real world conditions [17]. ToC is vital for intervention planning through identifying the underlying conditions and assumptions and acknowledging contextual effects [44]. Varying contexts allow for the presence of unmeasured confounding variables that can affect the causal statements for bundled interventions. These could lead to the occurrence of Type III error in the conclusions drawn under such competing factors.

Type III error is correctly rejecting the null hypothesis but for the wrong reason, which needs to be avoided [45]. This is exemplified by a situation where another program being operated within our treatment group had the positive effect, and our intervention had no effect. This error can be a consequence of contextual factors beyond the control of the intervention.

We recommend that key stakeholder and formative research input be incorporated to gain insight into the contextual attributes for addressing public health issues. Context must be measured hence data collection and appropriate metrics should be at all levels of the hierarchy. Analytical methods should adjust for background characteristics of the heterogeneous population and the process-driven participation metrics. To address the confounding problem associated with heterogeneous populations there is a need to measure as many variables as possible and adjust for them in the evaluation of the intervention [41]. Heterogeneous target population requires that the sample size be sufficiently large for significant conclusions to be drawn [46]. To avoid Type III error, documentation of competing events and the interactions of the intervention with context and the minimization of contamination are fundamental.

2.4 Discussion

This study highlighted the statistical considerations for the implementation and evaluation of hierarchically implemented bundled nutrition interventions. We acknowledged bundling and hierarchy as implementation innovates for nutrition interventions. Four statistical issues were identified requiring careful statistical thought for the betterment of evaluation for bundled interventions. They were about bundling, implementation heterogeneity, implementation hierarchy, and varying contexts. Their significance lies in the facilitation for contingency measures to ensure that implementation and intervention effectiveness remain the goals for bundled interventions.

We recommended sound ToC, the development of process-driven participation metrics, power and sample size optimization within the bundled components. Context and spillover measurement and the use of analytical methods that adjust for clustering and implementation dynamics covariates were also recommended. The rigorous data collection proposed may seem to be a burden for bundled interventions, but new technologies can help alleviate it. Tools such as the open data kit (ODK) allows for real-time monitoring, and corrective action to be undertaken on implementation.

These tools are becoming ubiquitous even for developing countries due to improved internet access and exposure to smartphones and electrical gadgets. Documentation of the processes involved in the implementation can aid in the specification of the causal pathways for bundled interventions.

Adaptability of bundled interventions to local contexts while minimizing contamination, and ensuring comparability enhances the generalization of their findings. Statistical modeling should adjust for contextual and hierarchical level-specific covariates in causal inference [47]. There is need for delivery capacity and reception optimization metrics for bundle components under constrained resources to ascertain their feasibility.

In order to ensure the translation of research to routine practice, implementation effectiveness should be distinguished from intervention effectiveness [48]. This helps in separating intervention failure from implementation failure which impact on the adoption for potentially effective bundled nutrition interventions in real world.

These highlighted statistical considerations may need to be addressed for the contribution of bundled nutrition intervention to ISN research. They can serve to improve on their implementation quality, evaluation, adoption, sustainability, and scaling.

3. PROCESS-DRIVEN METRICS AND PROCESS EVALUATION OF BUNDLED INTERVENTIONS: THE AGRICULTURE TO NUTRITION (ATONU) TRIAL

Abstract

Background

Bundled nutrition interventions examine causes of nutritional deficiencies through the additive and synergistic effects of their components. Their implementation is often heterogeneous due to contextual confounders that impact their effectiveness.

Objective

We propose process-driven participation metrics to capture implementation dynamics and apply them to a bundled nutrition intervention, the Agriculture to Nutrition (ATONU) intervention. We generate specific recommendations to improve implementation quality and evidence for impact of the intervention the primary outcome, women's dietary diversity.

Methods

A cluster randomized experimental design was used for the agriculture to nutrition (ATONU) intervention in Ethiopia and Tanzania. Villages formed the clusters. The aim of the intervention was to improve the nutritional welfare of vulnerable members in subsistence farming communities. The metrics were compliance, bundled intervention components received (BICR), and gender-specific engagement for men, women and joint. Beta-logistic and logistic models were used to determine the sources of

variation in the process-driven metrics. Further, generalized mixed models were applied to link the intervention and the outcome, the dietary diversity for women of reproductive age (WRA) at the end of the intervention.

Results

The implementation of ATONU among the villages in Ethiopia and Tanzania was heterogeneous in terms of content delivery and timing of delivery. Variation in compliance was greater within villages, and variation for BICR was greater between the villages. To improve compliance, focus should be on participants' mobilization and for BICR, the administration of the research staff must be revamped. The linear mixed model was a better fit than the Poisson mixed model for the dietary diversity score for WRA. Compliance was a significant determinant of the mechanism of impact of bundled intervention on the WRA's dietary diversity. Adjusting for clustering, compliance, livestock diversity, baseline dietary diversity score, and contextual factors is important for the process evaluation of the bundled nutrition intervention.

Conclusion

Bundled interventions are needed to improve nutritional outcomes. Their evaluation requires a focus on the individual participants and accounting for implementation heterogeneity in different settings. A considerable amount of participation variation is due to inter-household and intra-household factors. The linear mixed model with adjustments for clustering, process-metrics and contextual covariates can significantly explain the change in women's dietary diversity scores.

3.1 Introduction

Malnutrition is a multifactorial problem that requires holistic and multidimensional interventions [49]. Bundled nutrition interventions are nutritional methodologies for solving complex nutrition problems in communities. Their implementation in varying geographical locations introduces a hierarchical structure that impacts on the delivery and reception of bundled components. Observational studies based on aggregate metrics have been shown to be effective ways to improve on nutrition outcomes in women and children [50].

Public health interventions are frequently implemented at the cluster level to minimize costs and contamination, and for administrative convenience. Their metrics are often aggregated, however they seek to address population level changes of outcomes that are captured at the disaggregate level. Aggregate metrics though simple, neglect information on the implementation dynamics for bundled interventions. Decisions based on aggregate metrics for changes in populations are prone to the ecological fallacy problem, where inferences about individuals are deduced from inference about the group to which those individuals belong. Observational studies fail on the establishment of causal statements to link the interventions to the nutrition outcomes [50], because of the presence of unmeasured confounding variables. On the other hand, bundled interventions conducted in communal settings lack clear evidence of impact as they focus on distal instead of proximal measures for women's nutritional outcomes [51]. Gender inequities on food decisions and participation dynamics are potential causes for such effects. We posit that the engagement of participants with the components of bundled nutrition interventions is essential for their effectiveness.

The dietary diversity score is a key proximal indicator for women's nutritional adequacy and quality. It is defined as a function of several food groups eaten within the previous 1 or 7 days. The women's minimum dietary diversity (MDD-W) is defined in terms of the following ten food groups, (i) staples, (ii) pulses, (iii) seeds, (iv) dairy produce, (v) meats, (vi) poultry produce, (vii) green vegetables, (viii) fruits and vegetables containing Vitamin-A, (ix) non-green vegetables, and (x) non-Vitamin A rich fruits [27]. Rural communities in developing countries perennially face the problem of poor dietary diversity [52]. The MDD-W score for rural Ethiopia farming communities has been shown to be poor and beyond the solution of home gardening interventions [27, 53]. Diets for WRA are typically monotonous and of low quality for low and middle income countries (LMICs) [54], and have been found to be low on diversity [55]. Increasing dietary diversity could potentially reduce the burden of malnutrition [56].

Randomized controlled trials (RCT) for socially complex interventions have been acknowledged to be problematic in their evaluation [57]. Bundled nutrition interventions involving nutrition behavior change communication can be characterized as complex. They lack blinding, involve heterogeneous participants and may be implemented heterogeneously, and have difficulty in controlling for confounders [58]. These attributes may violate the conditions for them to be assessed as standard cluster randomized controlled trials (cRCTs) and thereby fail to guarantee attribution of causation to the interventions [59].

We examine determinants of how and why change occurs through the process, dynamics and conditions of intervention implementation. A lack of intervention effectiveness can be attributed to imprecise measurement [60], and poor evaluation makes evidence of interventions effect inconclusive [61]. As a result it may be difficult to get information to improve processes, to ascribe causality; and to establish ecological validity, i.e. to generalize research findings. There is a need for metrics that can be captured for contextual effects. Defining concepts and developing measurement tools are crucial for ascertaining causal relationships and generalization of bundled interventions' findings. These are crucial for their adoption, sustainability, and scalability.

There is need for appropriate metrics and relevant methodologies to monitor and evaluate bundled nutrition intervention [62]. These can illuminate their mechanism of impact and thus provide an evidence base for the generalizability of their findings. The understanding of their change process can offer feedback for the consolidation of their complex theory of change (ToC) framework and hypotheses for the determinants for positive nutrition outcomes. There is a need in delivery-system research for the understanding of the process underlying the intervention [63]. This requires process data which is difficult to collect. The availability of smart technologies such as open data kit (ODK) especially in developing countries can facilitate the capture and management of process data.

Process methods and metrics are needed for capturing participant engagement given heterogeneous implementation where compliance confounds intervention delivery and participant engagement. Metrics are needed for participant engagement distinguishing it from delivery. We propose process-driven participation metrics that (a) allow for the individual tracking of participants, (b) quantify compliance of a package of components, and (c) quantify gender inequities in participation. We demonstrate that these novel process-driven participation metrics can be used to improve the implementation and establish the process-outcome link of heterogeneously implemented bundled interventions. Our objectives are to:

(i) develop metrics that capture participation dynamics for bundled interventions.

(ii) identify factors that explain variation in household participation metrics for the Agriculture to Nutrition (ATONU) bundled intervention.

(iii) identify contextual factors that define the change process linking ATONU bundled intervention to WRA's dietary diversity.

3.2 Methods

3.2.1 ATONU intervention

The Food, Agriculture, Natural Resources Policy Analysis Network (FANRPAN) initiated ATONU to promote nutritional security for the vulnerable WRA and young children in sub-Saharan smallholder farming families. This was implemented as a cRCT in Ethiopian and Tanzanian villages during the period February 2017 to April 2018. It focused on behavior change communication and had the following five thematic components: family nutrition, dietary diversity, maternal infant and young children feeding (IYCF), women's empowerment, and home gardening. These were administered through group discussion meetings, home visits, and practical activities.

3.2.2 Study area

We did not have access to the outcome data for Tanzania, hence we focused our study on Ethiopia. The study area was a low resource smallholder farming rural area with varying agro-ecological zones, and social norms. Data was collected from 20 villages from the 4 study regions and in each village 40 households were targeted. The regions served as strata from which villages were randomly sampled and assigned to the treatment arms. Our focus here is on the treatment arm only.

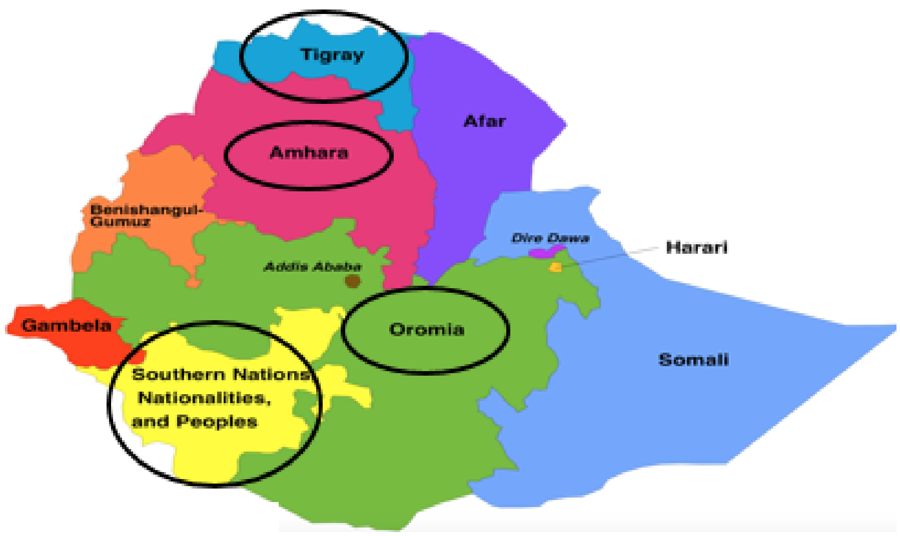

Figure 3.1. ATONU study regions in Ethiopia (circled)

3.2.3 Implementation dynamics for ATONU intervention

Heat maps were used to visualize the implementation dynamics of ATONU between the two countries and among the regions and villages.

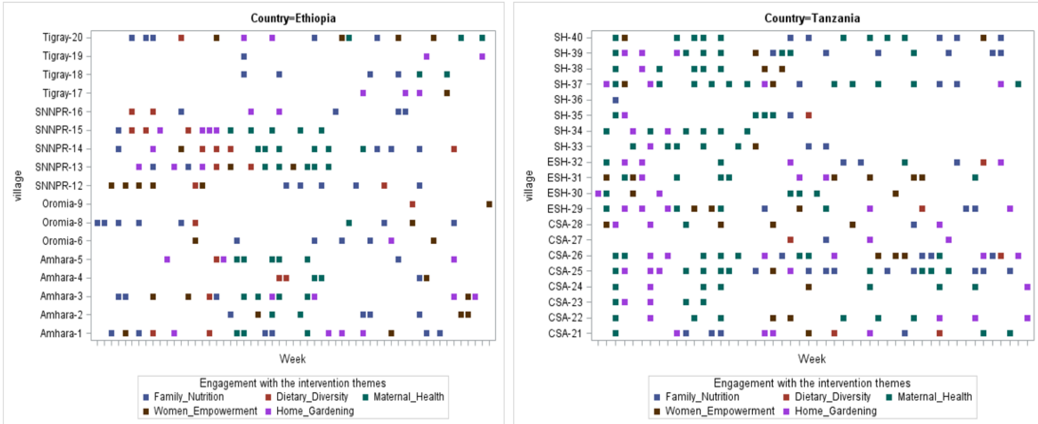

Figure 3.2. Implementation dynamics of the bundled components for ATONU in Ethiopia and Tanzania

Figure 3.2 shows that the implementation of the five bundle components between the two countries was heterogeneous. . If the implementation was done homogeneously, the display of the heat maps would relay a distinct and similar pattern within all the regions and villages. This however was not so, the delivery in Ethiopia for Tigray-20 was delayed and thereafter some consistency prevailed yet in Tigray-19 it was delayed and scantily as only two messages were delivered. This pattern is prevalent both between the two countries and also within the regions and villages, showing that there were peculiar contextual factors that determined the delivery-reception dynamics for the bundled intervention. The heterogeneity was in terms of delivery content and timing which could be attributed to staff turnover, contextual and background characteristics of the participants. Seasonal variation could not be identified as the intervention was conducted over a short period of time. Mobilization incentives of seeds and cooking activities were the dominant home gardening and maternal IYCF components.

3.2.4 Participation metrics

Conventional participation metrics at cluster level are typically attendance, coverage, and dose received. They focus on either but not both dimensions of participation, i.e. frequency and extent of involvement. They neglect substantial information on participation dynamics in relation to the bundled intervention as shown for coverage in comparison to retention in Tables 1 and 2 below.

Table 3.1.
Same coverage, good retention scenario.

Participant	Time 1	Time 2	Time 3	Time 4	**Retention**
1	X	X	X	X	100%
2	X	X	X	X	100%
3					0%
4					0%
Coverage	50%	50%	50%	50%	**50%**

The coverage in Table 3.1 is overall 50% and fails to capture the non-participation and thereby suppress the variation among participants. The retention metric captures the non-participation and hence allows for the variation in analysis. The bundled intervention may have impact on only 50% of the target population.

Table 3.2.
Same coverage, poor retention scenario

Participant	Time 1	Time 2	Time 3	Time 4	**Retention**
1	X	X			50%
2	X	X			50%
3			X	X	50%
4			X	X	50%
Coverage	50%	50%	50%	50%	

Table 3.2 shows 100% coverage but does not capture the extent of involvement thereby failing to reveal the non-participation in the paired time slots. On the other hand, retention levels of 50% reveal that there was non-participation but cannot distinguish it in terms of delivery times. The interventions may not have the intended impact on all the participants as they each received half of the bundle components.

The other dimension of disparity on coverage is when we factor in the gender inequities prevalent in patriarchal communities which impacts decisions to participation. Suppose the target group consists of 20 households in which an intervention is targeting participation of both the husband and wife. We propose that behavioral change in household on nutritional status requires mutual participation of the adults. We can have the participation scenarios depicted in Table 3.3 below.

Table 3.3.
Same household coverage, different gender composition scenarios

Case	Female only engagement	Male only engagement	Joint engagement	Coverage
1	0	0	10	50%
2	5	5	0	50%
3	10	0	0	50%
4	0	10	0	50%

The case 1 is ideal but shows that the bundle intervention would impart only 50% of the target populations, and the other cases shows no impact as only half of the target audience is receiving. The coverage situations shown in Tables 3.1 to 3.3 forms the basis for our argument for individualized process-driven participation metrics. Process metrics are valuable for the description of the functioning of interventions in real world. They support causal statements for bundled interventions, and inform and improve implementation quality. These metrics will allow for tracking of participation over the continuum of the intervention's lifespan, their engagement with the different components, and gender disparities. We proposed the metrics for compliance, bundle intervention components received (BICR), male participation, female participation, and joint participation.

