Emerging Topics in Statistics and Biostatistics

Ding-Geng (Din) Chen
Samuel O. M. Manda
Tobias F. Chirwa *Editors*

Modern Biostatistical Methods for Evidence-Based Global Health Research

Springer

Emerging Topics in Statistics and Biostatistics

Ding-Geng (Din) Chen • Samuel O. M. Manda •
Tobias F. Chirwa
Editors

Modern Biostatistical Methods for Evidence-Based Global Health Research

Springer

Editors
Ding-Geng (Din) Chen
College of Health Solutions
Arizona State University
Phoenix, AZ, USA

Samuel O. M. Manda
Department of Statistics
University of Pretoria
Pretoria, South Africa

Tobias F. Chirwa
School of Public Health, Division of
Epidemiology and Biostatistics
University of the Witwatersrand
Johannesburg, South Africa

ISSN 2524-7735 ISSN 2524-7743 (electronic)
Emerging Topics in Statistics and Biostatistics
ISBN 978-3-031-11011-5 ISBN 978-3-031-11012-2 (eBook)
https://doi.org/10.1007/978-3-031-11012-2

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The DELTAS Africa Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) training program is funded by the Wellcome Trust in partnership with the Alliance for Accelerating Excellence in Science in Africa (AESA). The consortium was established in 2015, with the overall aim of building a critical mass of biostatisticians and biostatistics research leadership in Sub-Saharan Africa. This is achieved through the development and strengthening of biostatistics capacity and resource at the 11 participating local institutions, in collaboration with four local research institutions and three northern university partners.

In celebrating the contributions, achievements, and progress of SSACAB scientists and their partners and collaborators, this book is organized to document the contributions from the consortium with a diverse mix of current scholarship and exposition of biostatistics methods and application for evidence-based global health in the Region. The volume features inspiring and informative chapters that reflect on the accomplishments of biostatistics research and its applications that offer solutions to local health problems. There are a total of 18 chapters to provide an overview of the emerging topics in biostatistical methods and their applications to Sub-Saharan Africa public health research and evidence-based management decision-making.

The structure of these 18 chapters is subsequently organized with the following five parts. As an introductory chapter, chapter "Sub-Saharan African Region Strategies to Improve Biostatistics Capacity: Exploring Collaborations Between Training and Research Institutions," describes the origins and contributions of SACCAB as well as its structure.

Part I (Data Harmonization and Analysis) contains three chapters (Chapters 2 to 4). In chapter "Diagonal Reference Modelling of the Effects of Educational Differences Between Couples on Women's Health-Care Utilization in Eritrea," Ghilagaber and Carlson adapted models developed in the social mobility literature to examine the effects of differences between couples' educational levels on women's health-related decisions (such as the propensity to deliver in health facilities). Both conventional modeling and Diagonal Reference Modeling (DRM) which account for origin (woman's education), destination (partner's education), and "mobility" (differences between couples' educational levels) are applied on data from Demographic and Health Surveys (DHS). Results from conventional models reveal strong effects of educational differences on women's health-related decisions, but such strong effects disappear when data is analyzed using DRM. In chapter "Sequential Probit Modeling of Regional Differences in the Effects of Education on Parity Progression Ratios in Ethiopia," Ghilagaber and Peristera proposed a sequential procedure to model differentials in parity progression in Eritrea based on data from its three Demographic and Health Surveys (1995, 2002, and 2010). Their preliminary results showed that the sequential model provides more insight than standard models with respect to correlates of parity progression ratios in particular and the fertility decision process in general. They found similarities and differences in the effect of covariates on parity progression among women in different surveys. Substantive explanations for the results will be attempted. In chapter "Propensity Score Approaches for Estimating Causal Effects of Exposures in Observational Studies," Twabi and Manda assessed causal effects of maternal health (including HIV infection) and breastfeeding practices on child health outcomes. They offered a statistical causal inference method to rigorously investigate the purported causal relationships of maternal HIV infection, nutritional status, and breastfeeding practices on child health outcomes from population-based nationally representative data from Demographic and Health Surveys in Malawi and Zambia.

