SUSAN SSACAB 2019

The Biostatistics Research Unit of the South African Medical Research Council is pleased to invite you to attend the Joint Conference of the Sub-Saharan African Network (SUSAN) of the International Biometrics Society (IBS) and DELTAS Africa Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB)

Cape Town,
South Africa
8 - 11 September 2019

Abstract Book

saMRC advancinglife

SSACAB

DELTAS Africa

IBS

Detailed programme

* Indicates a SSACAB funded student
# Indicates a student scholarship awardee

SUNDAY, 8 SEPTEMBER 2019				
08:00 – 09:00	**Registration**			
09:00 – 10:30	Infectious diseases monitoring and modeling using R **Venue: Auditorium**	Longitudinal and Incomplete Data **Venue: Sugar Bird**	Subgroup analyses in clinical trials: challenges and benefits **Venue: Hadeda**	Get your journal article written and published **Venue: Board Room**
10:30 – 11:00	**Tea break**			
11:00 – 12:30		Workshop continues		
12:30 – 13:30	**Lunch**			
13:30 – 15:00		Workshop continues		
15:00 – 15:30	**Tea break**			
15:30 – 17:00		Workshop continues		

MONDAY, 9 SEPTEMBER 2019			
08:00 – 09:00	Registration		
09:00 – 10:30	**Opening ceremony - Prof Samuel Manda: Director of Biostatistics Research Unit, SAMRC and Chair of the SC**		*Auditorium*
	09:00 – 09:30	Opening remarks	Dr Tarylee Reddy: Chair of the LOC
			Prof Tobias Chirwa: Principle Lead of SSACAB
			Dr Jupiter Simbeye Coordinator of SUSAN
	09:30 – 09:50	Official opening of the conference	Prof Glenda Gray: President of the SAMRC
			Prof Jeffrey Mphahlele: Vice President for Research of the SAMRC
	09:50 – 10:30	Opening keynote address	Prof Geert Molenberghs
10:30 – 11:00	**Tea break**		

11:00 – 12:30	Parallel sessions		
	STS 1: Biostatistical Methods in Clinical Trials		*Venue: Auditorium* *Chair: Din Chen*
		Penalized multivariate t mixed model for multiple longitudinal data analysis	Mohammad Arashi
		A generalized Bayesian nonlinear mixed effects regression model for zero inflated longitudinal count data in tuberculosis trials	Divan Aristo Burger
		Applications of the expected power (statistical assurance) for bioequivalence trials	Arne Ring
		Classifying gene expression data with mixture models	Michelle de Klerk

PS 1: Statistics for high-dimensional data		*Venue: Sugar Bird* *Chair: Ziv Shkedy*
11:00 – 11:15	An exploratory analysis of multidimensional binary data using correspondence analysis, non-metric multidimensional scaling and cluster analysis as applied to Biolog EcoPlate data	Sheroline Nombasa Ntushelo
11:15 – 11:30	Geospatial computing of large scale spatial data: Faster computation of the dense spatial matrix	Eustasius Musenge
11:30 – 11:45	High dimensional surrogacy: Modeling and computational aspects	Rudradev Sengupta
11:45 – 12:00	Improving chest x-ray classification using transfer learning	Paul Mwaniki*
PS 2: Applied statistics		*Venue: Cape Teal* *Chair: Henry Mwambi*
11:00 – 11:15	Comparison of antimicrobial resistant Neisseria gonorrhea to ceftriaxone and ciprofloxacin using proportional odds model for patients seen at the University Teaching Hospital in Lusaka, Zambia	Priscilla Kapombe*
11:15 – 11:30	A two parameter Rama distribution with applications	Edith Umeh
11:30 – 11:45	Order statistics approach to modeling and prediction of early mood swing	Ajibola Taiwo Soyinka
11:45 – 12:00	Prediction models for newborn complications at birth: comparing multinomial logit models and supervised machine learning algorithm Naïve-Bayes classification methods	Paul Mubiri
12:00 – 12:15	Causal inference for multiple outcomes for observational studies	Halima Twabi*
PS 3: Survival analysis		*Venue: Hadeda* *Chair: Samuel Manda*
11:00 – 11:15	Modelling cure fractions for complex hazard functions in a cancer study	Peter Koleoso
11:15 – 11:30	Prediction of family state occupancy in rural South Africa using multistate transition modeling	Jesca Mercy Batidzirai*
11:30 – 11:45	Modelling CD4 Count and mortality in a cohort of patients initiated on HAART	Nobuhle Mchunu*
11:45 – 12:00	Modelling recurrent events data for hypertension and diabetes control in a peri-urban area in South Africa	Charl Janse van Rensburg
12:00 – 12:15	Risk of hepatic injury among adult patients concomitantly on statin and statin interacting drugs: A self-controlled case-series analysis using an electronic health database	Henry Athiany
12:30 – 13:30	**Lunch**	

13:30 – 15:00	Parallel sessions	
	STS 2: Biostatistics capacity building in sub-Saharan African countries in the era of increasing Biomedical Research in the region	*Venue: Auditorium* *Chair: Misrak Gezmu*
	Moi-Brown Partnership for Biostatistics Training in HIV	Ann Mwangi
	Sharing experiences in developing and implementing the MSc Biostatistics program at Stellenbosch University	Taryn Young
	Building Biostatistics research leadership for Africa through pooling of limited and sparse resources: A case of DELTAS Africa	Tobias Chirwa
	Building Biostatistics Capacity in South Africa: The role of the SAMRC Biostatistics Unit	Tarylee Reddy
	Discussant	Samuel Manda
	Discussant	Misrak Gezmu
	PS 1: Applied statistics	*Venue: Sugar Bird* *Chair: Jupiter Simbeye*
13:30 – 13:45	Modelling the force of infection for hepatitis B among heterogeneous groups reporting at Tertiary Hospital, Ghana	Emmanuel Kweku Nakua
13:45 – 14:00	Factors responsible for safe sex practice among female adolescent, A case study of the University of Fort Hare students	Makhadimola Rosa Leshabane
14:00 – 14:15	Evaluations of error in variable regression methods for the analysis of ECG data	Arne Ring
14:15 – 14:30	Factors associated with modern family planning (MFP) use among women age 15-24 years in Uganda	Glorious Atukunda
14:30 – 14:45	Some new nonlinear growth models for biological processes based on hyperbolic sine function	Oluwafemi Samuel Oyamakin
14:45 – 15:00	Modeling proportion of infected tsetse flies over time in Kajiado and Narok districts, Kenya	Caroline Mugo*
	PS 2: Correlated data analysis	*Venue: Cape Teal* *Chair: Birhanu Ayele*
13:30 – 13:45	Hierarchical mixed effects model for type 2 diabetes	Haile Mekonnen Fenta
13:45 – 14:00	Intimate partner violence assessment: A comparative analysis of the standard logit model versus copula bivariate semiparametric logit in determining the factors associated with intimate partner violence from Kenya demographic and health survey in 2014	Thomas Achia
14:00 – 14:15	Linear mixed models with time-varying covariates: Application of disaggregation of within-subject and between-subject effects	Stella May Gwini

14:15 – 14:30	Semiparametric techniques for multilevel discrete survival data	Thambeleni Nevhungoni
14:30 – 14:45	Modeling the stability and determinant factors of household food insecurity: A pair copula construction approach	John Olaomi
PS 3: Statistics in sample surveys		*Venue: Hadeda* *Chair: Khangelani Zuma*
13:30 – 13:45	On the correct modeling of the association between an outcome and a mismeasured covariate in clustered cross-sectional surveys: A simulation-based study	Alexander Kasyoki*
13:45 – 14:00	Evaluating the effect of sampling weights on the predictors of contraceptive use in Uganda using 2016 UDHS Data	Edson Mwebesa
14:00 – 14:15	Feasibility of employing systematic random cluster selection with probability proportional to estimated size and without replacement, to obtain a nationally representative sample of 9204 clusters for estimating vaccination coverage in South Africa	Portia Mutevedzi
14:15 – 14:30	Inequalities in stunting among under five children in Tanzania: Decomposing the concentration indexes using demographic health surveys from 2004/5 - 2015/6	Edwin Musheiguza*#
14:30 – 14:45	Time interval to modern contraceptive use following child birth among reproductive women in Tanzania: Evidence from Tanzania Demographic Health Survey2015/16	Martin Mujuni Rwabilimbo*
14:45 – 15:00	Comparison of random survival forests split rules in selecting the determinants of under-five mortality using 2014 Kenya DHS data	Kennedy Wanyonyi*
15:00 – 15:30	**Tea break**	
15:30 – 16:15	**Invited talk: Prof Tom Louis** (Johns Hopkins Bloomberg School of Public Health)	*Auditorium* *Chair: Henry Mwambi*
17:00 – 19:00	**Welcome reception**	

TUESDAY, 10 SEPTEMBER 2019			
08:30 – 09:15	**Invited speaker: Prof Carl Lombard** (South African Medical Research Council and University of Stellenbosch, South Africa)		*Auditorium* *Chair: Samuel Manda*
09:15 – 10:30	Parallel sessions		
	STS 3: Biostatistical Methods in Public Health Research		*Venue: Auditorium* *Chair: Din Chen*
	Investigation of hepatitis B vaccination behaviour using unsupervised- and supervised statistical methods		Johan Ferreira
	On the electroencephalogram classification via the mixture of mean-mixture of normal autoregressive models		Mehrdad Naderi
	Spatial disease modelling with application examples		Henry Mwambi
	Spatial hotspot sampling for a rabies vaccination schedule in rural villages		Inger Fabris-Rotelli
	PS 1: Spatial statistics		*Venue: Sugar Bird* *Chair: Thomas Achia*
	09:15 – 09:30	Spatial and temporal patterns of acute febrile illness and confirmed typhoid fever infection and their associated correlates in Ndirande, Blantyre, Malawi: An analysis of STRATAA data	Jessie Khaki*
	09:30 – 09:45	Malaria morbidity in under-five children in Malawi: A Bayesian spatial-temporal Model	Vitumbiko Chijere Chirwa*
	09:45 – 10:00	Skewed random effects distribution in conditionally autoregressive spatial models for estimating HIV prevalence at local level in South Africa	Kassahun Abere Ayalew
	10:00 – 10:15	Spatiotemporal patterns of successful TB treatment outcomes among HIV co-infected patients in Kenya	Verrah Otiende
	10:15 – 10:30	The impact of correlated socio-demographic factors in spatial profile of diabetes in africa with missingness	Sakhile Mnguni*
	PS 2: Statistics for agriculture		*Venue: Cape Teal* *Chair: Legesse Kassa Debusho*
	09:15 – 09:30	Analysis of climatic variability and its effects on production of selected crops in Ada'a District, Ethiopia: Multivariate time series approach	Diribsa T. Bedada
	09:30 – 09:45	Effect of different potassium fertilizer rates and liming on recommended maize yield grown in western Kenya	Cyrus Githunguri
	09:45 – 10:00	Product profiling: Standard multivariate or multi-block statistical methods? An application to SA honeybush herbal tea	Marieta van der Rijst
	10:00 – 10:15	Effect of different zinc fertilizer rates and cropping systems on recommended maize and beans grown in western Kenya	Cyrus Githunguri

	PS 3: Statistics for clinical trials		Venue: Hadeda Chair: Carl Lombard
	09:15 – 09:30	Digital data capturing: Using REDCap for a multi-arm, multi-site randomised clinical trial	Ishen Seocharan
	09:30 – 09:45	Estimation of individual-level local average treatment effects in cluster randomised trials with non-adherence	Schadrac Agbla
	09:45 – 10:00	Permutation multivariate tests for treatment effect: Theory and recent developments	Stefano Bonnini
	10:00 – 10:15	Superiority and non-inferiority hypothesis testing with functional data endpoints	Arsene Brunelle Sandie
	10:15 – 10:30	The usefulness of prediction intervals in quantifying effect heterogeneity in randomized controlled trials	Hao Zhang
10:30 – 11:00	**Tea break**		
11:00 – 11:45	**Invited talks: Prof Lehana Thabane** (University of McMaster, Canada)		*Chair: Tobias Chirwa*
11:45 – 13:00	Parallel sessions		
	STS 4: Advances in Bayesian Spatio-temporal modelling of disease risk based complex household surveys in Sub-Saharan Africa (SSA)		*Venue: Auditorium* *Chair: Ngianga-Bakwin Kandala*
	Modelling and mapping prevalence of Female Genital Mutilation/C (FGM/C) among 0-14 years old girls in Kenya, Nigeria and Senegal		Ngianga-Bakwin Kandala
	Geographic Variation in the Prevalence of Childhood Diarrhoea and Malaria in Malawi and Zambia Using Remotely Sensed and Climatic Data, Controlling for Deprivation		Samuel Manda
	Incorporating Sampling Design in Spatial Estimation of Small Area Health Outcomes Sub-Saharan Africa		Sheyla Cassy#
	Dietary patterns and linkages to non-communicable diseases in Namibia		Laina T. Mbongo
	Cross sectional and spatial analyses of self-reported Tuberculosis in South Africa: Results from SABSSM 2017		Inbarani Naidoo
	PS 1: General biostatistics		*Venue: Sugar Bird* *Chair: Jim Todd*
	11:45 – 12:00	Factors associated with prevalence of pregnancy termination in Uganda: A population-based study	Edson Mwebesa*
	12:00 – 12:15	The association between childhood environmental exposures and the subsequent development of Crohn's disease in the Western Cape, South Africa	Mikateko Mazinu
	12:15 – 12:30	Developing and validating a scoring tool for mortality among neonates with oesophageal atresia and tracheoesophageal fistula at a tertiary level hospital in KwaZulu-Natal, South Africa	Yusentha Balakrishna
	12:30 – 12:45	Risk factors and spatial heterogeneity of childhood anaemia in four sub-Saharan African countries	Danielle Roberts

	12:45 – 13:00	Mortality rates, associated risk factors and causes of death in older children with neurological impairments in rural Kenya: A cohort study	Jonathan Abel*
	PS 2: Longitudinal data analysis		*Venue: Cape Teal* *Chair: Freedom Gumedze*
	11:45 – 12:00	A semi-parametric mixed model for longitudinally measured fasting blood sugar level of adult diabetic patients	Legesse Kassa Debusho
	12:00 – 12:15	Diagnostics for a two-stage joint survival model	Isaac Singini
	12:15 – 12:30	Right-censoring bias correction for growth curve linear mixed models	Dominique Laurent Couturier
	12:30 – 12.45	Temporal interactions of microbiota in longitudinal nasopharyngeal samples	Maia Lesosky
	12:45 – 13:00	Modelling the effect of rape on mental health status one year post enrollment: RICE study	Shibe Mhlongo
	PS 3: Statistics for missing data		*Venue: Hadeda* *Chair: Geert Molenberghs*
	11:45 – 12:00	Solutions for selective loss to follow-up in HIV-cancer cohorts in Malawi	Evaristar Kudowa
	12:00 – 12:15	Assessing the sensitivity and robustness of randomization test in analysis of repeated measures design with missing observations	Abimibola Oladugba
	12:15 – 12:30	Random Forests application in missing data and predictive modelling for hierarchical routine clinical data: A case study of childhood pneumonia in Kenya	Steven Wambua*#
	12:30 – 12:45	Efficient methodologies for handling missing data for longitudinal ordinal outcome	Omololu Aluko
	12:45 – 13:00	Evaluation of the robustness of imputation methods combined to backpropagation algorithm in frame of multiple non linear regression	Gbememali Castro Hounmenou
13:00 – 13:45	**Lunch**		
13:45 – 15:00	Parallel sessions		
	STS 5: Statistical Methods for the Analysis of Genomic Data		*Venue: Auditorium* *Chair: Bernard Omolo*
	A model-based approach to genetic association testing in malaria studies		Bernard Omolo
	An Integrated RNA and DNA Molecular Signature for Colorectal Cancer Classification		Bernard Omolo
	A full Bayesian hierarchical mixture model for the variance of gene differential expression		Samuel Manda
	protGear: A tool for protein microarray data pre-processing		Mwai Wambui#
	A comparison of cancer classification methods based on microarray data		Mohanad Mohammed*

PS 1: General biostatistics		*Venue: Sugar Bird* *Chair: Bernard Rachet*
13:45 – 14:00	Multilevel logistic regression for patients' satisfaction in Ethiopian public hospitals	Anteneh Yalew
14:00 – 14:15	Health spending and economic growth nexus: Evidence from Uganda	Smartson Ainomugisha*
14:15 – 14:30	analysis of student-lecturer interaction in online PHT 112 course at Maseno University	Mary Dancilla Wanjiru
14:30 – 14:45	Mortality rate and associated factors among preterm babies born in Moshi Municipality in northern Tanzania	Michael Johnson Mahande
14:45 – 15:00	Using software in teaching medical data	Lily Clements
PS 2: Bayesian biostatistics		*Venue: Cape Teal* *Chair: Eustacius Musenge*
13:45 – 14:00	Inform prior elicitation for Bayesian questionnaire validation using confirmatory factor analysis	Hao Zhang
14:00 – 14:15	Bayesian dynamic models for time-varying outcomes: Application to a patient cohort on ART	Lineekela Gabriel*
14:15 – 14:30	Bias estimation and sensitivity analysis in questionnaire studies with "embarrassing" questions	Christian Hansen
14:30 – 14:45	Modelling predictors of stroke disease in South Africa: Bayesian binary quantile regression approach	Lyness Matizirofa
PS 3: HIV and TB statistics		*Venue: Hadeda* *Chair: Tobias Chirwa*
13:45 – 14:00	Predictors of non-completion to Isoniazid preventive therapy among people living with HIV attended care and treatment clinics in Dar-es Salaam from 2013 to 2017	Masanja Robert*
14:00 – 14:15	Cost-effective, cross-sectional cohort estimation of HIV incidence rate in presence of misclassification error using doubly robust estimation	Kesaobaka Molebatsi
14:15 – 14:30	Changes in socioeconomic status among HIV patients after 10-years of antiretroviral therapy in Uganda: An estimation of the within-subject correlations	Godwin Anguzu
14:30 – 14:45	Mixed method estimation of population HIV viral suppression in the Western Cape, South Africa	Elton Mukonda
14:45 – 15:00	Time to HIV serostatus disclosure, pattern and factors among pregnant women living with HIV in Moshi urban.	Victoria Barabona*
15:00 – 15:30	**Tea break**	
18:30 – 21:00	Dinner	Gold Restaurant

WEDNESDAY, 11 SEPTEMBER 2019			
08:30 – 09:15	**Invited talks: Prof Andreas Ziegler (StatSol, Lubeck, Germany)**		*Auditorium* *Chair: Jupiter Simbeye*
09:15 – 10:30	Parallel sessions		
	STS 6: The >eR-Biostat: A new E-learning system for BioStatistics		*Venue: Auditorium* *Chairs: Ziv Shkedy, Khangelani Zuma*
	The E-learning system for generalized linear models (GLM)		Khangelani Zuma
	Development of a Master program using the >eR-Biostat		Tadesse Awoke Ayele
	Using R markdown to develop E-learning capacity		Olajumoke Evangelina Owokotomo
	An introduction to the >eR-Biostat E-learning system		Ziv Shkedy
	PS 1: Statistics for missing data		*Venue: Sugar Bird* *Chair: Tarylee Reddy*
	09:15 – 09:30	Bayesian approach in treating missing values with metabolomics	Jasmit Shah
	09:30 – 09:45	Predictors of perinatal deaths in the presence of missing data: A registry based study in northern Tanzania	Innocent Mboya*
	09:45 – 10:00	Robust estimation of single-index models with responses missing at random	Masego Otlaadisa
	10:00 – 10:15	Multiple imputation of clinician-level covariates applied in modelling quality of clinician-prescribed care: Sensitivity analysis of departure from Missing at Random (MAR) assumption	Susan Gachau*#
	10:15 – 10:30	Analytical methods used in handling missing data in estimating prevalence of HIV/AIDS for demographic and cross-sectional surveys: A systematic review.	Neema Mosha*
	PS 2: Disease mapping		*Venue: Cape Teal* *Chair: Inger Fabris Rotelli*
	09:15 – 09:30	Climatic changes and tsetse fly distributions in Kwazulu-Natal, South Africa: A Spatio-temporal Model	Nada Abdelatif
	09:30 – 09:45	Bayesian modelling of tick count data over time and space using discrete model approaches	Thabo Lephoto
	09:45 – 10:00	Spatial distribution of HIV prevalence among young people in Mozambique	Rachid Muleia
	10:00 – 10:15	Combining data from national surveys with facility-based HIV testing to obtain more accurate estimate of HIV prevalence in districts in Uganda	Joseph Ouma*
	10:15 – 10:30	Statistical measures for multivariate spatial autocorrelation	Timotheus Darikwa

	PS 3: Evidence synthesis and data harmonisation		*Veue: Hadeda* *Chair: Samuel Manda*
	09:15 – 09:30	A comparative study of palmar and digital dermatoglyphic patterns among type-II diabetic & non-diabetic adults: A meta-analysis	Ameet Jha
	09:30 – 09:45	An algorithm for data reconstruction from published articles - Application on insect life tables	Dorcas Kareithi
	09:45 – 10:00	A review on disease mapping and modelling of cardiovascular diseases in South Africa	Nomonde Gwebushe
	10:00 – 10:15	Impact of improved drinking water source and sanitation, and urban residence on child diarrhoea in Sub-Saharan Africa. A meta-analysis of demographic and health surveys	Rejoice Msiska
10:30 – 11:00	**Tea break**		
11:00 – 11:45	Parallel sessions		
	PS 1: Infectious disease modelling		*Venue: Auditorium* *Chair: Henry Mwambi*
	11:00 – 11:15	Probabilistic modeling for an integrated temporary acquired immunity with norovirus epidemiological data	Emmanuel deGraft Johnson Owusu Ansah
	11:15 – 11:30	Mathematical modeling of the interruption of the transmission of soil transmitted helminths infections in Kenya	Collins Okoyo*
	11:30 – 11:45	Analysis of a vaccination mathematical model of an infectious measles disease	Hope Mbachu
	PS 2: Survival and longitudinal analysis		*Venue: Sugar Bird* *Chair: Michael Johnson Mahande*
	11:00 – 11:15	Nonparametric cure rate estimation when cure is partially known	Wende Clarence Safari
	11:15 – 11:30	Discrete Time Hazard Parameterization for Child Survival in Malawi	Jessie Khaki*
	11:30 – 11:45	Bayesian Copula-Based analysis of joint model for multi-variable longitudinal and recurrent events	Tefo Baraki#
	PS 3: TB and HIV		*Venue: Cape Teal* *Chair: Kathy Baisley*
	11:00 – 11:15	The effect of cotrimoxazole prophylaxis on CD4 cell-count profiles in HIV-positive patients stabilized on anti-retroviral therapy: Analysis of data from the COSTOP trial	Tshwaraganang Modise
	11:15 – 11:30	Gender differential and social determinant of tuberculosis/HIV co-infected patients	Opeyemi Oyekola Ogungbola
	11:30 – 11:45	Copula bivariate based semiparametric logit models: Application to the joint modeling of psychoactive drugs and HIV risk among fisherfolk in Kenya	Thomas Achia

11:45 – 13:00	Parallel sessions	
	STS 7: Time-to-event analysis using population-based data	*Venue: Auditorium* *Chair: Bernard Rachet*
	A flexible excess hazard modelling approach within the relative survival setting	Bernard Rachet
	Modeling the crude probability of cancer death and number of life-years lost due to cancer using a pseudo-observation approach	Bernard Rachet
	Small area variations in excess prostate cancer specific survival in the Northern and Yorkshire region of England using Bayesian hierarchical spatial smoothing	Samuel Manda
	Statistical methods to link demographic and health survey data with census data	Ruth Vellemu*#
	Age, period and cohort analysis of young adult mortality due to HIV and TB in South Africa: 1997-2015.	Tshifhiwa Nkwenika
	PS 1: General biostatistics	*Venue: Sugar Bird* *Chair: Andreas Ziegler*
11:45 – 12:00	Hematological reference intervals for children in Bagamoyo District, Tanzania	Ummi Abdul Kibondo*
12:00 – 12:15	Comparing coverage and precision of Newcombe confidence intervals for proportion differences generated using different weighting systems	Miriam Wathuo
12:15 – 12:30	Fertility differential between Mombasa and Nairobi counties in Kenya	Lilian Owino
12:30 – 12:45	Progression of Cisplatin-associated Ototoxicity amongst Patients receiving Cancer Chemotherapy: Application of Tobit Regression	Birhanu Ayele
12:45 – 13:00	Developing Socio-Economic Status (SES) index for Basse region in The Gambia	Abdul Khalie Muhammad
	PS 2: Survival and longitudinal data analysis	*Venue: Cape Teal* *Chair: Lehana Thabane*
11:45 – 12:00	Joint multistate modelling of viral load dynamics and multistate CD4 count progression	Zelalem Dessie*
12:00 – 12:15	Exploring association between predominant breastfeeding and infection-related hospitalization, over time: an empirical comparison of models to account for clustering	Moleen Dzikiti (née Zunza)*
12:15 – 12:30	Comparison of COX proportional hazard model and accelerated failure time model with application to data on tuberculosis/HIV patients in Nigeria	Opeyemi Oyekola Ogungbola
12:30 – 12:45	Multi-state models for the analysis of Wheezing episodes in a birth cohort of african children	Patrick Hannan

12:45 – 13:00	A pairwise joint modelling approach for multivariate longitudinal immuno-epidemiological data from the Infant BCG study in Entebbe, Uganda	Lawrence Lubyayi*
PS 3: Causal inference		*Venue: Hadeda* *Chair: Nonhlanhla Yende*
11:45 – 12:00	Do parents' mental disorders affect child behaviour via home environments and/ or primary caregiver's personal and social performance? A mediation analysis	Md Jamal Uddin
12:00 – 12:15	Investigation of three instrumental variable methods in the presence of errors in explanatory variables: A comparative analysis	Maureen Nwakuya
12:15 – 12:30	Auxiliary variables in Latent Profile Analysis (LPA) and methods to estimate a distal outcome model	Esme Jordaan

13:00 – 13:45	**Lunch**		
13:45 – 15:00	Closing ceremony		*Auditorium*
	13:45 – 14:25	Closing keynote address	*Prof Jonathan Levin* *Chair: Taryn Young*
	14:25 – 15:00	Closing remarks	Tobias Chirwa/Samuel Manda
15:00 – 15:30	**Tea break**		

SPECIAL TOPIC SESSIONS

BIOSTATISTICAL METHODS IN CLINICAL TRIALS

Chair and organiser: Prof Din Chen

Sponsored by the South Africa DST-NRF-SAMRC SARChI Chair in Biostatistics, this special session is organized to discuss the current development in biostatistical methods and the applications to clinical trials research. Four experts are invited to present their findings. Specifically, Drs. Arashi and Taavoni will discuss the development of a penalized multivariate t-mixed model for multiple longitudinal data analysis and show the application to the Mayo Clinic Primary Biliary Cirrhosis sequential data followed by Dr. Burger and his co-authors to discuss the development of a generalized Bayesian nonlinear mixed-effects regression model to analyse the zeroinflated longitudinal count data in tuberculosis clinical trials. Continuing with biostatistical clinical trials, Dr. Ring will discuss the development of the expected power (statistical assurance) for bioequivalence trials and show examples on the use of the assurance concept in comparisons to power calculations followed by Ms. de Klerk and her co-authors to discuss on classifying gene expression data with mixture models.

Penalized multivariate t mixed model for multiple longitudinal data analysis

M Arashi, M Taavoni
Department of Statistics, Faculty of Mathematical Sciences, Shahrood University of Technology, Iran

In many biomedical studies or clinical trials, we come up with a data with more than one response variables on the same subject measured repeatedly over time. In analyzing such data, we adopt a multivariate linear mixed-effects longitudinal model. On the other hand, we often encounter with insignificant features in longitudinal modeling which do not have impact on modeling the response variable and must be eliminated from the study. In this paper, we consider the problem of simultaneous variable selection and estimation in a multivariate t linear mixed-effects model (MtLMM) for analyzing longitudinally measured multi-outcome data. The motivation behind this work comes from a cohort study of patients with primary biliary cirrhosis, where the interest is in eliminating insignificant variables using the smoothly clipped and absolute deviation penalty function in the MtLMM. The proposed penalized model offers robustness and flexibility to accommodate fat tails. An efficient alternating expectation conditional maximization algorithm is employed for the computation of maximum likelihood estimates of parameters. The calculation of standard errors is effected by an information-based method. The methodology is illustrated by analyzing Mayo Clinic Primary Biliary Cirrhosis sequential (PBCseq) data and a simulation study.
Keywords: Heavy-tailed distribution; Longitudinal data; Multivariate mixed-effects model; Penalized SCAD penalty.

A Generalized Bayesian Nonlinear Mixed Effects Regression Model for Zero Inflated Longitudinal Count Data in Tuberculosis Trials

Divan Aristo Burger, Robert Schall, Rianne Jacobs, Ding-Geng Chen
Department of Statistics, University of Pretoria, Pretoria, South Africa

In this presentation we investigate Bayesian generalized nonlinear mixed effects (NLME) regression models for zero inflated longitudinal count data. The methodology is motivated by and applied to colony forming unit (CFU) counts in extended bactericidal activity tuberculosis (TB) trials. Furthermore, for model comparisons we present a generalized method for calculating the marginal likelihoods required to determine Bayes factors. A simulation study shows that the proposed zero inflated negative binomial regression model has good accuracy, precision and credibility interval coverage. In contrast, conventional normal NLME regression models applied to log-transformed count data, which handle zero counts as left censored values, may yield credibility intervals that undercover the true bactericidal activity of anti-TB drugs. We therefore recommend that zero inflated NLME regression models should be fitted to CFU count on the original scale, as an alternative to conventional normal NLME regression models on the logarithmic scale.
Keywords: Bayesian, bactericidal activity, longitudinal, mixed effects, zero inflated.