The compliance metric tracked the individual participants' engagement with the intervention over its lifetime, i.e. retention.

$$Compliance = \frac{Count\ of\ messages\ received}{Count\ of\ messages\ delivered} \tag{3.1}$$

Compliance is a function of the process dynamics of delivery and context which influence decisions to participate. It captures the frequency of attendance and the extent of involvement in relation to delivery.This metric has similar traits to those of the compliance metric for clinical trials. The implementation heterogeneity shown in Figure 3.2 can be revealed through this compliance metric. However, it does not retain the timing of engagement with the bundle components.

The bundled intervention components received (BICR) metric quantifies the extent to which individual participants engaged with the bundle components. It is a function of content, the contextual effects on implementation, and the background characteristics of the participants. However, it does not preserve participation time order.

$$BICR = \frac{Count\ of\ bundle\ components\ received}{Expected\ count\ of\ bundle\ components\ delivered} \tag{3.2}$$

The gender coverage metrics are binary measures for joint, male, and female engagements with the bundle intervention. These ascertain the social drivers for participation. They are functions of the frequency dimension of participation and they do not capture retention and participation time. The female participation metric in 3.3 below illustrates the gender participation metrics.

$$Female\, participation = \begin{cases} 1 \quad \text{if woman attended in household attended at least one meeting,} \\ 0 \quad \text{otherwise.} \end{cases} \tag{3.3}$$

3.2.5 Variance decomposition and Mediation analysis

Errors bars were used to describe the variation in the process-driven metrics of compliance and BICR. Based on the model in Figure 3.3 below, we sought to develop a framework for the process evaluation for ATONU bundled nutrition. We sought to demonstrate the causal relationships among the intervention, context and outcomes, facilitating for the no confounding assumption [64] through utilizing as much data as possible from the intervention, context and background characteristics of the participants and adjustments for clustering. Conventional mediation analysis often use regression models that do not adjust for clustering. We argue for its accommodation because of the hierarchy structure for bundled nutrition interventions implementation. We identified the determinants of the process-driven participation metrics and the WRA dietary diversity scores for 24 hours and 7 days recall.

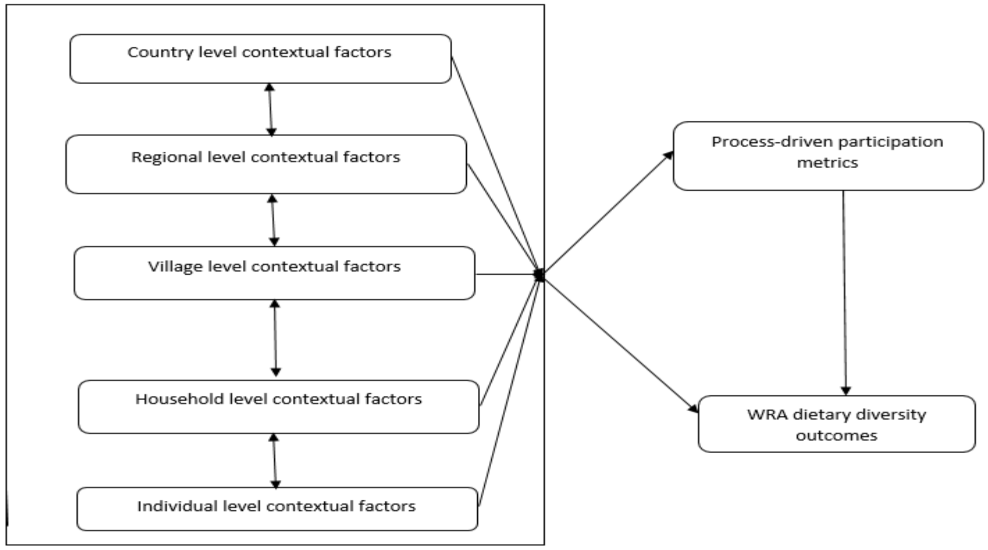

Figure 3.3. Mechanism of impact for hierarchical structured bundled ATONU intervention

Beta-logistic and traditional logistic models were used to investigate the link between context and demographic variables with the process-driven metrics. Linear and Poisson mixed models were used for the mediation analysis of the WRA dietary diversity outcomes.

The proposed metrics of compliance and BICR are proportions at the individual level and do not represent independent trials; they are not binomial variables. Transformation of these data such as the logit for standard linear analysis have the shortcomings in terms of parameter interpretation and such data are often heteroskedastic and deviate from normality [65]. We ascertained their contextual determinants and variance decomposition through the Beta-logistic model. This model treats the proportions for selecting options as dependent on exogenous variables with heterogeneous variance [66]. The Levene's and Brown-Forsythe test for homogeneity of variance amongst the villages for the compliance and BICR test were conducted. These are robust techniques that are insensitive to heavy-tailed and skewed distributions in contrast to the Bartlett test that depends on the normality assumption.

For our analysis we use the beta distribution, whose mean lies within (0, 1,) with a logistic link function. The scale parameter of the Beta-logistic model is inversely related to the variance of the response variable. A limitation of the model is that it does not allow for proportions equal to zero or one.

$$f(y_{ijk}) = \frac{\Gamma(a_{ijk} + b_{ijk})}{\Gamma(a_{ijk})\Gamma(b_{ijk})} y_{ijk}^{a_{ijk}-1}(1 - y_{ijk})^{b_{ijk}-1} + \epsilon_{ijk} \tag{3.4}$$

where y_{ijk} is the compliance and BICR response, $a_{ijk} = e^{\alpha_l' g(\mathbf{X})}$, $b_{ijk} = e^{\beta_l' h(\mathbf{X})}$ and $\mathbf{X} = [[X]_{ijk}, Z] = (Region_i,\ Village(region)_{j(i)}, Household_{ijk},\ Covariate)$

Participation in bundled nutrition interventions can be affected by individual, household, and communal level factors. We examine the following hypotheses on the contextual determinants the process-driven participation metrics of compliance, BIRC, and gender engagement. Gender participation in bundled interventions is defined in terms of communication, decision inequities within households, and communal values. The logistic model with adjustments for clustering was used to identify the determinants of gender participant metrics. The logit model has a binomial distribution and a logistic link function and seeks to model the logarithm of the odds ratio.

$$Logit(\pi_{ijk}) = \mu + \alpha_i + \beta_{j(i)} + \tau Z + \epsilon_{ijk} \tag{3.5}$$

where $\pi_{ijk} = \frac{p}{1-p}$, μ is the grand mean, α_i is the fixed region factor, $\beta_{j(i)}$ is the village nested in region random factor, Z is the contextual covariate defined at either i, j, k level and a random error $\epsilon_{ijkl} \sim N(0, \sigma^2)$.

We used the metrics to identify the determinants of participation by the target households, men and women. We examined the following hypotheses for the contextual determinants of the process-driven metrics and also for the outcomes for hierarchically implemented bundled intervention.

H1: High baseline livestock household wealth can either promote or impede participation in bundled nutrition interventions

Under rural and low resource settings, wealth is often associated with access to infrastructure and resources. Households in the higher wealth quintiles tend to have diverse foods to incorporate in their diets, thus they are less motivated to participate in behavioral change communication that promote nutrition status of their members. Livestock and crop diversity at home [27] often associated with wealthier households promotes dietary diversity, which may negate their participation levels. On the contrary wealthier folks might have more time to attend, as bundled components for behavioral change communication may be beneficial for them.

H2: Larger baseline family size can hinder participation in bundled interventions

Larger family size may involve more sharing of food and less resources per person. Under such conditions there are challenges on welfare priority and time management for household decision makers making their participation in interventions with multiple components to be inconsistent.

H3: High education for woman in household can hinder participation in bundled nutrition interventions.

When there is variation in the education status for women in the households, their uptake and importance of messages on nutritional needs for their families may be divergent. This can be an indirect measure of their self-efficacy, which can measure how they perceive the nutritional content delivered in line with their already acquired knowledge and experience.

H4: Female headed households are less likely to participate in bundled nutrition interventions.

Women are less inclined to seek out nutritional resources for their households for rural and limited resource settings due to marginalization and the social structure. Women-headed households have one less person to take care of household responsibilities, so their time burden is way too much to allow for their participation.

H5: Remoteness hinders participation in bundled nutrition interventions

Households that are located faraway from meeting places and markets tend to be low in their engagement with bundled nutrition interventions.

H6: Agro-ecological zones can both promote and hinder participate in

bundled nutrition interventions.

Agro ecological zones measures the elevation from sea level of the settlements for subsistence farmers which influence their agro-produce. Those at high elevation have commercial produce and are more susceptible to restricted diverse food production. They may have a high dependence on the market's availability, affordability and diversity for food, and may be more inclined to seek knowledge on nutrition education and behaviors.

H7: Baseline parity can hinder participation in bundled nutrition interventions.

Baseline parity is the number of infants within a household. They require adequate care-giving and stimulation for food consumption. These time constraints limit their caregivers' participation in bundled nutrition interventions. Baseline parity is associated with maternal age and hence can indirectly influence participation.

H8: Farm size can promote or hinder participation in bundled nutrition interventions.

This can be an indirect measure of wealth, household productivity, and food security. This economic indicator may allow for low participation when outsourcing labor is expensive for those with bigger farms. It can also lower participation among those with small farms as they are more inclined to offer labor to those with big farms when harvests have been adverse.

H9: Age of household head can hinder participation in bundled nutrition interventions

Old age tends to hinder participation in interventions and this can also be attributed to distance traveled and gender factors [67].

3.2.6 Determinants of change in female dietary diversity scores for ATONU bundled intervention

The response of women to nutrition interventions has been shown to vary along contextual factors [68]. Livestock ownership and market participation of WRA are associated with the adequacy of dietary diversity [55]. Gender has been shown to be a significant factor on dietary diversity and agro-ecological zones are insignificant [53]. Husbands support and more participation of women in household financial decisions enhances women's adequate dietary diversity [55]. Home vegetable gardening and food preparation and nutrition knowledge are positively associated with household dietary diversity [69]. Ownership of livestock and female headed households improve on dietary diversity for rural communities [70]. Family food security and farm production diversity facilitate dietary diversity [71]. Linear regression models for dietary diversity have also indicated low R^2 values showing that there are other potential determinants of dietary diversity that need to be discovered [72]. The hierarchical structure introduced in the ATONU bundled interventions calls for hierarchical defined mixed models use in analyzing their target outcome. We seek to compare linear and Poisson mixed models on how they ascertain the change process on WRA's dietary diversity outcomes in relation to ATONU intervention's participation dynamics and contextual factors.

The literature described above suggests that contextual and background characteristics of the participants are related to the WRA dietary diversity scores. We address these factors with measurements from lower hierarchical levels (disaggregate metrics) and also adjust for process-driven participation metrics for the assessment of dietary diversity scores for bundled nutrition interventions. We model the effects of these covariates on the bundled ATONU intervention outcomes using linear and Poisson (*with a log link function*) mixed models. The generalized mixed model is given by

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \tau Z + \epsilon_{ijk} \tag{3.6}$$

where y_{ijk} is the women's dietary diversity, μ is the grand mean, α_i is the fixed region factor, $\beta_{j(i)}$ is the village nested in region random factor, Z is the contextual covariate defined at either i, j, k level and $\beta_{j(i)} \sim N(0, \sigma^2_{\beta(\alpha)})$, $\epsilon_{ijkl} \sim N(0, \sigma^2)$.

3.3 Results

3.3.1 Variation decomposition for process-driven participation metrics

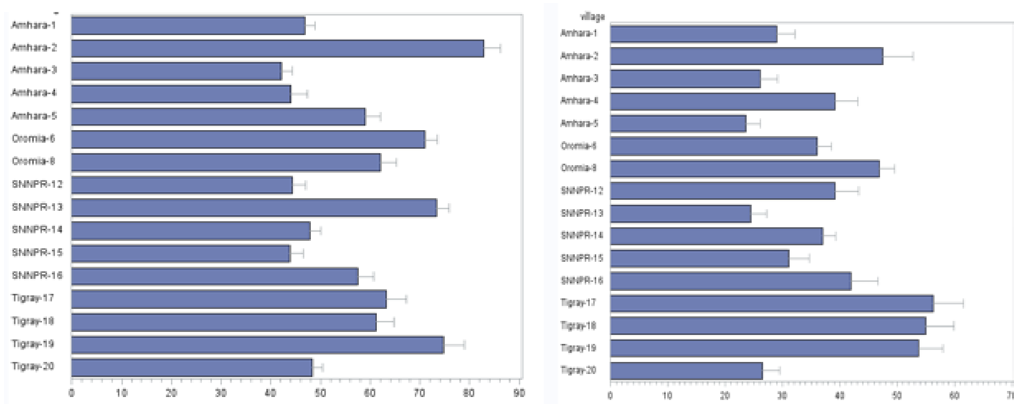

Figure 3.4. Error bars for the compliance and BICR metrics for ATONU intervention

Figure 3.4 shows that there is variation among the four regions of this study for the process-driven metrics of compliance and BICR. This variation is further distinguished both between and within the villages in the regions. Compliance metrics have a wider range in comparison to the BICR metric. Generally there was good compliance and a poor BICR among the villages.

Table 3.4.
Variance decomposition for the compliance and BICR metrics for ATONU

Source of Variation	Compliance	BICR
Between villages(nested in regions)	.047 (25.0%)	.027(82.7%)
Within Villages(nested in regions)	.141 (75.0%)	.006(17.3%)

Table 3.4 shows that the variation in the compliance metric was larger within villages and that for BICR was larger between villages. Compliance improvement requires focus on the participantsâĂŹ engagement and addressing disparities in participation. For the BICR there is need for supervision improvement for the research staff to ensure that they deliver all the bundle components in all the villages and for minimization of staff turnover.

The homogeneity of variance tests for both compliance and BICR gave p-values $<.0001$, indicating presence of heterogeneity among the villages. This supported the use of Beta-logistic model in the identification of the determinants of the participation dynamics for ATONU.

3.3.2 Determinants of participation and WRA dietary diversity scores for ATONU

Statistical models with adjustments for clustering were utilized to identify the covariates that influenced participation and the target outcomes of dietary diversity scores for WRA for ATONU.

Table 3.5.
Determinants of compliance for ATONU bundled intervention

Determinant	Estimate	Standard Error	p-value
Baseline wealth quintile (ref=5)	0	-	-
4	.1187	.0941	.2073
3	-.26685	.0956	.0054*
2	-.0120	.109	.9127
1	-.2980	.1304	.0226*
Family size	.0537	.0156	.0006*
Women's education (years)	.0092	.0129	.4733
Women headed household (ref=0)	0	-	-
1	.1391	.0899	.1219
Remoteness(minutes)	.0002	.0010	.8555
Baseline parity (ref=1 infant)	0	-	-
2 - 4 infants	.0300	.1237	.8084
More than 4 infants	.1575	.1180	.1824
Farm size (1 timad = 4 ha)	-.0163	.0126	.1943
Agro-ecological zone (ref=low altitude)	0	-	-
Medium altitude	.1117	.5147	.8282
High altitude	1.1936	.6349	.0603
Age of household head (years)	.0060	.0070	.3942

Table 3.5 shows the results based on the beta-logistic model with adjustments for contextual factors. It shows that family size and baseline wealth were significant determinants for compliance. The relative increase in the odds of compliance for a unit increase in family size was 1.0552; and that for the first and third quintiles for baseline wealth were 0.766 and 0.742, respectively.

Table 3.6.
Determinants of BICR for ATONU bundled intervention

Determinant	Estimate	Standard Error	p-value
Baseline wealth quintile (ref=5)	0	-	-
4	-.0790	.0333	.0339*
3	-.1649	.0344	<.0001*
2	-.1523	.0371	.0008*
1	-.1172	.0379	.0020*
Family size	.0015	.0050	.7580
Women's education (years)	-.0046	.0042	.2765
Women headed household (ref=0)	0	-	-
1	-.0324	.0293	.2693
Remoteness(minutes)	-.0003	.0004	.4124
Baseline parity (ref=1 infant)	0	-	-
2 - 4 infants	.0362	.0444	.4146
More than 4 infants	.0384	.0427	.36794
Farm size (1 timad = 4 ha)	.0011	.0026	.6776
Agro-ecological zone (ref=low altitude)	0	-	-
Medium altitude	-.0514	.1055	.6263
High altitude	.1718	.1300	.1863
Age of household head (years)	.0047	.0022	.0349*

Table 3.6 shows that baseline wealth and age of household head were significant factors in determining the number of bundle components that the participants received. The relative increase in the odds of BICR for a unit increase in the age of the household head was 1.0047; and for the first up to the fourth quintile of baseline wealth were 0.924, .848, .859, and 889, respectively.