Part II (Systematic Review and Statistical Meta-Analysis) is organized with four chapters (Chapters 5 to 8). In chapter "Evidence-Informed Public Health, Systematic Reviews and Meta-analysis," Abariga, Ayele, McCaul, Musekiwa, Ochodo, and Rohwer used systematic reviews, statistical meta-analysis, and illustrative examples relevant to Sub-Saharan Africa that can be used to inform public health decisions. They unpacked aspects that need to be considered when performing meta-analysis including statistical tests to use, assessment of heterogeneity, subgroup analysis, meta-regression, and sensitivity analysis. Furthermore, they covered emerging techniques in the meta-analysis, including network meta-analysis, multivariate meta-analysis, data synthesis when meta-analysis is not possible, and meta-analysis of diagnostic test accuracy (DTA) studies. In chapter "Statistical Meta-analysis and Its Efficiency: A Real Data Analysis and a Monte-Carlo Simulation Study," Chen gave an overview of meta-analysis on classical fixed-effects and random-effects to synthesize summary statistics as well as meta-regression to explain the between-study heterogeneity. A Monte-Carlo simulation study was designed to illustrate the relative efficiency of the MA using summary statistics to the MA using the original individual participant-level data. Real meta-data from 13 clinical trials to assess the Bacillus Calmette-Guerin vaccine in the prevention of tuberculosis was used to demonstrate the implementation of these meta-analysis models. In chapter "Meta-Analysis Using R Statistical Software," Onyango and Wao introduced a series of topics in systematic review and meta-analysis (SRMA). They used illustrative examples to demonstrate how SRMA is undertaken for one continuous and one dichotomous outcome. In chapter "Longitudinal Meta-analysis of Multiple Effect Sizes," Musekiwa and colleagues discussed the meta-analysis from multiple outcomes where multiple effect sizes are estimated and produced. These estimated effect sizes could be correlated because they are measured from the same studies. Additionally, the outcomes are often measured longitudinally, resulting in multiple effect sizes estimated repeatedly over time. This chapter proposes methods for statistical meta-analysis combining summary data from more than one longitudinal study with multiple effect sizes. The proposed methods were illustrated by an analysis of an example involving longitudinal meta-analysis of HIV studies assessing the effect of some antiretroviral drugs in improving viral load suppression and increasing CD4 count at weeks 4, 8, 12, 16, 20, 24, 32, 40, and 48 after start of treatment assignment.

Part III (Spatial-Temporal Modelling and Disease Mapping) consists of two chapters (Chapters 9 to 10). In chapter "Measuring Bivariate Spatial Clustering in Disease Risks," Darikwa and Manda compared a set of full Bayesian estimations for fitting a multivariate spatial disease model. They applied the models to age-gender all-cause mortality in South Africa and childhood illnesses in Malawi. The effect on the degree of spatial correlation after adjusting for socio-demographic factors previously associated with studies diseases is also assessed. In chapter "Bivariate Copula-Based Spatial Modelling of Health Care Utilisation in Malawi," Gondwe, Chipeta, and Kazembe constructed three joint models: first to analyze the distribution of mixed binary-continuous data, a second for a mixture of a count and continuous variables, and a third for a discrete set of count and binary variables. The models are applied to study ANC utilization among Malawian women using the 2015 Malawi Demographic and Health Survey (MDHS) data, drawn using a stratified cross-sectional survey design. The models allowed for simultaneous estimation of dependence and marginal distribution parameters of timing and frequency of healthcare utilization to understand factors influencing utilization. Covariates included demographics, socio-economic factors, and location. Various models were fitted and compared, assuming different spatial structures.

Part IV (Bayesian Statistical Modelling) is composed of four chapters (Chapters 11–14). In chapter "Bayesian Survival Analysis with the Extended Generalized Gamma Model: Application to Demographic and Health Survey Data," Liang and Ghilagaber extended the existing family of flexible survival models by assembling models scattered across the literature into a more knit-form and under the same umbrella. New special cases are obtained not only by constraining the shape and scale parameters of the extended generalized gamma (EGG) model to fixed constants but also by imposing relationships (such as reciprocal) between them. The models were illustrated using data on family initiation from Demographic and Health Surveys in some Sub-Saharan African countries. Preliminary results showed that the further extended family of distributions provided a wide range of alternatives for a baseline distribution in the analysis of survival data. In chapter "Dynamic Bayesian Modeling of Educational and Residential Differences in Family Initiation Among Eritrean Men and Women," Munezero and Ghilagaber proposed a dynamic Bayesian survival model in analyzing differentials in the timing of family initiation. Such formulation relaxed the strong assumption of constant hazard ratio in conventional proportional hazard models and allows covariate effects to vary over time. The inference is fully Bayesian and efficient sequential Monte Carlo (Particle Filter) is used to sample from the posterior distribution. They illustrated the proposed model with data on entry into first marriage among Eritrean men and women surveyed in the 2010 Eritrean Population and Health Survey. Results from the conventional proportional hazards model indicated significant differences in family initiation among all educational and residential groups. In the dynamic model, on the other hand, only one educational and one residential group among the women and only one residential group among the men differed from their respective baseline groups. In chapter "Bayesian Spatial Modeling of HIV Using Conditional Autoregressive Model," Ogunsakin and Chen proposed a generalized linear model (GLM) with Bayesian inference to build the Spatially Varying Coefficients model and compared it with the stationary model to evaluate the spatial association between the incidence of HIV and some socio-demographic risk factors in Nigeria. They found a nonlinear relationship between the incidence of HIV and age. The modeling of the socio-demographic predictors of HIV infection and spatial maps provided in this study could aid in developing a framework to alleviate HIV and identify its hotspots for urgent intervention in the endemic regions. In chapter "Estimating Determinants of Stage at Diagnosis of Breast Cancer Prevalence in Western Nigeria Using Bayesian Logistic Regression," Ogunsakin and Chen estimated the prevalence and investigated determinants of stage at diagnosis by constructing Bayesian logistic regression model from a generalized linear modeling using socio-economic, demographic, and medical factors. They established that age, higher educational level, being a westerner, as well as choosing nursing as a career were the major factors that motivate early stage at breast cancer diagnosis in this part of Nigeria and that delays in diagnosis reflect a lack of education. They recommend an intensive health education program in order to increase early-stage diagnosis for patients.