Applications of the expected power (statistical assurance) for bioequivalence trials

Arne Ring[1,2], Benjamin Lang[3], Ibrahim Dan Dije[4], Detlew Labes[5] and Helmut Schütz[6]
[1]University of the Free State, Bloemfontein, South Africa, [2]medac, Wedel, Germany,
[3]Boehringer Ingelheim, Biberach, Germany, [4]AIMS Senegal, Mbour-Thies, Senegal,
[5]Consultant, Berlin, Germany, [6]BEBAC, Vienna, Austria

Bioequivalence (BE) trials are performed to demonstrate that the pharmacokinetic properties of a generic drug formulation (T-Test) are similar to those of the originator formulation (R-Reference).
Traditional the sample size determination of BE trials is based on fixed values for the T/R ratio and the intraindividual coefficient of variation CV. The ratio is often assumed to be 1.0, while the CV is chosen using historic pharmacokinetic trials.
In contrast, the assurance does not only use a single value, but an appropriate distribution of the input parameters, in order to model the uncertainty of the assumptions. While the assurance has already widespread use for planning superiority trials, its application in bioequivalence trials has not been described yet.
We show examples on the use of the assurance concept in BE trials in comparisons to power calculations and demonstrate that this concept inflates the sample sizes only slightly under typical circumstances.
Keywords: Bioequivalence, statistical power, sample size determination, trial design, crossover trial.

Classifying gene expression data with mixture models

'Michelle de Klerk [2]Frans Kanfer, [3]Sollie Millard
'University of Pretoria [2]University of Pretoria, [3]University of Pretoria

Background

Due to the vast number of observations in gene expression data high dimensionalty can be a problem when trying to model DNA microarray data. Using mixture models is a great way to overcome this by creating clusters with a gene belonging to a specific cluster.

Methods

Different packages in R are used to identify the number of clusters that is in the dataset. Each cluster also has a mixing coefficient which indicates the weight each gene has in that specific cluster and the effect size which will give a good indication if the cluster is part of the control or the experimental group. Genes will be categorised by using posterior probabilities to identify which genes belong to which clusters.

Results

Comparing the results of the genes identified in the in cluster/s with the large effect sizes to other traditional statistical models it can be see the results of the genes highlighted with mixture models is more predictive.

Conclusion

This is very valuable in large datasets such as DNA microarray, used to profile thousands of genes simultaneously. With mixture models it isn't necessary to remove observations by first testing which observations/genes are significant. All data captured can be used in the statistical model.

BIOSTATISTICS CAPACITY BUILDING IN SUB SAHARAN AFRICAN COUNTRIES IN THE ERA OF INCREASING BIOMEDICAL RESEARCH IN THE REGION

Chair and organiser: Dr. Misrak Gezmu

In the last two decades, South-North collaborative biomedical research studies have increased and are being conducted in many sub Saharan African countries. The collaborative research projects require the participation of interdisciplinary research teams both from the South and North. Biostatisticians are critical members of the research teams and play important roles in conducting the research projects. Because of the shortage of biostatisticians from the South, the biostatisticians from the North become project biostatistical leaders. In-country biostatistical leaders are essential part of the local research team. As a project biostatistical leader, they will participate in the thinking process of developing the study design, choosing the analysis methods, assuring that the research questions addressed are those that offer the greatest relevance for the local people and interpreting the study results to help policy makers revise or change health policies. Many collaborative researchers have recognized the shortage of local biostatisticians, but training local biostatisticians is a challenge. Efforts made towards biostatistics capacity building in the region include conducting workshops, building biostatistics programs, training the trainers and biostatistics courses conducted with South-South and SouthNorth collaborations. In this session, we will hear from those involved in this effort of strengthening the local biostatistics resources. Speakers will discuss success and challenges encountered and funding opportunities used in building biostatistics capacity in the region.

Moi-Brown Partnership for Biostatistics Training in HIV

Ann Mwangi

Moi University in Eldoret Kenya and Brown University in Providence USA have collaborated over the past years on several training activities to build research capacity in Biostatistics at Moi. The success of the Brown Fogarty AITRP which supported graduate-level training in biostatistics and epidemiology for 4 Kenyan students and in 2014 sponsored a highly successful one-week workshop on Causal Inference and Missing Data at Moi. Based on this success, Brown was awarded a Fogarty D43 training grant that includes three components: degree training in biostatistics at Brown for Kenya trainees at Msc and PhD level; a biostatistics 'train the trainers' program for faculty at Moi University College of Health Sciences; and a biannual summer workshop on advanced methods in biostatistics , held in Eldoret. The achievement so far: The long term program has trained 4 masters' students and 2 PhD students; train the trainers program has enrolled 50 faculty at Moi College of Health Sciences during the past 3 years; the summer workshop has served several hundred statisticians from 11 countries in sub-Saharan Africa. There has also been growth of biostatistics staff at Moi university and the range of expertise of the staff has substantially increased through the collaboration. The current talk will provide a brief description of the program current status, enumerate the key successes and challenges and future plans.

Sharing experiences in developing and implementing the MSc Biostatistics program at Stellenbosch University

T Young
Stellenbosch University

In sub-Saharan Africa (SSA), the burden of communicable diseases such as HIV, Tuberculosis and malaria continues unabated despite the increasing efforts to fight these diseases through prevention and treatment. In addition, the burden of non-communicable diseases such as cancer, diabetes, hypertension, and heart disease in the region has increased dramatically. The creation of a sustainable, multidisciplinary health research enterprise to inform and contribute to strategies addressing these problems in SSA countries is of paramount importance for improving health, promoting development and advancing science in the region. A universal shortage of biostatisticians has been documented in many countries and a cadre of professional and academic biostatisticians is needed to play a leadership role in developing the discipline of biostatistics, contribute to multidisciplinary, collaborative research in the health sciences and train the future generation of biostatisticians.

Since 2017, the Faculty of Medicine and Health Sciences at Stellenbosch University have implemented an MSc Biostatistics program. Students can either have a health, or statistical background, or need to complete a statistical bridging course before enrolment into the program. The program is offered on a full-time basis over a period of two years. Students complete 12 modules, of which ten are compulsory and two elective. Modules are offered using a combination of face-to-face teaching and e-learning using Sun Learn, Stellenbosch University's online learning environment. Students also complete a practical work place internship and do a research assignment.
We work collaboratively with local and international partners and this presentation will share the successes and challenges of our program implementation.

Building biostatistics research leadership for Africa through pooling of limited and sparse resources: A case of DELTAS Africa

Tobias Chirwa
University of the Witwatersrand, Johannesburg, South Africa

Sub-Saharan Africa (SSA) has a complex burden of communicable and non-communicable diseases. Funding initiatives for health sciences research exist in the region. However, the generation of high quality research is compromised due to limited analytical and biostatistics expertise. Imbedding such training within existing research will result in local research leadership and high-level data analysis. This will generate a pool of African researchers who can lead high quality research that is locally relevant and globally competitive.

Efforts to develop biostatistics capacity are not new in the region and exist in few academic institutions. The few efforts were disjointed, did not emphasize biostatistics, and were not linked to local biomedical research. This called for the formation of the Sub-Saharan African Consortium for Advanced Biostatistical training (SSACAB) which aimed to have coordinated efforts with north-south and south-south partner institutions and build such capacity for local needs so that we can improve health outcomes by creating a cadre of epidemiologists and biostatisticians.

Led by University of the Witwatersrand (Wits), the consortium has developed partner institutional capacity by strengthening existing and supporting the development of new postgraduate (MSc/PhD) programmes and on completion train at least 90 masters and 15 PhD students. It has supported the organisation of national statistical associations, short courses and conferences to enhance biostatistical skills. The consortium has fully supported the development of an African Centre for Biostatistical Excellence. For competitiveness and international recognition, the consortium is in the process of seeking international accreditation through the Royal Statistical Society.

The ultimate goal is to create research centres of excellence to grow the discipline and nurture researchers with advanced biostatistics skills and expertise. The regional based training is relevant, cost-effective and sustainable through student fees.

Building Biostatistics Capacity in South Africa: The role of the SAMRC Biostatistics Unit

Tarylee Reddy[1] and Samuel Manda[1,2]
[1]Biostatistics Research Unit, South African Medical Research Council, South Africa
[2]Department of Statistics, University of Pretoria, South Africa

The Biostatistics Research Unit is one of several intramural research of the South African Medical Research Council (SAMRC). It is one of the founding units of the SAMRC and was established primarily to provide design and statistical data analysis to biomedical studies conducted by units in the SAMRC. Over the years, the Unit's role has evolved to cover multiple areas ranging from the development of generic statistical methodology, applied biostatistics to capacity building in biostatistics. The purposes of this presentation is to provide an overview of the Unit's capacity building initiatives, activities and alignments. It will focus on the recently completed cooperation between the National Treasury (IDC), Enabel (ex-Belgian Technical Cooperation) and the South African Medical Research Council (SAMRC) for Building Academic Partnerships for Economic Development (BAPED) for Biostatistics in South Africa. It will cover the aims, implementation, lessons and experiences and sustainability of the program.

BIOSTATISTICAL METHODS IN PUBLIC HEALTH RESEARCH

Chair and organiser: Prof Din Chen

Sponsored by the South Africa DST-NRF-SAMRC SARChI Chair in Biostatistics, this special session is organized to discuss the current development in biostatistical methods and the applications to public health research and evidence-based decision-making. Four experts are invited to present their findings. Specifically, Dr. Ferreira and his co-authors will discuss an investigation of hepatitis B vaccination behaviour using unsupervised- and supervised statistical methods followed by Drs Naderi and Bekker on the electroencephalogram classification via the mixture of mean-mixture of normal autoregressive models. Further to biostatistical methods in geospatial disease mapping, Dr. Mwambi and his co-authors will discuss the development of joint spatial disease modelling for the prevalence of HIV and HSV-2 in Kenya followed by Dr. Fabris-Rotell and her coauthors to discuss the spatial hotspot sampling for a rabies vaccination schedule in rural villages of Tanzania.

Investigation of hepatitis B vaccination behaviour using unsupervised- and supervised statistical methods

Johan Ferreira, Elham Mirfarah, Tanita Cronje
Department of Statistics, University of Pretoria, Faculty of Natural and Agricultural Sciences, Pretoria

Delay of childhood vaccines has increased in recent years and is believed to cluster in some communities. Such clusters could pose public health risks and barriers to achieving immunization quality benchmarks. Providing hepatitis B vaccine to all neonates within 24 hours of birth is the key preventative measure to control perinatal hepatitis B virus infection. In this talk, possible geographical clusters of vaccination behaviour with regards to percentage of infants vaccinated against hepatitis B across different facilities in counties of New York State between 2012 and 2018 will be explored using unsupervised- and supervised statistical methods. Possible socio-economic effects of results will be briefly discussed.
Keywords: clustering, hepatitis, supervised learning, vaccines, unsupervised learning

On the electroencephalogram classification via the mixture of mean-mixture of normal autoregressive models

Mehrdad Naderi, Andriette Bekker
Department of Statistics, Faculty of Natural & Agricultural Sciences, University of Pretoria, Pretoria, South Africa

Electroencephalogram (EEG) data is a result of a test that describes electrical activity in the brain. The data is collected by using small, flat metal discs attached to the scalp. Although the ordinary autoregressive (AR) model is mostly used to analyze EEG data, it is well known that the AR model with the assumption of asymmetry, as well as, heterogeneity are promising alternatives. This paper presents an AR model with the mixture of mean-mixture of normal innovations to analyze the EEG data. By presenting a hierarchical representation of the model, the EM-type algorithm is implemented to obtain the maximum likelihood estimates of model parameters.

Spatial Disease Modelling With Application Examples

Henry Mwambi, Elphas Okango and Melkamu Gishu
School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg Campus

The importance of understanding spatial distribution of disease prevalence and occurrence or incidence is becoming an important tool for decision making among them prioritisation of resources and designing intervention strategies, In this talk we demonstrate the use of spatial statistical modelling using a joint spatial model for the prevalence of HIV and HSV-2 in Kenya where the spatial unit of interests is the county. This will help to explain which of the 47 counties in that country have an increased burden of these two diseases and hence inform health policy makers to decide on target intervention counties, The second example is about the spatial spread CVDs in Ethiopia both overtime and spatially, In this example a series of Bayesian hierarchical models are used with increasing complexity. We focus on disease clustering both temporally and spatially. Such models are important in explaining when in time and where in space to put in place intervention strategies.
Keywords: Spatial distribution, mapping, prevalence, incidence, intervention strategies, disease clustering, spatio-temporal.

Spatial hotspot sampling for a rabies vaccination schedule in rural villages

[1]Inger Fabris Rotelli [2]Hayley Reynolds, [3]Alfred Stein, [4]Theodor Loots
[1]University Of Pretoria, [3]University of Pretoria & University of Twente᾽ Netherlands, [4]University of Pretoria

Rabies has been reported in Tanzania, mainly in the southern highland regions, since 1954. To date, rabies is endemic in all districts in Tanzania and efforts are being made to contain the disease. It was determined that mass vaccination of at least 70% of an animal population is most effective, in terms of profitability and cost, in reducing transmission of rabies. Current approaches for vaccination in Tanzanian villages have many administrative and logistical challenges. Spatial sampling of households in villages is proposed, where optimality is measured through the distance traversed by the veterinarian by foot for vaccinating at each sampled household. The walking distance is attained by incorporating a driving network between optimally determined stopping points from which the veterinarian then walks for executing vaccinations, all while ensuring the 70% coverage of the animal population. A systematic regular spatial sampling is found to be most optimal. The vaccination scheme proposed, provides a technique to effectively manage a vaccination campaign with predictable surety of obtaining the vaccination rate of 70%.

ADVANCES IN BAYESIAN SPATIO-TEMPORAL MODELLING OF DISEASE RISK BASED COMPLEX HOUSEHOLD SURVEYS IN SUBSAHARAN AFRICA (SSA)

Chair and organiser: Prof Ngianga-Bakwin Kandala

Sub-Saharan Africa faces a high disease burden in communicable diseases and an increasing burden in noncommunicable diseases with a strong spatial and temporal structure. More recently, increased funding for research from donor initiatives has generated high-quality household data volume, but there is a lack of capacity for advanced data analysis. Globally, the fields of geographical epidemiology and public health surveillance have benefited from combined advances in hierarchical model building and in geographical information systems. Exploring and characterising a variety of spatial patterns of diseases at the disaggregated fine geographical resolution has become possible (Banerjee et al, 2004). Insight into the sensitivity of the resulting inference to the choice of the structure of the different components of the hierarchical model has been gained through the use of simulation studies (Best et al, 2005) and numerous case studies worldwide with few in SSA. Baseline results on how to use the posterior distribution of relative risk estimates to detect areas of increased risks have been implemented and tested mostly in western countries. Extension of hierarchical disease mapping models to models that simultaneously consider space and time leads to a number of benefits in terms of interpretation and potential for detection of localised excesses to inform local policies. Such extension is accompanied by an increase of the complexity of the model structures that might be specified and the rationale for the many choices that have to be made is less clearly documented and tested in complex sampling in household surveys in SSA. In this session, we present and discuss classes of Bayesian hierarchical space time models that can be used to characterise the patterns of communicable and non-communicable disease burden in SSA (Kandala, 2014). Particular attention will be paid to the influence of the geographic location and time discretisation on the resulting inference and how the space-time consideration of patterns can strengthen the inference. Secondly, we show case applications of Global health issues in SSA using different household data sources from complex survey data such the Demographic and Health Surveys (DHS) and the Multiple Indicators Cluster Surveys (MICs) and use the above models for detecting space-time clusters in a number of scenarios. Finally, we seek papers that use these models to analyse the spatio-temporal variations of HIV, TB, HSV-2, malaria, hypertension, diabetes, malnutrition, and obesity) in selected SSA countries.

Modelling and mapping prevalence of Female Genital Mutilation/C (FGM/C) among 0-14 years old girls in Kenya, Nigeria and Senegal

Ngianga-Bakwin Kandala, Glory Atilola1, Chibuzor Christopher Nnanatu, Paul Komba, Lubanzadio Mavatikua, Zhuzhi Moore, Dennis Matanda.

Background

World Health Organisation defines Female Genital Mutilation/cutting (FGM/C), also known as female circumcision, as all forms of injury caused to the external female genitalia for non-medical reasons. FGM/C is a public health and human right issue, which is strongly anchored in customs and traditions, without any established benefit. The practice has both short and long term consequences ranging from haemorrhage to complications during child birth. It is estimated that about 200 million women and girls alive today globally, have undergone FGM/C at some point in their lives, with more cuttings being performed in Africa, the Middle East and Asia, with a report by UNICEF showing that in Africa, about 3 million girls are at risk of the being cut each year.

According to Kenya Demographic and Health survey, FGM/C prevalence among women and girls aged 15-49 was estimated at 37.6% in 1998, 32.2% in 2003, and 27.1% in 2008-9. Recent studies showed that in 2017, FGM/C prevalence among girls aged 0-14 years stood at 14.0% and 25.3% in Senegal and Nigeria, respectively. There are several concerted efforts geared towards bringing change that would end the practice. Consequently, change has been reported, but the change has been sluggish, thus requiring further examinations into where, when and how change is taking place using model-based statistical approaches to model, map and describe the characteristics of the hotspots where the practice is still rife.

Methods
Robust Bayesian Hierarchical spatial and spatio-temporal models which simultaneously accounted for unobserved effects of space and time, as well as space-time interactions, whilst controlling for other linear and non-linear covariates were employed. These models were developed and fitted on the available datasets in a coherent mixed models regression framework. Posterior inference was carried out using Markov Chain Monte Carlo (MCMC) techniques, while model fit and complexity assessments utilised Deviance Information Criterion (DIC) approach.

Results
There was an overall decline in the practice as found across the three countries. The Bayesian hierarchical modelling approach allowed us to jointly account for both individual-, household-, community-level factors, map and identify patterns and spatial variations in the practice, thus unmasking the hotspots across the three countries. Factors found to associate with higher risk of the practice included mother's FGM/C status, support for FGM/C continuation, household wealth index, level of education of mother, region and type of place of residence, marital status and religion.

Conclusion
Bayesian Hierarchical models which considered variations in space and time, as well as space-time interaction, offered us a greater insight in understanding the spatial and spatio-temporal structure of FGM/C prevalence. There has been a decline in FGM/C prevalence across the three countries, however, the menace of the practice is not yet over. More rigorous programmatic bespoke intervention approaches should be adopted and targeted on the identified hotspots.

Geographic Variation in the Prevalence of Childhood Diarrhoea and Malaria in Malawi and Zambia Using Remotely Sensed and Climatic Data, Controlling for Deprivation

Samuel Manda[1,2], Tshifhiwa Nkwenika[1],Jonthan Kamw[3], Robert Bergquist[4]

[1]Biostatistics Research Unit, South African Medical Research Council, Pretoria, South Africa,
[2]Department of Statistics, University of Pretoria, Pretoria, South Africa

Background
Satellite remote sensing and climatic data are increasingly being used in modelling and mitigating health risks. They impact transmissibility on a range of vector-borne diseases, such as malaria and diarrhoea and the, undermine nutrition through harmful impacts on food production and concomitant increases in food prices. However, challenges remain in the utility and robustness of the data and models that can be used in characterizing and monitoring the spatial and temporal patterns of diseases in many resource limited countries. In this study, spatial models are used to estimate and map epidemiological risks of fever, which

is an indicator of malaria and diarrhoea in Malawi and Zambia, two of the malaria prevalent sub-Saharan African countries using satellite remote sensing and climatic data, adjusting for poverty.

Methods
Data on childhood fever and diarrhoea were obtained from two recent Demographic and Health Surveys in each country; Zambia (2007 and 2013) and Malawi (2010 and 2015-16). Remote sensed and climatic data were obtained from satellite image site. A Bayesian multivariate spatial mode is fitted to the childhood fever and diarrhoea with exposure measurement errors to estimate diseases at appropriate administrative level in the two countries.

Results
Childhood fever and diarrhoea showed considerable variation across small areas in Zambia and Malawi. There were indicative associations between environmental, and climatic factors and poverty on childhood fever and diarrhoea.

Conclusion
This study has described methods that make use of satellite remote sensing and climatic data to better understand childhood disease outcomes collected in Demographic and Health Surveys, a rich source of health data in the region. By bringing newly developed multivariate spatial methods to multiple sources of datasets on exposure and human health outcomes, this study has shown how new Biostatistics Methods could be used to better understanding human health and associated influences.

Incorporating Sampling Design in Spatial Estimation of Small Area Health Outcomes Sub Saharan Africa

'Sheyla Cassy [2]Samuel Manda, [3]Filipe Marques
'DMI' Faculdade de Ciências' Universidade Eduardo Mondlane' Maputo' Mozambique [2]Biostatistics Research Unit' South Africa Medical Research Council' Pretoria' South Africa, [3]CMA' DM' Faculdade de Ciências e Tecnologia' Universidade Nova de Lisboa' Caparica' Portugal

Background
Several studies in the Sub Saharan Africa (SSA) use nationally representative household and population surveys for spatially smoothed estimation and prediction of health outcomes at subnational levels. These surveys often employ complex sampling survey designs that include multistage sampling and stratification, with disproportionate allocation of units. Most studies have accounted for the clustering effects in the analyses. However, few have considered sampling weights in the spatial smoothing of single health outcomes.

Methods
In this study, we considered multivariate spatial smoothing methods that integrate sampling weights for small areas estimation of multiple health outcomes. This was done under Bayesian Hierarchical Spatial Smoothing Models, using different approaches for adjusting survey weights We used the derived methods to estimate the geographical distribution of stunting, wasting and underweight among children under five at the district level based on the 2015 16 Malawi Demographic and Health Survey, and childhood malaria and diarrhea at the district level in Mozambique using 2015 Immunisation, Malaria and HIV/AIDS Key Indicator Survey (IMASIDA). The models were fitted within the R environment using the available packages.

Results
There was great variation in the prevalence of malaria and fever between districts in Mozambique and in the prevalence of stunting, wasting and underweight across districts in Malawi. The estimated spatial variations did not change much according to the sampling design adjustment methods.
Conclusion
This study incorporated sampling design in spatial estimation of small area health outcomes using available data sources from Malawi and Mozambique to provide precise and reliable estimates and produce appropriate statistical methods and models for this complex surveys. The results obtained were not significantly different for the different sampling design methods.

Dietary patterns and linkages to Non-Communicable Diseases in Namibia

Laina T. Mbongo
UNAM

The role of diet in promoting health and preventing disease is difficult to explain due to its complex network of foods and nutrients. Nutrition is essential to maintain life, but dietary composition can differ widely among individuals. Traditionally, analysis in nutritional epidemiology typically examined diseases in relation to a single or a few nutrients or foods. However, people do not eat isolated nutrients. Instead, they eat meals consisting of a variety of foods with complex combinations of nutrients. The high degree of inter-correlation among nutrients as well as among foods makes it difficult to attribute effects to single dietary components. The main objective of this study it to quantify the kind of diets people consume by establish dietary patterns and how the patterns relate to health outcomes. Two major approaches have long been applied to investigate and derive dietary patterns namely the investigator driven- a priori/hypothesis oriented approach and the data driven/exploratory- a posteriori (Castro, et al. 2016). Thus, a number of data reduction methods will be explored to determine the patterns and how they are associated with non-communicable diseases. The Namibia Household and Income Expenditure Survey (2015/2016) will be used to achieve the objectives of this study.

Cross sectional and spatial analyses of self-reported Tuberculosis in South Africa: results from SABSSM 2017

Inbarani Naidoo, Njeri Wabiri, Adlai Davids, Musa Mabaso and Khangelani Zuma
Human Sciences Research Council

Background
South Africa represents a challenging high burden tuberculosis (TB) country with TB being a notifiable disease across all provinces since 1919. Prompt detection, diagnosis and treatment of TB cases are part of infection prevention and control. The focus is to track every confirmed TB case and to identify transmission hotspots as they change over time. To this end, spatial approaches have been identified as an important tool for defining infectious disease distribution at different geographic scales and for effective disease control efforts. Studies show that TB distribution is heterogeneous due to its transmission dynamics at household and community levels. Therefore spatial and temporal scale is an important consideration in describing the distribution of TB. In South Africa studies have shown that the spatial distribution of TB and resistant TB is heterogeneously distributed within province level.

The mapped national profile of TB prevalence shows that the Northern Cape, Western Cape, North West and Eastern Cape are provinces with the highest TB prevalence (> 400 cases per 100 000). KwaZulu-Natal, Mpumalanga and Free State represent provinces with medium TB prevalence (274 cases per 100 1000). Gauteng and Limpopo have low TB prevalence with 178 cases per 100 000 (figure 1).

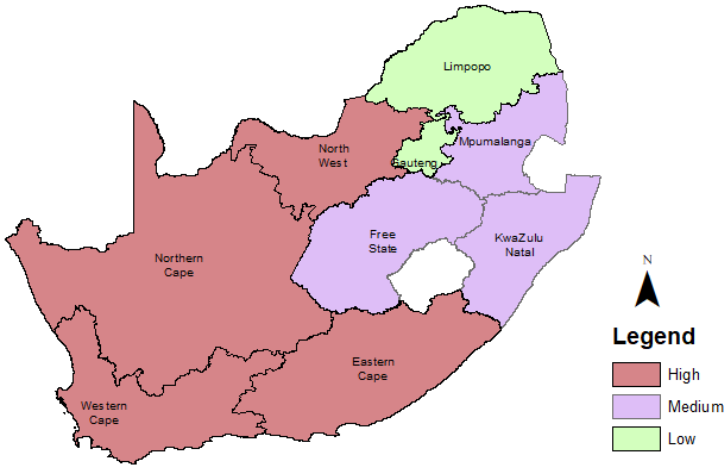

Figure 1: TB prevalence profile in South Africa

Spatial models rely on geographically referenced data linked to TB cases at different spatial scales, some of which are collected routinely in the national health laboratory database (NHLS). These routine data then need to be linked to patient identifiers including their geographic location information (place of residence). Hence the data require processing to enable spatial mapping. Attribute data associated with TB risk is also needed to profile TB hotspots in the country. Other authors have demonstrated the utility of linking qualitative data to TB transmission dynamics at community level. The aim of the study is to analyse the occurrence of self-reported TB diagnosis in South Africa at national, provincial and district level, drawn from the SABSSM 2017 survey with a view to contribute to the current knowledge of TB spatial mapping and epidemiology.

Methods

This is a secondary analysis of self-reported TB data collected during the 2017 South African national HIV population-based cross sectional household survey (SABSSM). A multi-stage disproportionate, stratified sampling approach was used and the survey population included all people living in South African households over the period 2016-2017. All analyses were performed in Stata SE version 15 using weighted data. The primary outcome was national level self-reports of a TB diagnosis among adults aged ≥ 15 years. Recent diagnosis was defined as being diagnosed in the previous 12 months. Respondents were asked whether or not they had ever been told by a doctor or other health professional that they had TB (yes/no/don't know). Preliminary univariate results are reported here with significance p= 0.05% and 95% confidence intervals. Further spatial analyses will be undertaken using vector and raster methods at different spatial scales within a geographical information system.

Results

Approximately 5.6% (n=39,885) of respondents indicated they had a TB diagnosis and among these 12.0% were recently diagnosed. Using ever diagnosed with TB as the outcome, statistically significant associations (p<0.001) of TB with race, age, employment status and education were found. The following highest ever diagnosed TB rates were found among Black Africans (86.3%), those aged 25-49 years age groups (63.6%)),the unemployed (71.2%) and those having secondary school education

(66.5%). There were also significant associations ($p<0.001$) based on localities, with urban areas (66.4%) having more self-reported TB compared to rural/traditional areas (29.3%) and farms (4.3%). The highest proportions of self-reported TB was evident in Eastern Cape (20.9%), Gauteng (19.1%) and KwaZulu-Natal (18.0%). The proportions of self-reported TB at district level ranged from 9.4% in OR Tambo district to 0.1% in Xhariep.

Conclusions

Univariate analyses showed significant provincial and district level differences for self-reported TB cases. This suggests that these data drawn from a cross sectional survey, lends itself to spatial analyses with selected covariates, which will be explored and presented for this conference paper and could add to better understanding the epidemiology of TB in South Africa.

STATISTICAL METHODS FOR THE ANALYSIS OF GENOMIC DATA

Chair and organiser: Prof Bernard Omolo

In this session, we will discuss Bayesian and non-Bayesian methods for the analysis of microarray, sequence and GWAS data. Emphasis will be placed on variable selection and the subsequent inferences. Data on malaria, coronary heart disease, cancer and Dengue hemorrhagic fever phenotypes will be used for illustrations.