Table 3.7.
Determinants of men's participation for ATONU bundled intervention

Determinant	Estimate	Standard Error	p-value
Baseline wealth quintile (ref=5)	0	-	-
4	.2891	.1937	.1417
3	.2956	.2112	.1676
2	-.1770	.2234	.4316
1	.3581	.2262	.1194
Family size	.1591	.0306	<.0001*
Women's education (years)	-.0127	.0244	.6043
Remoteness(minutes)	-.0028	.0022	.1989
Baseline parity (ref=1 infant)	0	-	-
2 - 4 infants	.8914	.2607	.0020*
More than 4 infants	.8343	.2505	.0025*
Farm size (1 timad = 4 ha)	-.0446	.0180	.0131*
Agro-ecological zone (ref=low altitude)	0	-	-
Medium altitude	-.1939	.7171	.7924
High altitude	2.6986	.8982	.0132*
Age of household head (years)	-.0135	.0129	.2930

Table 3.7 above shows that determinants for men's participation were family size, high elevation, farm sizes, and the number of infants in households. The relative increase in the odds of men's participation for a unit increase in family size was 1.1725, for high altitude agro-ecological zone (12.7458), farm size (0.9564); families with between 2 and 4 infants (2.4385), and for families with more than 4 infants (2.3032).

Table 3.8.
Determinants of women's participation for ATONU bundled intervention

Determinant	Estimate	Standard Error	p-value
Baseline wealth quintile (ref=5)	0	-	-
4	-.3940	.1884	.0414*
3	-.1111	.1992	.5795
2	-.0408	.2166	.8479
1	.0825	.2167	.7048
Family size	-.0069	.0285	.0151*
Women's education (years)	.0128	.0239	.5934
Remoteness(minutes)	.0043	.0020	.0373*
Baseline parity (ref=1 infant)	0	-	-
2 - 4 infants	-.5048	.2574	.0603
More than 4 infants	-.5880	.2479	.0251*
Farm size (1 timad = 4 ha)	-.0392	.0154	.0110*
Agro-ecological zone (ref=low altitude)	0	-	-
Medium altitude	.4953	.5828	.4152
High altitude	-.5989	.7218	.4260
Age of household head (years)	.0025	.0124	.8383

Table 3.8 shows that women were driven to participate in ATONU bundled nutrition intervention because of the demands of their family size, distance to the meeting place (*distance to the market was the proxy*), farm sizes, and the number of infants in their families. The relative increase in the odds of women's participation for a unit increase in family size was .9303, for distance to meeting place was 1.0041, farm size (1.0378), and for families with more than 4 infant children (.5554).

Table 3.9.
Determinants of joint participation for ATONU bundled intervention

Determinant	Estimate	Standard Error	p-value
Baseline wealth quintile (ref=5)	0	-	-
4	-.2591	.2890	.3741
3	.2429	.2898	.4058
2	-.3870	.3062	.2119
1	.6184	.2910	.0384*
Family size	.1359	.0388	.0005*
Women's education (years)	.0017	.0335	.9606
Remoteness(minutes)	.0037	.0026	.1574
Baseline parity (ref=1 infant)	0	-	-
2 - 4 infants	.7514	.3929	.0665
More than 4 infants	.4779	.3782	.2172
Farm size (1 timad = 4 ha)	-.0047	.0173	.7854
Agro-ecological zone (ref=low altitude)	0	-	-
Medium altitude	1.4384	1.8759	.4609
High altitude	4.7004	2.4273	.0816
Age of household head (years)	-.0298	.0172	.0847

Table 3.9 above shows that the relative increase in the odds for joint participation were 1.8560 for the first quintile of baseline wealth and 1.1456 for a unit increase in family size.

The Poisson mixed model was not a good fit for the WRA's dietary diversity scores with adjustments for clustering, associated covariates and the process-driven metrics. Its $\frac{\chi^2}{df}$ statistic was not approximately equal to one. The linear mixed model was a good fit based on the AIC.

Table 3.10.
Determinants of WRA end of the intervention 24-hour recall dietary diversity score for ATONU bundled intervention

Determinant	Estimate	Standard Error	p-value
Baseline wealth quintile (ref=5)	0	-	-
4	-.1297	.0961	.1772
3	-.0473	.0987	.6316
2	.0157	.1071	.8833
1	.1264	.1095	.2487
Family size	.0210	.0139	.1322
Women's education (years)	.0161	.0094	.0859
Remoteness(minutes)	-.0024	.0010	.0193*
Baseline parity (ref=1 infant)	0	-	-
2 - 4 infants	.0211	.1258	.8670
More than 4 infants	.2026	.1204	.0928
Farm size (1 timad = 4 ha)	.0206	.0069	.0029*
Agro-ecological zone (ref=low altitude)	0	-	-
Medium altitude	-.0008	.2152	.9971
High altitude	-.2439	.2651	.3792
Age of household head (years)	.0025	.0050	.6102
Women headed household (ref=0)	0	-	-
1	.1565	.0837	.0618
Compliance	.4331	.1557	.0055*
BICR	-.1802	.3421	.5985
Women participation(ref=0)	.0274	.0630	.6643
Men's participation(ref=0)	-.0258	.0657	.6949
Joint participation(ref=0)	.0065	.0850	.9389
Livestock diversity	.1082	.0234	<.0001*
Baseline dietary diversity score	.4483	.0301	<.0001*

Table 3.10 shows that for a unit increase in the distance to the market (a proxy for remoteness), farm size, compliance, livestock diversity, and baseline 24 hour recall dietary diversity score, the end of the intervention 24 hour recall dietary diversity score would change by -.0024, .0069, .4331, .1082, and .4483 units, respectively.

Table 3.11.
Determinants of WRA end of the intervention 7-days recall dietary diversity score for ATONU bundled intervention

Determinant	Estimate	Standard Error	p-value
Baseline wealth quintile (ref=5)	0	-	-
4	-.0547	.1413	.6986
3	.0411	.1451	.7770
2	.0500	.1576	.7513
1	.3322	.1613	.0396*
Family size	.0660	.0204	.0013*
Women's education (years)	.0384	.0144	.0079*
Remoteness(minutes)	-.0019	.0015	.2159
Baseline parity (ref=i infant)	0	-	-
2 - 4 infants	.1034	.1830	.5721
More than 4 infants	.2930	.1752	.0947
Farm size (1 timad = 4 ha)	.0361	.0103	.0005*
Agro-ecological zone (ref=low altitude)	0	-	-
Medium altitude	-.2150	.3950	.5980
High altitude	-.6561	.4867	.2073
Age of household head (years)	-.0013	.0076	.8642
Women headed household (ref=0)	0	-	-
1	.0977	.1233	.4281
Compliance	.5813	.2304	.0117*
BICR	-.6435	.5039	.2018
Women participation(ref=0)	-.0040	.0929	.9658
Men's participation(ref=0)	-.0067	.0971	.9451
Joint participation(ref=0)	-.0184	.1256	.8838
Livestock diversity	.1839	.0341	<.0001*
Baseline dietary diversity score	.4791	.0285	<.0001*

Table 3.11 shows that for a unit increase in the family size, women's education, farm size, compliance, livestock diversity, and baseline dietary diversity score, the end of the intervention 7 days recall dietary diversity score would change by .0660, .0384, .0361, .5813, .1839, and .4791 units, respectively. The relative increase in the odds of end of the intervention 7 days recall dietary diversity score for participants in the fifth quintile level of baseline livestock wealth was 1.3940.

3.4 Discussion

The ATONU bundled intervention was implemented heterogeneously in terms of content delivery and timing among the villages. This impacted the participation dynamics over the course of the intervention and the number of the bundled components received by the participants. The process-driven metrics of compliance and BICR had heterogeneous variance. The variance decomposition showed greater variation within villages for compliance and greater variation between villages for BICR. Compliance improvement requires participants' mobilization, and BICR improvement requires staff retention and delivery of all the bundled components.

The significant determinants for men's participation were farm size, high agro-ecological zone, having at least two infants in the household, and farm size. Women's participation was determined by the fourth quintile for baseline wealth, family size, remoteness, having at least 5 infants in household, and farm size. The joint participation was determined by the first quintile baseline livestock wealth and family size.

The Poisson mixed model was not a good fit for the end of the intervention WRA dietary diversity scores compared with the linear mixed model. The determinants for compliance were the third quintile of baseline livestock wealth, and family size; while those for BICR were the age of the household head and baseline livestock wealth for the range second to fifth quintiles. Remoteness, farm size, compliance, baseline dietary diversity score were the determinants for the end of the intervention dietary diversity score for 24 hour recall. The first quintile of baseline livestock wealth, family size, women's education, farm size, compliance, and baseline dietary diversity score were the determinants for the end of the intervention dietary diversity score for 7 days recall. The 24 hour recall and the 7 days recall dietary diversity scores are proxies for household food access and consumption, measured in terms of the variety of food types consumed.

The common determinants for both dietary diversity measures shows a positive contribution. Distance away from the meeting place (remoteness) has a negative effect on the dietary diversity for the 24 hour recall, while family size, baseline wealth and women's education have a positive effect on the 7 days recall dietary diversity metric. Compliance contributed positively to both dietary diversity measures while the gender and BICR metrics had insignificant effects.

The mediation analysis conducted for the ATONU bundled intervention showed that compliance was a significant determinant for both measures of dietary diversity for WRA. Adjustment for clustering, compliance, baseline WRA dietary diversity scores, livestock diversity, and the contextual and background characteristics' are important for linking the intervention to the end of the intervention WRA dietary diversity scores.

3.5 Conclusion

There are different context-sensitive profiles of engagement for bundled nutrition interventions. Process-driven metrics capture aspects of implementation that are missed by traditional metrics. Identifying at which level of the hierarchical implementation variation exists for these process metrics allows for the differentiation among strategies and decisions to improve implementation quality. Poor implementation can be attributed to staff turnover, supervision, context, and participants' decisions. We applied the metrics to identify the determinants of greater participation by the target households, men and women. The determinants of greater participation by target households included farm size, baseline parity, baseline wealth, and family size.

These should be targets for the improvement of implementation for future bundled interventions. These attributes are important for the establishment of the bundled nutrition interventions' complex ToC that can substantiate their causal statements.

Compliance had a significant effect on the WRA's dietary diversity scores, showing that to effectively ascertain the impact of bundled interventions on outcomes compliance has to be adequately measured and monitored. In spite of the BICR being an insignificant factor for the effects of the intervention on the WRA's dietary diversity scores, the low values shows that there is a need to promote adherence to the implementation of the intervention by the research staff to ensure delivery of all the bundled components for the full realization of its impact.

4. SIMULATION STUDY OF TIME SERIES MODELS GENERATED BY UNDERLYING DYNAMICS

4.1 Introduction

Time series analysis has been successfully applied in many areas of science and engineering. This has been necessitated when data records met strong statistical assumptions underlying traditional methods and were long enough for the results obtained by these methods to be reliable. In atmospheric and climate studies, however, observed records are often prohibitively short with only one record typically available, and the underlying assumptions for time series modeling are rarely met [14].

4.2 Motivating Example

Figure 4.1 below shows a typical atmospheric record - the vertical velocity of wind in a convective boundary layer, taken 29km across lake Michigan, 50m above the lake.

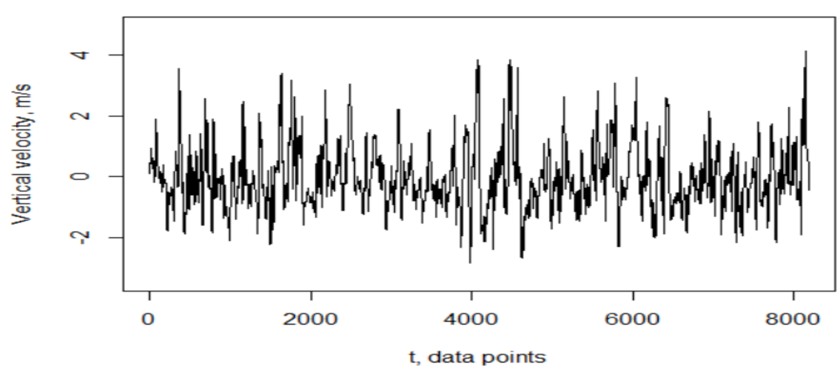

Figure 4.1. Record of 20-Hz vertical velocity measurements over Lake Michigan. Figure from [73]

For this realization of data, the routinely computed sample mean, variance, skewness, and kurtosis were -0.04, 1.06, 0.83, 4.10, respectively. The elevated skewness and kurtosis (*from values 0 and 3 specific for a normal distribution*) were attributed to the occurrence of coherent structures in turbulent flows [74], but to learn the extent one can trust such statistics, confidence intervals (CI) are needed. The extent to which sample statistics estimate the underlying population parameters is on its own an open-ended research problem [75]. To make inference on such numbers, a measure for precision would be required to account for the associated random error [76]. The establishment of the measure of precision depends on the assumptions made on the data generating mechanism for the underlying population.

The other challenge is the attainment of the accuracy level (coverage probability, say 0.90). This is attained only if the assumptions underlying the CI construction are met, a common one being that the model generating the series is linear. Atmospheric time series are produced by inherently nonlinear systems, hence the linearity assumption fails to be met. The actual coverage probability may differ from the target level (0.90), sometimes considerably. Moreover, the CIs for the skewness cannot be based on linear models, which imply zero skewness, inference made from such models would be unreliable. Thus, there is a need for nonlinear models, but finding an appropriate one among the conventional time series models is problematic.

We aimed at improving the reliability of statistical inference on atmospheric data through time series models generated by atmospheric underlying dynamics. The following were the objectives of our study;

1. estimating the subsampling confidence interval for the skewness of the vertical velocity of wind using time series generated from the underlying dynamic of atmospheric systems (*G-models*) and comparing them with those from conventional nonlinear time series models.

2. expanding on the G-models to incorporate more atmospheric mechanisms and compare and contrast their associated subsampling confidence intervals with the basic G-model at varying confidence levels.

4.3 Literature Review

4.3.1 Modern Statistical inference

The progress in Statistics has been stimulated and can be traced to the realization of what statisticians can provide to address the problems in real world application areas. This offers an indication for a mutual and symbiotic relationship between statistical theory and statistical applications [77]. Theory on one hand offers the framework, guidelines and arguments for statistical methodologies development, while applications aid in the justification of the postulated assumptions and relevance of inference derived through the statistical methods. A considerable number of statistical methodology have been developed through endeavors to solve problems in physical sciences and engineering. Response surface design was developed by George Box in his collaboration with chemical engineers, exploratory data analysis (EDA) was developed by John Tukey alongside telecommunication engineers, and sequential testing was postulated by Abraham Wald in his work with military engineers [77]. Interestingly, most of these developments relied on both Fisher and Neyman's considerations on statistical modeling [78].

R.A Fisher identified that specification, estimation, and distribution were the fundamental problems for modern statistical inference, but little attention has been directed towards addressing the specification challenge [79, 80]. Statistical models are fundamental for Statistics, hence the issue of their specification require utmost attention. In particular, the role of subject matter in statistical modeling is crucial to address the relevance of inference in statistical applications [81]. This is so because the specification problem centers on the choice of the mathematical form of the population from which the sample originates, i.e. addressing the question of how the observed data was generated [79].

One of the foundational problems on frequentist inference is the role of subject matter information in statistical modeling, in terms of their theoretical explanation [82]. According to Fisher, data generating mechanism (DGM) are important for addressing the specification problem, and that often may require knowledge beyond Statistics [79]. If there is some form of subject matter information available for a phenomenon of interest, statistical models should incorporate it [82]. Neyman's explanatory models make an attempt to explain the mechanism underlying the observed phenomena [81]. The two schools of thought on statistical model building of Fisher and Neyman creates an interesting position on how to address the DGM, the former tend to make assumptions on it and the latter acknowledges its presence and contribution. This scenario highlights the need for statistical theory and subject matter expertise to be considered in statistical modeling to enhance applications especially in situations where the data is generated under complexities, simplification of which may belittle the inference obtained.

Statistical models are often data driven, which may fall short in their relevance for application areas when expertise knowledge is not considered. The latter aspect is the basis for the argument for data-centric statistical model building, which require the incorporation of the scientific understanding of the application area and perturbations allowing for randomness to improve the relevance and reliability of inference for statistical applications. When statistical models are fully pre-specified, some of the deficiencies in statistical inference can be resolved [83].

4.3.2 Dynamical systems theory and nonlinear time series analysis

Dynamical systems theory is a branch of mathematics that consists of principles and tools for studying serial changes in physical or artificial systems. Lorenz emphasized on the importance of understanding the nonlinearity of atmospheric motion in modeling procedures [84]. Most of the systems in nature can best be described through nonlinear models [85]. The complexity of geophysical phenomena can be exemplified by temperature which requires high-dimensional physics-based models of the atmosphere instead of $AR(1)$ models to accurately describe it [86]. Nonlinear time series utilizes dynamical systems theory in the analysis of univariate observational data [86]. Our knowledge of the underlying systems is often restricted to the information we have from a single realization of data from a variable in the system, called a time series [85]; thus the state-space reconstruction of the underlying attractors for the system forms the foundation for nonlinear time series analysis [86]. The later can be co-opted in the time series model to explain the underlying dynamics, in the system under study, responsible for the generation of the time series data.