Part V (Statistical Applications) has four chapters (Chapters 15–18) to discuss the statistical methods and applications in longitudinal data, survival data, and missing data imputation. In chapter "Identifying Outlying and Influential Clusters in Multivariate Survival Data Models," Kaombe and Manda developed methods for group outlier and influence assessments for the time-independent clustered survival model. Appropriate extensions of martingale-based residuals in univariate survival model and the re-weighted minimum covariance determinant method in multivariate linear mixed-effects model have been defined for group outlier analysis for the clustered survival model. They adapted influence approximations based on the one-step Newton-Raphson method for maximum likelihood estimators in univariate survival analysis to develop a group influence method for the survival mixed model. They demonstrated the performance of the proposed methods through a simulation study and real data application. In chapter "Joint Modelling of Longitudinal and Competing Risks Survival Data," Masangwi, Muula, and Mukaka used a joint modeling framework to combine the three blocks in the analysis. The methods were applied to the malaria dataset from Malawi where longitudinal markers hemoglobin level and parasite count were considered. Time to treatment failure due to severe malaria and time to withdrawal were the survival outcomes. Different survival outcomes were observed, and they noted that when there is an association between longitudinal and survival outcomes in biomedical research, joint models should be considered as they performed better than the separate methods. But where there is no association, separate models for survival and longitudinal data analysis should be considered. In chapter "Stratified Multilevel Modelling of Survival Data: Application to Modelling Regional Differences in Transition to Parenthood in Ethiopia," Ghilagaber, Lagehäll, and Yemane presented a multilevel extension of the Cox proportional hazards model where a shared frailty term is included to account for clustering of women within households. The extended model is used to analyze regional differences in the intensity of transition to parenthood among 15,019 Ethiopian women aged 15–49 years old in the country's Demographic and Health Survey of 2016. They found that household frailty effects are fairly small in the nine regions, but the log-normal frailties were significant in the entire country and the two city administrations which are relatively heterogeneous with inhabitants from many ethnic groups. They also found regional differences in the effects of the background variables on the intensity of transition to parenthood, but the effects were generally stable across the three models in each region. In chapter "Application of Multiple Imputation, Inverse Probability Weighting, and Double Robustness in Determining Blood Donor Deferral Characteristics in Malawi," Kudowa, Mavuto, and Mukaka addressed missing data to a retrospective cohort involving blood donor data to estimate predictors of donor deferral status. The logistic regression model was fit on deferral status and the independent variables. Multiple Imputation by Chained Equation, Inverse Probability Weighting (IPW), and Double Robustness (DR-IPW) were applied to correct for the missingness. The estimates from these methods were compared with estimates from the CC method.

We sincerely thank all of the people who have given us strong support for the publication of this book on time. Our acknowledgments go to all the chapter authors (in the "List of Contributors") for submitting the excellent works to this book. We deeply appreciate the reviews of many reviewers (in the "List of Chapter Reviewers"). Their comments and suggestions have improved the quality and presentation of the book substantially. Last but not least, we are so grateful to Laura Aileen Briskman and Eva Hiripi (Editors, Statistics, Springer Nature) and Kirthika Selvaraju (Project Coordinator of Books, Springer Nature) for their full support during the long publication process. We look forward to receiving comments about the book from the readers. For any suggestions about further improvements to the book, please contact us by email.

Phoenix, AZ, USA — Ding-Geng (Din) Chen
Pretoria, South Africa — Samuel O. M. Manda
Johannesburg, South Africa — Tobias F. Chirwa

Contents