A Model-based Approach to Genetic Association Testing in Malaria Studies

Bernard Omolo[1], Morine Akoth[2], and John Odhiambo[2]
[1]University of South Carolina – Upstate, Spartanburg, USA
[2]Strathmore University, Nairobi, Kenya

In human genetics, heterozygote advantage (heterosis) has been detected in studies that focused on specific genes, but not in genome-wide association studies (GWAS). For example, heterosis is believed to confer resistance to certain strains of malaria in patients heterozygous for the sickle-cell gene HbS. Yet the power of allele-based tests can be substantially diminished by heterosis. Since GWAS (and haplotype-associations) also utilize allele-based tests, it is unclear to what degree GWAS could underachieve because heterosis is ignored. In this study, we propose a two-step approach to genetic association testing in malaria studies in a GWAS setting that may enhance the power of the tests, by identifying the underlying genetic model first before applying the association tests. We fit generalized linear models for the dominant, recessive, additive and heterotic effects and perform tests of significance using the MAX and the allelic tests, noting the minimum p-values across all the models and the proportion of tests that a given genetic model was deemed the best, using simulated data. Case-control genotype data on malaria from Kenya and the Gambia are used for validation. Results show that the allelic test returned a number of false negatives under the heterosis model, suggesting reduced power in testing genetic association. Thus, GWAS and haplotype associations should be treated with caution, unless the underlying genetic model had been determined.
Keywords: allelic test; case-control study; genome-wide association; malaria; maximum test

An Integrated RNA and DNA Molecular Signature for Colorectal Cancer Classification

Bernard Omolo[1], Mohanad Mohammed[2] and Henry Mwambi[2]
[1]University of South Carolina – Upstate, Spartanburg, USA, [2]University of KwaZulu-Natal, Pietermaritzburg, South Africa

Colorectal cancer (CRC) is the third most common cancer among women and men in the USA. The KRAS gene is mutated in 40% of the CRC cases and hence the RAS pathway activation has become a major focus of drug targeting efforts. However, nearly 60% of patients with wild-type KRAS fail to respond to RAS-targeted therapies, for example the anti-epithelial growth factor receptor inhibitor (EGFRi) combination therapies. Thus, there is a need to develop more reliable molecular signatures to better predict mutation status. In this study, we develop a hybrid (DNA mutation and RNA expression) signature and assess its predictive properties for the mutation status of CRC patients.
Publicly-available microarray and RNA-Seq data from 54 matched formalin-fixed paraffin embedded (FFPE) samples from the Affymetrix GeneChip and RNA-Seq platforms, were used to obtain differentially expressed genes between mutant and wild-type samples. For classification, the support-vector machines,

artificial neural networks, random forests, k-nearest neighbors and the naïve Bayes algorithms were employed. Compared to the genelist from each of the platforms, the hybrid genelist had the highest accuracy, sensitivity, specificity and AUC for mutation status and could therefore be useful in clinical practice, especially for colorectal cancer diagnosis and therapeutics.
Keywords: colorectal cancer, FFPE, microarray, RAS pathway signature, RNA-Seq

protGear: A tool for protein microarray data pre processing

'Kennedy Mwai Wambui [2]Nelson Kibinge, [3]Samson Kinyanjui, [4]Faith Osier, [5]Eustatsius Musenge
'University of the Witwatersrand [2]KEMRI Wellcome Trust, [3]KEMRI Wellcome Trust, [4]University of Heildelberg, [5]University of WItwatersrand

Background

Protein microarray technology is increasingly being used for antigen discovery in experiments where thousands of samples are analysed. However, data generated using this approach has been shown to contain systematic and non systematic sources of bias. Careful pre processing is essential to eliminate this technical bias while keeping the biological variation. This bias has been attributed to nonbiological sources, which is introduced by small variations in the experimental conditions during the experiment process. Unlike DNA microarray technology, however, protein microarrays have found more use just recently and present unique sets of data challenges.Guidelines for pre processing the protein microarray data before analysis, therefore, continue to evolve. Pre processing involves; background correction, quantifying within sample variation, batch correction4, normalisation among other steps. Background correction is the statistical process of excluding the background noise that may attach to the array spots. Normalisation helps to adjust for any bias which arises from variation in the microarray technology rather than from biological differences. In this project, we develop a generic pre processing tool for protein microarrays in R software with statistical functions for the above procedures.

Methods

The tool has been developed with different statistical functions to perform the pre processing data procedures which are organized into an R package. Our package implements procedures for quantifying within sample variation, background correction, batch effect removal and normalization. A function implementing Coefficient of Variation (CV) and pooled estimate of variance (PEV) for the technical repeats is used to estimate the within sample variation. Additionally, have included functions to examine within sample variability based on the CV filtering for technical repeat spots. CV based filtering provides an algorithm for the selection of less variable spots among technical repeat.Background fluorescence can be attributed to within slide or spot specific artefacts; thus, correction minimizes the resultant noise. protGear implements five different techniques for background correction. The techniques have been adopted from the Linear Models for Microarray Data (Limma). Data from these experiments have been shown to have a mean variance dependence that may mask the true biological variability. To assist in minimising this problem, different normalization and batch correction techniques have been adopted. On our package, we have included but not limited to robust linear (RLM) regression normalisation assumes the bias is linear dependent with the estimates; local regression normalisation assumes a non linear dependency; hyperbolic sine normalisation asinh and Variance Stabilization Normalisation (VSN) adopted from a DNA microarray approach that utilises parametric transformations and maximum likelihood approaches and log2 normalisation. These transformations have been incorporated with batch correction techniques using (ComBat) empirical Bayesian framework from SVA package for removing batch effects.

Results

The package is on the final phases of development and most of the functions have been tested using data from KILchip v1.0: A Novel Plasmodium falciparum Merozoite Protein Microarray to Facilitate Malaria Vaccine Candidate Prioritization6. The methods were evaluated in their ability to decrease variation among technical repeats, reduce the mean variance dependence of the measured intensities which is caused by technical errors and still maintain the biological difference. The CV based filtering of technical repeats significantly reduces the variation among the technical replicates. A CV of 20% was considered the cut off for agreeable technical repeats. A subtraction of the local background was selected as the best approach to handle the background noise for the KILchip v1.0 data after comparing with the other approaches. The subtraction of the local background approach was selected since it did not produce negative intensities as per the design of the experiment. The different normalisation techniques and batch correction were run on the KILchip v1.0 data. ComBat. The different techniques significantly reduced the mean variance dependence on the data. The package implements the pooled median absolute deviation (PMAD) and the pooled estimate of variance (PEV) together with graphical representation to check on the performance of the different methods. On top of ComBat batch correction, we incorporated a within function batch correction by including the batch parameter in the normalisation model for VSN and RLM normalization. VSN+Combat normalisation approaches were selected for the KILchip v1.0 array data.

Conclusion

protGear provides a rational data pre processing framework and platform for exploratory and visualization to facilitate separation of technical variation from biological variation. Most of the available open source packages have been implemented for DNA but there are limited standard pre processing guidelines for protein microarray data. Additionally, this tool allows the users to compare the performance of the different normalization techniques utilised in different protein micro array projects and select the suitable for their data. Logarithmic normalisation using a base n was less effective with a dataset with outliers which was the case with the KILChip v1.0 data2. Additionally, subtraction of purification tag intensities for lab experiments generated negative intensities which caused convergence problem for RLM, Loess and log transformations. We are working on approaches to help handle this statistically on our package. Finally, this tool is a reusable package developed in the R statistical programming language. It provides a unified platform for pre processing protein and peptide array data. It has been developed in such a way that other users can contribute in development after it has been published

A Comparison of Cancer Classification Methods based on Microarray Data

[1]Mohanad Mohammed [2]Henry Mwambi, [3]Bernard Omolo
[1]University of KwaZulu Natal [2]University of KwaZulu Natal, [3]University of South Carolina Upstate

Background

Cancer is among the leading causes of death in both developed and developing countries. Through gene expression profiling of tumors, the accuracy of cancer classification has been enhanced, leading to correct diagnoses and the application of effective therapies.

Methods

Here, we discuss a comparative review of the predictive ability of seven classification methods (support vector machines, with the radial basis kernel (SVM(RK)), linear kernel (SVM(LK)) and the polynomial kernel

(SVM(PK)), artificial neural networks (ANN), random forests (RF), k nearest neighbor (KNN), and naive Bayes (NB)), using publicly available gene expression data from cancer research.

Results

Results indicate that NB outperformed the other methods in terms of the accuracy, sensitivity, specificity, kappa coefficient, area under the curve (AUC), and balanced error rate (BER) of the binary classifier. We recommend NB as the \gold standard" for cancer classification using microarray data.

Conclusion

We used microarray data for classification of cancer. In the future research we are going to study the RNA Seq data which is discrete in nature and widely used for classification using discrete classification methods such as negative binomial linear discriminant analysis (NBLDA) and Poisson linear discriminant analysis (PLDA).

THE >eR-BIOSTAT: A NEW E-LEARNING SYSTEM FOR BIOSTATISTICS

Organisers: Ziv Shkedy and Khangelani Zuma

One of the main problems in high education at both under graduate and master levels in developing countries is the lack of high quality, R based, materials for courses in education programs. The >eR-Biostat initiative (https:// er-biostat.github.io/Courses) is focused on education programs (both undergraduate and master) in Biostatistics/ Statistics and for non-statisticians and aim to develop new E-learning system for courses at different education level. We believe that accessibility to free high quality education materials is crucial to ensure a high standard in education. We offer an "open sources", R based, education materials in statistics. Everybody (teachers and students) can download and use the courses for free. The >eR-Biostat initiative introduces a new, R based, learning system, the multi-module learning system, in which the students in the local universities in developing countries are be able to follow courses in different learning format, including e-courses taken online and a combination between e-courses and local lectures given by local staff members. R software and packages are used in all courses as data analysis tool for all examples and illustrations. The >eR-Biostat initiative provides a free, accessible and ready to use tool for capacity building in biostatistics/statistics for local universities in developing countries with current low or near zero capacity in these topics.

The E learning system for generalized linear models (GLM)

'Khangelani Zuma [2]Ziv Shkedy
'Human Sciences Research Council [2]University of Hasselt

Background

Generalized Linear models (GLMs) are a central tool in data analysis in any scientific discipline and in particular in biostatistics applications.

Methods

We present a new generalized linear models course which was developed as a part of the >eR Biostat Initiative.

Results

The generalized linear models course provides online materials for master students in biostatistics/statistics. The materials developed for the GLM course are R oriented and publicly available. Several types of course materials are presented: (1) slides for the course, (2) Ready to use R programs, which contain all data and related codes for all the examples and illustrations discussed in the course and (3) YouTube tutorials.

Conclusion

We present the different types of course materials for a specific module in the GLM course (parameters estimation for GLM) and present all the materials for the course which are currently available online in the website of the >eR Biostat initiative

Development of a Master program using the >eR Biostat

[1]Tadesse Awoke Ayele [2]Ziv Shkedy , [3]Khangelani Zuma
[1]University of Gondar [2]Hasselt University , [3]HSRC, South Africa

Background
We discuss the setup of the master program in Biostatistics and Epidemiology, established in 2015, in Gondar University, Ethiopia, and illustrate how the >eR Biostat initiative was used to support the master program.

Methods
The web page has been developed and the courses are uploaded together with the dataset and R programs.

Results
We discuss the curriculum stricture and show how courses in the master program can be linked to the >eR Biostat website. Further, we show how introductory courses, offer to the students in order to improve their knowledge in Biostatistics as recreation to master curses given as a part of the program, are linked to the >eR Biostat online curriculum.

Conclusion
The program has been used by students and instructors from different universities in Ethiopia.

Using R markdown to develop E learning capacity

[1]Olajumoke Evangelina Owokotomo [2]Ziv Shkedy
[1]Universiteit Hasselt, Hasselt, Belgium [2]Universiteit Hasselt

The >eR Biostat initiative is based on the underlying approach of "community based education development" in which different members of the community developed different courses (often independently). We discuss the R markdown platform as a tool for the development of new educational materials that can be added easily to the >eR Biostat website. In this way, future contributors to the > >eR Biostat initiative can add their own materials to the >eR Biostat curriculum and still keep the same presentation style for the courses.

An introduction to the >eR Biostat E learning system

[1]Ziv Shkedy
[1]Hasselt University

One of the main problems in high education in developing countries is the lack of high quality course materials for courses in education programs (at all levels). The >eR Biostat initiative is focused on education programs in Biostatistics/Statistics and aim to develop new E learning system for courses at different education levels (i.e under graduate and master levels). The E learning system, developed as a part of the >eR Biostat initiative, offers free online course materials for students and teachers in developing countries. For each course, the materials are publicly available and consist of several type of course materials: (1) notes for the course, (2) slides for the course, and (3) R programs, ready to use, which

contain all data and R code for the all examples illustrations discussed in the course and (4) YouTube tutorials. The >eR Biostat initiative introduces a new, R based, learning system, the multi module learning system, in which the students in the local universities in developing countries will be able to follow courses in different learning format, including e courses taken online and a combination between e courses and local lectures given by local staff members. The >eR Biostat is developed for both statisticians and non statisticians that need to use statistical methods for their research.The >eR Biostat initiative provides a free, accessible and ready to use tool for capacity building in biostatistics/statistics for local universities in developing countries with current low or near zero capacity in these topics.

TIME-TO-EVENT ANALYSIS USING POPULATION-BASED DATA

Chair and organiser: Prof Bernard Rachet

Time-to-event ('survival') analysis provides valuable information on the disease of interest, at population or individual level, depending on the metric chosen. The estimation method varies according to that metric. Furthermore, analyses using population-based data require specific methods to account for the competing risks of death because of the lack of reliable information on the cause of death. Methods developed for relative survival data setting are employed, instead of the more conventional approaches within the cause-specific setting. Concepts and estimation methods used in relative survival setting will be described and their use and challenges will be illustrated through different applications.

Measures in population-based cancer survival: overview and estimation with a focus on net survival

Bernard Rachet
Cancer Survival Group, Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, United Kingdom

Crucial conceptual clarification has occurred in the population-based cancer survival field in the last decade, in parallel with important methodological development.
I will first define various measures of survival (including crude mortality, net survival, number of life years lost). These measures describe survival data either on a probability scale or on a time scale. The clinical or population-health purpose of each measure is described, and their advantages and drawbacks discussed. Two data settings (cause-specific and relative survival settings) will be presented and their challenges for the estimation of the measures will be detailed.
I will then focus on how net survival is estimated within the relative survival data setting. The non-parametric Pohar Perme estimator will be explained. Net survival can also be estimated from an excess hazard model. Flexible excess hazard models will be described though it presents some methodological challenges. Advantages and limitations of both estimation approaches will be discussed.

Direct modelling of alternative survival indicators using pseudo-observations in the relative survival setting

Dimitra-Kleio Kipourou,[1] Maja Pohar Perme,[2] Bernard Rachet,[1] Aurélien Belot[1]
[1]Cancer Survival Group, Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, United Kingdom
[2]Faculty of Medicine, Institute for Biostatistics and Medical Informatics, University of Ljubljana, Slovenia

Background
In population-based cancer studies, net survival is a crucial measure for population comparison purposes. Alternatively, the crude probability of death and number of life years lost due to death according to different causes, represent complementary measures for reflecting different dimensions in terms of prognosis, treatment choice, or development of a control strategy.
Objective: When the cause-of-death information is available, both measures can be estimated and modelled in the competing risks setting using hazard-based regression models or the pseudo-observation approach. We extended the pseudo-observation approach in order to model directly these measures when the cause of death information is unavailable or unreliable (i.e. in relative survival setting).

Method
In a simulation study we assessed the performance of the proposed approach in estimating the regression parameters and examined models with different link functions that can provide an easier interpretation of the parameters. We also illustrated the proposed method using cervical cancer data from the England population-based cancer registry.

Results
Our approach based on pseudo-observations in relative survival setting demonstrated nice frequentist properties on estimating both measures regardless the choice of link function and by assuming a simple independent working covariance structure.

Conclusions
Pseudo-observations can be used as an alternative to standard methods as they provide covariate effects directly affecting the indicators of interest in the relative survival setting, allowing for an easier interpretation when using the identity or log link function. However, the method might be prone to a longer computational time compared to other methods.

Statistical methods to link Demographic and Health Survey data with census data

[1]Ruth Vellemu [2]James Chirombo
[1]University of Malawi [2]Malawi Liverpool Wellcome Trust

Background
Demographic and health surveys (DHS) are large nationally representative surveys held every five years to provide estimates on health and demographic outcomes such as child health, malaria prevalence, HIV prevalence, and nutrition among others at the district level. These estimates are widely used by different programmes for monitoring and evaluation purpose. Typically, DHS surveys have sample sizes between 5,000 30,000 households. One of the general challenges facing national surveys is declining response rates which may affect the reliability of estimates. In a Malawian setting, it is also not possible to obtain estimates at sub district level as the DHS is powered to provide district estimates. To obtain estimates for subpopulations of interest, it may be necessary to combine the information from surveys with other data sources with wider coverage and higher spatial resolution such as population and housing census (PHC) data. A combined dataset can provide comprehensive estimates of useful health indicators at a very fine spatial scale for improved decision making at the local level.

Methods
Bayesian Hierarchical models for geostatistical and areal data were used to combine different data sources to provide small area estimates of key health indicators at sub district level in Malawi. A final model, implemented in a Bayesian framework with local area effects, was then developed to generate estimates at sub district level.

Results
Results show that combining survey with census data provides robust estimates at high spatial resolution compared to when using only one dataset.

Conclusion

The project improves our understanding of methods used in combining information from different sources and provides a means for performing small area estimation of population parameters of interest to different researchers. Estimates of health indicators at high spatial resolution is an important undertaking important for disease monitoring and surveillance purposes.

Age, Period and Cohort Analysis of young adult Mortality due to HIV and TB in South Africa: 1997 2015

'Tshifhiwa Nkwenika [2]Samuel Manda, [3]samuel Manda
'Biostatistics Unit' South African Medical Research Council' Pretoria' South Africa. [2]Biostatistics Unit' South African Medical Research Council' Pretoria' South Africa., [3]Department of Statistics' University of Pretoria' Pretoria' South Africa

Background
Young adult mortality is very important in South Africa with the impact of Human Immunodeficiency Virus /Acquired Immune deficiency Syndrome (HIV/AIDS), Tuberculosis (TB), injuries and emerging non communicable diseases (NCDs). Investigation of temporal trends for adult mortality associated with TB and HIV is often based on age, gender and birth cohort separately. This has hindered an insightful interpretation of the mortality trends and effects by age, period, and birth cohort simultaneously. The overall aim of this study was to estimate age effect across period and birth cohort; period effect across age and birth cohort; and birth cohort effect across age and period.

Methods
Mortality data and mid population estimates were obtained from Statistics South Africa for the period 1997 to 2015. Three year age, period and birth cohort intervals for 15 64 years, 1997 2015 and 1934 2000 respectively were used. Age Period Cohort (APC) analysis using the Poisson distribution was used to compute effects of age, period and cohort.

Results
The results showed a concave down association between age and TB mortality, with a peak at 36 38 years. There was a concave down relationship between TB cause specific mortality in the studies mortality data period between 1997 and 2015. There was a downward trend between TB mortality and the effect of birth cohort from 1934 to 2000. There was an inverse flatter U shaped association between age and HIV mortality, and was more pronounced at 36 38 years. The estimated relative risks showed approximately linear relationship between HIV mortality and effect of period from 1997 to 2015. An inverted U shape relationship between birth cohort and HIV related mortality with a slight increase at later birth cohort was observed.

Conclusion
Despite the limitations of official published mortality data, we still found a notable age and period effect according to APC analysis, which were similar to the results of previous studies. However, the pattern of period effect associated with HIV mortality differed from the one published by Statistics South Africa. Age Period Cohort Model of HIV and TB mortality offers a more robust assessment of effect of age, period and birth cohort, which would not be possible using traditional Poisson regression model on the death counts separately.

Oral List

NO	PRESENTING AUTHOR	ABSTRACT TITLE
1	Nada Abdelatif	Climatic changes and Tsetse Fly Distributions in Kwazulu-Natal, South Africa: A spatio-temporal model
2	Jonathan Abel	Mortality rates, associated risk factors and causes of death in older children with neurological impairments in rural Kenya: A cohort study
3	Thomas Achia	Copula bivariate based semiparametric logit models: Application to the joint modeling of psychoactive drugs and HIV risk among Fisherfolk in Kenya
4	Thomas Achia	Intimate partner violence assessment: A comparative analysis of the standard logit model versus copula bivariate semiparametric logit in determining the factors associated with intimate partner violence from Kenya demographic and health survey in 2014
5	Schadrac Agbla	Estimation of individual-level local average treatment effects in cluster randomised trials with non-adherence
6	Smartson Ainomugisha	Health spending and Economic growth nexus: Evidence from Uganda
7	Omololu Aluko	Efficient methodologies for handling missing data for longitudinal ordinal outcome
8	Godwin Anguzu	Changes in socioeconomic status among HIV patients after 10-years of antiretroviral therapy in Uganda: An estimation of the within-subject correlations
9	Henry Athiany	Risk of hepatic injury among adult patients concomitantly on statin and statin interacting drugs: A self-controlled case-series analysis using an electronic health database
10	Kassahun Abere Ayalew	Skewed random effects distribution in Conditionally Autoregressive Spatial models for estimating HIV prevalence at local level in South Africa
11	Birhanu Ayele	Title: Progression of cisplatin-associated ototoxicity amongst patients receiving cancer chemotherapy: Application of Tobit Regression
12	Yusentha Balakrishna	Developing and validating a scoring tool for mortality among neonates with oesophageal atresia and tracheoesophageal fistula at a tertiary level hospital in KwaZulu-Natal, South Africa
13	Victoria Barabona	Time to HIV serostatus disclosure, pattern and factors among pregnant women living with HIV in Moshi urban
14	Tefo Baraki	Bayesian copula-based analysis of joint model for multi-variable longitudinal and recurrent events
15	Jesca Mercy Batidzirai	Prediction of family state occupancy in rural South Africa using multistate transition modeling
16	Diribsa T Bedada	Analysis of climatic variability and its effects on production of selected crops in Ada'a District, Ethiopia: Multivariate time series approach
17	Stefano Bonnini	Permutation multivariate tests for treatment effect: Theory and recent developments
18	Vitumbiko Chijere Chirwa	Malaria morbidity in under-five children in Malawi: A Bayesian spatial-temporal Model
19	Lily Clements	Using Software in Teaching Medical Data
20	Dominique Laurent Couturier	Right-censoring bias correction for growth curve linear mixed models

21	Timotheus Darikwa	Statistical measures for multivariate spatial autocorrelation
22	Legesse Kassa Debusho	A semi-parametric mixed models for longitudinally measured fasting blood sugar level of adult diabetic patients
23	Zelalem Dessie	Joint multistate modelling of viral load dynamics and multistate CD4 count progression
24	Moleen Dzikiti nee Zunza	Exploring association between predominant breastfeeding and infection-related hospitalization, over time: An empirical comparison of models to account for clustering
25	Haile Mekonnen Fenta	Hierarchical mixed effects model for type 2 diabetes
26	Lineekela Gabriel	Bayesian dynamic models for time-varying outcomes: Application to a patient cohort on ART
27	Susan Gachau	Multiple imputation of clinician-level covariates applied in modelling quality of clinician-prescribed care: Sensitivity analysis of departure from Missing at Random (MAR) assumption
28	Cyrus Githunguri	Effect of different potassium fertilizer rates and liming on recommended maize yield grown in western Kenya
29	Cyrus Githunguri	Effect of different zinc fertilizer rates and cropping systems on recommended maize and beans grown in western Kenya
30	Nomonde Gwebushe	A review on disease mapping and modelling of cardiovascular diseases in South Africa
31	Stella May Gwini	Linear mixed models with time-varying covariates: Application of disaggregation of within-subject and between-subject effects
32	Patrick Hannan	Multi-state models for the analysis of Wheezing episodes in a birth cohort of African children
33	Christian Hansen	Bias estimation and sensitivity analysis in questionnaire studies with "embarrassing" questions
34	Gbememali Castro Hounmenou	Evaluation of the robustness of imputation methods combined to backpropagation algorithm in frame of multiple non linear regression
35	Charl Janse van Rensburg	Modelling recurrent events data for hypertension and diabetes control in a peri-urban area in South Africa
36	Ameet Jha	A comparative study of palmar and digital dermatoglyphic patterns among type-II diabetic & non-diabetic adults: A meta-analysis
37	Esme Jordaan	Auxiliary variables in Latent Profile Analysis (LPA) and methods to estimate a distal outcome model
38	Priscilla Kapombe	Comparison of Antimicrobial Resistant Neisseria gonorrhea to Ceftriaxone and Ciprofloxacin using Proportional Odds Model for patients seen at the University Teaching Hospital in Lusaka, Zambia
39	Alexander Kasyoki	On the correct modeling of the association between an outcome and a mismeasured covariate in clustered cross-sectional surveys: A simulation-based study
40	Jessie Khaki	Discrete time hazard parameterization for child survival in Malawi
41	Jessie Khaki	Spatial and temporal patterns of acute febrile illness and confirmed typhoid fever infection and their associated correlates in Ndirande, Blantyre, Malawi: An analysis of STRATAA data
42	Ummi Abdul Kibondo	Hematological reference intervals for children in Bagamoyo District, Tanzania
43	Peter Koleoso	Modelling cure fractions for complex hazard functions in a cancer study
44	Evaristar Kudowa	Solutions for selective loss to follow-up in HIV-Cancer cohorts in Malawi

45	Thabo Lephoto	Bayesian modelling of tick count data over time and space using discrete model approaches
46	Maia Lesosky	Temporal interactions of microbiota in longitudinal nasopharyngeal samples
47	Lawrence Lubyayi	A pairwise joint modelling approach for multivariate longitudinal immuno-epidemiological data from the Infant BCG study in Entebbe, Uganda
48	Michael Johnson Mahande	Mortality rate and associated factors among preterm babies born in Moshi Municipality in northern Tanzania
49	Lyness Matizirofa	Modelling predictors of stroke disease in South Africa: Bayesian binary quantile regression approach
50	Mikateko Mazinu	The association between childhood environmental exposures and the subsequent development of Crohn's disease in the Western Cape, South Africa
51	Hope Mbachu	Analysis of a vaccination mathematical model of an infectious measles disease
52	Innocent Mboya	Predictors of perinatal deaths in the presence of missing data: A registry based study in northern Tanzania
53	Nobuhle Mchunu	Modelling CD4 Count and mortality in a cohort of patients Initiated on HAART
54	Shibe Mhlongo	Modelling the effect of rape on mental health status one year post enrollment: RICE study
55	Sakhile Mnguni	The impact of correlated socio-demographic factors in Spatial profile of Diabetes in Africa with Missingness
56	Tshwaraganang Modise	The effect of cotrimoxazole prophylaxis on CD4 cell-count profiles in HIV-positive patients stabilized on anti-retroviral therapy: Analysis of data from the COSTOP trial
57	Mohanad Mohammed	A comparison of cancer classification methods based on microarray data
58	Kesaobaka Molebatsi	Cost-effective, cross-sectional cohort estimation of HIV incidence rate in presence of misclassification error using doubly robust estimation
59	Neema Mosha	Analytical methods used in handling missing data in estimating prevalence of HIV/AIDS for demographic and cross-sectional surveys: A systematic review.
60	Rejoice Msiska	Impact of improved drinking water source and sanitation, and urban residence on child diarrhoea in Sub-Saharan Africa. A meta-analysis of demographic and health surveys
61	Paul Mubiri	Prediction models for newborn complications at birth: comparing multinomial logit models and supervised machine learning algorithm Naïve-Bayes classification methods
62	Caroline Mugo	Modeling proportion of infected tsetse flies over time in Kajiado and Narok districts, Kenya
63	Abdul Khalie Muhammad	Developing Socio-Economic Status (SES) index for Basse region in The Gambia
64	Martin Mujuni Rwabilimbo	Time interval to modern contraceptive use following child birth among reproductive women in Tanzania: Evidence from Tanzania Demographic Health Survey2015/16
65	Elton Mukonda	Mixed method estimation of population HIV viral suppression in the Western Cape, South Africa

66	Rachid Muleia	Spatial distribution of HIV prevalence among young people in Mozambique
67	Eustasius Musenge	Geospatial Computing of large scale spatial data: Faster computation of the dense spatial matrix
68	Edwin Musheiguza	Inequalities in stunting among under five children in Tanzania: Decomposing the concentration indexes using demographic health surveys from 2004/5 - 2015/6
69	Portia Mutevedzi	Feasibility of employing systematic random cluster selection with probability proportional to estimated size and without replacement, to obtain a nationally representative sample of 9204 clusters for estimating vaccination coverage in South Africa
70	Paul Mwaniki	Improving chest x-ray classification using transfer learning
71	Edson Mwebesa	Evaluating the effect of sampling weights on the predictors of contraceptive use in Uganda using 2016 UDHS data
72	Emmanuel Kweku Nakua	Modelling the force of infection for hepatitis B among heterogeneous groups reporting at Tertiary Hospital, Ghana
73	Thambeleni Nevhungoni	Semiparametric techniques for multilevel discrete survival data
74	Tshifhiwa Nkwenika	Age, period and cohort analysis of young adult mortality due to HIV and TB in South Africa: 1997-2015
75	Sheroline Nombasa Ntushelo	An exploratory analysis of multidimensional binary data using Correspondence Analysis, Non-metric multidimensional scaling and Cluster analysis as applied to Biolog EcoPlate data
76	Maureen Nwakuya	Investigation of three instrumental variable methods in the presence of errors in explanatory variables: A comparative analysis
77	Opeyemi Oyekola Ogungbola	Comparison of cox proportional hazard model and accelerated failure time model with application to data on Tuberculosis/HIV patients in Nigeria
78	Opeyemi Oyekola Ogungbola	Gender differential and social determinant of tuberculosis/HIV co-infected patients
79	Collins Okoyo	Mathematical modeling of the interruption of the transmission of soil transmitted helminths infections in Kenya
80	Abimibola Oladugba	Assessing the sensitivity and robustness of randomization test in analysis of repeated measures design with missing observations
81	John Olaomi	Modeling the stability and determinant factors of household food insecurity: A pair copula construction approach
82	Verrah Otiende	Spatiotemporal patterns of successful TB treatment outcomes among HIV co-infected patients in Kenya
83	Masego Otlaadisa	Robust estimation of single-index models with responses missing at random
84	Joseph Ouma	Combining data from national surveys with facility-based HIV testing to obtain more accurate estimate of HIV prevalence in districts in Uganda.
85	Lilian Owino	Fertility differential between Mombasa and Nairobi counties in Kenya
86	Emmanuel de-Graft Johnson Owusu-Ansah	Probabilistic modeling for an integrated temporary acquired immunity with norovirus epidemiological data
87	Oluwafemi Oyamakin	Some new nonlinear growth models for biological processes based on hyperbolic sine function
88	Arne Ring	Evaluations of error in variable regression methods for the analysis of ECG data

89	Masanja Robert	Predictors of non-completion to Isoniazid Preventive Therapy among People Living with HIV attended Care and Treatment Clinics in Dar-es Salaam from 2013 to 2017
90	Danielle Roberts	Risk factors and spatial heterogeneity of childhood anaemia in four sub-Saharan African countries
91	Wende Clarence Safari	Nonparametric cure rate estimation when cure is partially known
92	Arsene Brunelle Sandie	Superiority and non-inferiority hypothesis testing with functional data endpoints
93	Rudradev Sengupta	High dimensional surrogacy: Modeling and computational aspects
94	Ishen Seocharan	Digital data capturing: Using REDCap for a multi-arm, multi-site randomised clinical trial
95	Jasmit Shah	Bayesian approach in treating missing values with metabolomics
96	Isaac Singini	Diagnostics for a two-stage joint survival model
97	Ajibola Taiwo Soyinka	Order statistics approach to modeling and prediction of early mood swing
98	Halima Twabi	Causal inference for multiple outcomes for observational studies
99	Md Jamal Uddin	Do parents' mental disorders affect child behaviour via home environments and/or primary caregiver's personal and social performance? A mediation analysis
100	Edith Umeh	A two parameter rama distribution with applications
101	Marieta van der Rijst	Product profiling: Standard multivariate or multi-block statistical methods? An application to SA honeybush herbal tea
102	Ruth Vellemu	Statistical methods to link demographic and health survey data with census data
103	Steven Wambua	Random Forests application in missing data and predictive modelling for hierarchical routine clinical data: A case study of childhood pneumonia in Kenya
104	Mwai Wambui	ProtGear: A tool for protein microarray data pre-processing.
105	Mary Dancilla Wanjiru	Analysis of student-lecturer interactionin online PHT 112 course at Maseno University
106	Kennedy Wanyonyi	Comparison of random survival forests split rules in selecting the determinants of under-five mortality using 2014 Kenya DHS data.
107	Miriam Wathuo	Comparing coverage and precision of Newcombe confidence intervals for proportion differences generated using different weighting systems
108	Anteneh Yalew	Multilevel Logistic Regression for Patients' Satisfaction in Ethiopian Public Hospitals
109	Hao Zhang	Inform prior elicitation for Bayesian questionnaire validation using confirmatory factor analysis
110	Hao Zhang	The usefulness of prediction intervals in quantifying effect heterogeneity in randomized controlled trials

Climatic Changes and Tsetse Fly Distributions in Kwazulu-Natal, South Africa: A Spatio-temporal Model

[1]Nada Abdelatif, [1]Samuel Manda, [2]Kesh Govinder, [3]Chantel de Beer, [4]John Hargrove
[1]South African Medical Research Council, [2]University of KwaZulu-Natal, [3]Agricultural Research Council, [4]SACEMA,

Background
Tsetse flies are the vectors of sleeping sickness in humans and nagana in cattle in Africa. African trypanosomiasis adversely affects rural development and poverty alleviation in Sub Saharan Africa. This is due to a reduction in the supply of meat and milk and also animal draft power for crop production. Changing climate has and is resulting in increased adverse effects on health, such as changing patterns of infection. Tsetse flies are strongly affected by climatic conditions, where changes in climate alter the transmission seasons of the disease and geographic range of the tsetse fly itself.