Attempts to model nonlinear non-normal time series has led to the development and utilization of new models such as the newer exponential auto-regressive (NEAR) and product auto-regressive models (PAR) [87], which depend on the AR(1) model's characterization that is often used in atmospheric modeling. Such models may acknowledge the nonlinearity and non-normality of the observed data, and consequently can give similar statistics but do not exhibit the fundamental theory underlying the operations effecting the generation of the observed data in dynamic systems. This overshadows the reliability of their statistical inference. The building of phenomenon-specific models as derivatives of the governing physical laws and associated properties and controlling variables can enhance the modeling of systematical variables [88]. We seek to postulate nonlinear statistical models that can explain the underlying processes for atmospheric phenomena, which are typically complex in their comprehension as they are generated from the interaction of nonlinear atmospheric processes.

4.3.3 Atmospheric systems and statistical inference

Mathematical models underlying phenomena in physical science and engineering are a source of prior knowledge about the problems that are in need of being solved [89]. They help in the description of the science of the problems we intend to address using statistical methods. The amalgamation of such mathematical models in statistical procedures and the use of statistical techniques in estimating the parameters of the mathematical models can aid the statistical modeling and interpretation of data realizations on physical phenomena [89]. In essence, scientifically justified statistical methodology are pivotal for understanding the often complex underlying dynamics responsible for the generation of the observed physical science data. Statistical research aims to develop tools for use at the frontiers of science which can be heightened through collaborations, as statisticians acquire/comprehend application area knowledge and offer statistical expertise [90]. These endeavors can help facilitate agreement between statistical significance and substantive significance which thereby aid the relevance of statistical applications in scientific research.

The atmosphere is a complex nonlinear system with mechanisms such as rotation, topography, shear, and stratification, constituting its underlying dynamics. A dynamical system can be mathematically defined by the triple, (Ω, ϕ, T), where $\Omega \subseteq R^d$ is the state space, ϕ is an evolution operator, and T denotes the set of possible times [91]. Atmospheric processes are essential to the determination of the state of the climate, and to climate change studies. Statistical inference are conclusions drawn on unknown population parameters based on probability models of data generating processes, based on sampled data [83]. On the contrary scientific inference depend on the accumulated subject-matter knowledge acceptable by members of the field, which plays a crucial role in their acceptance of new findings [46]. Inferential problems for atmospheric data can be attributed to the need for both their deterministic and statistical properties to be incorporated in their modeling. This allows for an understanding of the atmospheric system through physical thinking applied to statistical analysis of the observed data. Statistical modeling of climate phenomena should be preceded by consideration of the nonlinearity property of their underlying dynamics. Classical time series models hinge on unrealistic assumptions on the data generating mechanisms (DGMs) for atmospheric data, yielding misleading inference. In particular, the usual assumptions for time series of linearity and stationarity are often violated in practice [91]. Alternative time series models for classical time series models should capture the underlying theory and provide potentially better forecasts for the observed series [92]. According to [92] such models must exhibit the following features;

(i) They must be interpreted and based on potentially realistic theory.

(ii) Must exhibit the stability condition that is necessary for the stationarity of their associated time series.

(iii) All their components must at least be potentially observable.

Atmospheric data are non-normal and high-order moments such as skewness and kurtosis are required for their description [8]. Skewness measures the asymmetry of distribution, while kurtosis measures the peakedness of a distribution function. Much of the information that has been acknowledged as missing from the first and second moments maybe found in the third and fourth moments, and especially if they are tied to the physics underlying the observed data [93]. Higher order moments can be used to ascertain the levels of normality for atmospheric data, in particular, their skewness has been shown to be significantly different from the zero value for normality [94]. These nonlinearities in the underlying data generating mechanisms (DGM) for atmosphere data promotes misleading inference from traditional time series models that assume linearity [95]. On the other hand, statistical advances have shown that slight deviations from normality are a source of great concern [96]. Subsampling methods which work under weak assumptions are a useful option for finding the standard errors for high-order moments [73]. The variability of non-normal data depends on their underlying distributions [8].

Sampling distribution is fundamental to statistical inference as it allows for relating sample statistics to population parameters [97]. Hence, efforts to make inference on atmospheric data using subsampling methods with approximating models that do not infuse the physics of the original data can be questionable. On the other hand, resampling methods though flexible may under-perform in handling atmosphere data whose observed realizations are commonly too short for asymptotic inference. Knowledge of higher order statistical moments plays a crucial role in validating the approximating models for extreme events [8], a chief characteristic of atmosphere data. They also serve to assist in the analysis of the coherent structures (CSs) of atmosphere and climate data that are characteristically non-normal [95]. A coherent structure is said to be a connected turbulent fluid mass with phase-correlated vorticity over its spatial extent [98]. The CSs are responsible for the heat and moisture exchange that is responsible for the transportation of mass and momentum, which heightens the measures for skewness and kurtosis [95]. Coherent structures occur in localized regions of persistent vorticity, and they strongly influence heat exchange and turbulent flows between locations [74]. Fully developed turbulence is prevalent at boundary layers [93], and investigations of atmospheric phenomena there need to take into account its presence and impact on the assumptions for their DGM.

Confidence interval (CI) provides information on the amount of random error associated with an observed statistic (*precision*) and on the probability of how it relates to the corresponding parameter in the population from which the sample under investigation was drawn (*accuracy*) [99]. The trade-off between precision and accuracy is that an increase in precision entails a decrease in accuracy, and vice versa. Some of the advantages of confidence intervals include their link with p-values for hypothesis testing, they give information about precision, and estimates are in units that are readily comprehensible with the research context [100].

Statistical modeling seeks to complement mathematical modeling of atmospheric phenomena, as their forecasting ability hinge on the computing power, quality of data, and the challenge of initial conditions for the complex equations. Many strides are being made for the treatment of physical processes in atmospheric models and the exploration of advanced statistical methods. We seek to highlight the importance of subsampling methods for inference on atmospheric data using G-models as time series models whose DGM are inherited from their governing equations.

4.3.4 Subsampling Confidence intervals

It is a resampling procedure without replacement from the original sample n, yielding samples of smaller size b, where $b \ll n$ [101]. This techniques works in complex situations without asserting unverifiable assumptions on the data generating mechanisms (DGM). The record at hand of length n is divided in $n - b + 1$ subsamples or blocks of consecutive observations, all of the same length b, that retains the dependence structure of the series [102]. The technique of randomization which is at the heart of most simulation and resampling techniques can affect the resultant inference based on their assumption of the randomness of the data. In order to capture physical meaningful relationships, there is a need for procedures such as subsampling that allow for the capture of the complex dependent structure between observations. Subsampling allows for samples to be taken from the true unknown distribution function F of the original sample. This technique contrasts Efron's bootstrap method in that it uses b instead of n on sample size, and also that bootstrap samples are from an empirical distribution $\hat{F}$ associated with the original sample. Subsampling can be used on dependent data which are identically distributed (ID) and for extreme events that are independent and identically distributed (IID). In contrast, bootstrap requires distribution of data to be both identically and independently. The scenario above, does not give subsampling any superiority over bootstrap, but rather opportunities for us to experiment with it in more varied situations. The blocking in subsampling can capture the dependence in the original data, which allows it to work for stationary time series data.

Subsampling has been proposed to be a method for estimating parameters for the sampling distribution of statistics based on sub-series [103]. The performance of such parameter estimates for fixed n depends on the sub-series length b. Suppose we are interesting in inference on a parameter θ, typically a summary or shape measure for an observed time series realization, using the subsampling procedure. We postulate that $\hat{\theta}_n$ is an arbitrary statistic that is consistent for θ at the convergence rate of τ_n, then for large n, $\tau_n(\hat{\theta_n} - \theta)$ tends to some well-defined asymptotic distribution, say J [104]. The distribution of J needs not to be normal or its shape to be known but that its existence be acknowledged, and the main hypothesis in subsampling is that the subsampling empirical distribution converges weakly to J, the limiting/asymptotic distribution. The subsampling estimator for J will be the associated empirical distribution of $\tau_b(\hat{\theta_{i,b}} - \hat{\theta_n})$, where $\hat{\theta_{i,b}}$ is the subsampling value for the statistic of interest that was obtained from the i subsample of size b.

Subsampling confidence intervals were developed in [101], and in particular for stationary time series to address the problem of estimating variance of a statistics based on values of that statistic computed from sub-series. The use of overlapping blocks is more efficient, in comparison to non-overlapping sub-series, but they are both L_2 consistent and almost sure convergent [102]. Biased reduction ensures that the estimate is closer to the parameter of interest, at time series statistics are often heavily biased. The asymptotic consistency of the subsampling estimator of J has been shown [101] and it allows for the construction of confidence intervals for θ using its quantiles instead of those for the unknown J [104].The following assumptions facilitates the construction of subsampling confidence intervals for unknown parameters θs of time series of asymptotically correct coverage when.

(i) $b \to \infty$

(ii) $\frac{b}{n} \to 0$

(iii) $\tau_b \to \infty$

(iv) $\frac{\tau_b}{\tau_n} \to 0$

Under these assumptions, the weak convergence in distribution hypothesis is satisfied, where τ_n is the convergence rate, given by n^β, for $0 < \beta < 1$. The sampling distributions for the sub-samples and that of the the original sample are close to each other. If β is 0.5, this satisfies the "square root law" for the standard error. This is not so for atmosphere data as their limiting distribution is non-normal. Variance is of order $O(\frac{b}{n})$, hence it requires that the first two assumptions above, be satisfied. The following weak conditions, which can be relaxed, to work alongside the assumptions for subsampling confidence interval above;

(i) The observed time series is strictly stationary.

(ii) The observed time series is strong mixing.

(iii) The rate τ_n is known.

Upon relaxation the first condition allows for asymptotic stationarity, the second whittles to the weak dependence condition [104]; while the last condition is important for the practical considerations for subsampling confidence interval construction. The use of subsampling methodology in the derivation of a consistent estimator for τ_n, has facilitated the relaxation the third condition above [105]. The latter estimate is then used for the subsampling confidence interval construction with the actual coverage that is as near to the target coverage as possible. Overall, the subsampling method does not require any specific knowledge of the structures of the time series other than its attributes of asymptotic stationary and strong mixing.

The estimator for the statistic of interest T_n, $\hat{\theta}$ depends on the unknown distribution F. The difficult part to subsampling procedure is the determination of the underlying subsampling distribution, F. Monte Carlo simulations for time series data require models that can preserve the dependence structure in the data for reliable inference to be made. In the case under review, valid confidence intervals for skewness and kurtosis for nonlinear time series cannot be obtained using linear models [95]. The two issues that have to be addressed concurrently for subsampling confidence intervals' efficacy are the short record of realizations and the approximation of the underlying data generating mechanisms (DGMs) for atmosphere data.

4.3.5 The challenge of short record length for atmosphere data

Subsampling confidence interval construction depends on the block size for their accuracy, and they also have to contend with the challenge of shortness of atmosphere data realizations. They tend to fail to satisfy the conditions for the assumptions for subsampling confidence interval, and so in practice, approximating models are needed (those sharing statistical properties with the series under study) to assess the actual coverage of the subsampling confidence intervals. In order to satisfy the convergence in distribution assumption for subsampling methods, a convergence rate is needed. It ensures that the target coverage is attained in the computations of the confidence interval. The empirical convergence rate $\tau_n = n^\beta$ was introduced [106], where the value of the exponent β was different from the theoretical one.

Atmospheric data records are usually short in length, and single realizations that can contain very specific attributes. Monte Carlo simulation has been used to address the challenge of the short record length and models with similar statistical properties help in the selection of the optimal block size [107]. Plots of block size b against coverage are profound in the determination of the optimal fixed b for subsampling confidence interval construction. The use of approximating models that exhibit some of the statistical properties of the original data as sampling distributions for the subsampling procedure has also been shown to be helpful in ensuring that the target coverage is attained [107].

4.3.6 Time series modeling challenge for atmospheric data

The primary purpose for time series analysis is to develop statistical models that can describe the sampling data, which is an often data-driven endeavor. The "confusion factor" postulated in [108] shows the challenge of model computation agreement with observations, at the expense of the sufficiency of the model's representation of the physical processes underlying the data. The distribution of $\tau_b(\hat{\theta}_{i,b} - \hat{\theta}_n)$ in subsampling confidence interval is empirically derived from the subsamples data, that has nothing to do with the original data. In such instances, it may yield some of the statistical properties of the data, but falls short in accounting for the influence of the physics of the atmospheric data under investigation. [108] proposed that models of low complexity would be appropriate in geophysical simulations to reach scientific conclusions.

We seek to employ a new form of time series models that retain the physics of atmosphere data in the construction of the confidence interval for their skewness. They are characteristically simple, have the conservative property, and are able to incorporate mechanisms peculiar to atmospheric dynamics for their expansion which further retain the atmospheric reality.

Time series serve to offer some information about the systems that generate them, whose comprehension is pivotal for predictions to be made on the time-dependent variables under consideration. The assumptions made on the underlying DGM goes a long way in giving credit on the inference made in time series analysis. The governing equations and field records helps in advancing our understanding of atmospheric dynamics [13]. The assumption of normality do not hold for atmospheric data, which is often non-normal and non-linear, hence inference made from classical time series analysis can be misleading. The underlying dynamics for atmospheric data are nonlinear [73], which has to be captured in the approximating models. $AR(1)$-based nonlinear models satisfying some of the statistical properties have been employed, but they have nothing to do with the physics of atmospheric data.

Low-order models (LOMs), are a system of finite ordinary differential equations (ODEs) popularized by [109] that approximates the partial differential equations (PDEs) underlying the DGM for atmospheric data. These however fail to retain the conservative properties of the original PDEs in their endeavor to realistically model atmospheric dynamics, due to mathematical problems encountered in their establishment. This problem was solved through the establishment of G-models, which are physically sound, proposed by [13]. G-models have been shown to capture some of the statistical properties of atmospheric data, and their allowance for the incorporation of more mechanisms peculiar to them to improve their capture of the reality of the original data have been documented.

4.3.7 Related Works

We seek to investigate the nonlinear atmospheric data on vertical velocity of wind in a convective boundary layer, Figure 4.1 data. The convective boundary layer is the part of the atmosphere that is most directly affected by the solar heating of the earth's surface. Buoyancy is an atmospheric mechanism that is generated by the heating from the surface, and it is responsible for the vertical transportation of heat, pollutants, moisture and momentum. Buoyancy is responsible for the generation of convective turbulence which is an important aspect for global climate modeling and for the dynamics of many atmospheric phenomena. The treatment of turbulence as a random process raise profound statistical questions [110]. Efforts to construct the 90% subsampling confidence interval for the skewness parameter of these data have brought eye-opening results depending on the underlying approximating model and tuning parameters involved.

Subsampling confidence intervals were developed [101] for stationary time series to address the problem of the estimating variance of a statistic based on its values computed from sub-series. This procedure allows for the construction of confidence intervals from single records of time series. The use of overlapping blocks was found to be more efficient, in comparison to non-overlapping sub-series, but they are both L_2 consistent and almost sure convergent [102]. Bias reduction ensures that the estimate is closer to the parameter of interest, as time series statistics are often heavily biased. The estimator for the statistic of interest T_n, $\hat{\theta}$ depends on the unknown distribution F. The difficult part to subsampling procedure is the determination of the underlying subsampling distribution, F. Monte Carlo simulations for time series data require models that can preserve the dependence structure in the data for reliable inference. Valid confidence intervals for skewness and kurtosis for nonlinear time series cannot be obtained using linear models [95].

Subsampling confidence interval construction depends on the block size for their accuracy, which in turn depends on the coverage level. In order to satisfy the convergence in distribution assumption for subsampling methods, a convergence rate to is needed to ensure convergence in distribution for the estimate of the parameter of interest. This ensures that the target coverage is attained in the computations of the confidence interval for accurate interpretation of the results. Plots of block size b against coverage have been used to determine the optimal fixed b for use in subsampling confidence interval construction. The sampling distributions for the sub-samples and that of the the original sample are close to each other. If β is 0.5, this satisfies the "square root law" for the standard error. This is not so for atmosphere data as their limiting distribution is non-normal. Variance is of order $O(\frac{b}{n})$, requiring that the first two assumptions above, be satisfied.