Methods
A Bayesian spatiotemporal model was used to assess environmental and climatic factors, such as temperature, relative humidity and the normalized difference vegetation index, and to evaluate which of those significantly influence presence of tsetse flies. The model will then be used to provide probabilistic predictions of whether a change in the environmental and climatic factors will result in future increases in tsetse populations.

Results
This study will examine the important climatic and environmental factors that influence tsetse distribution in South Africa, and whether future changing climate will result in changes to the population.

Conclusion
This study is important in determining whether there is a need for a tsetse control programme in South Africa, since trypanosomiasis affects resource poor farmers in KwaZulu Natal.

Mortality Rates, Associated Risk Factors and Causes of Death in Older Children With Neurological Impairments in Rural Kenya: A Cohort Study

'Jonathan Abuga [2]Jonathan Abuga Abel , [3]Symon Muchiri Kariuki, [4]Samson Muchina Kinyanjui, [5]Michael Boele Van Hensbroeke, [6]Charles Newton Newton

'Kemri-Wellcome Trust Research Programme' Kilifi' Kenya [2]Kemri-Wellcome Trust Research Programme' Kilifi' Kenya, [3]Kemri-Wellcome Trust Research Programme' Kilifi' Kenya, [4]Kemri-Wellcome Trust Research Programme' Kilifi' Kenya, [5]University of Amsterdam' Amsterdam' Netherlands, [6]Kemri-Wellcome Trust Research Programme' Kilifi' Kenya

Background

Neurodevelopmental disability significantly contributes to the burden of disease in low and medium-income countries. There is no empirical evidence of premature mortality and associated factors in older children with neurological impairments (NI) in Sub-Saharan Africa. We hypothesized that there is an increased risk of premature mortality in children with neurodisability compared with normally developing children. The current study estimates the risk of premature mortality in older children with NI, identifies risk factors and causes of death in a rural setting along the coast of Kenya.

Methods

A cohort of children aged 6-9 years identified from an epidemiological survey conducted in 2001, NI cases and controls, and an age-matched comparison group from the general population were prospectively observed for all-cause mortality over a 17-year period in the Kilifi Health and Demographic Surveillance System. The mortality rates, standardized mortality ratio (SMR), and hazard ratios for risk factors were estimated and causes of death identified using the standard World Health Organization (WHO) verbal autopsy interviews. Overall, cause-specific and sex- and age-adjusted mortality rates were computed for the cases, controls and the general population. We did indirect standardization using mortality rates from the general population to compute the standardized mortality ratio. We fitted Cox proportional hazard regression models using R packages for survival analysis to compute hazard ratios and investigate the contribution of age, sex, medical and developmental history at the baseline on mortality among NI cases. We checked the proportional hazards assumption and assessed the appropriate functional forms for continuous variables in the final model. All statistical analyses were done in R environment for statistical computing and graphics.

Results

The overall mortality rate in NI cases was 309·8 (95%CI:126·7-492·9) deaths per 100, 000 person-years of observation (PYO) compared to 80·8 (95%CI:64·3-97·3) deaths per 100, 000 PYO in controls and 100·0(95%CI:50·7-61·0) deaths per 100, 000 PYO in a comparison group from the general population. The risk of mortality was increased by the severity of neurological impairment with a mortality rate of 482·1 (0-1150·3) deaths per 100, 000 PYO in severe NI versus a mortality rate of 240·1 (95%CI:48·0-432·3) deaths per 100, 000 PYO in mild NI. The overall risk of death adjusted for age and sex was higher in NI cases compared with controls (HRADJ=4·24, 95%CI:2·26-7·94, p=0·002) and an SMR of 3·11 (95%CI: 1·64-5·41) was obtained after indirect standardization. In multivariable risk factor analysis, developmental delay increased the risk of mortality in children with NI (HRAJD=5·56, 95%CI:1·11-27·88, p=0·037) while better nutritional status decreased the risk (HRADJ=0·64, 95%CI:0·48-0·86, p=0·002). Infections such as HIV/AIDS and accidents were common among those who died.

Conclusion

There is increased mortality in children with NI compared with the general population. Developmental delay was associated with increased the risk of premature mortality in children with NI while better nutritional status was associated with decreased risk. Childhood developmental outcomes and nutritional status should be assessed in children with neurological impairments and tailored interventions started to prevent premature mortality. This study, however, was conducted in a single geographical area in rural Kenya and this may limit the generalizability of these findings to wider African populations. Further research is required from Sub-Saharan Africa where there is a paucity of evidence.

Copula bivariate based semiparametric logit models: Application to the joint modeling of psychoactive drugs and HIV risk among fisherfolk in Kenya

'Thomas Achia [2]Daniel Kwaro, [3]Lily Nyagah, [4]George Mgomela, [5]Sammy Khagayi, [6]Mary Schmitz, [7]Anne Adega, [8]Hellen Awuoche, [9]Emily Zielinski-Guttierez

'US Centers for Disease Control' Nairobi' Kenya [2]Kenya Medical Research Institute' Center for Global Health Research' Kisumu' Kenya, [3]Kenya National AIDS Control Council, [4]US Centers for Disease Control' Dar es Saalam' Tanzania, [5]Kenya Medical Research Institute' Center for Global Health Research' Kisumu' Kenya, [6]US Centers for Disease Control' Nairobi' Kenya, [7] Kenya Medical Research Institute' Center for Global Health Research' Kisumu' Kenya, [8]Kenya Medical Research Institute' Center for Global Health Research' Kisumu' Kenya, [9]US Centers for Disease Control' Nairobi' Kenya

Background

The advantages using multivariate techniques to model the joint variation in outcomes are well known especially where the dependence between the outcome variables can be computed. Applications of copulas for modeling the dependence between bivariate continuous data abound in the literature but there are only a few studies that discuss applications to bivariate binary data. In this study, we carried out a joint modeling of HIV and the use of psychoactive drugs, simultaneously analyzing two different responses emanating from the same individual, using copulas to model dependence structure. The types of psychoactive drugs considered in this study were khat, cannabis, alcohol, heroin and kuber. While an overall reduction in HIV prevalence has been achieved nationally, the continued high prevalence among fisherfolk in Kenya is an area of concern. Substance abuse and addiction have been inextricably linked with HIV/AIDS since the beginning of the epidemic.

Methods

Data for this study were obtained from a cross-sectional household bio-behavioral survey conducted in eight islands beaches in 2017-2018. The outcome variable of interest in this study was a bivariate outcome based on two dichotomous variables: the use of mind-altering drugs; and HIV serostatus of the individual. The correlation between these two variables was assessed using Kendall's tau, r_τ, statistics. A survey weighted bivariate logistic regression model was fitted using the Generalised Joint Regression Modelling (GJRM) package in R to examine how covariate information affects these associations. The dependence structure of the bivariate responses was further modeled using the bivariate normal, Clayton, survival Clayton, Gumbel, rotated Gumbel, and survival Gumbel Copulas. AIC and BIC statistics were used to select the appropriate copula model for the data. The linearity assumption was relaxed for some covariates and their effects modelled using penalized regression splines.

Results

Among the 1696 adult participants enrolled, 531 were HIV positive, with a weighted HIV prevalence of 33.7% [95% confidence interval (CI): 31.0-36.4%], and 388 used psychoactive drugs, with a weighted prevalence of 24.5% [95% CI: 22.1-26.9%]. There was a significant dependence between HIV risk and the use of psychoactive drugs [$r_\tau=0.09, p<0.001$]. The Claytons copula was used to model the dependence structure of the outcome variables [Akaike Information Criterion=2603.8, degrees of freedom=19]. HIV risk was associated with sex, marital status and use of mind altering drugs. The use of mind altering drugs was associated with sex, marital status, level of education, age at first sex, number of partners in the last 12 months, and transactional sex. HIV risk and use of mind-altering drugs both had a non-linear relationship with age, peaking in the 40-49 and 20-29 year age categories, respectively. Individuals that

were separated or divorced were at higher risk of using psychoactive drugs compared to those that had never married [AOR=2.20, 95% CI=1.23-3.92]. Fishermen were at higher risk of HIV compared to all other cadre of fisherfolk [AOR=2.13, 95% CI=1.48-3.08].

Conclusion

The study demonstrates the application of statistical models to analyze bivariate outcomes from complex surveys, using copulas to capture the dependence structure in the outcomes. Links between the use of mind altering drugs and HIV risk have been the subject of a large amount of empirical research, yet still there seems little agreement about the causal connections between these two important public health concerns. The findings of this study suggest that the simultaneous impact of correlated outcomes can be addressed for the commonality between different responses, which is otherwise overestimated when examined separately.

Intimate partner violence assessment: A comparative analysis of the standard logit model versus copula bivariate semiparametric logit in determining the factors associated with intimate partner violence from Kenya demographic and health survey in 2014

'Thomas Noel Ochieng Achia [2]Elias Obudho
'University of the Witwatersrand [2]University of Naiobi

Background
Intimate partner violence is a common source of physical, psychological, and emotional morbidity. An assessment of factors associated with emotional, physical, and sexual violence is an important part of the development of interventions to curb the menace of intimate partner violence.The study compared the performance of the standard logit model with the Bivariate logit model, assessing the performance of these models empirically.

Methods
Data on emotional, physical, and sexual violence meted against women was obtained from the 2014 Kenya Demographic and Health Survey. The impact of selected associated variables on a bivariate outcome variables based on two dichotomous variables: the experience of sexual violence; and the experience of severe physical violence was assessed. A survey weighted bivariate logistic regression model was fitted using the R package SemiParBIVProbit. The dependence structure of the bivariate outcome variables was modeled using Copulas, with AIC and BIC used to select the appropriate copula model for the data. In this approach,the linearity assumption was relaxed for some covariates and their effects modelled using penalized regression splines.

Results
The Bivariate Semiparametric Logit model performed better than the univariate alternatives. Joe's copula was employed to model the dependence structure of the data. The odds Physical and sexual violence declined significantly with socioeconomic status, partner education and media exposure. The risk of severe physical and sexual violence increased linearly with the woman's age, leveling off at 30 years of age.

Conclusion
The study demonstrates the application of statistical models to analysis bivariate outcomes from complex surveys, using copulas to capture the dependence structure in the outcomes.

Estimation of individual level local average treatment effects in cluster randomised trials with non adherence

'Schadrac C. Agbla [2]Karla DiazOrdaz, [3]Bianca DeStavola
'London School of Hygiene and Tropical Medicine [2]London School of Hygiene and Tropical Medicine,
[3]UCL GOS Institute of Child Health

Background

There have been extensive methodological contributions towards the estimation of causal treatment effects when using data from cluster randomised trials (CRTs) in the presence of non adherence. A popular causal estimand is the local average treatment effect (LATE). When treatment assignment is binary and inference at the individual level, LATE answers the question: what is the effect of receiving the active treatment (relative to the control treatment) in individuals who would comply (or adhere) to randomised treatment?Individual level LATE is identifiable assuming that – (i) individuals' outcome within clusters are not affected by the treatment assigned to individuals in other clusters (referred to as cluster non interference), – (ii) for any individual, there are no other possible outcomes under a given treatment that would have occurred except the outcome that has been observed under that treatment (known as consistency), – (iii) randomised treatment is a valid instrumental variable (IV) and – (iv) the effect of randomisation on treatment received is in the same direction for all units (known as monotonicity). By valid IV it is meant that the IV does not share common causes with the outcome, has a strong effect on treatment received (referred to as IV relevance) and has no direct effect on the outcome except through treatment received (so called exclusion restriction). Estimation of LATE is often performed via the two stage least squares (TSLS) method. In the absence of covariate adjustment, TSLS is equivalent to the Wald estimator, which is the ratio of the intention to treat effect on the outcome (effect of random treatment assignment on outcome) to the intention to treat effect on treatment received (effect of random treatment assignment on treatment received). However, to our knowledge, the performance of such ratio when covariates are adjusted for (which we refer to as conditional Wald estimator with a slight abuse of terminology) has not been fully exposed.

Methods

We used simulations to assess the performance of these methods in estimating LATE at the individual level using data from CRTs where non adherence to treatment is binary (the individual either received the assigned treatment or not) and only occurs in the active treatment group (one sided non adherence). Simulations focus on continuous outcomes and vary by cluster size and homogeneity of the outcome (as measured by the intracluster correlation coefficient, ICC), type of adherence (whether adherence is at the cluster level or at the individual level), and strength of covariates' effect sizes (small or large effect).For the Bayesian multilevel mixture modelling, uninformative normal priors were used for all coefficients and vague inverse gamma prior for level 1 variance. Both uniform and Half Cauchy priors were used for level 2 variance.

Results

When randomised treatment is strongly associated with the treatment received (i.e. when the IV is strongly associated with treatment received or in other words, the IV is relevant), the (conditional) Wald estimator, TSLS with HWR SEs and TSLS with Moulton corrected SEs show similar performance and tend to have slightly low coverage in the presence of high ICC but perform well when ICC is low. The Bayesian

multilevel mixture modelling outperforms the (conditional) Wald estimator, TSLS with HWR SEs and TSLS with Moulton corrected SEs. Covariate adjustment has little impact on the bias and coverage for the (conditional) Wald estimator and TSLS but reduces the bias and leads to a coverage that is conservative for the Bayesian multilevel mixture modelling.

Conclusion

The Bayesian multilevel mixture modelling outperforms the (conditional) Wald estimator, TSLS with HWR SEs and TSLS with Moulton corrected SEs and moreover offers flexibility in relaxing level 1 or level 2 variance heterogeneity. Simulations only focus on a relevant instrument. However, treatment assignment is expected to be associated with treatment received in well conducted trials. We assume so far that exclusion restriction holds, but it is possible to relax it when using TSLS and Bayesian methods. Monotonicity holds by design in one sided non adherence trials. Design such as double blinding may enhance the plausibility of the identification assumptions.

Health spending and Economic growth nexus: Evidence from Uganda

'Smartson Ainomugisha [1]Birooma Godfrey Birooma
'Makerere University

Background
Over the past decade, Uganda has experienced a steady increase in healthcare expenditures (both private and public) along with improved economic growth. Understanding the relationship between health spending and economic growth is pivotal for health policy makers; if positive, then the policies that increase the health care expenditure can be used as an effective way to stimulate economic growth. On the contrary, if the health spending cannot effectively boost the sustainable economic growth, then the relevant health policies are not so significant. This paper aimed at examining the relationship between health spending and economic growth in Uganda. Economic growth is proxied by GDP (current US $) and health spending is expressed by health expenditure per capita (current US$). In this study, short-run, long-run and casual relationships were investigated.

Methods
Firstly, the augmented Dickey-Fuller (ADF) unit root test was used to explore the stationary characteristics of health expenditure per capita and GDP time series to avoid spurious results. Error Correction Analysis was employed to investigate the short-run relationship between the study variable whereas the Johansen cointegration technique based on the likelihood ratio test (LR) was used to examine the long-run relationship. The granger casuality test was used to investigate the casual relationship between health expenditure per capita and economic growth. Data obtained from World Bank development indicators for the years 2000-2015 was used for the panel study.

Results
The series of GDP and health expenditure (in level) were nonstationary, however the series became stationary at first differencing thus were integrated of order one [I(1)] at 10% level of significance. The estimated results of the VECM show that coefficient of estimated lagged error correction term is negative suggesting that error correction is happening in the model thus there is a short-run relationship; however the relationship is not statistically significant. The Johansen cointegration empirical evidence revealed that there is a long-run relationship between economic growth and heath spending though not significant. The granger casuality test results indicate that there exists a unidirectional casual relationship and economic growth causes public spending on health. The results also revealed that health spending does not cause economic growth.

Conclusion
The findings of the study indicated that there exists a link between health spending and economic growth in Uganda both in the short-run and long-run though the relationship is not significant. The findings also revealed that health spending does not cause economic growth in Uganda. The government should actively look for optimizing policy related to health care expenditure, such as by standardizing the detail statement of the health care expenditure, improving the efficiency of health spending and so on to promote economic growth.

Efficient methodologies for handling missing data for longitudinal ordinal outcome

'Omololu Aluko [2]Henry Mwambi, [3]Birhanu Ayele
'Stellenbosch University [2]University of KwaZulu-Natal' South Africa, [3]Stellenbosch University

Background

As of 2011, approximately 33 million people were recorded to be HIV positive. Antiretroviral therapy (ART) and similar clinical strategies have been improved survival rates and turned HIV infection into a chronic disease. However, persistent loss of immunity from HIV positive individual results in increased development of comorbid diseases had significantly affected mortality. Some of these people live in resource-limited regions experience serious lung complications. The study was established as a collaborator multi-R01 consortium created to address missing links in the knowledge of axial emphysema disease. Axial emphysema is defined as the abnormal permanent enlargement of the airspaces distal to the terminal bronchioles accompanied by the destruction of the alveolar wall and with obvious fibrosis. In recent times, axial emphysema is highly common with people living with HIV. This was best visualized on the computed tomography (CT) scanning because spirometry may be inadequate in evaluating the degree of deterioration. This is a repeated measurement of ordinal longitudinal data with the missing outcome. The deterioration levels were classified ``normal to evenly distributed'' representing 1 to 4 where highest indicated worst condition. In particular, models for ordered categorical data tend to be more parsimonious than their unordered counterparts, thus resulting in more efficient inferences which facilitate the interpretation of parameters. Adjacent-categories logits, the continuation-ratio logits and the cumulative logits are the most popular methods of modeling ordinal data. The demerits of these models are also explained. In our study, we focused primarily on comparing marginal and generalized linear-mixed effects (GLMM) model families fitted to ordinal incomplete longitudinal data. Marginal models were fitted to generalized estimating equations (GEE) and GLMM were fitted to mixed-effects proportional odds and ordinal negative binomial (ONB) models, using dif

Methods

This section presents the methods used to handle incomplete data. we describe random-effects models using ordinal negative binomial (ONB) and mixed-effects proportional odds (mixed-effects) models, both models belong to generalized linear mixed model (GLMM). We fitted marginal models to multiple imputation generalized estimating equations (MI-GEE) which is valid under missing at random (MAR) assumption.

Results

We have results for simulation study and empirical data analysis. In the simulation study, ONB produces small bias estimate when the sample size and percentage of dropout were small meaning ONB method performs better than the mixed-effects and MI-GEE in terms of bias. The standard error obtain by ONB methods was marginally smaller than mixed-effects and MI-GEE. The results obtain from simulation study was similar to the empirical example. \\ In the empirical data analysis, The negative effect of Age means the prevalence of axial emphysema is lower a among HIV-infected persons with ``no burden'' of emphysema as compared to those who had emphysema ``burden''. However, all the models provided similar results for the effect of Age meaning that each unit increase in Age, the odds of prevalence of axial emphysema burden decreased by 5% (for example, it is e^{0.0581} =0.9436 for mixed-effects)

Conclusion

This study explored the performance of ONB, mixed-effects, and MI-GEE. ONB and mixed-effects belong to the same model family, as against the MI-GEE. But this did not discredit or limit our findings. All the models performed excellently in handling longitudinal ordinal data, but ONB performed best and recommended because of its flexibility approach.

Changes in socioeconomic status among HIV patients after 10-years of antiretroviral therapy in Uganda: An estimation of the within-subject correlations

'Godwin Anguzu [2]Andrew Flynn, [3]Joseph Musaazi, [4]Frank Mubiru, [5]Martha Atwiine, [6]Allen Kabagenyi, [7]Barbara Castelnuovo, [8]Agnes N Kiragga

'Infectious Diseases Institute, Makerere University, Kampala, Uganda [2]University of Colorado, Aurora, USA, [3]Infectious Diseases Institute, Makerere University, Kampala, Uganda, [4]Infectious Diseases Institute, Makerere University, Kampala, Uganda, [5]Infectious Diseases Institute, Makerere University, Kampala, Uganda, [6]School of Statistics and Applied Economics, Makerere University, Kampala, Uganda, [7]Infectious Diseases Institute, Makerere University, Kampala, Uganda, [8]Infectious Diseases Institute, Makerere University, Kampala, Uganda

Background

Understanding the long-term socioeconomic benefits of antiretroviral therapy (ART) is essential for cost-benefit policy decisions regarding providing this lifesaving intervention in settings with competing needs. Current literature indicates ART yields short-term gains in recipients' employment and income. However, whether these benefits persist in long-term is unknown, and clinical correlates of long-term economic benefits of ART are also poorly understood.

Methods

We used data from a cohort of people living with HIV (PLHIV), started on ART between 2004 and 2005 and followed up for ten years, at the Infectious Diseases Institute in Kampala, Uganda. We compared socioeconomic status (SES) indicators (employment status, housing type, highest education level and monthly household income) at baseline and at 10 years of ART. We used McNemar chi-square tests for dichotomous SES indicators (employment status categorized as employed/not-employed and housing type categorized as mud house/brick house), while symmetry chi-square tests was used for SES indicators categorized in more than two levels (education level categorized as Primary/Secondary/Tertiary and household income categorized as No income/<18USD/≥18USD). Factors associated with change in employment status were examined using GEE Poisson models with robust standard errors and exchangeable correlation matrix to account for within-subject correlations of outcome at two time points (baseline and after ten years of ART). We performed post-estimation of within-subject correlation matrix to understand the level of inter-subject variability. Analyses were performed using STATA 14.1, USA.

Results

This analysis included 345 PLHIVs who had data on SES at both baseline and 10-years of ART. Majority 70% were females with median baseline age 33 years and inter-quartile range 22 to 52 years. Percentage employed increased by 45% from 57.4% at baseline to 83.3% after 10 years on ART (p-value = <0.01). The proportion of those with monthly household income of ≥50,000 USHS, more than doubled from 35.1% at baseline to 81.4% after 10 years on ART (p-value <0.01). Attainment of secondary and tertiary education increased to 44.0% from 41.2% and 14.4% from 10.1% respectively (p-value<0.01 on each). Nonetheless there was improvement in housing type over the ten-year period but this was not statistically significant (p-value=0.11). Improvement in employment status was similar for males and females (28.2% vs 24.9% absolute change) however; females demonstrated greater improvement in household monthly income

than male (51.7% vs 34% absolute increase). No physical disability versus physical ability (Adjusted rate ratio [ARR] 1.75; 95%CI 1.31-2.34; p<0.01), age ≥35 years versus age <35 years (ARR 1.19; 95%CI 1.01-1.40; p=0.03), and male sex (ARR 1.23; 95%CI 1.13-1.35; p<0.01) were independently associated with positive change in employment status, after adjusting for each of the above factors and body mass index (BMI). The within-subject correlation of being employed was negligible (model post-estimation correlation = 0.10).

Conclusion

Long-term ART significantly improved employment status and income status of PLHIVs. Improvement in employment status was associated with male sex, older age and physical ability. Accounting for the with-in subject correlation of employment status did not matter because of the long period between the two time points compared.

Risk of hepatic injury among adult patients concomitantly on statin and statin interacting drugs: A self-controlled case-series analysis using an electronic health database

'Henry Athiany [2]Dorothea Nitsch, [3]Mariam Molokhia, [4]Samuel Mwalili, [5]Anthony Wanjoya, [6]George Orwa

'Jomo Kenyatta University of Agriculture and Technology (JKUAT) [2]London School of Hygiene and Tropical Medicine, [3]London School of Hygiene and Tropical Medicine, [4]Jomo Kenyatta University of Agriculture and Technology, [5]Jomo Kenyatta University of Agriculture and Technology, [6]Jomo Kenyatta University of Agriculture and Technology

Background

Statin drugs are among the most widely prescribed drugs in the United Kingdom (UK), and are recommended for preventing cardiovascular events among high risk groups of patients such as diabetics. Due to the aging population, the use of these drugs for prevention of cardiovascular diseases, and other co-morbidities, is on the rise among the adult patients. Despite their benefits, rare but serious side-effects such as myopathy and liver injury have been associated with statins. Individuals on statin drug therapy may have these comorbidities, with indications for other drugs, some of which are contraindicated in statin drug therapy, such as the cytochrome P450 3A4 (CYP3A4) inhibitor drugs. Consequently, statins may be used in combination with these drugs. However, little is known on whether co-prescription of statin drugs and Statin Contraindicated CYP3A4 Inhibitor Drugs (SCCID) translate into additional risk of liver injury. This research aimed at modelling the association between concomitant exposure to statin drug and statin interacting drugs and the risk of liver injury among the UK adult patients, using an improved conditional Poisson model.

Methods

This study utilized data from The Health Improvement Network (THIN) database, a large longitudinal computerized practice database comprising of patient records retrieved from participating practices across the UK from 1994 to 2006. The self-controlled case-series (SCCS) design was used to assess the association between exposure to co-prescription of statin drugs and SCCID and the risk of liver injury. Using the SCCS, differences between individuals were less of a problem as comparisons were made within persons. Thus, time-invariant factors that were different between individuals over the period of the study such as gender, individual frailty, and severity of the underlying diseases amongst others, recorded or not recorded were implicitly controlled for. This generally minimized the problem of residual confounding by indication and/or contraindication. Baseline risk of first liver injury event that varied by patients age, was allowed for by including age as a time-varying covariate in the models fitted. The validity of the case-series method is based on stronger assumptions than required by other methods. To achieve this, we only considered the first liver injury event in the follow-up period. In addition, we carried out sensitivity analyses by excluding individuals who were recorded to have died within three or within six months from the first liver injury event date. Moreover, individuals whose follow-up ended within these durations were also excluded.

Results

A total of 4,935 patients with at least one record of liver injury diagnostic code were identified, of whom 2,486(50.5%) were eligible for data analysis. The rate of liver injury was significantly increased in the periods of exposure to co-prescription of statin drugs and SCCID (Rate Ratio ((RR) 2.32 95% CI 1.11-4.84),

statin drugs alone (RR 1.43 95% CI 1.23-1.67) and SCCID alone (RR 2.43 95% CI 1.80-3.28), compared to baseline rate. Further analyses examining co-prescription of statin drugs and SCCID compared to either statins alone or SCCID alone did not show evidence of increased risk of liver injury. This study had good face validity, given the setting where most patients in the UK were assessed, treated, monitored and followed-up for long periods of time. Clinical data used in this research has been validated for pharmacoepidemiology research, and has demonstrated a high validity in the study of ischemic cerebrovascular events, and hepatitis C virus. The results are therefore more likely to be generalizable even to other populations including the African settings where similar databases have been developed.