The main problem encountered in subsampling confidence intervals (CIs) construction for the higher order moments, in particular for skewness of atmosphere data has been on coverage probabilities. It has been noted that the actual coverage tend to be considerably different from the target coverage, attributed to the availability of a single record of data of limited length. A single record cannot adequately answer a scientific question on its own, calling for at least meta-analytic thinking [111]. A calibration function $h : 1-\alpha \rightarrow 1-\lambda$, where $1-\alpha$ is the nominal confidence level and $1 - \lambda$ is the actual confidence level [107] can be applied. Attempts to use non-linear time series models face the daunting task of choosing models that can adequately capture non-linearity that is inherent in the DGMs of atmospheric data. Initially, the nonlinear approximating models were borrowed from traditional time series analysis, which allowed for the construction of subsampling CIs with the required coverage using calibrations [106]. Their data generating mechanisms (DGMs), however were considerably different from those of real atmospheric dynamics (*though some statisti-*

cal properties might be similar, thus motivating the choice of the models). The model 4.1 postulated by [112] was used in subsampling confidence interval construction

$$X_t = Y_t + a(Y_t^2 - 1) \tag{4.1}$$

where Y_t is an $AR(1)$ process, and for $a = 0.145$, the first four moments of X_t were close to those of the observed vertical velocity of wind data [13]. The $AR(1)$ with $\phi = 0.83$ served to fairly imitate the dependence structure as characterized by autocorrelation functions. Model 4.1 is an AR(1)-based nonlinear model, and using the calibration $h(0.95) = 0.9$, gave a 90% subsampling confidence interval for skewness of $(0.41, 1.24)$ [73]. The need to ensure that the actual coverage meets the target coverage led to the incorporation of a convergence rate function in the nonlinear models used to approximate the underlying dynamics of atmospheric data to improve inference, using subsampling methods. Consideration of the convergence rate $\tau_n = n^\beta, \beta \in (0, 1)$ [102] on model 4.1 (*referred below as approximating Model A*) for $\beta = 0.42$ gave a markedly improved 90% subsampling confidence interval $(0.56, 1.10)$ for the skewness of the vertical velocity of wind data, in terms of precision. Both methods served to show that there was nonlinearity in the vertical velocity of wind time series, through indicating a positive skewness.

One could then presume that Model A might be adequate for fixing subsampling confidence intervals, but there is no guarantee that other statistical properties of the data and the model do not differ to considerably affect the intended applications. The "confusion factor" postulated in [108] shows the challenge of model computation agreement with observations, at the expense of the sufficiency of the model's representation of the physical processes underlying the data. The confusion factor is the probability that an insufficient theory leads to similarities between model results and observational data. In particular, for nonlinear time series model the justification for model selection can be limited to the satisfaction of some and not necessarily all statistical properties of concern for an investigation to be generalized. The use of nonlinear time series methods to field measurements has been marred by controversy because of their exclusion of the fundamentals of dynamical systems theory from their theoretical basis [113]. The model in equation 4.1 was utilized because of the similarity of the first four moments from it to those in the observed data set, which may not hold in different data sets of the same variable under consideration. This may create a disconnection between model and the underlying theory of the application areas for time series as model parameters maybe subjective to the observational data, i.e. data-driven, for the output to be consistent. The distribution of $\tau_b(\hat{\theta}_b - \hat{\theta}_n)$ for the modified model (4.1) in subsampling confidence interval is empirically derived from the subsamples data, that has nothing to do with the original data. In such instances, it may yield some of the statistical properties of the data, but falls short in accounting for the influence of the physics of the atmospheric data under investigation. Using Model A at $a = 0.145$, and $\beta = 0.5$ the theoretical convergence rate, for various block sizes indicates under-coverage in the constructed subsampling confidence intervals [106]. Estimating the skewness does require long records, and a simple way to improve coverage is to increase the record length, which is possible via Monte Carlo simulations with approximating models. This can be lead to the actual coverage probabilities being closer to the target when the empirical convergence rate of $\beta = 0.42$ is applied, in comparison to $\beta = 0.5$ as shown in Figure 4.2.

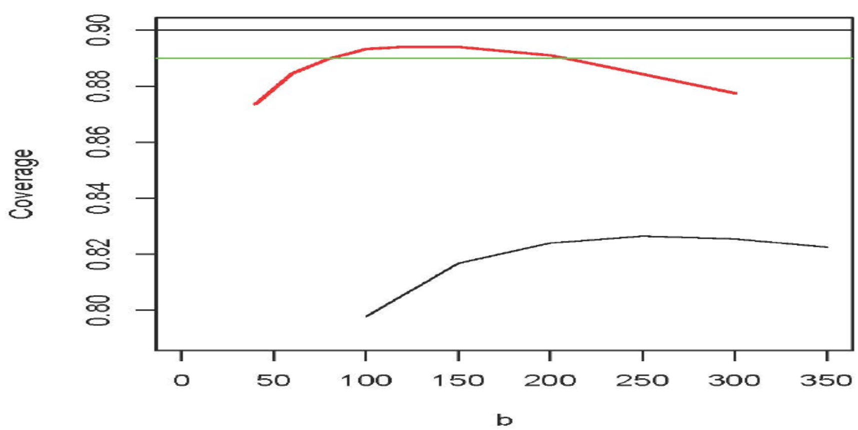

Figure 4.2. Actual coverage probabilities of 90% subsampling CIs with $\beta = 0.42$ (in red) and $\beta = 0.5$ (in black) using Model A for the skewness of nonlinear time series. Figure is adjusted and adopted from that in [106]

It has been proposed that models of low complexity can be appropriate in geophysical simulations to reach scientific conclusions [108]. We seek to employ a new form of time series models that retain the physics of atmosphere data in the construction of the subsampling confidence interval for their skewness. G-models have the property of retaining the physics of the underlying atmospheric dynamics and hence the statistics obtained from them are a near reflection of the reality for the vertical wind velocity under investigation. They have been used as physically sound low-order models in problems of atmospheric dynamics [13, 114], and have drawn increasing attention in various physical and mathematical studies [115–119].

4.4 G-Models and subsampling confidence interval for atmosphere data

Turbulent dynamical systems occur in systems exemplified by the atmosphere and the ocean, and have large dimensional phase space [120]. These are responsible for the behaviors exhibited by atmospheric and oceanic phenomena, i.e. the underlying dynamics determining the measures on such phenomena. Atmospheric dynamics offer an important advantage in providing the governing equations that generates the data on phenomena we seek to model [13]. This is a reservoir of subject matter knowledge that we can potentially tap into for statistical modeling. The governing equations for atmospheric dynamics consists of partial differential equations (PDEs) [121], that are problematic to solve due to the butterfly effect attributed to sensitivity to initial and boundary conditions.

Simple models can advance our understanding of the atmosphere, but there is little hope of establishing such models that can simulate all atmospheric processes from the global to the micro-physical scale, at least in the foreseeable future [122]. Attempts to handle them through approximations have led to the establishment of finite systems of ordinary differential equations (ODEs) called low-order models (LOMs) [109]. Here, we are seeking to represent a high dimension model with a simple model, and we are transitioning from PDEs, to ODEs. LOMs have been used for studying atmospheric phenomena, and their nonlinear Volterra gyrostats equivalence possess fundamental properties of the PDEs, promoting their use as a basis for the the development of G-models in particular [123]. Inasmuch as ODEs are a subset of PDEs, the reverse does not hold as they are derivatives in multiple variables, the curse of dimensionality maybe apparent in this endeavor. This is so, because the nonlinearity in the governing PDEs causes LOMs to contain more unknowns that equations, which creates a need for increasing the LOMs' row dimension [124]. LOMs are an important tool for geophysics fluid dynamics and need to retain the following features of the original system, quadratic nonlinearity, and in the absence of forcing and dissipation, conservation of energy and of space phase volume [125].

Gyrostat models (G-models) are a form of LOMs that exhibit sound physical behavior that were developed to solve the problem of loss of upholding the conservative properties of the PDEs by the ODEs [13]. The loss of conservative properties is due to the truncation employed in the Galerkin method for the construction of the LOMs. The statistical properties of dynamical systems have been noted to be simple and predictable, in particular the geometric Lorenz flow satisfies the almost sure invariance principle (ASIP) because of the attractor present in it, which in turn implies that they satisfy the central limit theorem [126]. Consequently, G-models satisfy the central limit theorem, and exhibit the physical ergodic invariant probability measure possessed by the Lorenz model [13], asserting their prospect as alternative time series models for atmospheric dynamics [106]. The latter attribute of invariant probability measure can be due to the fact that the flows described by Lorenz equations have a basin that covers Lebesgue almost every point of the topological basin of attraction, and are expansive [127].

A gyrostat is a mechanical system of bodies whose motion is explained by Volterra equations, without changing the mass distribution of the system [125]. The Volterra gyrostat is the basic G-model that consists of a system of mechanical and allowing for fluid dynamical components of atmospheric dynamics [128] as shown below in (4.2).

$$\begin{aligned} \dot{x_1} &= px_2x_3 + bx_3 - cx_2, \\ \dot{x_2} &= qx_1x_3 + cx_1 - ax_3, \\ \dot{x_3} &= rx_1x_2 + ax_2 - bx_1, \end{aligned} \tag{4.2}$$

where $p+q+r = 0$, and the linear terms called linear gyrostatic terms do not affect the conservation of energy or the conservation of phase space volume. They exhibit some form of energy, the quadratic integral motion, that ensures that they retain the physical behavior of the underlying atmospheric dynamics upon increasing the order of approximation for the Galerkin method [128]. These models are simple, and unlike the large numerical models often in use in climate numeric modeling, can also be used in data simulations which allows for their potential use in resampling methodologies [13]. They have been used in problems of atmospheric dynamics [13, 114], and have drawn increasing attention in various physical and mathematical studies [115–119].

Subsampling procedures are extremely flexible making them one of the most intuitive method for statistical inference [129]. They can handle dependent data as they hinge on a weak set of assumptions. The convergence in distribution assumption for subsampling, allows for the use of the often consistent estimator of the asymptotic distribution for the subsampling confidence interval construction using its associated quantiles for the parameters of interest to one's investigation [129]. The adoption of such models as alternatives for time series analysis may allow for the realistic representation of the underlying dynamics generating the data under investigation. The simplest G-model (r=b=c=0) with added forcing and linear friction terms is the G-model equivalence of the Lorenz model. The state vector **X**, $[X_i], i = 1, 2, 3$ for the Lorenz model consists of fluid velocity, horizontal and vertical temperature gradients for modeling thermal convection [110]. The Lorenz gyrostat given below fails to be a suitable approximating model for subsampling confidence interval for atmosphere data in spite of its well-defined statistical properties, and possession of the Rayleigh-Bénard convection (RBC), responsible for the generation of the original data [13].

$$
\begin{aligned}
\dot{x_1} &= -x_2x_3 \qquad\qquad - \alpha_1 x_1 + F, \\
\dot{x_2} &= \quad x_1x_3 \quad - x_3 - \alpha_2 x_2, \\
\dot{x_3} &= \qquad\qquad\quad x_2 - \alpha_3 x_3,
\end{aligned} \tag{4.3}
$$

Model (4.3)'s simulated records gave a skewness value of zero, which points to Gaussian distribution, but the observed sample's skewness value was 0.83. This result shows the inadequacy of the Lorenz systems of equations as approximations for the data generating mechanism for nonlinear atmospheric time series data.

Time series model specification must allow for the capture of the underlying data generating mechanism's salient features to facilitate relevance of inference made from them [130]. One intricate feature of G-models is their allowance for the incorporation of mechanisms of atmospheric such as stratification, rotation, topography, shear, magnetohydrodynamic effects as linear gyrostatic terms, to capture the physics of the underlying dynamics [13]. These facilitates their capture of the atmospheric reality, and the use of such models for time series heightens their usefulness for this particular application, and appeal amongst atmospheric scientists in particular and physical scientists in general for their scientific inference on atmospheric data. The introduction of one pair of linear gyrostatic terms in model (4.3) as shown in model (4.4), herein called Model B, below for a value of 0.35 for the constant c, yielded the values of 0.81 and 4.2, for skewness and kurtosis, respectively. The term X_3 represents the vertical velocity of wind time series in Figure 4.1. The summary statistics were closer to those for the observed data, and were a considerable improvement from those obtained from the nonlinear AR(1) derived Model A in 4.1.

$$
\begin{aligned}
\dot{x_1} &= -x_2x_3 \qquad\qquad + cx_3 \; - \alpha_1 x_1 + F, \\
\dot{x_2} &= \quad x_1x_3 \; - x_3 \qquad\qquad - \alpha_2 x_2, \\
\dot{x_3} &= \qquad\qquad\quad x_2 \; - cx_1 - \alpha_3 x_3,
\end{aligned} \tag{4.4}
$$

Further an introduction of another pair of linear gyrostatic terms in model (4.4) resulted in a G-model (4.5), herein called Model C, with a value of 1 for the constant d, yielded the values of 0.83 and 4.3, for the skewness and kurtosis, respectively. This new G-model retains the physics of the observed data with more mechanisms explaining it, which is an exclusive advantage of G-models, gaining from the knowledge already available from the underlying governing equations for atmospheric dynamics. This was facilitated by the fact that in addition to the Rayleigh-Bernard convection principal mechanism, the dynamics over Lake Michigan involves a hoist of other mechanism accounted for through terms associated coefficients c and further d in the model below. G-models allows for the incorporation of these mechanisms which serve to make them capture the reality of the underlying dynamics without loss of the physical properties.

$$\begin{aligned}
\dot{x_1} &= -x_2x_3 \qquad\qquad + cx_3 - dx_2 - \alpha_1 x_1 + F, \\
\dot{x_2} &= \quad x_1x_3 - x_3 \qquad\qquad + dx_1 - \alpha_2 x_2, \\
\dot{x_3} &= \qquad\qquad x_2 - cx_1 \qquad\qquad - \alpha_3 x_3,
\end{aligned} \tag{4.5}$$

The first four statistical moments from the two G-models were similar to those from the original data and asserts the nonlinearity exhibited in them. The results obtained using model (4.4), shows that G-models allows for the incorporation of mechanisms explaining the underlying dynamics for the observed data, hence increases their explanation from the atmospheric science theory and their capturing of the reality of the associated physical behavior.

We proceed to incorporate these models in the construction of the subsampling confidence intervals for the vertical wind velocity data, as the data generating mechanism approximations. Firstly, the models were used to determine the block sizes that would ensure that the coverage was as close as possible to the accuracy level we intend to make inference at. Once, the best possible block size was determined, we proceeded to ensuring that the actual coverage was as close as possible to the target coverage. Upon attaining proximity of the target coverage, the corresponding subsampling intervals were constructed and investigated for their behavior in terms of precision, and accuracy as we changed the confidence level for the inference.

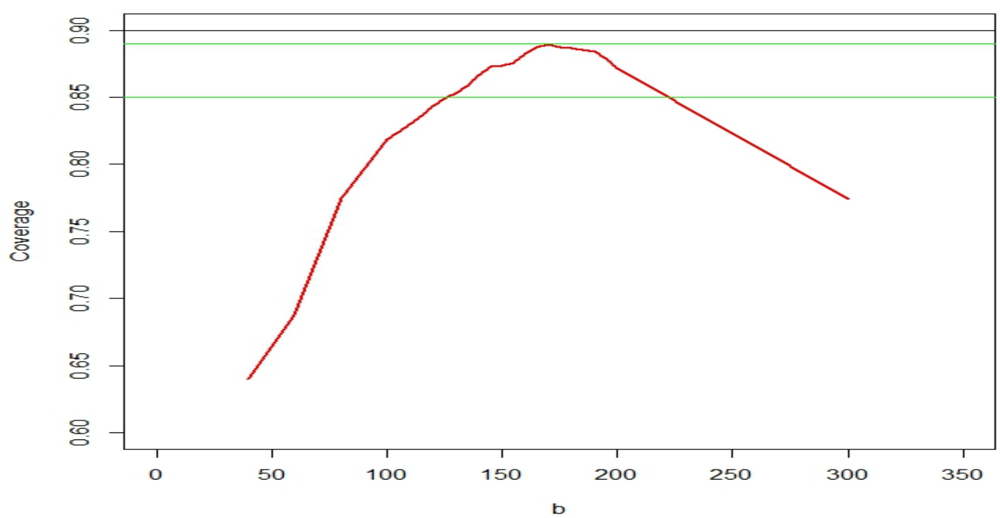

Figure 4.3. Actual coverage probabilities of 90% subsampling CIs with $\beta = 0.65$ using Model B for the skewness of nonlinear time series

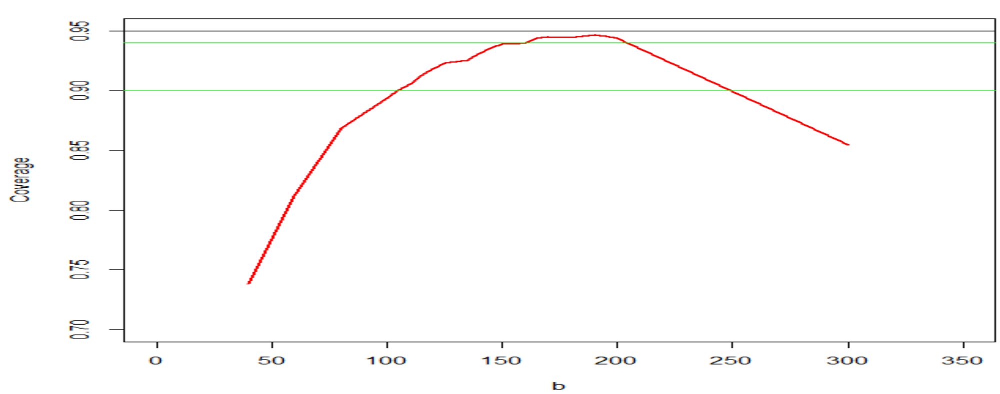

Figure 4.4. Actual coverage probabilities of 95% subsampling CIs with $\beta = 0.61$ using Model B for the skewness of nonlinear time series

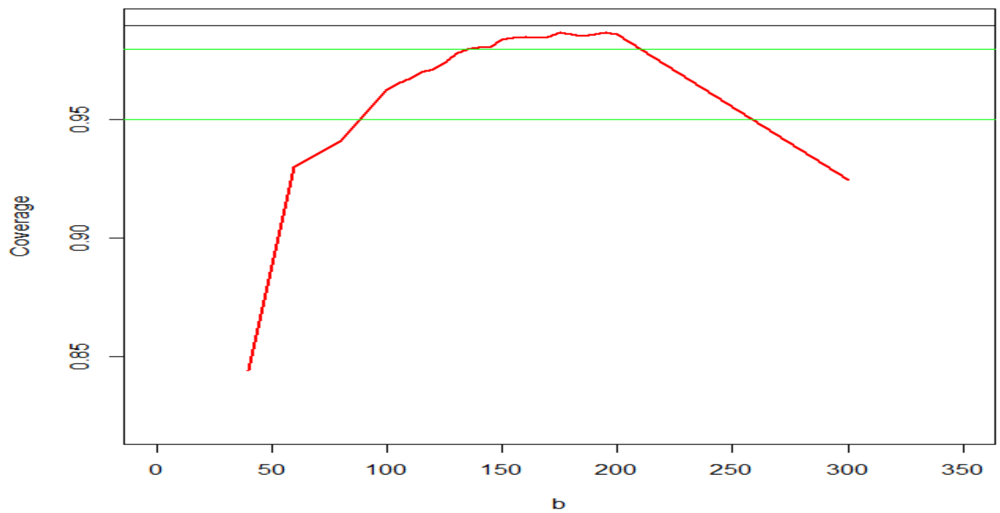

Figure 4.5. Actual coverage probabilities of 99% subsampling CIs with $\beta = 0.57$ using Model B for the skewness of nonlinear time series

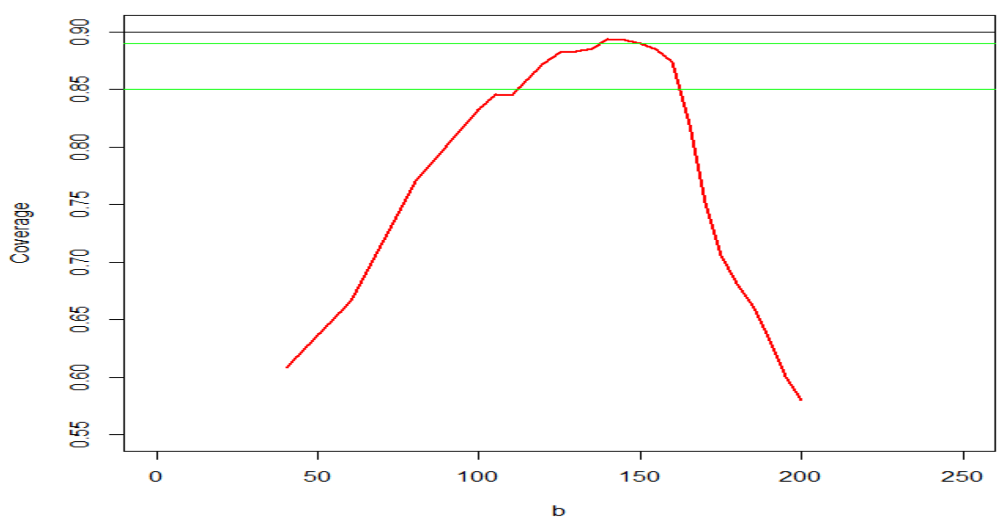

Figure 4.6. Actual coverage probabilities of 90% subsampling CIs with $\beta = 0.74$ using Model C for the skewness of nonlinear time series

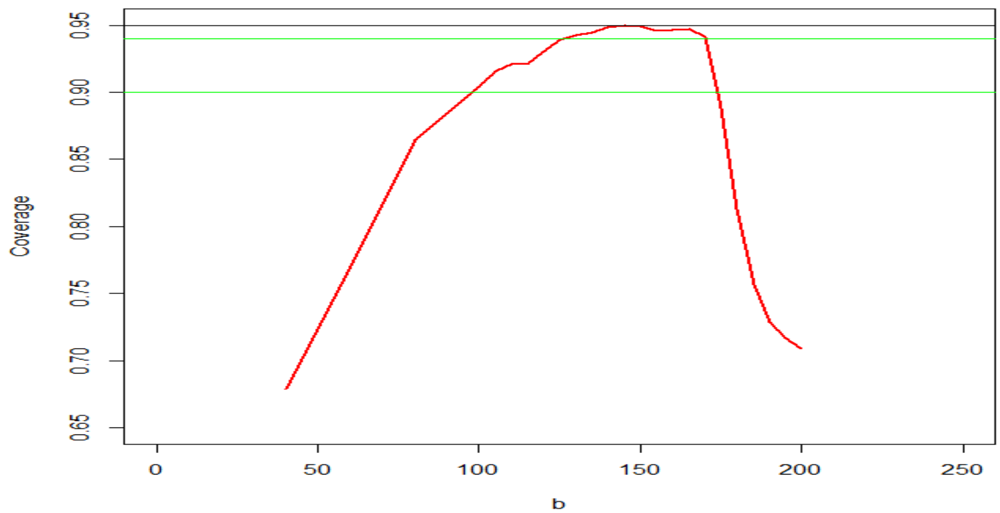

Figure 4.7. Actual coverage probabilities of 95% subsampling CIs with $\beta = 0.71$ using Model C for the skewness of nonlinear time series

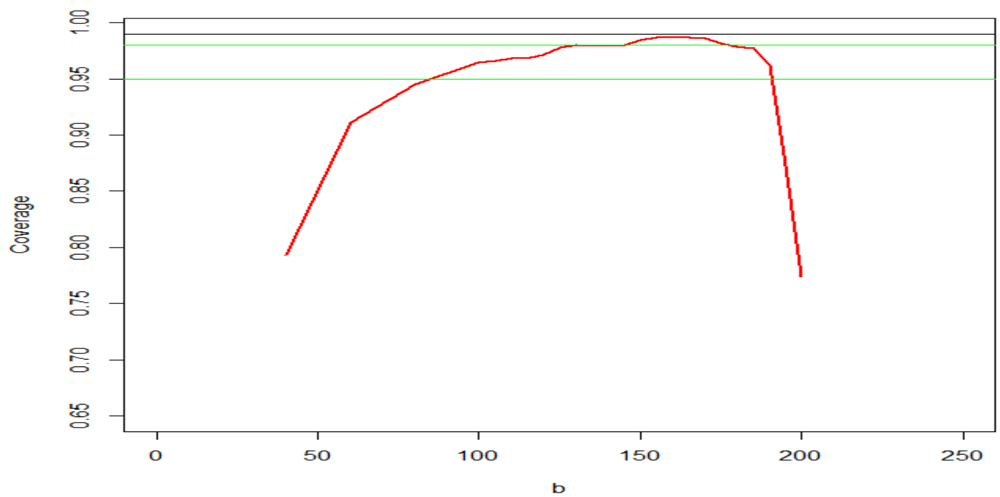

Figure 4.8. Actual coverage probabilities of 99% subsampling CIs with $\beta = 0.67$ using Model C for the skewness of nonlinear time series

Using the plots in figure 4.3 - Figure 4.8, we were able to determine the block sizes b that would lead to the construction of subsampling confidence intervals with actual coverage that were almost the same as the target coverage. These values occurred in distinctly short ranges. The constructed subsampling confidence intervals for the vertical wind velocity based on simulations of its data from the G-models in 4.4 and 4.5 were shown in Table 4.1 below. These were compared with the 90% intervals constructed from subsampling nonlinear AR(1) modified model with a convergence rate function.

4.5 Discussion

Table 4.1.
Subsampling confidence intervals

Model	Confidence Level	b	β	subsampling CI
A	90%	100	0.42	$(0.560, 1.100)$
B	90%	170	0.65	$(0.650, 1.000)$
B	95%	190	0.61	$(0.634, 1.015)$
B	99%	175	0.57	$(0.621, 1.028)$
C	90%	140	0.74	(0.677,0.972)
C	95%	145	0.71	$(0.670, 0.980)$
C	99%	160	0.67	$(0.654, 0.995)$

Table 4.1 shows that the interval obtained for the 90% confidence level using the G-model became narrower than the one obtained using the classical nonlinear time series model with a convergence rate tweak. This indicated that it became more precise as the random error associated with it had become smaller. For models B and C, there was a general downward trend in the convergence rate as the confidence levels increased. Model C also showed that as the confidence level increased, the block size

increased, but for Model B there was no obvious trend. Upon adding more mechanisms underlying vertical wind velocity in model C, through a pair of linear gyrostats, the precision of the subsampling confidence interval for skewness increased, i.e. became narrower. The above observation was similar for each confidence level between the two models B and C. The actual coverage were closer to the target coverage for Model C in comparison to those for Model B, as the confidence levels increased.

The simulation study conducted in this research served to demonstrate that subsampling techniques may be developed to obtain valid statistical inference in a variety of problems, where traditional time series analyses are hindered due to nonlinear data generating mechanisms and limited records. This involved the incorporation of G-model approximations to the underlying DGM for atmospheric dynamics in the subsampling procedure.

5. SUMMARY

5.1 Handling complexity through Statistics

The statistical challenge of complexity is not only limited to the big data, and small and difficult to obtain data, but also to the subject-matter knowledge, developments in application areas, and the dynamics associated with data generation. Simple linear thinking though relevant is proving to be limited in the reliability of inference in the presence of the complexity trait of data. This calls for a revamp of both statistical theory and methodology as we acknowledge and seek to model it. Complexity is a phenomenon not limited to specific application areas but also to the interactions of emerging scientific fields, which require inference to be made. Most of these emerging fields have an organic/systematic view for which simple models based on rigid assumptions may fail to objectively address the scientific problems they are trying to solve. This can create a gap between statistical and scientific inference, hence the need for an objective endeavor for Statistics to comprehend complexity.

5.2 Statistical input for bundled interventions implementation and evaluation

Acknowledging the statistical concerns on bundling, hierarchy, heterogeneity of population and implementation, and the varying contexts in which bundled interventions can be used to resolve public health issues is critical for their evaluation. These careful statistical thoughts serve to streamline the focus for statistical inference on bundled interventions, so as to avoid the challenges of ecological fallacy, and Type III error which impact on the decisions for practice from research findings on adoption, sustainability, and scaling. Such input aid in widening the theory of change for the mechanism of impact, a crucial component of process evaluation that has not been well documented for in the Medical Research Council (MRC) guidelines, pivotal for linking the intervention to the outcome. Contextual considerations alongside adjustments for confounders in the analysis for bundled interventions will improve on the inference reliability, and allow for the replicability of such interventions with a consideration for their adaptability for their comparability.

In the evaluation of the ATONU bundled intervention, which consisted of five behavioral messages to improve on women's dietary diversity, its implementation was shown to be heterogeneous. Heterogeneity of implementation was attributed to contextual factors and the background characteristics of the participants, hence a suggestion to capture the interaction of the participants with the intervention at individual level was proposed. This would help in tracking their retention, contact with the various bundle components, and measure gender coverage to ascertain the source of the variations in the outcome in relation to the implementation dynamics. This focus on the participants from the implementers' based fidelity widens the measurement of process dynamics on delivery-reception interactions. Bundling emphasizes the importance of participants engagement for the success of the intervention on effectiveness. We measured how the individual participants responded to the heterogeneous scheduling of the bundle components in the presence of the intervention's hierarchy structure. We developed the process-driven metrics for compliance, BICR, and gender engagement. These metrics captured the implementation dynamics that were missed by traditional metrics. They offer an insight into how the participants interact with the bundled intervention, were emphasis for successful implementation is on engagement with the bundled components over the continuum of the intervention study.

Identify at which level in the hierarchical implementation variation existed for the process-driven metrics facilitated for the differentiation of the strategies or decisions to improve implementation quality. A considerable amount of variation in the postulated quantitative metrics of compliance and intervention components received

(BICR) were attributed to within villages source, indicating the presence of both inter- and intra-household variation. The latter is not adequately accounted for in cluster level analysis as public health interventions are often conducted at communal level. There is a need to mobilize participants to improve on BICR proportions while ensuring that all the bundled components are adequately delivered by the research staff. There was a domination of between village variation over the within village for compliance. This could be attributed to the impact of training and retention of staffers,delivery, and administration on implementation of the bundle components over the course of the intervention's lifespan. Staff retention, competence, and adherence to the implementation protocol needs to be emphasized in the theory of change framework for bundled interventions. The moderately high amount of within villages variations points to both inter- and intra-household variations bring to light that some of the noise in the outcome could be attributed to the individuals' characterization within the villages. The engagement with the components of the bundle metric ICR, showed the reverse composition, with a domination of the within village variation, which can be attributed to the individual background characterization, and decision-making inequities within patriarchal communities.

The process-driven metrics were used to identify the determinants for greater participation by target households, men and women. These determinants were defined at the hierarchical levels of the bundled intervention showcasing the need to adjust beyond the clustering through inclusion of covariates that are context-defined. These findings were crucial for the improvement of future implementations of such interventions.

Participation has been known to mediate the effects of interventions on outcomes, so in this research we proposed to utilize the new metrics as mediators alongside other demographic factors in evaluating the impact of the intervention on the dietary diversity score for women of reproductive age (WRA) in Ethiopia. The accuracy of inference from bundled interventions is steeped in the critical analysis of the implementation dynamics to avoid Type III error where the implementation is not properly done. Adaptation of bundle components facilitate for potential context-intervention interactions that should be accounted for in the interpretation of the findings, and where possible recourse be found to avoid them. Sample size and power issues need to be addressed at the lowest level of the hierarchical structure of bundled interventions to ensure the validity of inference across the structures in order to avoid ecological fallacy challenge. Compliance had a significant effect on the women's dietary diversity. In spite of the BICR being an insignificant factor for the effects of the intervention on the WRA's dietary diversity, its observed low values showed the need to promote adherence to the implementation of the intervention by the research staff to ensure delivery of all the bundled components for the full realization of its impact. Similarly the gender metrics gave different factors influencing participation which are crucial for the need differential mobilization for participation by gender on the different bundled components to accentuate the success of the implementation effectiveness which translate to intervention effectiveness.

Data management and warehousing is crucial when implementation heterogeneity is evident as much noise can be introduced into the data through revisits, delivery decisions and strategies and the intervention process evolve. The use of technologies for data collection such as the open data kit (ODK) allows for good management of data in low resource settings, and also for the collection of more data at disaggregate levels. These data bring out as much information as maybe deemed necessary for the evaluation of bundled interventions especially on the often unmeasured confounding variables which need to be adjusted for in their statistical analysis. The process metrics captured were able to capture implementation dynamics, which have been missed by the conventional participation metrics. On the other hand, such innovations for data management as ODK allows for the capturing data on as many intermediate outcomes and contextual factors whose adjustments for are paramount to enhancing the process-outcome link for bundled interventions. This is more defined under low resource settings where there are competing factors that promotes time diversion in the participation dynamics of the respondents and if not adjusted for may lead to

Type III error which is consequential on the adoption, sustainability and scaling of such interventions for practice and policy.

The linear mixed model with adjustments for the process-driven metrics and contextual measures was used to explain the link between the women's dietary diversity outcomes and ATONU bundled interventions. The dietary diversity scores measured after 24 hours and 7 days, respectively, for the women of reproductive age in Ethiopia can be mediated by distance to market, a measure of how far the participants had to travel to attend the intervention's activities, baseline dietary diversity scores, and compliance. Almost all the variation in the outcomes of interest were attributed to within village sources, which include inter- and intra-household variation. The latter agrees well with the evidence shown that a considerable amount of the variation in the process-driven metrics was due to inter- and intrahousehold sources. The linear mixed models shows that adjusting for both clustering, handling the hierarchical structure complexity in the bundled intervention, and process dynamics, handling implementation heterogeneity attributed to the bundling complexity through process-driven metrics can enhance the process evaluation of bundled interventions. This may give an impetus for reliable effectiveness assessment, which may allow for adoption, sustainability, and scaling of bundled interventions into practice. In spite of women's dietary diversity measure being a perceived count, the Poisson mixed model could not adequately fit the data from the bundled ATONU intervention. This can be attributed to the underlying nonlinear attributes of the data generating mechanisms during the implementation process which heightens unpredictability in the intervention-outcome linkage.