Conclusion

Whereas results from this study indicate that periods of exposure to co-prescription of statin drugs and SCCID, statin drugs alone, and SCCID alone were significantly associated with increased rates of liver injury event, compared to the baseline period; there was no evidence to support the hypothesis that co-prescription of statin drugs and SCCID increased the risk of liver injury than when compared to statin drugs alone or SCCID alone. These results may be a reflection of the differences in patient behavior regarding drug adherence when co-prescribed statin drugs and SCCID, or induction of alternate metabolic pathways. There may also be high awareness among patients, and practitioners of the risks of co-prescribing mitigating against a further increase in risk. To reduce the potential for such clinically significant drug-drug interactions, care should be taken when prescribing and dispensing medications to patients especially those with multiple conditions.

Skewed random effects distribution in Conditionally Autoregressive Spatial models for estimating HIV prevalence at local level in South Africa

[1]Kassahun Ayalew, [2]Samuel Manda
[1]Centers for Disease Control and Prevention, [2]South African Medical Research Council

Background
Local level public health estimates are increasingly being sought from data that are originally meant to provide reliable estimates at national and regional levels. In terms of modeling, the Conditionally Autoregressive (CAR) spatial models have often been used to estimate reliable estimates by reducing the instability associated with the sparse spatial data.

Methods
However, the CAR spatial model specification that are implemented in several statistical software and geographical information systems packages are premised on the assumption that the spatial random effects are normally distributed. The normality assumptions could be violated in some situations especially if there are outliers in the data and this assumption may lack robustness and flexibility. Misspecification of the distribution of the random effects may result in estimates that are biased and could have implications on inferences about the parameters of interest.

Results
Therefore, in this study we develop and validate flexible spatial model by relaxing this stringent assumption by assuming that the random effects follow a skew elliptical distribution specifically skew t and skew normal distributions.

Conclusion
HIV prevalence data from a nationally representative survey (2016) and antennal HIV survey (2017) are used in this study. The intrinsic CAR model is generalized for modelling asymmetric spatial random component. This study develops and validates flexible spatial model for modelling data that drifts away from symmetry. Thus, epidemiologists and practicing statisticians can use this generalized approach for conducting spatial analysis and making appropriate decisions using the results estimated using this method. The findings from this study will help governmental and non governmental originations, and the private sector to know the level of the epidemics at lower administrative level, and thus prioritize and plan appropriate public health programs tailored to each community and evaluate the combined impact of national and local public health programs.

Progression of Cisplatin-associated Ototoxicity amongst Patients receiving Cancer Chemotherapy: Application of Tobit Regression

Jessica Paken[1], Cyril D. Govender[1], Mershen Pillay[1], Birhanu T. Ayele[2], Vikash Sewram[1, 3, 4]

[1]Discipline of Audiology, School of Health Sciences, University of KwaZulu-Natal, [2]Division of Epidemiology and Biostatistics, Faculty of Medicine and Health Sciences, Stellenbosch University, [3]African Cancer Institute, [4]Division of Health Systems & Public Health, Faculty of Medicine and Health Sciences, Stellenbosch University

Cisplatin is one of the anti-cancer drug used for the treatment of a variety of cancers. However due to its expansive toxicity profile, patients receiving cisplatin can experience high frequency hearing loss, a side effect known as ototoxicity. A prospective cohort study was conducted to investigate cisplatin-associated ototoxicity in female patients with cervical cancer in a tertiary hospital in South Africa. Patients subsequently underwent audiological assessment prior to commencement of chemotherapy, and at the beginning of the fourth cycle followed by subsequent assessments at one, three and six months after their last chemotherapy cycle. The outcome variable, pure-tone air conduction thresholds measured at different frequencies is right censored. Hence, use of the standard models might bias estimates of the coefficients and lead to a wrong conclusion. We used Tobit regression to analyze the data. In this presentation, we will demonstrate application of the Tobit regression, assess progression of hearing loss and discuss factors predisposing patients to cisplatin ototoxicity.

Developing and validating a scoring tool for mortality among neonates with oesophageal atresia and tracheoesophageal fistula at a tertiary level hospital in KwaZulu-Natal, South Africa

'Yusentha Balakrishna [2]Agneta Odera, [3]Nasheeta Peer, [4]Mahomed Sheik-Gafoor
'Biostatistics Unit, South African Medical Research Council, Durban, South Africa [2]Department of Paediatric Surgery, Inkosi Albert Luthuli Central Hospital, Durban, South Africa; Tenwek Hospital, Bomet, Kenya, [3]Non-communicable Diseases Research Unit, South African Medical Research Council, Durban, South Africa, [4]Department of Paediatric Surgery, Inkosi Albert Luthuli Central Hospital, Durban, South Africa

Background

Oesophageal atresia (OA) and tracheoesophageal fistula (TOF) are uncommon but well-recognized life threatening congenital anomalies of the trachea and oesophagus. While the developed world reports survival rates of more than 90%, many developing countries, particularly in Africa, are still struggling with suboptimal outcomes.

Methods

Demographics, clinical presentations, associated anomalies, surgical approaches, post-operative outcomes, as well as characteristics of the mother, were collected from the charts of neonates diagnosed with OA/TOF at a local tertiary level hospital. Significant predictors of mortality were identified using forward stepwise logistic regression (with entry threshold of 0.05) and were used to develop a weighted scoring algorithm by rounding the beta coefficients to the nearest integer. K-fold cross validation was implemented to evaluate the performance of the scoring tool.

Results

Birth weight, associated anomalies, ventilator dependence, OA/TOF type and sepsis state were found to be significant predictors of mortality and were included in the scoring tool. Scores could range from -5 to 9. A score of 0 or higher best distinguished those at greater risk of mortality with sensitivity of 88.5% and specificity of 91.3%.

Conclusion

The developed scoring tool will alert physicians in identifying neonates with OA/TOF at high risk for mortality and stimulate treatment plans to be adjusted accordingly.

Time to HIV serostatus disclosure, pattern and factors among pregnant women living with HIV in Moshi urban

'Victoria Barabona [2]Prof.Sia Msuya, [3]Dr. James Ngocho
'KCMUCo [2]KCMUCo, [3]KCMUCo

Background

Mother to child transmission is still a public health concern in sub-Saharan Africa including Tanzania. In order to eliminate mother to child transmission among pregnant women viral load suppression has to increase to greater than 90%.In 2017 viral load suppression increased to only 66.7% in Kilimanjaro and the main problem was poor drug adherence which was associated with stigma and discrimination. To address this burden, effective HIV serostatus disclosure is paramount

Methods

We conducted a retrospective cohort study from year 2002 to 2015. A total of 340 pregnant women were followed from when they were in their third trimester to 24 months after delivery. Categorical and continuous variables were summarized using descriptive statistics. Chi-square statistics was used to check difference between groups. Time to HIV serostatus disclosure was estimated using survival analysis. Crude estimates with 95% CI for factors associated with time to HIV serostatus disclosure was estimated using Mult -level survival analysis.

Results

Median time to HIV serostatus disclosure was 12 months with IQR of 6 to 23 since diagnosis. HIV serostatus disclosure to partner was higher (57%) compared to family members (22.53%) friends (15.36%) and others (5.12%).Pregnant women who disclosed to their partners used shorter time compared to those who disclosed to family members, friends and others but the difference was not statistically significant (p=0.94). Pregnant women who were single and had mother to child transmission knowledge used longer time to disclose their status compared to those who were married and had no mother to child transmission knowledge respectively [β(CI):0.15(0.01-0.29); 0.32(-0.58- -0.07)]. On the other side pregnant women aged 36 and above with formal employment status used shorter time to disclose their status compared to those with younger age with no formal employment respectively [β(CI): -0.26(0.56-0.86); -0.27(0.59-0.99)].

Conclusion

Time to HIV serostatus disclosure ranges from a day to two years and it was positively associated with age, marital status, employment status and knowledge of Mother to child transmission. However, multivariable regression analysis is warranted to confirm the finding

Bayesian Copula-Based Analysis of Joint Model for Multi-variable Longitudinal and Recurrent Events

'Tefo Baraki [2]Samuel Manda, [3]Renate Meyer, [4]Jose Romeo
'Botswana International University of Science and Technology [2]Biostatistics Research Unit' South African Medical Research Council' Department of Statistics' University of Pretoria, [3]Department of Statistics' University of Auckland' Australia, [4]Department of Statistics' University of Santiago-Chile' Chile

Background
Longitudinal studies in public health often collect multiple repeated measurements and multiple recurrent events data on the same subjects, and often these processes are correlated. For example, in child health studies, childhood malnutrition measurements defined by height for age, weight for height and weight for age z scores and episodes of common childhood infections such as diarrhea, pneumonia and fever are measured from birth to 60 months. Childhood malnutrition and infections are known to have a complex and bi directional relationship, with malnourished children having a substantially higher risk of multiple recurrent infections and repeated infections contributing to malnutrition. In analyses of such data, several approaches have been used to model the common dependence of one repeated measurement and one recurrent event, in the same subject using joint models. However, the methodological work on how best to model dependence within and between multivariate recurrent events and multivariate repeated measurements processes, based on joint modelling is sparse and subject of ongoing research. We set out to model the dependency structures using Bayesian copula constructions.

Methods
In this work, we propose a Bayesian Copula Based Approach to model dependence within and between multiple repeated measurement processes and multiple recurrent events processes based on a joint model. We will conduct a simulation study to show the viability of the proposed strategy to model correlations within and between the various processes. The methods will then be applied to multiple recurrent infections and multiple repeated malnutrition measurements data for children below the age of 60 months in Botswana.

Results
Preliminary results show association between recurrent episodes of diarrhea, pneumonia and fever, and repeated measurements of height for age, weight for height and weight for age z scores among children aged under five years.

Conclusion
In this study, we propose a Bayesian Copula joint model to describe the interplay between multiple childhood malnutrition measurements and multiple recurrent infections in a longitudinal study. The model allows us to assess how the changing malnutrition status in children influences the risk of multiple recurrent childhood infections. The knowledge of how the changing malnutrition status in children influences the risk of multiple recurrent childhood infections could inform pediatricians on how best to provide care to children aged below 60 months. Our proposed method will greatly facilitate the measuring dependence among several data processes in the application of joint modelling models health data analysis.

Prediction of Family State Occupancy in Rural South Africa Using Multistate Transition Modeling

'Jesca Mercy Batidzirai [2]Samuel Manda, [3]Henry Mwambi, [4]Frank Tanser
'School of Mathematics' Statistics & Computer Science' University of KwaZulu- Natal' Pietermaritzburg' South Africa [2]South African Medical Research Council' Pretoria' South Africa, [3]School of Mathematics' Statistics & Computer Science' University of KwaZulu- Natal' Pietermaritzburg' South Africa, [4]School of Nursing and Public Health' University of KwaZulu- Natal' Durban' South Africa

Background

In family formation and dissolution studies, transition probabilities and lengths of stay in the same state (Never Married, Married, Separated and Widowed) are of interest to demographers. In health studies, literature regards marital state transitions as one of the key determinants of psychological and physical health. In an attempt to reduce these health problems, it is important to understand the dynamics of marriage formation and dissolution together with factors leading to them. This study, therefore, investigates the durations of staying in particular marital states as well as determinants of transitions between various marital states. In addition, prediction of marital state occupation is done for subjects who posses particular attributes. This will provide an insight about social population dynamics and would be important for planning towards giving social and emotional support among long-term widows and children raised by non-biological parents.

Methods

Multistate models were employed to estimate durations in particular marital states through a 1-year step transition model. For a given set of covariates, prediction of marital state occupancy over a period of time was done. Population-based data from rural KwaZulu-Natal of South Africa was utilized and a discrete-time first-order time homogeneous Markov model was constructed to specifically estimate the likelihood of transition for a subject with some specified covariates using a multinomial logistic regression as a basis.

Results

Results show that age at first sex had a slightly positive effect on the Never Married to Married transition (OR = 1:18;CI = (1:16; 1:20)) while having a negative effect on marital separation (OR = 0:97;CI = (0:95; 0:98)). Males had a lower likelihood of being separated than females (OR = 0:26;CI = (0:16; 0:43)) and a higher likelihood of remarrying after being widowed (OR = 1:05;CI = (1:04; 1:06)). The odds of marital separation were higher among subjects who had been married for not more than 10 years than those who had been married for more than 10 years (OR = 1:81;CI = (1:72; 1:90)). Assuming a time-homogeneous transition matrix, a 20 year old subject who was never married had a 73:17% likelihood of being still single and 15:76% likelihood of being married and 9:02% likelihood of being separated in the next 10 years. A 45 year old widowed female was 84:9% likely to remain widowed in the next10 years while a 45 year old widowed male had a 57:01% chance of remaining widowed in the next 10 years. The likelihood of a never married subject to remain never married within a year was 99:43% while that of a married subject to be separated was 0:16%. For those who had been widowed, they had an 8:45% chance of remarrying within a year. The sojourn time in a Never Married, Married, Separated and Widowed state was 9 years, 7 years, 8 years and 10 years, respectively. In the long run, subjects have a 57%, 34%, 25% and 16% probability of being never married, married, separated and widowed, respectively (steady state probabilities 0:57; 0:34; 0:25and 0:16, respectively).

Conclusion

Over a 10-year period, the likelihood of entering a first-time marriage for adolescents was low while that of remarriage was lower for females than males. The mean duration time in a Widowed state was the highest compared to other states. Compared to females, males had a lower chance of exiting a marriage but higher chance of remarrying after a marital dissolution. Subjects who had a late sexual debut were less likely experience a marital separation than those who had an early sexual debut. These results point towards a possible need for programs such as those aimed at raising the age at first sexual debut, as means of trying to reduce psychological problem scaused by marital dissolution. A time homogeneous model was used in this study. However, a non home geneous model may be considered where a time dependent transition matrix can be used.

Analysis of Climatic Variability and its Effects on Production of Selected Crops in Ada'a District, Ethiopia: Multivariate Time Series Approach

'Diribsa T. Bedada, [2]Diribsa Tsegaye
'Stellenbosch University [2]Stellenbosch University

Background

Climate change affects all economic sectors to some degree, but the agricultural sector is perhaps the most sensitive and vulnerable. In the last three decades, Ethiopia has been affected by climate-related hazards.

Methods

Agriculture, the most dominant sector in the national economy, has been most at risk because of its dependence on seasonal rainfall. Anticipated climate change has negatively impacted the agricultural sector due to increased temperatures and decreased or greater variability in precipitation, leading to increased food insecurity.

Results

This study was carried out with the general objective of examining climate variability and its effects on selected crop production for the last three decades in Ada'a district of East Showa Zone of Oromia Regional State. The relevant data were obtained from the National Meteorological Agency (NMA) and district's agricultural office. Data on selected crop production, total rainfall and average temperature for the period of 1985 to 2015 were used. The vector autoregressive (VAR) model is employed for modelling. The cointegration relations among the variables were identified by applying Johansen's cointegration tests, while potential causal relations were examined by employing Granger's causality tests. Moreover, the short run interactions among the variables were determined through the application of impulse response analysis.

Conclusion

The results of the research imply the existence of short-term adjustments and long-term dynamics in crop productivity, total rainfall, minimum and maximum temperature. The result of Johansen test indicates the existence of one cointegration relation between the variables. The result shows that a Vector Error Correction (VEC) model of lag two with one cointegration equations best fits the data.

Permutation multivariate tests for treatment effect: theory and recent developments

'Stefano Bonnini [1]Getnet Melak Assegie
'University of Ferrara

Background

When the effect of a treatment is evaluated according to several outcomes, a suitable multivariate test must be applied. When the number of response variables is very large, especially in the presence of a small number of patients, typical multivariate parametric solutions (such as Hotelling T-square test) are not possible. The difficulty of the problem is much higher in the very common case of one-sided hypotheses, e.g. in the typical location problem represented by stochastic dominance. Even when a parametric solution is available, for example within the family of likelihood based inferential methods, the strong assumptions required about the dependence structure of the response variables, make this solution not robust with respect to the diparture from the family of distributions assumed as underlying multivariate distribution of the response and not powerful, because of the large number of nuisance parameters to be estimated (e.g. correlations or covariances in the case of normal distribution).As a motivating example, we consider a randomized controllled trial to test the effectiveness of a specific myotensive technique, which should increase the tone of foot muscles and improve the postural system of healthy subjects who frequently practice sports.

Methods

In the randomized controlled trial of myotensive podal technique effect on postural system, considered as a motivsting example, 14 healthy subjects, 20 to 30-year-old football players, were involved. The myotensive muscle technique was applied to 7 of them (randomly selected) and the results were compared to a control group which consisted of the other 7 subjects to whom the treatment application was simulated but, in fact, not applied. For all the 14 subjects pre- and post-treatment data were collected using rasterstereography and baropodometry devices. In order to evaluate the treatment effect, 85 numeric response variables, concerning static and dynamic rasterstereography and baropodometry, were taken into account and the differences between observed values after the treatment and before the treatment were computed. Hence, for such a complex problem, the nuber of outcomes is much larger than the sample sizes, thus the Hotelling T square test or simular parametric methods cannot be applied. We propose a nonparametric solution that belongs to the family of combined permutation tests (Bonnini S, 2015; Bonnini S, Corain L, Marozzi M, Salmaso L, 2014; Pesarin F, Salmaso L, 2010). This is nonparametric because does not require any assumption about the population multivariate distribution.As a consequence, this methodology does not require the dependence structure of variables to be explicitly defined, as well as in the likelihood based methods. Through a Monte Carlo Simulation study we want to prove the good power behavior of the test and the specific conditions under which the power of the test is maximum.

Results

The methodology of combined permutation tests is suitable, especially (but not only) in the presence of a large number of variables and small sample sizes. The power of the test depends on the combining function and the number of marginal variables under H1. The test is consistent because the power increases with the sample sizes. It is more powerful than the Hotelling T square test when the number of variables is low and the only available solution when the number of variables is too large. The power of

the test also depends on the strength of the dependence between the single outcomes and on the number of outcomes itself.The main advantage of the method is that the dependence between response variables doesn't need to be explicitly defined, modeled or estimated, because it is implicitly taken into account by the joint use of the adopted permutation strategy and of a suitable combining function.

Conclusion

The proposed methodology of combined permutation test for treatment effect in multivariate problems is distribution free and powerful even when the probabilistic assumptions of parametric tests are true. It is an exact, unbiased and consistent testing procedure, with good performance even in the presence of small sample sizes.This method is suitable when the testing problem can be broken down into several sub-problems (multiaspect test, multistrata test, multiple comparisons,…). Hence it is appropriate even for multivariate problems and also when the type of alternative hypothesis is not the same for all the single outcomes. The method is suitable for complex alternatives like stochastic ordering, umbrella hypotheses, multivariate one-sided tests. It can be applied with multivariate numeric, categorical and mixed variables.The power is increasing function of the number of variables for which the alternative hypothesis is true.

Malaria morbidity in under-five children in Malawi: A Bayesian spatial-temporal Model

'Vitumbiko Chijere Chirwa [2]Eustasius Musenge, [3]Kennedy Otwombe
'University of the Witwatersrand School of Public Health [2]University of the Witwatersrand School of Public Health, [3]University of the Witwatersrand School of Public Health

BACKGROUND

Malaria remains a leading cause of morbidity, hospital admissions, and mortality in under-five children in Malawi. Despite a decrease in Malaria prevalence nationally from 43% in 2010 to 24% in 2017, Malaria still remains one of the major public health challenges in Malawi. Most studies done on under-five malaria morbidity in Malawi have focused only on spatial variations at one point in time; these studies do not show how the spatial patterns of malaria morbidity have changed over time. There is little statistical work that has been done to explore the spatial and temporal variations on under-five malaria. This study aimed at modelling factors associated with malaria risk in Malawian under-five children while accounting for spatial and temporal variations. A better understanding of the spatial-temporal variations of malaria morbidity in under-five children in Malawi is essential for policy makers and implementers to think and develop significant interventions to help lessen under-five malaria morbidity by prioritizing those specific geographical areas which still have high malaria burden despite the passing of time

METHODS

This study used the Malawi Malaria Indicator Surveys (MMIS) data that were done in 2010, 2012, 2014, and 2017. The population of interest is under-five children. The outcome was whether a child was malaria positive or negative (a binary outcome). For exploratory level analysis, maps showing malaria prevalence per district, for each year, were plotted to show the hotspots and coldspots of malaria. These maps were plotted using ArcMap software. In terms of modeling malaria at individual child level, four logistic regression models using Bayesian hierarchical framework were fitted: 1) a model with fixed effects only and no random effects, 2) a model with unstructured spatial random effects, 3) a model with both unstructured and structured spatial random effects, and 4) a model that incorporated the spatial and temporal random effects. At geographical area level, a Poisson model was fitted that accounted for individual level scores from the logistic regression, as well as the spatial and temporal effects. For model comparison, this study used the Deviance Information Criterion (DIC) and the effective number of parameters (pD). The model that had the smallest DIC was considered as the better fitting model. The modeling technique was based on the Integrated Laplace Approximation (INLA) methods. The models were built using R-INLA package in R-studio software.

RESULTS

Out of 1822 children in 2010, 785(43%) were found malaria positive; for 2012, 585(28%) out of 2112 children were found to be malaria positive; for 2014, 640(33%) out of 1928 children were found to be malaria positive; and for 2017, 564(24%) out of 2305 children were found malaria positive. A spatial exploratory distribution of malaria, per year, showed that most hot spots of malaria were in the central region of Malawi, followed by the southern region. For all the years except 2017, the global Moran's index for spatial auto-correlation showed significant clustering of malaria prevalence. The logistic model which took into account the spatial and temporal effects was considered as the better fitting model because it had the smallest DIC value as compared to the rest of the models (DIC=7781.55, pD=36.47).

Socioeconomic and demographic factors [OR (95% Credible Interval)] that had a significant protective effect on malaria risk were: having a, Insecticide Treated mosquito net [0.79(0.79,89)], living in a household with higher wealth index [0.55(0.47,0.64)], and mothers who had a higher education level [0.46(0.36,0.57)] Children in rural areas were 3.62(3.01,4.37) times more likely to have malaria than children in urban areas. Chances of having malaria also increased with age. Both the logistic and Poisson models showed significant spatial and temporal effect variations on malaria.

CONCLUSION

This study showed that accounting for spatial as well as temporal random effects is important in order to understand spatial patterns of malaria risk over time; hence necessitating the use of spatial-temporal techniques. This study further shows that, to help lessen malaria burden, there is need to scale up the use of insecticide treated nets in household, and also to target the less privileged households. Central region, as well as rural areas in general, need more interventions.

Using Software in Teaching Medical Data

'Lily Clements [2]Roger Stern
'University of Southampton [2]University of Reading

Background

A popular software used in Africa to teach medical data is Stata, a statistics package which allows individuals to analyse their data either through a menu-driven front-end, or by writing commands. Another option to consider teaching medical data is the free, open-source statistics language, R. It hosts an increasing library of commands for medical statistics. While several developers have built menu-driven front-ends for R, R-Instat is a suitable front-end. It includes general menus alongside specific menus for tailored analyses. One under construction is for medical statistics. This paper considers the case where software is used in teaching medical data, and queries the extent to which R, together with R-Instat, can rival Stata. To compare these software we consider two datasets. The first is a large "real world" dataset from Tanzania. Called a Demographic Sentinel Surveillance (DSS), it looks at population changes from 1994-2013. This data requires a large amount of cleaning before analysis. The second dataset is based on kidney transplant survival times, this in contrast is smaller and ready for analysis.

Methods

Stata has an intuitive language for preparing and analysing data. Its depth allows it to be useful for cleaning datasets, such as the DSS dataset. In the analyses the user can define their data for a specific area of medical statistics, for example our second dataset can be defined as survival data by the "stset" command. This generates new columns to be used in later analyses and means that in future commands Stata automatically selects the correct columns adding an ease-of-use. When considering data cleaning in R, we look at a new suite of packages called the Tidyverse. This aims to make the analyses more approachable for the user through its consistent and coherent structure of reusing the same set of functions, known as verbs. However, it is arguably not intuitive at a first glance. Once the data is clean, R offers hundreds of packages containing different commands for analysing medical data. Unlike with the Tidyverse, there is not as much of a consistent formula when coding with these different packages. We then look at the front-ends in Stata and R-Instat. Stata offers comprehensive menus and dialogs which have been built around the Stata functions. Each selection outputs the analyses alongside the appropriate Stata command, for if the user wished to convert to coding. The extensive menu options allow for real-depth for users who are not proficient in coding. In contrast, R-Instat has built the menu and dialogs to fit into one of the software aims of teaching good statistical practice. As a result, the analyses for medical data are more-or-less confined to one menu, and hence the dialogs can be placed in a logical order which correspond to the successive stages in an analysis. This is beneficial to the user as they can find their relevant statistics. Further, it allows for decluttering among the standard statistics menus allowing them to be simpler and easier to navigate.

Results

The DSS data was analysed in both Stata and R. Stata gave a cleaner, more intuitive code when preparing the data through its quick, simple commands. However, the logical structure of the Tidyverse in R is arguably easier to use than Stata once it is understood. Analysing the DSS and kidney transplant data is arguably easier in Stata due to its consistent commands across the software – an advantage of not having a diverse range of user contributed packages means that it is written in the same format. This ease and consistency is enhanced due to its use of the define commands. However, since R is open-source, it has a

wide selection of options, and if something is not available then it can be easily written into the software. Both Stata and R have the depth to analyse medical data, so we can next consider the front-ends for each of these software and how these can fit within teaching a course. Stata offers comprehensive menu and dialogs. While this also creates more of a cluttered design which is arguably harder to navigate, it allows for the resulting commands from the dialog to be given in a cleaner format than in R-Instat. This is particularly important for individuals who may consider migrating from using the menus to coding, and hence could be valuable within a course. R-Instat puts the dialogs relating to medical data in one menu, which can allow for a course to be based on these dialogs where it runs down this single menu in a consistent and coherent order. Since R-Instat focuses on teaching good statistical practice, the dialogs are based around this philosophy and hence fit for a course setting. This results in dialogs which are much less cluttered than in Stata, and its sub-dialogs contain more advanced options if individuals wish to explore further.

Conclusion

While we have only considered two datasets for a course setting, they contain contrasts suitable for teaching. The transplant dataset is more typical of a current course since it is ready for analysis and hence allows for the focus to be on the analysis itself. Meanwhile, the DSS data can help broaden a course to include more on data cleaning, a large area of dealing with “real-life” datasets which often gets omitted from teaching. Further, it has over 142,000 cases. This type of big data is often absent from a course setting despite being increasingly used in the “real-world” setting. Overall we found that Stata is a “hard act to beat”. Hence organisations could well continue with Stata if they have access to the software for all their students to use individually and where students will continue to be able to have access to the package once they graduate. However, giving students experience in more than one package can be useful. Hence, adding R, perhaps with a front-end such as R-Instat, could enhance some courses. The availability of the option of using R, perhaps with R-Instat, is important in cases where purchasing software licences is not feasible, since R and R-Instat are both free and open source.

Right-censoring bias correction for growth curve linear mixed models

'Dominique-Laurent Couturier [2]Maria-Pia Victoria-Feser
'Cancer Research UK - Cambridge Institute [2]University of Geneva

Background

Tumour growth inhibition studies typically involve analysing tumour sizes measured regularly over a period of time. The aim is usually to detect differences in growth rate between experimental conditions. Many methods have been considered. Some summarise each growth curve into a single measure, like the area-under-the-curve statistics or the growth slope on the log scale, and compare the location parameter of these statistics between different experimental conditions by means of Welsh tests. Others consider mixed/longitudinal models, taking into account the time and within-tumour dependence of the observations to provide a parametric fit on all collected data. As animals are culled when their tumour size exceeds a legal upper limit or when the discomfort level is considered too high, such data are often right censored, leading to biased growth estimates.

Methods

Our objective is to develop a method allowing one to correct the bias of growth curve linear mixed models in the presence of right-censoring due to a fixed upper tumour size limit.

Results

Simulations show that the iterative bootstrap bias corrected estimator we developed for random intercept and slope mixed models allows us to obtain unbiased growth rate estimates as well as confidence intervals showing coverages close to the nominal value.

Conclusion

The proposed simulation-based bias correction method provides unbiased estimates when fitting growth curve linear mixed models to data showing right-censoring. Furthermore, it correctly assesses the differences in growth rate between experimental conditions.

Statistical Measures for Multivariate Spatial Autocorrelation

'Timotheus Darikwa [2]Samuel Manda, [3]'Maseka Lesaoana
'University of Limpopo [2]Medical Research Council' Pretoria' South Africa & School of Mathematics' Statistics and Computer Science' University of Kwazulu-Natal' Pietermaritzburg' South Africa & Department of Statistics' University of Pretoria' Pretoria' South Africa , [3]University of Limpopo

Background
Bivariate spatial autocorrelation methods have been developed and successfully applied to bivariate health outcomes that are spatially correlated. However, spatial autocorrelation statistics measures for more than two possible correlated spatial health outcomes are not developed. This paper is concerned with developing a multivariate spatial autocorrelation extension of the Moran's Index.

Methods
o achieve this a canonical correlation approach is proposed. Three spatially correlated multivariate normal datasets were simulated using conditional distribution from Gaussian data. The new approach to multivariate spatial autocorrelation was applied to the data for three variables of a given simulated dataset. Monte Carlo simulation of the data with different spatial weights and sample sizes was done and evaluated based on the root-mean square error and bias. The method was then applied to three cardiovascular-related mortality from cerebrovascular heart disease (CVA), ischaemic heart disease (IHD) and hypertensive heart disease (HHD) for three variables that are spatially dependent on each other.

Results
Monte Carlo simulation of the data with different spatial weights and sample sizes confirms the appropriateness of the new method its generalis ability to both univariate and multivariate approaches. Spatial dependency was found between CVA and IHD, CVA and HHD (p-value less than 0.05) but none between IHD and HHD (p-value>0.05). The multivariate spatial autocorrelation extension showed that CVA is spatially dependent on both IHD and HHD.

Conclusion
A new method multivariate spatial autocorrelation method that is easy to use and interpret has been developed to detect multivariate spatial dependency. The effectiveness of the method has been established through some simulation studies and its usefulness was illustrated using cardiovascular mortality in South Africa. The method can easily be extended to more than three variables.

A semi-parametric mixed models for longitudinally measured fasting blood sugar level of adult diabetic patients

'Legesse Kassa Debusho ²Tafere Tilahun Aniley
'University of South Africa ²University of South Africa

Background

Background: At the diabetic clinic of Jimma University Specialized Hospital, health professionals provide regular follow-up to help people with diabetes live long and relatively healthy lives. Based on patient condition, they also provide interventions in the form of counselling to promote a healthy diet and physical activity and prescribing medicines. The main purpose of this study is to estimate the rate of change of fasting blood sugar (FBS) profile experienced by patients over time. The change may help to assess the effectiveness of interventions taken by the clinic to regulate FBS level, where rates of change close to zero over time may indicate the interventions are good regulating the level.

Methods

Methods: In the analysis of longitudinal data, the mean profile is often estimated by parametric linear mixed effects model. However, the individual and mean profile plots of FBS level for diabetic patients are nonlinear and imposing parametric models may be too restrictive and yield unsatisfactory results. We propose a semi-parametric mixed model, in particular using spline smoothing to efficiently analyze a longitudinal measured fasting blood sugar level of adult diabetic patients accounting for correlation between observations through random effects.

Results

Results: The semi-parametric mixed models had better fit than the linear mixed models for various variance structures of subject-specific random effects. The study revealed that the rate of change in FBS level in diabetic patients, due to the clinic interventions, does not continue as a steady pace but changes with time and weight of patients.

Conclusion

Conclusions: The proposed method can help a physician in clinical monitoring of diabetic patients and to assess the effect of intervention packages, such as healthy diet, physical activity and prescribed medicines, because individualized curve may be obtained to follow patient-specific FBS level trends.

Joint Multistate Modelling of Viral Load Dynamics and Multistate CD4 count Progression

'Zelalem Dessie [2]Temesgen Zewotir, [3]Henry Mwambi, [4]Delia North
'University of KwaZulu Natal [2]UKZN, [3]UKZN, [4]UKZN

Background
HIV patients may experience a succession of clinical stages before the disease diagnosis and the patient health status may be followed-up by using long term viral load dynamics. In this work, we present a joint multistate models for clinical progression events which takes into account viral load biomarker, in order to study possible factors that affects the transition intensities between sequential events

Methods
The data is from an ongoing prospective cohort study conducted amongst adult women who are HIV-infected patients in Kwazulu-Natal, South Africa. Participants were enrolled into the acute HIV Infection phase, and then followed-up during chronic infection and up to ART initiation. Joint modelling of longitudinal and multi-state processes were applied

Results
We presented a joint multistate models for longitudinal viral load dynamics and CD4 count transitions. The model confirmed that viral load dynamics significant effects the intensities of transitions of HIV/AIDS disease progression. The analysis showed that patients with higher educational levels, higher RBC indices score and higher physical health score was significantly associated with lower viral load trajectories. Patient with many sex partner, older age and higher liver abnormality score were significantly associated with higher long term viral load trajectories. Moreover, having a high weight, higher education levels, higher QoL scores, having high RBC parameters and patients with middle age were significantly increased the intensities of immunological recovery transitions. Furthermore, patients having a high liver abnormality score and with many sex partners was significantly reduced the intensities of immunological deterioration transitions

Conclusion
It can be concluded that joint multistate model approach provides the wide-ranging information about the progression opens to much more precise knowledge of diseases and specific dynamic predictions. The tools are available in terms of methods and software, so hopefully this article has helped to familiarize applied researchers (for medical research) with some of these issues as well as with the interpretation of models

Exploring association between predominant breastfeeding and infection-related hospitalization, over time: an empirical comparison of models to account for clustering

'Moleen Dzikiti nee Zunza [2]Moleen Dzikiti nee Zunza
'Stellenbosch University [2]Stellenbosch University

Background

Longitudinal studies investigating changes in a specific outcome that is measured repeatedly over time are common. Such studies are appropriate for investigating the change of the outcome, and the effect of associated covariates. The credibility of study findings depend on the appropriateness of the analysis method used. It is important to assess the sensitivity of the study findings to different possible methods of analysis to ensure appropriate conclusions are drawn from the study. We performed six models using an empirical data set obtained from a real life non-trial clinical care setting to explore if there exists significant differences in the change of a dichotomous outcome (infection-related hospitalization), over the first year of life, between infants who were predominantly breastfed and those who were not. Data are from the Mother Infant Health study, a longitudinal cohort study that compared infection-related hospitalization among HIV-exposed uninfected infants and HIV-unexposed infants. The primary objective of this study was to assess the sensitivity of the findings to different models used to account for dependency of a binary outcome measured repeatedly over time.

Methods

We estimated the effect of predominant breastfeeding on infection-related hospitalization using generalized linear mixed models with (1) a random intercept and (2) a random slope. We used the logit link function and fitted the model using the Gauss-Hermite quadrature approximation. The fixed effects quantify the subject-specific effects of predominant breastfeeding on the odds of infection-related hospitalization, in children who share the same propensity to infection-related hospitalization. The random effects quantify the variation between subjects (random intercept model) and variation between subjects and variation in infant age (random slope model). The final models included fixed effects (predominant breastfeeding, infant HIV-exposure status, mother's educational level, and infant age).We estimated the effect of predominant breastfeeding on infection-related hospitalization rates using generalized estimating equations: with 1) an exchangeable correlation structure that assumes the outcome measurements within the same child are equally correlated; 2) autoregressive correlation structure of order 1 (AR1) which assumes that the correlation between outcome measurements is a function of the number of time points apart. In this structure the correlation between adjacent outcome measurements is ρ, irrespective of the position of the pair and the correlation is ρ^d for any pair of observations that are d units apart and 3) unstructured correlation structure which assume a unique correlation between any pair of measurements. For each model, we used the binomial family with a logit link function and computed the empirically-corrected standard error. The fixed effects quantify the marginal effects of predominant breastfeeding on the odds of infection-related hospitalization. The final model included fixed effects (predominant breastfeeding, infant HIV-exposure status, mother's educational level, and infant age). We also fitted the conventional logistic regression model.

Results

We analyzed data from 203 mothers and their infants who had at least one follow-up visit. Eighty four

infants were non-predominantly breastfed and 119 were predominantly breastfed. We observed 34 infection-related hospitalizations overall. Of the infants who were hospitalized due to infection, most were hospitalized once, except for four who were each hospitalized twice.We found a nearly identical odds ratio (95% CI) of 0.71 (0.26 to 1.95) across all models, including the logistic regression model that assume independent correlation structure, except for the generalized estimating equations with autoregressive (AR1) working correlation structure that had an odds ratio of 0.81 (0.24 to 2.73). Overall the width of the confidence intervals were wider for the generalized estimating equations, with the generalized estimating equations with autoregressive (AR1) working correlation structure having the widest confidence interval. We found a clinically important but statistically insignificant reduced odds of infection-related hospitalization among predominantly breastfed infants compared with non-predominantly breastfed infants, irrespective of the method of analysis used. The distinction between the methods was less important for the longitudinal binary outcome data with insufficient replications of the outcome.

Conclusion

The method of analysis we used had insignificant effect on the parameter estimates and this could have been driven by insufficient replication of infection-related hospitalization outcome in a cluster. Our estimates from the six models are likely to be unstable and need to be interpreted with caution.

Hierarchical Mixed Effects model for Type 2 diabetes

'Haile Mekonnen Fenta [2]Demeke Lakew Workie, [3]Dereje Tesfaye Zikie, [4]Seyifemicael Amare Yilma
'Bahir Dar university [2]Bahir Dar University, [3]Bahir Dar university, [4]Debre Tabor University

Background
Different statistical models are available for the analysis of longitudinal data and the linear mixed model is commonly used to understand changes over time. This study seeks to obtain the best model for a type 2 diabetes to estimate the rate of change of patients ' Fandom Blood Sugar (FBS)over time.

Methods
Data were obtained on 95 patients for five consecutive times they visit the hospital for review and the mixed effects model was used.

Results
The likelihood ratio test revealed that linear distribution trend in the mean of FBS was preferred compared with squared and cubic. The compound symmetry with heterogeneity was selected as the best variance-covariance structure. For final model building,step wise selection approach was used and it showed that treatments (time, gender, the presence of other diseases, types of treatment, blood pressure(systolic/diastolic) level were significantly affecting the level of FBS.

Conclusion
For planned intervention and management of diabetes' patient's health care providers should note these and lay emphasis on them in controlling diabetes.

Bayesian Dynamic Models for Time-Varying Outcomes: Application to a Patient Cohort on ART

[1]Lineekela Gabriel, [2]Lawrence Kazembe
[1]University of Namibia [2]University of Namibia

Background

Patients' adherence to a prescribed medication regimen is one of the most significant barriers to successful antiretroviral therapy (ART). The extent of the impact of poor adherence on resulting health measures is often unknown, and typical analyses ignore the time-varying nature of adherence.

Methods

This study uses Bayesian dynamic linear models (DLMs) to model longitudinally measured viral load as a function of time-varying adherence and other time-varying covariates. The objectives of this study are to review the dynamic models of HIV and investigate the effects of clinical factors on viral load over time. Data for patients initiated on ART between January 2014 and December 2016, in Namibia, is used to analyze the dynamics of viral load as adherence changes within patients over time.

Results

Baseline CD4 count, baseline weight, age at start of ART and sex are the non-dynamic covariates which were measured at the start of ART, while adherence to ART is the dynamic covariate which recorded at follow up visits and is believed to be time-varying.

Conclusion

We use non-informative priors that have minimal effect on the model with the parameters assumed to be gamma distributed. Patients with a good adherence rate tend to achieve viral suppression within 6 months of treatment and for as long as good adherence is maintained then viral load stays undetectable.

Multiple imputation of clinician-level covariates applied in modelling quality of clinician-prescribed care: Sensitivity analysis of departure from Missing at Random (MAR) assumption

'Susan Gachau [2]Matteo Quartagno, [3] Edmund Njeru Njagi, [4]Nelson Owuor, [5]Mike English, [6]Philip Ayieko

'Kenya Medical Research Institute-Wellcome Trust Research Programme' Nairobi Kenya' School of Mathematics' University of Nairobi' Kenya [2]Institute of Clinical Trials and Methodology' University College London' London' United Kingdom, [3]Department of Non-Communicable Disease Epidemiology' London School of Hygiene and Tropical Medicine' London' United Kingdom, [4]School of Mathematics' University of Nairobi' Kenya, [5]Kenya Medical Research Institute-Wellcome Trust Research Programme' Nairobi Kenya' Nuffield Department of Medicine' University of Oxford' United Kingdom, [6]Department of Infectious Disease Epidemiology' London School of Hygiene and Tropical Medicine' London' United Kingdom

Background

Routine data are often used to monitor the quality of clinicians' prescribed care in many health care settings. However, routine data are prone to missing information which can lead to biased and misleading inferences. Multiple imputation (MI) assuming that data are Missing at Random (MAR) is one of the widely used missing data handling method. However, missing data can be imputed assuming a Missing Not at Random (MNAR) mechanism. MAR and MNAR mechanisms cannot be distinguished from observed data alone and sensitivity analyses is therefore recommended to assess robustness of inference under MAR assumption. In this study, we conduct sensitivity analyses within the pattern-mixture models framework to assess departures from MAR assumption.

Methods

We analysed clustered routine paediatric data collected from 12 Kenyan hospitals between March and November 2016. The outcome of interest was Paediatric Admission Quality of Care (PAQC) score which summarizes the quality of inpatient paediatric care in low and middle income settings. We adjusted for patients, clinician and hospital level factors in the analysis model. Missing data occurred in patients' sex in level 1 and clinicians' sex and cadre in the second level of hierarchy. Multilevel multiple imputation within the joint modelling framework was used to fill-in missing data assuming MAR and MNAR mechanism respectively. Bayesian and delta adjustment approaches using prior distributions and shift parameters respectively were used to impute missing data under MNAR mechanism. A random effects model was used to analyse complete case records and imputed data sets.

Results

Overall, 2127 children aged 2 to 59 months were admitted by 378 clinicians across the 12 hospitals. Estimates from complete case analysis differed from estimates obtained after multiple imputation assuming MAR mechanism. However, MI estimates under MAR were close to MI estimates assuming MNAR mechanism. The similarities were observed in both sensitivity analyses methods.

Conclusion

Eliciting and incorporating experts' opinions is transparent means for assessing departures from the MAR assumption. From our study results, imputing from prior distributions in a Bayesian approach and altering the distribution of imputed values in the delta adjustment method led to parameter estimates close to those under MI assuming MAR mechanism. Therefore, our inferences were robust to plausible departures from a MAR assumption.

Effect of different potassium fertilizer rates and liming on recommended maize yield grown in western Kenya

'Cyrus Githunguri [2]Joseph Miriti, [3]Elias Thuranira, [4]Mary Koech, [5]Keziah Ndungu, [6]Vincent Woyengo
'Kenya Agricultural and Livestock Research Organization [2]Kenya Agricultural and Livestock Research Organization, [3]Kenya Agricultural and Livestock Research Organization, [4]Kenya Agricultural and Livestock Research Organization, [5]Kenya Agricultural and Livestock Research Organization, [6]Kenya Agricultural and Livestock Research Organization

Background

Despite the importance of maize, yields per acre have continued to decline in Kenya which is majorly attributed to low soil fertility. Potassium (K) is the third major primary essential nutrient after nitrogen and phosphorus which is required by plants in large quantities. Potassium regulates the uptake of nitrates from the soil, has a balancing effect on phosphorus uptake and strengthens stalks of plants which enhances the plants resistance to fungal and bacterial attacks and lodging. It also influences very many physiological processes in plants such as transpiration, synthesis of carbohydrates, proteins and translocation of the synthesized food. Potassium deficiency in Kenya may manifest in some soils depending on the management practices. Several fertilizer companies in Kenya include K in their product. Since K is a very expensive nutrient, these trials sought to assess the K response in maize production in selected western Counties. The broad objective of the study was to determine the crop response to potassium in western Kenya. The specific objective of the study was to demonstrate the effect of potassium and liming on the yield of maize in selected western Kenya sites.

Methods

The trials were established in KALRO Njoro, Nakuru County; Baraton University; Nandi County; KALRO Kitale; Trans Nzoia County; KALRO Kakamega; Kakamega County; and Bungoma Agricultural Training Centre in Bungoma County during the 2017 long rains season. Lime was applied at two levels (with and without lime) and K at 0, 40, 80,120, 160 and 200 kg K20 ha-1. All the plots received a blanket application of nitrogen (N) and phosphorus (P) at the recommended rates of 60 kg N and 60 kg P2O5 ha-1. The treatments were arranged in a split- plot in a randomized complete block design replicated four times. Lime was allocated to the main plot and the various K rates to the sub-plots. The trial was established in plot sizes of 6 m x 4 m with a harvested area of 9 m2. Potash was applied at different rates as K2O in form of Muriate of Potash as follows: i) absolute control (nil); ii) recommended N and P minus K (NP); iii) NP + 40kg K2O ha-1; NP + 80 kg K2O ha-1; iv) NP + 120 kg K2O ha-1; v) NP + 160 kg K2O ha-1; and vi) NP + 200 kg K2O ha-1. Prior to land preparation, soil sampling was conducted for analysis of selected physical and chemical properties. Basal N and P was applied as Mavuno fertilizer (10-26-10) by banding at 60 kg N and 60 kg P2O5 ha-1. Maize variety H6218 was planted at 0.75 m between rows × 0.25 m within the row. Recommended agronomic practices were followed. At six weeks after crop emergence, calcium ammonium nitrate, was used for topdressing. At harvest the grain was oven-dried to attain a moisture content of 12.5%. Data was subjected to analysis of variance (ANOVA) to determine the effect of treatments using SAS. Means were separated using least significant difference at 5% level of significance.

Results

Plots without liming that were supplied with 200 kg K ha-1 produced significantly higher ($p \leq 0.05$) maize

yield at KALRO Njoro in Nakuru County than those supplied with Nil, NP_0K, NP_80K, and NP_160K kg ha-1. However, plots without liming that were supplied with NP_40K and NP_120K kg ha-1 did not produce significantly ($p \leq 0.05$) different yield than those supplied with 200 kg K ha-1. On the other hand, where lime was supplied there were no significant differences in maize yield obtained with different K rates. In Nandi County, though not significantly different from the other K rates, maize supplied with NP_160 kg K20 ha-1and with or without lime produced significantly ($p \leq 0.05$) higher yields than those without fertilizer. In Trans Nzoia, Kakamega and Bungoma Counties sites, maize response to K was not significant.

Conclusion

There could be some interactive effects between K and lime that could be inhibiting uptake of K by maize plants in the KALRO Njoro site. Potassium was adequate in the Trans Nzoia, Kakamega and Bungoma Counties trial farms and as such application of K based fertilizers is not necessary in these sites, considering the high cost of K. The application of K and liming has the potential to produce positive responses on maize grain yields. The lack of differences between lime and no lime application may be due to the method of lime application, which may have negatively affected plant germination. As such, the method of lime application is likely to be a key factor in maximising the expected interactions between liming, K and NP. Finally, potassium application above NP_40 kg K+/ha in Nakuru and Nandi Counties County may not be necessary.

Effect of different zinc fertilizer rates and cropping systems on recommended maize and beans grown in western Kenya

'Cyrus Githunguri [2]Joseph Miriti, [3]Elias Thuranira, [4]Mary Koech, [5]Keziah Ndungu, [6]Vincent Woyengo
'Kenya Agricultural and Livestock Research Organization [2]Kenya Agricultural and Livestock Research Organization, [3]Kenya Agricultural and Livestock Research Organization, [4]Kenya Agricultural and Livestock Research Organization, [5]Kenya Agricultural and Livestock Research Organization, [6]Kenya Agricultural and Livestock Research Organization

Background

Micronutrients like zinc (Zn) are involved in the key physiological processes of photosynthesis and respiration and their deficiency can hamper these crucial physiological processes and thus curtailing yield gain. Dry beans and maize are sensitive to Zn deficiency. Micronutrient requirements depend on plant uptake, soil availability and growing seasonal conditions. Several soil conditions could lead to development of Zn deficiencies, which can reduce yields and delay crop maturity. Zinc has important functions in protein and carbohydrate metabolism. The main function of zinc is catalytic, playing a role in building and activating plant enzymes and as such its deficiency can be critical. Currently several different Zn sources, including ZnSO4, ZnCO3, ZnO, Zn (NO3)2 and ZnCl2 are being used as fertilizers. It is therefore important to explore management options that could enhance the performance, quality and yields of maize and bean crops using zinc. As such, the objective of the study was to demonstrate the effect of zinc fertilizer on the performance of maize and beans grown in Nandi, Trans Nzoia, and Kakamega counties in western Kenya.

Methods

The study was conducted under field conditions in Nandi, Trans Nzoia, and Kakamega counties during the main rain season of 2017. Different zinc fertilizer rates on recommended maize and beans grown in western Kenya were tested under optimum (with recommended rate of fertilizer) and low input (zero fertilizer) management. The different zinc fertilizer rates and two cropping systems were laid out in a randomized complete block design (RCBD) in a factorial arrangement with four replications. The zinc fertilizer was assigned as the main plot while the cropping system formed the sub-plot. During the study, all recommended cultural practices were carried out and the crops harvested at physiological maturity. Grain samples were oven-dried at 650C to attain 12.5% moisture content. Data was analysed following analysis of variance procedure in SAS software. Treatment means were separated using least significant difference at 5% level of significance.

Results

There were no significant differences ($p \leq 0.05$) in beans and maize grain yield planted under different zinc fertilizer rates and with or without Zn fertilizer across the three counties. Similarly, there were no significant differences ($p \leq 0.05$) in maize grain yield planted either as a mono crop or intercropped with beans across the three counties. However, beans planted as a mono crop produced significantly higher grain yield ($p \leq 0.05$) than those intercropped with maize across the three counties.

Conclusion

There could be some interactive effects between zinc and cropping system that could be inhibiting uptake of Zn by beans and maize plants in the three sites. Lack of significant response to Zn application could be an indication that Zn was adequate in these farms. This study therefore propose that application of Zn

based fertilizers in maize and bean crops is not necessary in the Nandi, Trans Nzoia, and Kakamega Counties sites. This study also suggests that when maize is the targeted crop it should be grown either as a monocrop or intercropped with beans. However, when bean production as the targeted enterprise, then it should be grown as a monocrop to achieve maximum yields. Finally, due to the nutritional importance of Zn there is need to conduct plant uptake of this element by both bean and maize grains.

A Review on Disease Mapping and Modelling Of Cardiovascular Diseases in South Africa

'Nomonde Gwebushe, [2]Samuel Manda, [3]Nada Abdelatif
'SAMRC [2]SAMRC, [3]SAMRC

Background

Cardiovascular disease (CVD) has been the leading cause of death indeveloped countries for most of the last century. Most CVD deaths, however, occur in low- and middle-income, developing countries (LMICs) and there is great concern that CVD mortality and burden are rapidly increasing in LMICs as a result of population growth, ageing and health transitions. In sub-Saharan Africa (SSA), where all countries are part of the LMICs, the pattern, magnitude and trends in CVD deaths remain incompletely understood, which limits formulation of data-driven regional and national health policies. South Africa is experiencing a pandemic of CVDs, metabolic syndrome, diabetes and cancers, which are largely due to changes in lifestyle and socioeconomic inequities. At the same time the pervasiveness of risk factors for non-communicable diseases is a public health concern. While we are developing a better understanding of these risk factors, background risks such as diabetes and obesity often do not receive the emphasis that they deserve. Furthermore, co-morbidity between cardiovascular conditions such as hypertension, CHD, stroke and hypercholesterolemia is common; however, little is known about environmental and geographic overlaps in these conditions.

Methods

Database and literature search was done, South African full-text articles that carried out descriptive, clustering or spatial methods of disease mapping were retrieved and included in this review. Tools and spatial modeling approaches employed in these studies were discussed and analyzed.

Results

There were 1377 articles with the keywords which were extracted from the databases. From these articles 10 were duplicates. 1367 articles were screened and only 69 that were found eligible. They were further assessed and only 6 that were included for this review.

Conclusion

There are few studies examined the spatial variation of the cardiovascular risks in South Africa. These studies have been limited by poor availability of subnational level data and the use of inappropriate statistical techniques to analyze such data.

Linear mixed models with time-varying covariates: application of disaggregation of within-subject and between-subject effects

'Stella May Gwini [2]Dean Mckenzie, [3]Ali Cheetham, [4]Joshua Garfield, [5]Sue Cotton, [6]Murat Yucel, [7]Nicholas Allen, [8]Dan Lubman

'Barwon Health, Victoria, Australia [2]Epworth Healthcare, [3]Turning Point, Eastern Health, Melbourne, VIC, Australia, [4]Turning Point, Eastern Health, Melbourne, VIC, Australia, [5]The National Centre of Excellence in Youth Mental Health, Melbourne, VIC, Australia, [6]Monash Institute of Cognitive and Clinical Neurosciences, Monash University, Melbourne, VIC, Australia, [7]Department of Psychology, University of Oregon, Eugene, USA, [8]Turning Point, Eastern Health, Melbourne, VIC, Australia

Background

In many longitudinal studies, data are collected on both time invariant covariates (i.e. measurements are made only once throughout the study, e.g. gender) and time-varying covariates (TVCs; i.e. measurements collected multiple times that may vary at each time point). TVCs are useful for understanding temporal responses of the dependent variable to changes in the covariates. Often researchers ignore the TVC and replace it with an estimate such as mean, peak measurement or mode, despite existing evidence and statistical methods for incorporating the TVC into the analyses. Much of this under-utilization is attributed to the limited literature on statistical methods that clearly illustrate how TVCs can be incorporated into generalized linear models (GLMs), as opposed to survival analyses where advanced statistical methods for incorporating TVCs are more commonly applied. One method appropriate for GLMs is the direct disaggregation of the within- and between-subject effects for time-varying covariates. This level of analyses allows exploration of the relationship between the dependent variable and the covariate at both the individual level as well as the group level, thus reducing the likelihood of the ecological fallacy (i.e. making inferences about individuals from group aggregated data) and maintaining the advantage of mixed model analyses at being able to examine individual effects. With the increased emphasis on precision medicine and individualized medicine, it has become imperative to increase awareness of these methods. The research presented herein considers the application of disaggregated within- and between-subject effects within a longitudinal study in the presence of a continuous outcome and a continuous TVC, and illustrates the usefulness of this disaggregation when between-subject and within-subjects effects differ.

Methods

While the most commonly used methods of disaggregating these effects are person-mean centring and detrending approaches, this paper presents the former where the between-subject effect will be estimated using the subjects' average of all measurements collected over time. The within-subject effect will be estimated using the difference between the subject's mean and actual measures at each time point. The method will be illustrated using both simulation data and original data from a longitudinal study of 121 opioid-dependent individuals. The study assessed the effect of monthly illicit opioid use on anhedonia (i.e. inability to feel pleasure in normally pleasurable activities). Anhedonia was assessed using the Temporal Experience of Pleasure Scale tool while the timeline follow-back tool was used for opioid use assessment. The tools were administered at baseline, then monthly until 6th months and finally at 12 months. Within-subject and between-subject effects were disaggregated following recommendations by Curran and Bauer (2011) . Linear mixed models with and without the effect disaggregation were used to examine whether illicit opioid use influenced anhedonia. Performance measures (i.e. bias, precision and coverage) from the simulation study will also be reported.

Results

Without disaggregating effects, the results suggest a direct relationship between illicit opioid use and anhedonia ($p<0.001$). However, disaggregation of effects indicated that the observed 'overall' effect among these opiate users was more likely to be a within-subject effect ($p<0.001$) and not a between-subject effect ($p=0.422$). Further results from the simulation study will be presented.

Conclusion

These findings illustrate that ignoring disaggregation of effects for TVCs may result in compounded conclusions which may not reflect the actual relationships for individuals. The current paper clearly illustrates use of disaggregated effects within a longitudinal study, however the methods are applicable to any multilevel study.

Multi-state models for the analysis of Wheezing episodes in a birth cohort of African children

'Patrick Hannan [2]Maia Lesosky, [3]Polite Nduru, [4]Heather Zar
'University of Cape Town [2]University of Cape Town, [3]University of Cape Town, [4]University of Cape Town

Background

Wheezing is common in young children. Data from high income countries, has shown that by the age of six, approximately 50% of children will have experienced at least one episode of wheezing in their life. Furthermore, childhood wheezing may be associated with reduced lung function and increased risk of asthma in later life. Determining the epidemiology of wheezing is complex given that wheezing is a symptom, reflecting airway obstruction that may result from many different causes. Furthermore, some children may develop recurrent wheezing, however traditional analyses of recurrent childhood wheezing rarely utilise methods that allow for recurrent outcomes. There are limited data from low- and middle-income countries (LMICs) on the natural history and determinants of recurrent wheezing in early childhood. This study aimed to use multi-state models to estimate the rate of transition among various states of wheeze in children from birth to the age of three years. This study also aimed to investigate the association between possible risk factors for childhood wheeze and the estimated transition intensities.

Methods

This study was a secondary analysis of data from 1143 children from birth to three years, born to mothers enrolled in the Drakenstein child health study, in the Drakenstein area of the Western Cape, South Africa. A time-inhomogeneous multi-state Markov model with three states (never wheeze, lower respiratory tract infection associated (LRTI) wheeze, and, wheeze not associated with LRTI) was constructed. Baseline transition intensities were assumed to be piecewise constant for each time period, with transition intensities changing at (0 - 180 days, 180 - 336 days, 336 - 728 days, and 728 - 1100 days). The multi-state model allowed for four possible transitions: 1) from never wheeze to wheeze not associated with LRTI or from 2) never wheeze to LRTI associated wheeze or from 3) wheeze not associated with LRTI to LRTI-associated wheeze and from 4) LRTI-associated wheeze to wheeze not associated with LRTI. Direct transitions were not permitted between non-adjacent states, therefore, the transition intensities for those transitions were set to zero. Transition intensities between wheeze states as well as the associations of risk factors with transition intensities were estimated using multivariable proportional hazards models.

Results

Of the 1143 children included in the study, 471 (41%) experienced at least a single episode of wheezing, and 225 (20%) experienced more than one episode of wheezing in the first three years of life. A total of 944 episodes of wheezing were recorded in the 36 months of follow-up time. LRTI-associated wheeze (0.0003) and wheeze not associated with LRTI (0.00042) were equally likely to be the first wheezing event. However, recurrent wheezing events were more likely to be associated with an LRTI. Male gender was associated with an increased likelihood of transitioning to the LRTI-associated wheeze state from never wheeze (HR: 1.736; 95% CI: 1.24-2.432) and had significantly higher likelihood of subsequent recurrent wheezing. Children exposed to maternal smoking prenatally had a significantly higher risk of transition (HR: 1.572; 95% CI: 1.151 - 2.147) to the wheeze state compared to unexposed children. The baseline transition rate for all wheeze transitions decreased significantly as age increased.