5.3 G-models and inference on atmospheric data

The findings in this research give pointers to the potential for G-models as substantive time series models, that are appropriate for handling atmospheric data. Firstly, we were able to extend the basic G-model with one pair linear gyrostatic terms, and were able to obtain estimates for skewness and kurtosis that were very close to those for the original data. Further, we computed the subsampling confidence intervals for the wind velocity data using simulated data from the G-models, and obtained more precise intervals than those previously computed.

All the intervals constructed confirmed the nonlinearity of the underlying atmospheric dynamics responsible for the generation of the observed vertical wind velocity data under study. As the confidence level increased, the block size did not exhibit a distinct trend for Model B, while for Model C it exhibited an upward trend. On the other hand, the convergence rate exhibited a steady downward trend for both models as the confidence level increased. Upon adding an extra pair of linear gyrostatic terms to model B, we obtained model C that also yielded similar statistical properties with the observed data for the extra parameter $d = 1.00$. Comparing the subsampling confidence intervals obtained through the use of these two approximating models, the precision increased from those obtained with model B to those from model C. The addition of an extra pair of linear gyrostatic terms to model B ensured that the block size dropped considerably, while the convergence rate increased substantially. The two attributes showed a potentially inverse relationship for G-models as approximating models for subsampling confidence intervals.

G-models help avoid the dilemma of choosing among the many data-driven nonlinear time series models which though giving a semblance of the statistical properties, have nothing to do with the physics underlying the data being investigated. G-models share some fundamental physics with the original system which helps to (a) better align statistical properties of series generated by the model with those of observed series beyond the first moment and autocorrelation function, (b) avoid the difficult task of finding an appropriate approximating model based entirely on statistical characteristics estimated with questionable accuracy, and (c) run meaningful Monte Carlo simulations, particularly when estimators are more sensitive to properties of the DGMs. The subsampling confidence intervals are narrower than those previously computed, showing that these models allow for the improvement of precision in the estimation. This heightens the reliability of the inference on atmospheric data. The fact that G-models are derived from the governing equations for atmospheric dynamics necessitate their potential appeal and uptake amongst geosciences researchers. This also opens opportunities for wider applications of the subsampling procedure in climate and weather inference.

5.4 Limitations

The data used in the assessment of bundled interventions was of low quality due to profound implementation challenges in the study regions, to the extent that we had to limit our analysis to data from Ethiopia and not incorporating data from Tanzania. Denominator challenges were a cause of concern in the development of the metrics to avoid biases in inference.Expected values were used for the BICR metric instead of the actual values to minimize on the variance distortion on it. How a metric handles variation is key to decision making based on it. The composition of the bundle message was also not objectively documented for prior to the evaluation stage of the intervention, whose composition and administration could also have been a source of the heterogeneity experienced in the implementation. Many implementation decisions were made during the course of the intervention's lifespan due to contextual factors but were not well documented to be acknowledged on the interpretation of the conclusions that can be obtained for ATONU. This can have the consequential effects of Type IV error where the interpretations may be based on wrong parameters that may impact adversely on the adoption of ATONU and decisions on its sustainability and scaling for practice and policy.

5.5 Future research on bundled interventions

Extensive assessments of bundled interventions under low resource settings is needed to strengthen their theory of change (ToC). The input of all stakeholders in the ToC should be amalgamated to ensure that the evaluation of bundled interventions will meet their respective aims, and ultimately address the public health issues at stake. This will help provide the guidelines for their replicability in other settings and for comparisons to be made. Identification of key and redundant bundle components is essential for the sustainability of bundled interventions on adoption, and can be of profound economic benefit for the implementers. The statistical considerations highlighted point to the need for ethics to be upheld accordingly in the administration and implementation of such interventions to enhance their relevance in the emerging fields such as implementation science in nutrition, translation, and scaling.

We seek to evaluate the bundled intervention, Engaging Fathers For Effective Child nutrition in Tanzania (EFFECT) that has a well-documented ToC and utilize process participation metrics and contextual factors. This interventions seeks to address the household power dynamics and decision-making on the roles that fathers can play in infant and young child feeding (IYCF) to address the challenges of malnutrition and further to investigate their role in early child development alongside nutritional engagement through the EFFECT+ bundled intervention. We seek to utilize composition data analysis technique in the explanation of the effect of bundled nutrition interventions on a host of outcomes based on process-driven participation metrics. We seek to further develop the process-driven participation metrics to have a time-dimension for longitudinal studies analysis.

5.6 Future research on subsampling and G-models

We seek to expand our exploration of the viability of G-models as alternative time series models through using expanded G-models with additional linear gyrostats, and other G- models for the construction of the subsampling confidence interval of skewness for nonlinear atmospheric time series. These G-models allows for the aligning of the statistical properties of the observed data with those from themselves beyond the second order moments. We would also want to investigate how subsampling confidence intervals come up for kurtosis using G-models in comparison to the AR(1)-derived nonlinear time series model. This allows for the assessment of the peakedness of their distribution, a crucial aspect for tail distribution of nonlinear data.

We seek to investigate the limiting behavior of subsampling CIs as $b \to n$, a vital assumption for the subsampling procedure application, taking note of the fact that b is integral for the accuracy of the intervals.We seek to incorporate other G-models for inference on atmospheric data for skewness and kurtosis. We would also want to investigate the limiting behavior of the models as the value of $\beta \to 0.5$ for the determination of the block size and the corresponding precision in the subsampling confidence intervals.

The perceived inverse relationship between the block size and the convergence rate in the G-models as approximating models for subsampling confidence interval estimation may allow for holding the convergence rate at $\beta = 0.5$, and obtaining the block size that ensures that similar statistical properties are obtained and then use it for the interval computation. This will help on the investigation of the behavior of the confidence intervals as $b \to \infty$.

Vertical wind velocity is a function of both time and position, spatial considerations for its modeling may allow for alternative perspectives of making inference on it based on location. The data has been shown to be stationary, there is a possibility for the use of G-models as intrinsic models for spatial modeling.

Methodology ties in with computational considerations, we seek to develop a time-efficient program that will facilitate for subsampling confidence intervals using G-models.

We seek to employ G-models for statistical inference on other atmospheric phenomena accounting for bay, ocean and land effects of atmospheric dynamics in the generation of their time series. Shear aids our understanding of sediment dynamics in coastal areas and beaches for sediment analysis. The underlying dynamics contributing to the generation of such phenomena data are nonlinear, accounting for this attribute in their modeling affords reliable inference to be obtained.

Future works will explore other types of G-models as approximating models for inference on atmospheric data. We will adopt this technique for inference on linking theory and data for Astro-Statistics and modeling pharmacokinetics for the absorption, metabolism, and excretion of drugs in living organisms.Biological systems and diseases have been explained through mathematical modeling based on systems of PDEs, we can also employ the concept that was used in the development of G-models to develop time series models for analyzing phenomena that includes HIV-AIDS, Tuberculosis, and malaria with adjustments for confounding variables on interventions to address them in low and middle income countries (LMIC).

REFERENCES

[1] Trisha Greenhalgh and Chrysanthi Papoutsi. Studying complexity in health services research: desperately seeking an overdue paradigm shift, 2018.

[2] Xiao-Li Meng. A trio of inference problems that could win you a nobel prize in statistics (if you help fund it). *Past, present, and future of statistical science*, 2014.

[3] Chad M Schafer. A framework for statistical inference in astrophysics. *Annual Review of Statistics and Its Application*, 2:141–162, 2015.

[4] Susan Holmes and Wolfgang Huber. *Modern statistics for modern biology*. Cambridge University Press, 2018.

[5] Thomas Homer-Dixon. Complexity science. *Oxford Leadership Journal*, 2(1):1–15, 2011.

[6] John Protevi. Deleuze, guattari and emergence. *Paragraph*, 29(2):19–39, 2006.

[7] Kate Churruca, Chiara Pomare, Louise A Ellis, Janet C Long, and Jeffrey Braithwaite. The influence of complexity: a bibliometric analysis of complexity science in healthcare. *BMJ open*, 9(3):e027308, 2019.

[8] Maxime Perron and Philip Sura. Climatology of non-gaussian atmospheric statistics. *Journal of Climate*, 26(3):1063–1083, 2013.

[9] Purnima Menon, Namukolo M Covic, Paige B Harrigan, Susan E Horton, Nabeeha M Kazi, Sascha Lamstein, Lynnette Neufeld, Erica Oakley, and David Pelletier. Strengthening implementation and utilization of nutrition interventions through research: a framework and research agenda. *Annals of the New York Academy of Sciences*, 1332(1):39–59, 2014.

[10] Reinette Tydeman-Edwards, Francois Cornel Van Rooyen, and Corinna May Walsh. Obesity, undernutrition and the double burden of malnutrition in the urban and rural southern free state, south africa. *Heliyon*, 4(12):e00983, 2018.

[11] Neil C Campbell, Elizabeth Murray, Janet Darbyshire, Jon Emery, Andrew Farmer, Frances Griffiths, Bruce Guthrie, Helen Lester, Phil Wilson, and Ann Louise Kinmonth. Designing and evaluating complex interventions to improve health care. *Bmj*, 334(7591):455–459, 2007.

[12] Rabiul Karim, Lene Lindberg, Sarah Wamala, and Maria Emmelin. MenâĂŹs perceptions of womenâĂŹs participation in development initiatives in rural bangladesh. *American journal of men’s health*, 12(2):398–410, 2018.

[13] Alexander Gluhovsky and Kevin Grady. Effective low-order models for atmospheric dynamics and time series analysis. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 26(2):023119, 2016.

[14] Michael Ghil, MR Allen, MD Dettinger, K Ide, D Kondrashov, ME Mann, Andrew W Robertson, A Saunders, Y Tian, F Varadi, et al. Advanced spectral methods for climatic time series. *Reviews of geophysics*, 40(1):3–1, 2002.

[15] Taren Swindle, Geoff M Curran, and Susan L Johnson. Implementation science and nutrition education and behavior: Opportunities for integration. *Journal of nutrition education and behavior*, 51(6):763–774, 2019.

[16] Graham F Moore, Rhiannon E Evans, Jemma Hawkins, Hannah Littlecott, GJ Melendez-Torres, Chris Bonell, and Simon Murphy. From complex social interventions to interventions in complex social systems: Future directions and unresolved questions for intervention development and evaluation. *Evaluation*, 25(1):23–45, 2019.

[17] Alison Tumilowicz, Marie T Ruel, Gretel Pelto, David Pelletier, Eva C Monterrosa, Karin Lapping, Klaus Kraemer, Luz Maria De Regil, Gilles Bergeron, Mandana Arabi, et al. Implementation science in nutrition: concepts and frameworks for an emerging field of science and practice. *Current Developments in Nutrition*, 3(3):nzy080, 2018.

[18] Karl E Peace, Anthony V Parrillo, and Charles J Hardy. Assessing the validity of statistical inferences in public health research: An evidence-based,âĂŸbest practicesâĂŹ approach. *Journal of the Georgia Public Health Association*, 2008.

[19] Hugo Sax, Benedetta Allegranzi, Ilker Uckay, E Larson, J Boyce, and Didier Pittet. âĂŸmy five moments for hand hygieneâĂŹ: a user-centred design approach to understand, train, monitor and report hand hygiene. *Journal of Hospital infection*, 67(1):9–21, 2007.

[20] Marin Schweizer, Eli Perencevich, Jennifer McDanel, Jennifer Carson, Michelle Formanek, Joanne Hafner, Barbara Braun, and Loreen Herwaldt. Effectiveness of a bundled intervention of decolonization and prophylaxis to decrease gram positive surgical site infections after cardiac or orthopedic surgery: systematic review and meta-analysis. *Bmj*, 346:f2743, 2013.

[21] Stephen A McClave, Beth E Taylor, Robert G Martindale, Malissa M Warren, Debbie R Johnson, Carol Braunschweig, Mary S McCarthy, Evangelia Davanos, Todd W Rice, Gail A Cresci, et al. American society for parenteral and enteral nutrition. guidelines for the provision and assessment of nutrition support therapy in the adult critically ill patient: Society of critical care medicine (sccm) and american society for parenteral and enteral nutrition (aspen). *JPEN J Parenter Enteral Nutr*, 40(2):159–211, 2016.

[22] SW Aboelela, PW Stone, and EL Larson. Effectiveness of bundled behavioural interventions to control healthcare-associated infections: a systematic review of the literature. *Journal of Hospital Infection*, 66(2):101–108, 2007.

[23] Amisha V Barochia, Xizhong Cui, David Vitberg, Anthony F Suffredini, Naomi P OâĂŹGrady, Steven M Banks, Peter Minneci, Steven J Kern, Robert L

Danner, Charles Natanson, et al. Bundled care for septic shock: an analysis of clinical trials. *Critical care medicine*, 38(2):668, 2010.

[24] William H Brown, Samuel L Odom, and Maureen A Conroy. An intervention hierarchy for promoting young children's peer interactions in natural environments. *Topics in early childhood special education*, 21(3):162–175, 2001.

[25] Neha Kumar, Phuong Hong Nguyen, Jody Harris, Danny Harvey, Rahul Rawat, and Marie T Ruel. What it takes: evidence from a nutrition-and gender-sensitive agriculture intervention in rural zambia. *Journal of Development Effectiveness*, 10(3):341–372, 2018.

[26] Dimitri Batras, Cameron Duff, and Ben J Smith. Organizational change theory: implications for health promotion practice. *Health promotion international*, 31(1):231–241, 2016.

[27] Ramya Ambikapathi, Nilupa S Gunaratna, Isabel Madzorera, Simone Passarelli, Chelsey R Canavan, Ramadhani A Noor, Tshilidzi Madzivhandila, Simbarashe Sibanda, Semira Abdelmenan, Amare Worku Tadesse, et al. Market food diversity mitigates the effect of environment on womenâĂŹs dietary diversity in the agriculture to nutrition (atonu) study, ethiopia. *Public health nutrition*, 22(11):2110–2119, 2019.

[28] Dixon Chibanda, Ruth Verhey, Epiphany Munetsi, Frances M Cowan, and Crick Lund. Using a theory driven approach to develop and evaluate a complex mental health intervention: the friendship bench project in zimbabwe. *International journal of mental health systems*, 10(1):16, 2016.

[29] Erica Breuer, Lucy Lee, Mary De Silva, and Crick Lund. Using theory of change to design and evaluate public health interventions: a systematic review. *Implementation Science*, 11(1):63, 2015.

[30] Mary J De Silva, Erica Breuer, Lucy Lee, Laura Asher, Neerja Chowdhary, Crick Lund, and Vikram Patel. Theory of change: a theory-driven approach to enhance the medical research council's framework for complex interventions. *Trials*, 15(1):267, 2014.

[31] Dean Rickles. Causality in complex interventions. *Medicine, Health Care and Philosophy*, 12(1):77–90, 2009.

[32] Kristen M Hurley, Aisha K Yousafzai, and Florencia Lopez-Boo. Early child development and nutrition: a review of the benefits and challenges of implementing integrated interventions. *Advances in nutrition*, 7(2):357–363, 2016.

[33] Frederick J Dorey. In brief: Statistics in brief: Statistical power: What is it and when should it be used?, 2011.

[34] Aolin Wang, Roch A Nianogo, and Onyebuchi A Arah. G-computation of average treatment effects on the treated and the untreated. *BMC medical research methodology*, 17(1):3, 2017.

[35] Jessica Datta and Mark Petticrew. Challenges to evaluating complex interventions: a content analysis of published papers. *BMC public health*, 13(1):568, 2013.

[36] Sharon Schwartz. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. *American journal of public health*, 84(5):819–824, 1994.

[37] Jerry A Jacobs. Gender inequality and higher education. *Annual review of sociology*, 22(1):153–185, 1996.

[38] Robert L Wears. Advanced statistics: statistical methods for analyzing cluster and cluster-randomized data. *Academic emergency medicine*, 9(4):330–341, 2002.

[39] Sumedha Sharma, Olalekan O Adetoro, Marianne Vidler, Sharla Drebit, Beth A Payne, David O Akeju, Akinmade Adepoju, Ebunoluwa Jaiyesimi, John Sotunsa, Zulfiqar A Bhutta, et al. A process evaluation plan for assessing a complex community-based maternal health intervention in ogun state, nigeria. *BMC health services research*, 17(1):238, 2017.

[40] Michelle L Bell, Antonella Zanobetti, and Francesca Dominici. Evidence on vulnerability and susceptibility to health risks associated with short-term exposure to particulate matter: a systematic review and meta-analysis. *American journal of epidemiology*, 178(6):865–876, 2013.

[41] Thomas A Glass, Steven N Goodman, Miguel A Hernán, and Jonathan M Samet. Causal inference in public health. *Annual review of public health*, 34:61–75, 2013.

[42] Sabu S Padmadas. Community-based bundled interventions for reproductive and child health in informal settlements: evidence, efficiency, and equity. *The Lancet Global Health*, 5(3):e240–e241, 2017.