Conclusion

Multi-state models provide a novel and flexible framework that characterises the transition of children between various states of childhood wheeze, including recurrent wheezing. Multi-state models successfully predicted the progression of children through discrete states of wheezing and produced results consistent with existing literature on childhood wheezing, while accounting for recurrent events and interval-censored data. These findings highlight the unique information that can be gleaned from using multi-state models to evaluate childhood wheeze. This work was supported by the Bill & Melinda Gates Foundation (OPP 1017641), the National Institutes of Health, USA (H3Africa 1U01AI110466-01A1), the SA Medical Research Council and the National Research Foundation, SA.

Bias estimation and sensitivity analysis in questionnaire studies with "embarrassing" questions

'Christian Hansen, [2] Deborah Watson-Jones
'Statistics and Data Management [2]MRC/UVRI & LSHTM Uganda Research Unit, Entebbe, Ugand

Background
Questionnaire studies that ask about sensitive issues such as respondents' sexual behaviours or sexual past are typically subject to reporting bias because not all respondents are likely to provide truthful answers. The resultant bias can affect study findings considerably, and assessing the impact on central parameter estimates may therefore be of importance to the interpretation of study findings.

Methods
The EDCTP funded RHASA study (Reproductive Health of Adolescent girls in sub-Saharan Africa) which completed in Mwanza, Tanzania in July 2014, asked a group of 17-18 year-old secondary school attendees whether they had had sexual intercourse.

Results
Of the respondents, 58% reported that they had never had intercourse although some of these were diagnosed with sexually transmitted infections which suggests some level of reporting bias.

Conclusion
Using the RHASA study as an example we illustrate a method for deriving point and interval estimates of the magnitude of such reporting bias, and use Bayesian methods to evaluate plausible ranges for parameters of primary interest after accounting for the bias.

Evaluation of the robustness of imputation methods combined to backpropagation algorithm in frame of multiple non linear regression

'Hounmenou Castro Gbêmêmali, ²Gneyou Kossi Essona
'LaBEF et IMSP, UAC, Benin ²FST, UL, Togo

Background

Missing data management for prediction purposes is an important issue in data analysis. This study aims to evaluate the efficiency of imputation methods combined with the backpropagation algorithm in a nonlinear regression context.

Methods

The evaluation is conducted through a simulation study including sample size (50, 100, 200, 300 and 400) containing different missing data rates (10, 20, 30 40 and 50\%) by considering three missingness mechanisms (MCAR, MAR and MNAR). Four imputation methods (Last Observation Carried Forward, Random Forest, Amelia and MICE) were used to impute datasets before making prediction with backpropagation. 3-MLP model was used by varying the activation function (Logistic-Linear, Logistic-Exponential, TanH-Linear and TanH-Exponentiel), the number of nodes in the hidden layer (3-15) and the learning rate (20-70\%).

Results

The analysis of the performance criterion (R^2, r and $RMSE$) the network produced good performances when it's trained with TanH-Linear activation as function, 11 nodes in the hidden layer and a learning rate of 50\%. MICE and Random Forest were more appropriate for data imputation. These methods can support up to 50\% of missing rate at an optimal sample size of 200.

Conclusion

In breeding for instance, the knowing of production is necessary for specialists who need simple and accurate techniques to predict the production of meat, eggs, milk etc. But data collected in this case are small (due to the cost of experimentation) and seldom complete. Missing data is one of the most important problems for researchers. It occurs because of human error, equipment failure, death of animal during the experiment, etc. Analysis of imcomplete datasets results in problems such as biased parameter estimates, inflation of standard errors, loss of information, and weak generalizability of results. Apart from Kohonen network, most of statistical analysis methods assume the absence of missing data, and are only able to include observations for which every variables are measured. To overcome this situation, rows with missing values can be deleted (deletion) but it leads to a loss of precision with weak sample size. To avoid this situation, the best imputation methods combined with Backpropagation algorithm following the sample size and the optimun hyper parameters choose were purposre.

Modelling recurrent events data for hypertension and diabetes control in a peri-urban area in South Africa

'Charl Janse van Rensburg [2]Leisha Genade, [3]Elize M. Webb, [4]Liz E. Wolfaardt
'South African Medical Research Council [2]University of Pretoria, [3]University of Pretoria, [4]University of Pretoria

Background

It is well-known that the prevalence of non-communicable diseases is on the rise globally, including South Africa. This study took place at a research site in a peri-urban area close to Johannesburg, Gauteng province. Participants were enrolled and followed up from September 2012 to October 2017. The aim was to understand when individuals with type 2 diabetes and/or hypertension, who received dedicated treatment at a health facility outside that of the Department of Health, became controlled and for how long they could maintain control.

Methods

In this study, 704 participants with uncontrolled hypertension and 66 with uncontrolled type 2 diabetes were included for modelling. The two groups were analysed separately. Time until the first controlled measurement was initially described using Kaplan-Meier curves. A frailty model was used to model the recurrent events, adjusting for risk factors such as age, gender and others. The use of a multi-state model was also explored.

Results

Eighty percent of the hypertension patients became controlled at some point, compared to only 50% in the Diabetic group. The median times until first control was 434 days and 1220 days for the hypertension and diabetes groups respectively. The median (IQR) proportion of visits participants who had hypertension control, was controlled was 0.45 (0.29 – 0.60). Final results to be presented.

Conclusion

Conducting research in peri-urban areas present many challenges to both the clinical side of the study, as well as to the research side. Patients in the hypertension group seemed to become controlled, but were only controlled for half the visits. The diabetics group did not seem to do as well.

A comparative study of palmar and digital dermatoglyphic patterns among type-II diabetic & non-diabetic adults: A meta-analysis

'Ameet Kumar Jha [2]Sujatha D'costa, [3]Daniel Josh Kanhai, [5]Grace Waldron White
'Texila American University [2]American University of Antigua, [3]Texila American university, [5]Georgetown Public Hospital Corporation

Background

Dermatoglyphic patterns are the epidermal ridges seen on the surface of palm, sole & digits. These ridges play a significant role in assessing various diseases in mankind. Diabetes in today's world is a challenge and serious threat as lifestyle disorder. It is very important to know about the early diagnosis and undertake the preventive measures to overcome the threat. It can be very important & significant to ascertain the person at higher risk for becoming diabetic beforehand. The aim & objective of this review is to find out the significance finding/results in the literature which shows the association between dermatoglyphic & diabetes mellitus. The aim of this review was to systematically identify, review and appraise available literature that evaluated an association of different dermatoglyphic variables with kidney diseases.

Methods

An intense systematic literature search was conducted using keywords 'Dermatoglyphic', stressed induced diabetes, diabetes from Medline(PUBMED), Google Scholar, EBSCO, HINARI etc. The review is performed based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement. Dermatoglyphics patterns atd, dat, adt angles, absolute finger ridge count (AFRC), total finger ridge count (TFRC), a-b ridge count, mainline index, and pattern index line were studied.

Results

The mean values of TFRC and AFRC were higher in male and lower in the female diabetic group. The mean values of a-b ridge count were lower in male and higher in the female diabetic group and a significant difference was found. The mean values of atd and adt were higher in the diabetic group.

Conclusion

In this way, there have been a number of studies showing that various dermatoglyphic markers are associated with the Diabetes. This alliance could be explained and justified if the risk towards developing Diabetes mellitus in future life could be connected with the fetal development of epidermal ridges. These connections can be a highly significant tool or of diagnostic aid in clinical point of view if could be rationalize or substantiated as these markers could be used for screening out an individuals who can be at higher risk of becoming a sufferer. Dermatoglyphics provide a simple, inexpensive, anatomical, and non-invasive means of determining the diseases which have a strong hereditary basis and can be employed as a method of screening for diabetes mellitus of the high-risk population on early detection, thus reducing the morbidity and mortality.

Auxiliary variables in Latent Profile Analysis (LPA) and methods to estimate a distal outcome model

'Esme Jordaan ²Martin Schwellnus
'Biostatistics unit' SAMRC and Statistics and Population studies' University of the Western Cape ²SEMLI' UP and IOC Research Centre' SA

Background

Exercise Associated Muscle cramping in runners (EAMC) is a serious condition affecting about 14% of runners during their running career. The risk factors for EAMC in distance runners include the runners training history. These observations include the continuous items number of years of being a recreational runner, the number of years of participating in distance races, the number of times per week training, the training distance per week and the average training speed. The first objective is to uncover an underlying unobserved categorical variable that divides the runner population into mutually exclusive and exhaustive latent profiles indicating subgroups (called classes) of runners w.r.t. their training history. Secondly, we are interested in using the latent categorical variable for further analysis and exploring the relationship between that variable and the auxiliary observed variable EAMC.

Methods

In LPA, also known as "gaussian (finite) mixture model", true class membership is unknown due to measurement error but can be inferred from a set of measured items. Employing a sufficient number of latent classes result in mutually exclusive and exhaustive classes with no residual covariance among the observed items. A major decision in LPA is the optimal number of classes which could be evaluated using various model fit indices. To model the relationships of the categorical latent variable with auxiliary variables, in this case a distal outcome, various approaches are available; the direct approach where the auxiliary variables are included in the model, referred to as the 1-step model, the pseudo-class (PC) method (Wang et al.), the 3-step approach (Vermunt, 2010 and Asparouhov-Muthén, 2014), the Lanza method (Lanza et al., 2013) and the modified BCH method (Bakk and Vermunt, 2014). The advantages and disadvantages of the relevant approaches are explored in the analysis of the data from a prospective cohort study of about 75 204 Two Oceans 21 km and 56 km runners who registered over 3 years from 2014 to 2016 and completed a questionnaire on their training and EAMC history.

Results

The model fit indices indicated that the 4 class model was the best fit to the data and resulted in a meaningful interpretation of the classes. When modeling a distal outcome the results varied depending on assumptions and other important considerations: (1) The scale of the distal outcome, continuous or categorical, (2) classification-error corrected, (3) susceptibility to class changes (4) controlling for covariates (5) the distal outcome has unequal variances across classes, (6) low entropy.

Conclusion

Many methods have been proposed for mixture modeling with auxiliary variables. It is important to consider the assumptions and considerations when selecting a method for analysis. When choosing a method, the pros and cons for each auxiliary setting should be taken into account.

Comparison of Antimicrobial Resistant Neisseria gonorrhea to Ceftriaxone and Ciprofloxacin using Proportional Odds Model for patients seen at the University Teaching Hospital in Lusaka, Zambia

'Priscilla Kapombe [2]Patrick Musonda, [3]Lungowe Sitali
'University of Zambia [2]University of Zambia, [3]University of Zambia

Background

Neisseria gonorrhea, the causative agent of Gonorrhea, has developed antibiotic resistance to the "last-line" Cephalosporin's, Quinolones and Macrolides which is cause for concern. In Zambia, despite recommendations of discontinued use, Ciprofloxacin is used for treatment. The lack of an active surveillance system, appropriate and structured data management and analysis tools magnifies the problem because resistance patterns cannot be monitored. The aim of this study was to ascertain the effectiveness of Ciprofloxacin a Quinolone in comparison with Ceftriaxone a Cephalosporin, in treatment of Gonorrhea using Antimicrobial Susceptibility testing; and to identify possible risk factors associated with drug resistance.

Methods

Methods: Study design was parallel non-inferior quasi experimental study. Patients at the University Teaching Hospital (UTH) presenting with discharge and symptoms in line with Gonorrhea who gave consent, were recruited. Fishers Exact Test for associations was used. Data was analyzed using Ordinal logistic regression as the Susceptibility was at 3 levels; Susceptible, Intermediate or Resistant with an assumed Ordinal nature. Proportionality assumption was checked, and when violated Partial Proportional Odds Model was used instead.

Results

A total of 104 isolates were obtained. The overall proportion of patients who had Susceptible, Intermediate and Resistant results were: 49 (47.1%), 55 (52.9%) and 0 (0) for Ceftriaxone and 70 (68.0%), 10 (9.7%) and 22 (22.3%) for Ciprofloxacin respectively. Adjusted estimates in partial Proportional Odds model showed that, Males were 4.1 (95% CI; 1.8, 9.4, p-value=0.001) times more likely to have Intermediate or Susceptible results compared to Resistance than females, or they were more likely to have Susceptible compared to Resistance or Intermediate result compared to females. Ciprofloxacin was 70% less likely than Ceftriaxone of having susceptible or intermediate results compared to resistance and this could be as high as 90% and as high as 40% p-values <0.001).

Conclusion

Level of Ciprofloxacin resistance (22.3%) detected from the Susceptibility testing, shows it is not an effective treatment for Gonorrhea as the cut off point for an alternative antibiotic regimen is 5%. We recommend an alternative antibiotic treatment. Ceftriaxone remains a satisfactory option for first-line treatment of Gonorrhea at UTH. Risk factors identified to be associated with resistance in this study were being female and use of Ciprofloxacin.

On the Correct Modeling of the Association Between an Outcome and a Mismeasured Covariate in Clustered Cross-Sectional Surveys: A Simulation-Based Study

'Alexander Muoka [2]Henry Mwambi, [3]George Agogo, [4]Oscar Ngesa
'University of KwaZulu Natal [2]University of KwaZulu Natal, [3]Yale University, [4]Taita Taveta University˙ Kenya

Background
Most statistical analyses on the association between an outcome and the associated risk factors depend on data collected from cross-sectional surveys. These surveys make use of data collection tools that are prone to measurement error. Cross-sectional surveys can also collect clustered data where the observations have a within-cluster dependency structure. Failure to adjust for covariate measurement error and take into consideration the within-cluster dependency structure can bias the association between the covariate and the outcome and therefore yield misleading results and conclusions.

Methods
This study uses simulated data to model the association between an outcome and a covariate adjusting for measurement error in the simulated mismeasured covariate using regression calibration and putting into consideration the within-cluster correlation structure of the data.

Results
The main finding of the study is that although regression calibration can adjust for covariate measurement error regardless of whether the within-cluster correlation structure is taken into account or not, the standard errors of the coefficient estimates can be overestimated or underestimated depending on the cluster sizes.

Conclusion
From the findings of this study, we conclude that it is important to model the correlation structure after adjusting for covariate measurement error in clustered data.

Discrete Time Hazard Parameterization for Child Survival in Malawi

[1]Jessie Khaki [2]Professor Samuel Manda
[1]University of Malawi [2]South African Medical Research Council

Background
Statistical analyses of child mortality often treat time to death as continuous even if mortality durations are measured and reported in discrete time. Moreover the underlying hazards have been modeled as parametric which may be restrictive with their assumed functional forms. Mortality outcomes of 17,395 children in the 2015-2016 Malawi Demographic and Health Survey were used.

Methods
Child survival was assessed in five age-groups. The baseline log odds of child mortality was modeled by using the general, linear, quadratic, and cubic logit models. Non-parametric specifications of the baseline discrete death hazard were also modeled using locally weighted scatter plot smoother, restricted cubic splines, and B-spline. The Akaike Information Criteria (AIC) was used to assess the best fitting hazard function specification.

Results
The general logit model was the best fitting model for estimating discrete hazard function of child survival. The hazard of child mortality was highest in first 30 days of life and reduced for the higher age-groups. This trend was also observed from the non-parametric models. Predictors of childhood mortality risks were birth order, preceding birth interval and maternal age.

Conclusion
The general logit model is recommended for assessing mortality trajectories within five years of birth in Malawi. A higher proportion of deaths occur in the first 30 days of life. This may point to the effect of the maternal factors (such as breastfeeding and mother's immunity) on health outcome of a child within first months of their life.

Spatial and Temporal Patterns of Acute Febrile Illness and Confirmed Typhoid Fever Infection and Their Associated Correlates in Ndirande, Blantyre, Malawi: An Analysis of STRATAA Data

'Jessie Khaki [2]James Meiring, [3]Marc Henrion, [4]Melita Gordon
'University of Malawi [2]Malawi-Liverpool-Wellcome Trust Clinical Research Programme, [3]Malawi – Liverpool - Wellcome Trust Clinical Research Programme, [4]Malawi-Liverpool-Wellcome Trust Clinical Research Programme

Background
The Strategic Typhoid Alliance across Africa and Asia is a multi-site study being implemented in Malawi, Nepal and Bangladesh. The primary objective of the consortium is to estimate the burden of Typhoid fever in three urban sites. In Malawi, STRATAA is being conducted in Ndirande township in Blantyre. The main objective of this study is to investigate the spatio-temporal trends of typhoid in Ndirande. Specifically, the study aims to identify and describe seasonal and spatial patterns in typhoid incidence and transmission. The study further aims to develop a spatio-temporal statistical model and interrogate the risk factors for typhoid.

Methods
A total of 97,000 individuals in Ndirande were enumerated in a census in 2016 and 2018. Data were further collected through the passive surveillance which took place between September 2016 and September 2018. A healthcare utilization survey of 740 households was also conducted between April 2017 and April 2018. The data were complimented by a sero-incidence survey of 13,000 individuals in Ndirande. We constructed time plots and incidence maps to investigate spatial patterns and used time plots to explore time trends of both confirmed typhoid and fever cases. We further tested for spatial, temporal, and spatiotemporal clusters of typhoid and fever cases within the data using formal statistical tests. We compared the performance of the log Gaussian cox process (LGCP) model with that of the ordinary logistic, Cox and Poisson regression models. Using the best performing model, we interrogated risk factors for typhoid and fever and looked for spatio-temporal clusters of cases in Ndirande.

Results
The study presents results on spatial and temporal patterns of typhoid and fever cases within Ndirande township in Blantyre district and highlights the spatial, temporal and spatiotemporal clusters identified in the data. We further present and discuss risk factors for both typhoid and febrile cases in Ndirande.

Conclusion
This is a first look at the modeling results from a spatio-temporal statistical analysis of the Malawi STRATAA passive surveillance study. We describe the spatio-temporal modelling framework adopted for the analysis of the data and discuss the risk factors associated with typhoid and fever and also implications of the results.

Hematological Reference Intervals for Children in Bagamoyo District, Tanzania

'Ummi Abdul Kibondo [2]Said A. Jongo, [3]Ali T. Motoro, [4]Ali M. Ali
'Ifakara Health Institute' Tanzania [2]Ifakara Health Institute' Tanzania, [3]Ifakara Health Institute' Tanzania,
[4]Ifakara Health Institute' Tanzania;[2]. University of California San Francisco' USA

Background
Clinical trials are designed to establish safety and effect of the various interventions. Normal reference values for clinical laboratory parameters to ensure safety monitoring of the study participants is of paramount. However, most of the available reference values used in most of African clinical trials which involving children have been adopted from western populations. Therefore, the aim of the study was to establish hematological reference values for children aged 6 months to 10 years in Bagamoyo district.

Methods
Data used were collected from the cross sectional study conducted by Ifakara Health Institute, Tanzania within IDEA project at Bagamoyo District from 2010-2013. The calculation of reference intervals was performed using the 2.5 and 97.5 centile of the distribution with 90% CI based on Clinical and Laboratory Standards Institute guidelines (NCCLS 2000). Partition test between different groups was done by using quintile regression.

Results
A total of 734 children with mean age of 4.46± 2.36 were analysed. The study included 19 hematological parameters. There was no statistical significance difference for each parameter between male and female children. However, the difference observed for all 19 parameters between children aged six months to 5 years and those aged 6 to 10 years. Children from the current study had lower limit of heamoglobin, hematocrit levels, Platelet counts (PLT) and mean corpuscular volume than those of international range. On the other hand, white blood cell count reference values and upper limit of PLT were higher in Bagamoyo children than those of the international interval.

Conclusion
The study found that, if hematology results were assessed based on international intervals, approximately 67% of Bagamoyo children would have been considered are out of normal range. Therefore, study results will help to ensure routinely proper patients management, screening, establishment of toxicity tables and safety monitoring during clinical trials.

Modelling Cure Fractions for Complex Hazard Functions in a Cancer Study

'Peter Koleoso , [1]Angela Chukwu
'University of Ibadan

Background

In modelling recurrence of a particular disease, some patients could be cured and never have a recurrence. Hence, cure models can be described as special types of models in survival analysis where it is assumed that there are proportions of patients who will never experience event of interest and the survival curve will reach a plateau. Survival times are often generated through simulations by using the exponential distribution or Weibull distribution. Previous research in population-based cancer studies have shown that estimation of cure fractions by fitting cure fraction models using parametric distributions are very sensitive to the choice of parametric distribution especially when there is high excess hazard rate shortly after diagnosis. It is unclear when the simpler models using Weibull distribution are adequate and when the use results in bias. The reasons given above warranted the simulation study to answer relevant questions on the limitations of fitting cure fraction models using Weibull distribution and to investigate the level of biasness of using the distribution to model cure when the assumptions of the distribution are not satisfied.

Methods

The simulation study involved simulating from six scenarios with each representing different hazards. The motivating scenarios used for the simulations were used to assess the performance of mixture cure model with Weibull distribution in estimating cure proportions and median survival of the uncured, when fitted to data sets generated from more complex underlying hazard function (two-component Weibull distribution). The two-component Weibull distribution is more biologically plausible than the standard Weibull model. The two-component Weibull distribution has scale parameters; $\lambda 1$, $\lambda 2 > 0$, shape parameters; $\gamma 1$, $\gamma 2 > 0$ and mixing parameter p; $0 \leq p \leq 1$. For each of 1000 repetitions, 1000 survival times were simulated from the chosen scale ( ) and shape (γ) parameters of standard Weibull and two-component Weibull distributions for the first four scenarios. For the first scenario, survival times were simulated from a Weibull distribution with chosen parameters and a mixture cure model with Weibull distribution was fitted to the simulated result. This is to show that, by fitting a true model, the right estimates will be obtained. Likewise, survival times were simulated for chosen parameters of second, third and fourth scenarios and mixture cure models with Weibull distribution were fitted to the simulated results, so as to assess the performance of mixture cure model with Weibull distribution when fitted to data set generated from two-component Weibull distribution. The bias of the estimates was obtained and coverage was assessed.

Results

When cure proportion is 0.20, the models fitted for the scenarios converged. For scenario 1, mixture cure model with Weibull distribution produced unbiased estimates for the cure proportion (0.0003), the median survival time (0.0028) and the 90th percentile survival time (-0.0008). It also produced good coverage probabilities for cure proportion (95.3), the median (95.1) and 90th percentile survival times (95.5). The Weibull model produced unbiased estimate for the cure proportion with good coverage probability (94.9) for scenario 2. The cure model produced unbiased estimates of cure proportion with good with good coverage probabilities for scenarios 3, 4, 5 and 6. The estimates of median and 90th percentile survival times are generally highly biased with poor coverage probabilities. When cure proportion is 0.50, the models fitted also converged for each of the scenarios. The cure model in scenario

1 produced unbiased estimates for the cure proportion (-0.0002), the median survival time (0.0015) and the 90th percentile survival time (-0.0065). Also produced are the good coverage probabilities for cure proportion, the median and 90th percentile survival times. The same model produced unbiased estimate for the cure proportion (0.0005) with good coverage probability (94.3) for scenario 2. The cure model produced unbiased estimates of cure proportion with good coverage probabilities for scenarios 3, 4, 5 and 6. The estimates of median and 90th percentile survival times are generally highly biased for these scenarios, with poor coverage probabilities. When cure proportion is 0.90, only scenarios 1, 2, 5 and 6 converged.

Conclusion

From the results of fitting mixture cure models with a Weibull distribution to the data sets generated in the six scenarios, the estimates of the cure proportion, median and 90th percentile survival times for scenario 1 were generally unbiased with good coverage probabilities. This is the case in all the motivating cure proportions (0.2, 0.5 and 0.9). The model also produced unbiased estimates of cure proportion for all the scenarios and good coverage probabilities. For scenarios 2 to 6, general biased estimates and poor coverage probabilities were obtained for the median survival time and the 90th percentile survival time. More flexible models than the standard Weibull model could be used in this scenario to estimate the median survival and the time at which 90% of the patients in the uncured group would have died.

Solutions for selective loss to follow-up in HIV-cancer cohorts in Malawi

'Evaristar Kudowa [2]Maganizo Chagomerana, [3]Marie-Josephe Horner, [4]Jess Edwards, [5]Steady Chasimpha, [6]Charles Dzamalala, [7]Satish Gopal

'University of North Carolina Project-Malawi [2]University of North Carolina Project-Malawi, [3]University of North Carolina at Chapel Hill, [4]University of North Carolina at Chapel Hill, [5]Malawi Cancer Registry, [6]Malawi Cancer Registry, [7]Malawi Cancer Consortium & Regional Center of Research Excellence for Non-Communicable Diseases' Lilongwe

Background

Survival analysis is a statistical method that is used in clinical and public health research to assess time to an event of interest. Accurate estimation of survival helps to assess treatment efficacy and monitor disease progression over time. However, survival estimates among HIV+ cancer patients in sub-Saharan Africa (SSA) are rare and often biased due to high rates of loss to follow-up (LTFU). In the presence of informative (non-random) censoring and high rates of LTFU, standard approaches of estimating survival may lead to biased estimates. With informative censoring, the probability of censoring depends on the value that would be observed in absence of censoring. We applied inverse probability of censoring weights (IPCW) to routinely collected HIV cohort data in Malawi with the aim of correcting survival estimate bias introduced by LTFU.

Methods

We estimated overall survival (OS) among incident cancer cases who were HIV+ and newly initiated on antiretroviral therapy (ART) between 2000 and 2010 in Malawi. Cases were identified by probabilistically linking two large HIV cohorts to the national cancer registry, and follow-up data were captured from the HIV cohorts. We used unadjusted risk ratios (RR) to assess associations between patient characteristics and LTFU. LTFU patients were defined as those who missed a clinic appointment with unsuccessful tracing for ≥180 days, transferred out of HIV cohort, or discontinued ART. We defined follow-up time as the time from cancer diagnosis to death, LTFU or administrative censoring on 31st, October 2010. OS was estimated using unweighted and weighted methods. IPCW modeled the probability of being LTFU given the observed data and weighted observed individuals by the inverse of probability of not being LTFU. This created a pseudo-population that would be observed had there been no LTFU. Variables in the weighted method included age, sex, WHO stage for HIV, and cohort type. Sensitivity analysis using best- and worst-case scenarios was applied to evaluate the robustness of our estimates.

Results

Among 883 HIV+ cancer cases, median age at cancer diagnosis was 35 years (IQR: 30-41). The most common cancers were Kaposi sarcoma (89%) and cervical cancer (5%). Median follow-up time was 5 years (IQR: 1-6), and 257 (29%) cases met the LTFU definition. Male gender (RR: 1.69, CI: 1.37-2.07) and HIV stage 4 (RR: 2.24, CI: 1.74-2.89) were associated with increased risk of LTFU. Both weighted and unweighted 2-year OS were 91%. Sensitivity analysis estimated best- and worst-case 2-year OS at 91% and 83%, respectively.

Conclusion

The IPCW method failed to significantly correct for bias due to LTFU, resulting in consistently overestimated survival among HIV+ cancer patients in Malawi. This is likely due to limited covariate availability in routinely collected cohort data which are needed to generate appropriate censoring weights. In SSA and other low-resource setting where comprehensive patient tracing among LTFU groups can be impractical, intensive tracing of representative samples of LTFU patients may lead to more accurate survival estimates.

Bayesian Modelling of Tick Count Data Over Time and Space Using Discrete Model Approaches

'Thabo Lephoto [2] Henry Mwambi, [3]Oliver Bodhlyera
'University of KwaZulu Natal [2]University of KwaZulu Natal, [3]University of KwaZulu Natal

Background
There is a vast amount of Geo-referenced data in many field of study. Geo-referencing is usually by point referencing; that is, latitudes and longitudes or Arial referencing which include districts, counties, states, provinces and other administrative units. The availability of Geo-referenced data has necessitated the development and application of spatial statistical methods in analysis of geographically correlated data. In this study we look at tick counts or abundance variability across 44 regions over time in Virginia, United States

Methods
In this study the joint spatial semi parametric models will be applied to relax linearity assumptions in the covariates. The spatially varying models will also be applied to allow the effects of the covariates to vary spatially. The study will apply different count models to account for the effects of dispersion in the data. The analysis will be carried out using the R or RStudio software.

Results
Preliminary results showed that the abundance of larva is in July to October, nymph in March to September and adult counts in March to July. Also, more larva counts are found in woods, while nymphs are mostly found on the grass and woods, and adults' counts on edge.

Conclusion
Preliminary results imply that the distribution of the life-stage of a tick depends on time and the space as there were variations due to time and variations due to environment.

Temporal interactions of microbiota in longitudinal nasopharyngeal samples

'Maia Lesosky [2]Brian Rambau, [3]Heather Zar, [4]Mark Nicol, [5]Polite Nduru
'University of Cape Town [2]University of Cape Town, [3]University of Cape Town, [4]University of Cape Town, [5]University of Cape Town

Background
In the aetiology of bacterial pneumonia, it is generally accepted that infection is preceded by nasopharyngeal (NP) colonization. Several studies support that NP flora develop and establish early in childhood or in first year of life. Temporal organism interactions have not been explored, especially regarding whether observed colonization patterns are resultant of the host's immunity, synergism or antagonism between bacterial species interactions.

Methods
We sought to identify, describe and quantify the temporal interactions existing between selected key bacteria colonizing the nasopharynx in young children (up to 1 year of life), and to compare these patterns in children who go on to develop pneumonia versus those that do not. We applied multi-state models to examine the patterns of transition among states of colonisation while accounting for key confounders.

Results
There were 760 individuals included in the analysis, with a total of 16,346 NP samples available and a median 364 person-days (IQR 346 – 365 person-days). There were temporally sustained positive interactions between S pneumoniae with H influenzae, S pneumoniae with M catarrhalis, and H influenzae with M catarrhalis. Moreover, the magnitude of association generally decreased with child age. S aureus had consistent negative interactions with other organisms. Associations varied by lower respiratory tract infection status.