[43] Aulo Gelli, Elisabetta Aurino, Gloria Folson, Daniel Arhinful, Clement Adamba, Isaac Osei-Akoto, Edoardo Masset, Kristie Watkins, Meena Fernandes, Lesley Drake, et al. A school meals program implemented at scale in ghana increases height-for-age during midchildhood in girls and in children from poor households: A cluster randomized trial. *The Journal of nutrition*, 2019.

[44] Rishma Maini, Sandra Mounier-Jack, and Josephine Borghi. How to and how not to develop a theory of change to evaluate a complex intervention: reflections on an experience in the democratic republic of congo. *BMJ global health*, 3(1):e000617, 2018.

[45] Charles E Basch, Elena M Sliepcevich, Robert S Gold, David F Duncan, and Lloyd J Kolbe. Avoiding type iii errors in health education program evaluations: a case study. *Health education quarterly*, 12(3):315–331, 1985.

[46] Raymond Hubbard, Brian D Haig, and Rahul A Parsa. The limited role of formal statistical inference in scientific inference. *The American Statistician*, 73(sup1):91–98, 2019.

[47] Ana V Diez Roux. Estimating the neighborhood health effects: The challenges of casual inference in a complex world. 2004.

[48] Enola Proctor, Hiie Silmere, Ramesh Raghavan, Peter Hovmand, Greg Aarons, Alicia Bunger, Richard Griffey, and Melissa Hensley. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(2):65–76, 2011.

[49] R Rukmani, R Gopinath, G Anuradha, R Sanjeev, and Varun Kumar Yadav. Women as drivers of change for nutrition-sensitive agriculture: Case study of a novel extension approach in wardha, india. *Agricultural Research*, pages 1–8.

[50] Marie T Ruel, Agnes R Quisumbing, and Mysbah Balagamwala. Nutrition-sensitive agriculture: What have we learned so far? *Global Food Security*, 17:128–153, 2018.

[51] Elizabeth L Fox, Claire Davis, Shauna M Downs, Werner Schultink, and Jessica Fanzo. Who is the woman in women's nutrition? a narrative review of evidence and actions to support women's nutrition throughout life. *Current developments in nutrition*, 3(1):nzy076, 2018.

[52] Marie T Ruel. Is dietary diversity an indicator of food security or dietary quality? a review of measurement issues and research needs. *Food and Nutrition Bulletin*, 24(2):231–232, 2003.

[53] Gamuchirai Chakona and Charlie Shackleton. Minimum dietary diversity scores for women indicate micronutrient adequacy and food insecurity status in south african towns. *Nutrients*, 9(8):812, 2017.

[54] Seid Aliwo, Melkitu Fentie, Tadesse Awoke, and Zemichael Gizaw. Dietary diversity practice and associated factors among pregnant women in north east ethiopia. *BMC research notes*, 12(1):123, 2019.

[55] Melaku Desta, Mohammed Akibu, Mesfin Tadese, and Meskerem Tesfaye. Dietary diversity and associated factors among pregnant women attending antenatal clinic in shashemane, oromia, central ethiopia: A cross-sectional study. *Journal of nutrition and metabolism*, 2019, 2019.

[56] Ali Sié, Charlemagne Tapsoba, Clarisse Dah, Lucienne Ouermi, Pascal Zabre, Till Bärnighausen, Ahmed M Arzika, Elodie Lebas, Blake M Snyder, Caitlin Moe, et al. Dietary diversity and nutritional status among children in rural burkina faso. *International health*, 10(3):157–162, 2018.

[57] Bruce Lindsay. Randomized controlled trials of socially complex nursing interventions: creating bias and unreliability? *Journal of advanced nursing*, 45(1):84–94, 2004.

[58] Feras Ali Mustafa. Notes on the use of randomised controlled trials to evaluate complex interventions: Community treatment orders as an illustrative case. *Journal of evaluation in clinical practice*, 23(1):185–192, 2017.

[59] Richard OâĂŹReilly and Evelyn Vingilis. Are randomized control trials the best method to assess the effectiveness of community treatment orders? *Administration and Policy in Mental Health and Mental Health Services Research*, 45(4):565–574, 2018.

[60] Hopin Lee, Alix Hall, Nicole Nathan, Kathryn L Reilly, Kirsty Seward, Christopher M Williams, Serene Yoong, Meghan Finch, John Wiggers, and Luke Wolfenden. Mechanisms of implementing public health interventions: a pooled causal mediation analysis of randomised trials. *Implementation Science*, 13(1):42, 2018.

[61] Marie T Ruel, Harold Alderman, Maternal, Child Nutrition Study Group, et al. Nutrition-sensitive interventions and programmes: how can they help to accelerate progress in improving maternal and child nutrition? *The lancet*, 382(9891):536–551, 2013.

[62] Patrick Webb and Eileen Kennedy. Impacts of agriculture on nutrition: nature of the evidence and research gaps. *Food and nutrition bulletin*, 35(1):126–132, 2014.

[63] Jeffrey A Alexander and Larry R Hearld. Methods and metrics challenges of delivery-system research. *Implementation Science*, 7(1):15, 2012.

[64] Hopin Lee, Robert D Herbert, and James H McAuley. Mediation analysis. *Jama*, 321(7):697–698, 2019.

[65] Francisco Cribari-Neto and Achim Zeileis. Beta regression in r. 2009.

[66] Min Yi, Nancy Flournoy, and Eloi Kpamegan. A ballooned beta-logistic model. *Statistics in Biopharmaceutical Research*, 9(1):98–106, 2017.

[67] Jacob J Oleson, Patrick J Breheny, Jane F Pendergast, Sandi Ryan, and Ruth Litchfield. Impact of travel distance on wisewoman intervention attendance for a rural population. *Preventive medicine*, 47(5):565–569, 2008.

[68] Hitomi Komatsu, Hazel Jean L Malapit, and Sophie Theis. Does womenâĂŹs time in domestic work and agriculture affect womenâĂŹs and childrenâĂŹs dietary diversity? evidence from bangladesh, nepal, cambodia, ghana, and mozambique. *Food policy*, 79:256–270, 2018.

[69] Justus Ochieng, Victor Afari-Sefa, Philipo Joseph Lukumay, and Thomas Dubois. Determinants of dietary diversity and the potential role of men in improving household nutrition in tanzania. *PloS one*, 12(12):e0189022, 2017.

[70] A Taruvinga, V Muchenje, and A Mushunje. Determinants of rural household dietary diversity: The case of amatole and nyandeni districts, south africa. *Int J Dev Sustainability*, 2(4):2233–2247, 2013.

[71] Laura Adubra, Mathilde Savy, Sonia Fortin, Yves Kameli, Niamké Ezoua Kodjo, Kamayera Fainke, Tanimoune Mahamadou, Agnès Le Port, and Yves Martin-Prével. The minimum dietary diversity for women of reproductive age (mdd-w) indicator is related to household food insecurity and farm production diversity: Evidence from rural mali. *Current developments in nutrition*, 3(3):nzz002, 2019.

[72] Liv E Torheim, Fatimata Ouattara, Modibo M Diarra, FD Thiam, Ingrid Barikmo, Anne Hatløy, and Arne Oshaug. Nutrient adequacy and dietary diversity in rural mali: association and determinants. *European journal of clinical nutrition*, 58(4):594, 2004.

[73] A Gluhovsky. Statistical inference from atmospheric time series: detecting trends and coherent structures. *Nonlinear Processes in Geophysics*, 18(4):537–544, 2011.

[74] Jori E Ruppert-Felsot, Olivier Praud, Eran Sharon, and Harry L Swinney. Extraction of coherent structures in a rotating turbulent flow experiment. *Physical Review E*, 72(1):016311, 2005.

[75] Avijit Hazra. Using the confidence interval confidently. *Journal of thoracic disease*, 9(10):4125, 2017.

[76] Abdelhamid Attia. Why should researchers report the confidence interval in modern research. *Middle East Fertil Soc J*, 10(1):78–81, 2005.

[77] Guest Edited By Diane Lambert. Statistics in the physical sciences and engineering. In *Statistics in the 21st Century*, pages 212–287. Chapman and Hall/CRC, 2001.

[78] Johannes Lenhard. Models and statistical inference: The controversy between fisher and neyman–pearson. *The British journal for the philosophy of science*, 57(1):69–91, 2006.

[79] Aris Spanos et al. Where do statistical models come from? revisiting the problem of specification. In *Optimality*, pages 98–119. Institute of Mathematical Statistics, 2006.

[80] Ronald A Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594-604):309–368, 1922.

[81] Eric L Lehmann. Model specification: the views of fisher and neyman, and later developments. In *Selected Works of EL Lehmann*, pages 955–963. Springer, 2012.

[82] Aris Spanos et al. Foundational issues in statistical modeling: Statistical model specification and validation. *Rationality, Markets and Morals*, 2(47), 2011.

[83] Christopher Tong. Statistical inference enables bad science; statistical thinking enables good science. *The American Statistician*, 73(sup1):246–261, 2019.

[84] Dragutin T Mihailović, Gordan Mimić, and Ilija Arsenić. Climate predictions: the chaos and complexity in climate models. *Advances in Meteorology*, 2014, 2014.

[85] KP Harikrishnan, Ranjeev Misra, and G Ambika. Nonlinear time series analysis of the light curves from the black hole system grs1915+ 105. *Research in Astronomy and Astrophysics*, 11(1):71, 2011.

[86] Elizabeth Bradley and Holger Kantz. Nonlinear time-series analysis revisited. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(9):097610, 2015.

[87] Jan G De Gooijer et al. *Elements of nonlinear time series analysis and forecasting*. Springer, 2017.

[88] Marian Anghel, Yehuda Ben-Zion, and Ramiro Rico-Martinez. Dynamical system analysis and forecasting of deformation produced by an earthquake fault. In *Computational Earthquake Science Part I*, pages 2023–2051. Springer, 2004.

[89] F Din-Houn Lau, Niall M Adams, Mark A Girolami, Liam J Butler, and Mohammed ZEB Elshafie. The role of statistics in data-centric engineering. *Statistics & Probability Letters*, 136:58–62, 2018.

[90] Bruce G Lindsay, Jon Kettenring, David O Siegmund, et al. A report on the future of statistics. *Statistical Science*, 19(3):387–413, 2004.

[91] Analisi Statistica dei Sistemi Dinamici. Statistical analysis of dynamic systems.

[92] CWJ Granger. New classes of time series models. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 27(3/4):237–253, 1978.

[93] John L Lumley. Coherent structures in turbulence. In *Transition and turbulence*, pages 215–242. Elsevier, 1981.

[94] Vladimir Petoukhov, Alexey V Eliseev, Rupert Klein, and Hermann Oesterle. On statistics of the free-troposphere synoptic component: an evaluation of skewnesses and mixed third-order moments contribution to the synoptic-scale dynamics and fluxes of heat and humidity. *Tellus A: Dynamic Meteorology and Oceanography*, 60(1):11–31, 2008.

[95] Alexander Gluhovsky and Ernest Agee. Estimating higher-order moments of nonlinear time series. *Journal of Applied Meteorology and Climatology*, 48(9):1948–1954, 2009.

[96] Alexander Gluhovsky and Ernest Agee. On the analysis of atmospheric and climatic time series. *Journal of applied meteorology and climatology*, 46(7):1125–1129, 2007.

[97] Nikolaos Pandis. Statistical inference with confidence intervals. *American journal of orthodontics and dentofacial orthopedics*, 147(5):632–634, 2015.

[98] AKM Fazle Hussain. Coherent structures and turbulence. *Journal of Fluid Mechanics*, 173:303–356, 1986.

[99] Julius Sim and Norma Reid. Statistical inference by confidence intervals: issues of interpretation and utilization. *Physical Therapy*, 79(2):186–195, 1999.

[100] Geoff Cumming and Sue Finch. Inference by eye: confidence intervals and how to read pictures of data. *American Psychologist*, 60(2):170, 2005.

[101] Dimitris N Politis and Joseph P Romano. Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, pages 2031–2050, 1994.

[102] Dimitris N Politis, Joseph P Romano, and Michael Wolf. *Subsampling*. Springer Science & Business Media, 1999.

[103] Edward Carlstein et al. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The annals of statistics*, 14(3):1171–1179, 1986.

[104] Efstathios Paparoditis and Dimitris N Politis. Resampling and subsampling for financial time series. In *Handbook of financial time series*, pages 983–999. Springer, 2009.

[105] Patrice Bertail, Dimitris N Politis, and Joseph P Romano. On subsampling estimators with unknown rate of convergence. *Journal of the American Statistical Association*, 94(446):569–579, 1999.

[106] A Gluhovsky and T Nielsen. Brief communication" improving the actual coverage of subsampling confidence intervals in atmospheric time series analysis". *Nonlinear Processes in Geophysics*, 19(5):473–477, 2012.

[107] Alexander Gluhovsky, Martin Zihlbauer, and Dimitris N Politis. Subsampling confidence intervals for parameters of atmospheric time series: block size choice and calibration. *Journal of Statistical Computation and Simulation*, 75(5):381–389, 2005.

[108] AJ Mannucci, BT Tsurutani, OP Verkhoglyadova, and X Meng. On scientific inference in geophysics and the use of numerical simulations for scientific investigations. *Earth and Space Science*, 2(8):359–367, 2015.

[109] Edward N Lorenz. Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2):130–141, 1963.

[110] Javier Jiménez. Coherent structures in wall-bounded turbulence. *Journal of Fluid Mechanics*, 842, 2018.

[111] Melissa Coulson, Michelle Healey, Fiona Fidler, and Geoff Cumming. Confidence intervals permit, but don't guarantee, better inference than statistical significance testing. *Frontiers in psychology*, 1:26, 2010.

[112] DH Lenschow, Jakob Mann, and Leif Kristensen. How long is long enough when measuring fluxes and other turbulence statistics? *Journal of Atmospheric and Oceanic Technology*, 11(3):661–673, 1994.

[113] Thomas Schreiber. Interdisciplinary application of nonlinear time series methods. *Physics reports*, 308(1):1–64, 1999.

[114] A Gluhovsky. Energy-conserving and hamiltonian low-order models in geophysical fluid dynamics. *Nonlinear Processes in Geophysics*, 13(2):125–133, 2006.

[115] Alexander Bihlo and Johannes Staufer. Minimal atmospheric finite-mode models preserving symmetry and generalized hamiltonian structures. *Physica D: Nonlinear Phenomena*, 240(7):599–606, 2011.

[116] Andre N Souza and Charles R Doering. Transport bounds for a truncated model of rayleigh–bénard convection. *Physica D: Nonlinear Phenomena*, 308:26–33, 2015.

[117] G Chen, Nikolay V Kuznetsov, Gennady A Leonov, and TN Mokaev. Hidden attractors on one path: Glukhovsky–dolzhansky, lorenz, and rabinovich systems. *International Journal of Bifurcation and Chaos*, 27(08):1750115, 2017.

[118] Marco Bianucci. Large scale emerging properties from non hamiltonian complex systems. *Entropy*, 19(7):302, 2017.

[119] Andrew J Majda and Di Qi. Strategies for reduced-order models for predicting the statistical responses and uncertainty quantification in complex turbulent dynamical systems. *SIAM Review*, 60(3):491–549, 2018.

[120] Themistoklis P Sapsis and Andrew J Majda. Statistically accurate low-order models for uncertainty quantification in turbulent dynamical systems. *Proceedings of the National Academy of Sciences*, 110(34):13705–13710, 2013.

[121] Fedor Mesinger and Akio Arakawa. Numerical methods used in atmospheric models. 1976.

[122] Penelope Maher, Edwin P Gerber, Brian Medeiros, Timothy Merlis, Steven Sherwood, Aditi Sheshadri, Adam Sobel, Geoffrey Vallis, Aiko Voigt, and Pablo Zurita-Gotor. The value of hierarchies and simple models in atmospheric.

[123] Alexander Gluhovsky and Ernest Agee. An interpretation of atmospheric low-order models. *Journal of the atmospheric sciences*, 54(6):768–773, 1997.

[124] Edward N Lorenz. Low order models representing realizations of turbulence. *Journal of Fluid Mechanics*, 55(3):545–563, 1972.

[125] Alexander Gluhovsky and Christopher Tong. The structure of energy conserving low-order models. *Physics of Fluids*, 11(2):334–343, 1999.

[126] Vitor Araujo, Stefano Galatolo, and Maria José Pacifico. Statistical properties of lorenz-like flows, recent developments and perspectives. *International Journal of Bifurcation and Chaos*, 24(10):1430028, 2014.

[127] Vitor Araujo, M Pacifico, Enrique Pujals, and Marcelo Viana. Singular-hyperbolic attractors are chaotic. *Transactions of the American Mathematical Society*, 361(5):2431–2485, 2009.

[128] Alexander Gluhovsky. A gyrostatic low-order model for the el ñino-southern oscillation. *Complexity*, 2017, 2017.

[129] Dimitris N Politis. The impact of bootstrap methods on time series analysis. *Statistical Science*, pages 219–230, 2003.

[130] JM Marriott and AR Tremayne. Alternative statistical approaches to time series modelling for forecasting purposes. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 37(2):187–197, 1988.