Conclusion
We have developed a flexible framework that characterised interactions between organisms while accounting for key time varying confounders. We have demonstrated that multi-state models provide a useful approach to supplement classical modelling approaches.

A pairwise joint modelling approach for multivariate longitudinal immuno-epidemiological data from the Infant BCG study in Entebbe, Uganda

'Lawrence Lubyayi [2]Emily Webb, [3]Alison Elliott, [4]Jonathan Levin
'University of the Witwatersrand, Johannesburg [2]London School of Hygiene and Tropical Medicine,
[3]MRC/UVRI and LSHTM Uganda Research Unit, [4]University of the Witwatersrand, Johannesburg

Background

My PhD project aims to investigate and to extend appropriate statistical modelling methods for longitudinal data, and their application to longitudinal multiple-outcome data from immuno-epidemiological studies. The aim of this particular analysis was to investigate the pairwise modelling approach of multivariate longitudinal data to understand the dynamics of vaccine response to Bacille Calmette-Guerine (BCG) in infants, using data from the Infant BCG cohort Study (IBS).Many challenges arise when analysing multivariate longitudinal data. Naturally, the multiple outcomes at multiple time points have complex random-error (association) structures that should be properly accounted for during analysis. This study provides an illustration of appropriate approaches for multivariate longitudinal immuno-epidemiological data.

Methods

The IBS was an observational longitudinal study, carried out under the Immunomodulation and Vaccines Programme at the MRC/UVRI and LSHTM Uganda Research Unit. The study aimed to examine the timing, magnitude and quality of the initial response to BCG immunization and determine the effect of prenatal exposure to maternal latent tuberculosis infection. Infants were followed up to age one year in order to assess the establishment of BCG-specific immunological memory. Blood samples were taken at selected time points throughout infancy. Infants were randomly assigned to two separate sampling strategies: 150 infants (75 from mothers with latent Mycobacterium tuberculosis infection (LTBI), and 75 from mothers without LTBI) to give blood at 1, 6, 14 and 52 weeks while the remaining infants were assigned to give blood at 4, 10, 24 and 52 weeks. Immunological parameters, including 17 cytokine/chemokine and antibody responses to vaccine-specific antigens, were assessed at each of these time points. The current analysis focuses on a joint model for the analysis of 7 cytokine responses, which are predominantly of T-cell origin. A general linear mixed model is used to describe the longitudinal profile of each cytokine response. In order to account for the correlation between the cytokine responses, the univariate models are combined into a multivariate mixed model by specifying a joint distribution for the random effects. Due to the complexity of model fitting as a result of the high number of cytokine responses, a pairwise modelling approach, where all possible pairs of bivariate mixed models are fitted, is used to obtain parameter estimates.

Results

Analysis is still ongoing and we anticipate that results from the approach described, in the methods section above, will be ready by the meeting date. From this analysis, we hope to get an understanding of the evolution of the infant cytokine responses following immunisation with BCG. We also hope to get an idea of which cytokines are more related in terms of their evolutions over time.

Conclusion

The pairwise joint modelling approach for multivariate longitudinal data has utility for a wide variety of immuno-epidemiological data. It reduces the complexity of analysis of multivariate repeated measures of a relatively high dimension and allows for proper accounting for association structures via specification of a joint distribution for random effects.

Mortality rate and associated factors among preterm babies born in Moshi Municipality in northern Tanzania

'Michael Johnson Mahande , [1]Rose Chengo, [1]Frida Mowo
'Kilimanjaro Christian Medical University College

Background
Globally, approximately 15 million babies are born before term each year. Of these, more than 1 million die within the first 28 days of their life. Understanding the mortality rate and its predictors during neonatal period among preterm babies is crucial to help designing interventions to avert the situation. This study aimed to determine the neonatal mortality rate and associated factors among preterm babies born in Moshi Municipality, Tanzania.

Methods
A prospective cohort study was conducted in three hospitals in Moshi Municipality from December 2016 to May 2017. All live births at gestational age of <37 weeks and those of <24 hours were studied. Babies who died prior to gestation age assessment and those whose mother did not consent were excluded. Cox regression model was used to estimate maternal and fetal factors associated with neonatal mortality. A p-value of <0.05 was considered statistically significant.

Results
A total of 311 of preterm babies were recruited from 265 mothers and were followed for 28 days. The neonatal mortality rate was 6.5deaths per 1,000 preterm live births (95% CI: 4.83-8.61). It was higher among extremely preterm babies compared to very preterm ones (HR: 38.24; 95% CI: 16.62-87.96) versus (HR: 8.01; 95% CI: 3.96-16.20) respectively. Apgar score of <7 at 1st minute (HR: 14.03; 95% CI: 7.27-27.06), respiratory distress syndrome (HR: 8.14; 95% CI: 4.27-15.54) and antepartum hemorrhage (HR: 3.32; 95% CI: 1.49-7.39) were significantly associated with neonatal mortality.

Conclusion
Preterm birth complication is the major cause of neonatal death in the study setting. Interventions to address the identified risk factors may reduce neonatal mortality among preterm babies.

Modelling predictors of stroke disease in South Africa: Bayesian binary quantile regression approach

'Lyness Matizirofa [2]Edmore Ranganai
'University of Johannesburg [2]UNISA

Background

Stroke is currently the second prevalent cause of death worldwide. The burden of stroke is emerging as a root of preventable death and disability in adults in Sub-Saharan Africa. South Africa (SA) is experiencing an epidemiological transition due to sociodemographic and lifestyle changes leading to an increase of non-communicable diseases, which in turn may result in an upswing of stroke cases. Stroke variable is binary; therefore, binary Bayesian quantile regression (BQR) technique was used. The purpose of this paper is to address important gaps in the stroke literature, modelling predictors of stroke and estimating linear quantile models when predictors are measured with error. The objectives of this study is to assess the effects of predictors of stroke for different quantiles for adults stroke patients and to estimate linear quantile regression models when predictors are measured with error was achieved through BQR analysis. BQR was applied to stroke data collected between 2014 and 2018 in SA. QR will provide a complete relationship between the outcome and covariates, since it allows examining the relationship in different parts of the distribution.

Methods

A cross-sectional study design was used to model the predictors of stroke disease incidences in SA. Reason being that , this is a descriptive epidemiologic study in which the exposure and disease status of the population are determined at a given point and aimed at obtaining immediate knowledge about the predictors of stroke. Research data was retrieved from the South African public and private sampled hospitals .The research instrument was validated before use. Computed tomographic (CT) or magnetic imaging scan (MRI) confirmed stroke cases. The proportion of 33% of the 203 private hospitals and 67% of the 407 public hospitals was sampled for the study. Stratified sampling technique was used to calculate the proportions accordingly. The proportions were calculated based on the total number of public and private hospitals in SA. Research data were collected for stroke patients who were 18 years and above. The sample size of stroke patients was 35730.Permission to conduct this research was obtained from individual hospitals and the committee of research on human subjects of the University of South Africa. The ethical clearance number is 2017/SSR-ERC/001.Data analysis was done in R version 3.5.4. The convergence of the MCMC chains was checked using trace and density plots of the different marginal distributions. The BQR focuses on conditional quintiles of Y given X rather than the conditional mean of Y given X, which can obtain robust analysis and leads to richer view of how covariates influence the response. We considered 10%, 25%, 50%, 75% and 95% quantiles to allow making comparison of the effects of predictors of stroke in the lower and upper quantiles.

Results

Demographic information indicates that ,64% of the stroke patients were married and 4% were divorced. Additionally, 34% blacks and 9% were Indians. Majority of patients were residing in urban areas (95%).The mean age was 54 years. Half of the stroke patients were in the age group 18-54 (55%) which is an indication of young strokes. The BQR estimates and 95% Credible Intervals for some predictors of stroke shows that ,Cholesterol and diabetes has positive statistically significant estimated quantile regression estimates

across all quantiles .The Bayes estimate effect on stroke is larger at the upper and central locations than the lower locations. Heart-problems has a negative statistically significant estimated conditional quantile functions with stroke across the quantiles

Conclusion

Gender, hypertension, cholesterol, diabetes, heart-problems and race were significant risk factors of stroke. However, modifiable risk factors such as hypertension, cholesterol and diabetes were identified as the major risk factors for this study. These factors can be managed by avoiding gaining excessive weight, reducing salt, and sugar intake, and by exercising regularly. The majority of South African children were found to be physically inactive due playing indoor games. Younger age stroke incidences due to hypertension was high, since SA has high unemployment rate among the youth group, it is highly likely that these young adults are suffering job stress, which may lead to increased hypertension. These study findings call for programs and policies meant to curb unhealthy lifestyles, promoting physical activity and health nutrition among adults. We recommend setting up of wellness centers at schools, tertiary institutions, work places and communities where people can be tested regularly on hypertension, cholesterol and diabetes level. Health professional should then raise awareness of the dangers of these risk factors. Further, government may increase value added tax on cigarettes, alcohol and other unhealthy foods in an effort to discourage consumption of unhealthy food and smoking. Moreover, Cholesterol and diabetes has positive effect across all quantiles whilst heart-problems has negative effect across all quantiles. The Bayes estimate effect on stroke is larger at the upper and central locations than the lower locations. Heart-problems has a negative statistically significant estimated conditional quantile functions with stroke across the quantiles.

The association between childhood environmental exposures and the subsequent development of Crohn's disease in the Western Cape, South Africa

'Mikateko Mazinu [2]Abigail Basson, [3]Esme Jordan
'SAMRC [2]UWC, [3]SAMRC

Background

Environmental factors during childhood are thought to play a role in the aetiolgy of Crohn's Disease (CD). However, the association between age at time of exposure and the subsequent development of CD in South Africa is unknown. The aim of this study was to investigate the association between a number of pre-defined environmental risk factors during the 3 age intervals: 0-5, 6-10 and 11-18 years, with the risk of CD development. To address the research question, logistic regression additive effect models was presented to evaluate the association

Methods

This is a case control study of all CD patients seen at two large inflammatory bowel disease referral centers in the Western Cape, between September 2011 and January 2013. Numerous environmental exposures during 3 age intervals were extracted using an investigator-administered questionnaire. Multiple logistic regression models were conducted to assess environmental risk factors and their impact on CD for separate age intervals. Risk factors that were significant for ($P < 0.05$) for a specific age interval were included in the three final models. Exact logistic regression was used for modeling with small cell sizes.

Results

This study included 194 CD patients and 213 controls. On multiple logistic regression analysis, a number of childhood environmental exposures during the 3 age intervals were significantly associated with the risk of developing CD. During the age interval 0-5 years, an increased risk-association was observed in subjects who never consumed unpasteurized milk (OR = 8.02; 95% CI, 3.19-23.28) and second-hand cigarette smoke exposure (OR = 1.71; 95% CI, 1.01-2.94). During the age interval 6-10 years, never having consumed unpasteurized milk (OR = 5.84; 95% CI, 2.73-13.53) significantly increased the risk of developing future CD. During the age interval 11-18 years, an independent risk-association was identified for; never having consumed unpasteurized milk (OR = 2.60; 95% CI, 1.17-6.10) and second-hand cigarette smoke exposure (OR = 1.93; 95% CI, 1.13-3.35).

Conclusion

This study demonstrates that both limited microbial exposures and exposure to second-hand cigarette smoke during childhood is associated with future development of CD.

Analysis of a vaccination mathematical model of an infectious measles disease

[1]Mbachu, Hope Ifeyinwa,[2]Inyama, Simeon Chioma, [3]Omame, Andrew
[1]Department of Statistics, Imo State University, Owerri, Imo State, Nigeria
[2, 3]Department of Mathematics, Federal University of Technology, Owerri, Imo State, Nigeria

Analysis of a vaccination Mathematical model of an infectious Measles disease was carried out. SVEIR epidemic model was investigated and incidence rate was considered. Its formulation and analytical study showed two equilibrium points (disease free equilibrium (DFE) and endemic equilibrium (EE)). Model proved positivity of solutions and obtained the basic reproduction number for determining whether disease dies out completely or not. The local stability of disease-free equilibrium was proved, and determined by the basic reproduction number. Lyapunov function method proved DFE and EE globally asymptotically stable. Numerical simulations were carried out. Simulation result showed the number of susceptible, infected and vaccinated individuals is consistent.

Predictors of perinatal deaths in the presence of missing data: A registry based study in northern Tanzania

'Innocent B. Mboya [2]Michael J. Mahande, [3]Joseph Obure, [4]Henry G. Mwambi
'University of University of KwaZulu-Natal [2]Kilimanjaro Christian Medical University College (KCMUCo),
[3]Kilimanjaro Christian Medical Center (KCMC), [4]University of KwaZulu-Natal

Background

More than 5 million perinatal deaths (number of stillbirths and deaths in the first week of life) occur each year globally, Sub-Saharan Africa being among the regions with the highest burden. Despite all efforts put forward during MDG era, perinatal and newborn deaths continues to increase relative to under-five deaths especially in resource limited countries. We aimed to determine predictors of perinatal death in northern Tanzania using a zonal hospital birth registry data between 2000 and 2017.

Methods

This was a hospital-based cohort study based on the KCMC birth registry (established since year 2000) located in KCMC zonal hospital, in northern Tanzania. The hospital has an average of 4,000 deliveries per year which is about 70,000 deliveries for about 17 years. Data was analyzed using Stata version 15.1 StataCorp LLC. Multiple imputation by chained equation (MICE) method was used to impute the missing data on covariates using data augmentation technique. Generalized estimating equations (GEE) was used to estimate the marginal effects of covariates on perinatal death using the log link function. Exchangeable correlation structure was used to account for dependence of observations within a mother.

Results

Among 32,189 recorded deliveries by year 2010, 1,334 (4.1%) were perinatal deaths. Highest risk of perinatal deaths were observed among women with pre-eclampsia (RR 1.78, 95%CI 1.45, 2.16), <4 antenatal care visits (RR 1.28, 95%CI 1.23, 1.46), preterm birth (RR 2.79, 95%CI 2.40, 3.23), anemia (RR 1.47, 95%CI 1.04, 2.08), abruption placenta (RR 32.48, 95%CI 19.97, 52.84) and those who delivered a low birth weight baby (RR 5.51, 95%CI 4.77, 6.37). Demographic characteristics associated with perinatal deaths included mother's age, urban residence, occupation and maternal body mass index.

Conclusion

About 4% of all deliveries at KCMC between 2000 and 2017 were perinatal death. Both maternal conditions and demographic characteristics are associated with high risk of death. We recommend early identification and care of high risk pregnancies especially during antenatal, labor and delivery as well as postnatal period.

Modelling CD4 Count and Mortality in a Cohort of Patients Initiated on HAART

'Nobuhle Mchunu [2]Tarylee Reddy, [3]Nonhlanhla Yende-Zuma, [4]Henry Mwambi
'UKZN [2]SAMRC, [3]CAPRISA, [4]UKZN

Background

Longitudinally measured data and time-to-event or survival data are often associated in some ways, and are traditionally analyzed separately (Asar et al., 2015). However, separate analyses are not applicable in this case because they may lead to inefficient or biased results. To remedy this, joint models optimally incorporate all available information (longitudinal and survival data) simultaneously (Wulfsohn & Tsiatis, 1997). Furthermore incorporating all sources of data improves the predictive capability of the joint model and lead to more informative inferences for the purpose of decision-making (Seyoum & Temesgen, 2017). The primary goal of this analysis was to determine the effect of repeatedly measured CD4 counts on mortality. The standard time-to-event models require that the time-dependent covariates of interest are external; where the value of the covariate at a future time point is not affected by the occurrence of the event. This requirement would not be fulfilled in this setting, since the repeatedly measured outcome is directly related to the mortality mechanism. Hence, a joint modeling approach was required.

Methods

This thesis undertook a retrospective review of medical records of patients that accessed HIV services at the CAPRISA AIDS Treatment program (CAT) between June 2004 to December 2013, HIV infected patients who were, 14 years or older were enrolled. We conducted two sets of analysis: separate models for mortality and CD4 count were fitted, followed by the joint model for both outcomes. Statistical analysis was done using SAS (version 9.4.; SAS Institute Inc., Cary, NC, USA) and R version 3.5.1.

Results

Out of the 4014 patients, 1457 (36.30%) were males. There were more patients presenting without TB at ART initiation, 3042 (75.78%) compared to those with prevalent TB, 972 (24.22%). Results from the multivariable random effects model showed that the patient's gender, age, baseline viral load and baseline CD8 cell count had statistically significant influences on the rate of change in CD4 cell count over time. The joint model found a significant association between the CD4 cell count and the risk for death (HR= 0.73, 95% CI: 0.68-0.77, p < 0:0001), with a higher CD4 count being associated with a lower risk of death.

Conclusion

CD4 cell count proved to be a significantly associated with mortality, after adjusting for age, gender and other potential confounders. An area of future work is the use of dynamic prediction of mortality using CD4 count measurements.

Modelling the effect of rape on mental health status one year post enrollment: RICE study

'Shibe Mhlongo [2]Carl Lombard, [3]Naeemah Abrahams
'South African Medical Research Council [2]South African Medical Research Council, [3]South African Medical Research Council

Background
Post rape psychological trauma is well described in literature, mostly through cross-sectional studies. The Rape Impact Cohort Evaluation (RICE) study is a comparative cohort study that aims to advance our understanding on the medium-term and long-term health consequences of rape and of particular interest is the mental health impact. The study is ongoing and seeks to recruit rape exposed women at post rape care services and non-exposed women from family clinics. The data is collected at baseline, 3, 6, 9 and 12 months and thereafter every 6 months until 36 months

Methods
The depression score from the Centre for Epidemiology Studies Depression (CES-D) scale, is the outcome that will be used to model the mean scores over time and used to compare the profile of the two exposure groups, 12 months post enrolment. The longitudinal analysis was carried out using a mixed linear spline model and the missing data techniques applied include full maximum likelihood, multiple imputation and inverse probability weighting.

Results
A total of 1042 participants that had completed their 12 months follow up visit was analysed, 509 (49%) in the rape exposed group and 533 (51%) in the non-exposed group. The mixed linear spline model shows a decline in mean CES-D score for this time period: -0.168 units/day (95% CI: -0.182 to -0.154 units/day) in the raped exposed group compared to -0.011 units/day (95% CI: -0.023 to 0.001 units/day) in the non-exposed group. The estimated fixed effects from all three missing data methods shows very little differences

Conclusion
Rape impacts on women's mental health, there are long term persistent effects, but much smaller than the immediate effects. The initial findings provide solutions for setting up a practical regression model for the modelling of depression in the two exposure groups given the missing data structure and trends over time. This should also serve as basis for continuing research and be incorporated in finding interventions for managing trauma in rape survivors.

The impact of correlated socio-demographic factors in Spatial profile of Diabetes in Africa with Missingness

'Sakhile Mnguni [2]Sakhile Mnguni, [3]Sileshi Melesse, [4]Henry Mwambi

'University of KwaZulu Natal [2]University of KwaZulu-Natal, [3]University of KwaZulu Natal, [4]University of KwaZulu Natal

Background

Diabetes is a serious, long-term (chronic) disease that occurs either when the pancreas does not produce enough insulin (a hormone that regulates blood glucose), or when the body cannot effectively use the insulin it produces. There were 1.5 million deaths worldwide in 2010 directly caused by diabetes. From around 450 million people living with diabetes in the world, more than16 million people are from the African region. The aim of this study is to estimate the prevalence of diabetes in African countries, with several contributing factors.

Methods

A meta-data for the year 2016 was extracted from the World Databank and World Health Organisation(WHO) database. African Shapefile will be used in R, with library packages map tools to geographically show the Diabetes prevalence with the impact of socioeconomic factors, health resources, and lifestyle factors. Also, Moran Index, Geary's C and Gestis-Ord G will be used in R for diabetes pattern to show evidence of spatial autocorrelation.

Results

I am still working on results now. I will finish my results and discussions before the end of July 2019.

Conclusion

Hence they will be no conclusions.

The effect of cotrimoxazole prophylaxis on CD4 cell count profiles in HIV positive patients stabilized on anti retroviral therapy: analysis of data from the COSTOP trial

[1]Tshwaraganang Modise, [2]Zacchaeus Anywaine
[1]Department of Health Sciences, Witwatersrand University School of Public Health, [2]Africa Health Research Institute, Somkhele, KwaZulu-Natal, South Africa

Background
Cotrimoxazole (CTX) prophylaxis has been successfully used for many years in preventing opportunistic infections in patients living with HIV. There is lack of data from controlled trials on the effects of CTX on CD4 cell count recovery among patients stabilized on anti retroviral therapy. This study evaluated the effects of discontinuing CTX on CD4 profiles of HIV positive patients stabilized on antiretroviral therapy in Uganda.

Methods
The study used data from the COSTOP double blind placebo controlled trial which ran in two sites, Entebbe and Masaka, in Uganda from 2011 to 2014. The study enrolled participants aged 18 years and above who were stabilized on ART, with a CD4 cell count of at least 250 cells/〖mm〗^3 and who had taken CTX prophylaxis for at least six months before enrollment. Participants were randomized to continue taking CTX prophylaxis (960 mg once per day) or to stop taking CTX and take a matching placebo. Participants were seen monthly for the first 3 months and thereafter every three months. The CD4 cell count was measured at each visit. The data were analyzed using linear mixed models for longitudinal data. A random intercept model was fitted to compare average CD4 count profiles between participants randomized to receive placebo and participants randomized to continue CTX. A further analysis was carried out using a random slopes model to determine the average difference in the slopes between participant randomized to CTX and those randomized to discontinue CTX. The model included treatment arm as the fixed effect component, adjusting for baseline CD4 count, study site, age, sex and duration on treatment. All participants with at least one CD4 cell count measure after baseline and who had taken study drug for at least one month were included in an Intention to treat (ITT) analysis.

Results
The analysis included data for 2132 participants with at least one post enrolment CD4 result, 1068 randomized to CTX and 1064 randomized to placebo. Participants who received placebo had a significantly higher final CD4 count [mean difference=37.3 cells/〖mm〗^3 (95% CI 31.4;43.2, p<0.001)] compared to those who continued CTX. Among the random intercept models, the best fitting model included a treatment arm by time interaction which showed that the increase in CD4 counts was significantly higher by 0.24 cells/〖mm〗^3 per week for participants randomized to placebo compared to the CTX arm (95% CI 0.15;0.34, P<0.001). The estimated mean increase for participants in the placebo arm was 0.55 cells/〖mm〗^3 per week (28.6 cells/〖mm〗^3 over a one year period) while for participants in the CTX arm the estimated mean increase was 0.34 cells/〖mm〗^3 per week (17.7 cells/〖mm〗^3 over a one year period). In the corresponding random slopes model the weekly difference in slopes was reduced slightly to an increase of 0.21 cells/〖mm〗^3 for participants randomized to placebo compared to the CTX arm (95 % CI 0.09;0.33, p<0.001).

Conclusion

This study showed that discontinuing CTX prophylaxis leads to a significantly higher rate of increase in CD4 cell count compared to continuing CTX prophylaxis. This finding is consistent with the main results from the COSTOP trial, which showed that continued CTX prophylaxis is beneficial for ART stable HIV infected patients who are at high risk for severe bacterial infections or malaria, but that this benefit needs to be weighed against the risk of haematologic toxicity, and patients eligible for CTX prophylaxis should have access to CD4 and complete blood count haematological side effects monitoring.

A Comparison of Cancer Classification Methods based on Microarray Data

'Mohanad Mohammed [2]Henry Mwambi, [3]Bernard Omolo
'University of KwaZulu-Natal [2]University of KwaZulu-Natal, [3]University of South Carolina - Upstate

Background
Cancer is among the leading causes of death in both developed and developing countries. Through gene expression profiling of tumors, the accuracy of cancer classification has been enhanced, leading to correct diagnoses and the application of effective therapies.

Methods
Here, we discuss a comparative review of the predictive ability of seven classification methods (support vector machines, with the radial basis kernel (SVM(RK)), linear kernel (SVM(LK)) and the polynomial kernel (SVM(PK)), artificial neural networks (ANN), random forests (RF), k-nearest neighbor (KNN), and naive Bayes (NB)), using publicly-available gene expression data from cancer research.

Results
Results indicate that NB outperformed the other methods in terms of the accuracy, sensitivity, specificity, kappa coefficient, area under the curve (AUC), and balanced error rate (BER) of the binary classifier. We recommend NB as the \gold standard" for cancer classification using microarray data.

Conclusion
We used microarray data for classification of cancer. In the future research we are going to study the RNA-Seq data which is discrete in nature and widely used for classification using discrete classification methods such as negative binomial linear discriminant analysis (NBLDA) and Poisson linear discriminant analysis (PLDA).

Cost-effective, cross-sectional cohort estimation of HIV incidence rate in presence of misclassification error using doubly robust estimation

'Kesaobaka Molebatsi [2]Lesego Gabaitiri, [3]Lucky Mokgatlhe, [4]Eric Tchetgen Tchetgen
'University of Botswana [2]Statistics Department' University of Botswana, [3]Statistics Department' University of Botswana, [4]Statistics Department' The Wharton School' University of Pennsylvania

Background

In our previous work, we have established that accurate estimates of HIV incidence rate can be obtained by combining error-prone self-reports of date of prior HIV negative test, with validated individuals' testing history under a cross-sectional cohort study design, while appropriately accounting simultaneously for misclassification error and selection bias. In order to also account for potential selection bias into the validated sample, we have previously used inverse probability weighting (IPW) methods, which depend on correct specification of the unknown probability of selection (PS) into the validated subsample. It is well known that if model PS is mis-specified, such estimator will generally be biased.

Methods

We propose an augmented inverse probability weighting (AIPW) estimator that additionally requires fitting a regression model (RM) for the underlying population full data with the advantage that the resulting estimator is guaranteed to remain consistent and asymptotically normal if either model PS or RM is correctly specified. Our approach therefore extends the theory of double robustness to the cross-sectional cohort study design. To investigate large sample properties of AIPW, we perform extensive Monte Carlo simulations and compare validated-sample only versus pooled-sample versions of our estimators over a variety of settings. We then apply our methods to estimate HIV incidence rate among individuals who tested HIV negative 1.5 and 5 years, respectively, prior to Botswana Combination Prevention Project baseline household survey.

Results

Results from our extensive Monte Carlo simulations show that our estimator is more efficient and robust to selection bias, misclassification error and model misspecification.

Conclusion

Our methods clearly demonstrate that highly accurate estimates of HIV incidence can be obtained by carefully simultaneously accounting for multiple potential sources of bias under the cost-effective cross-sectional cohort study design.

Analytical methods used in handling missing data in estimating prevalence of HIV/AIDS for demographic and cross-sectional surveys: A systematic review

'Neema R. Mosha [3]Prof.Taryn Young, [4]Dr'Omololu Aluko, [5]Prof.Rhoderick Machekano, [6]Prof.Jim Todd
''. Mwanza Intervention Trials Unit'National Institute for Medical Research' Mwanza-Tanzania' [2].
Stellenbosch University' Faculty of Medicine and Health Sciences' South Africa [3]Stellenbosch University'
Faculty of Medicine and Health Sciences' South Africa, [4]Stellenbosch University' Faculty of Medicine and
Health Sciences' South Africa, [5]Stellenbosch University' Faculty of Medicine and Health Sciences' South
Africa, [6]London School of Hygiene and Tropical Medicine' Keppel Street-London-UK

Background

Demographic or cross-sectional surveys usually experience a problem of missing data due to unit or item non-response. However, few studies report the proportion of missing data and even fewer describe the methods used to adjust for missing data. Most of the published articles providing HIV prevalence estimates are based on the use of complete case analysis, which can produce biased results and reduction of study precision. The availability of advanced methods such as imputations, double robust methods, instrumental variables and Heckman's selection models have proved to produce a less biased and precise estimates on estimating HIV prevalence with missing values problem. With the increase in number of demographic and cross-sectional surveys conducted, and reported in peer-reviewed journals, it is important for editors and authors to be clear on how missing data is addressed. The main objective of this review was to determine the analytical methods used in handling missing data in estimating prevalence of HIV/AIDS for population, demographic and cross-sectional surveys.

Methods

We searched for studies from population, demographic and cross-sectional surveys published from January 2000 to April 2018 from the following electronic databases: Pub Med/Medline, Web of Science, Latin American and Caribbean Sciences Literature, Africa -Wide Information and Scopus, and by reviewing references from included articles. All potential studies were imported to Covidence, titles and abstracts were screened, and full text assessed by two independent reviewers using pre-specified criteria. Disagreements were resolved through discussion. A tool adapted from Hoy et al ,2012 was used to assess the risk of bias and a piloted data extraction tool was used to extract data from the included studies. Data was analysed through quantitative approach, variables were presented and summarized using flow charts, tables and graphs. Further analysis including meta-analysis will be done on the collected data.

Results

A total of 3426 citations where identified, 194 duplicates were removed, 3232 screened and 69 full text studies obtained. Twenty-five studies were included for analysis while 44 studies excluded (Not surveys (16), Did not measure HIV prevalence (14), Methodological study (8), No missing data methods used during analysis (3) and Duplicates (3)). Of the 25 included studies (Demographic Health Surveys 14 (55%), Cross sectional surveys 7 (28%), 3 (12%) population surveys and 1 (4%) ANC survey). The age of the participants ranged from 15 to 64, and most studies done in sub-Saharan Africa. All studies used more that 1 method of analysis, three methods was the maximum number. Complete case analysis used as a primary method for all the studies. The most used method to adjust for missing data was Multiple Imputation 11(44%), and Heckman's selection models 9(36%). Single Imputation and Instrumental variables accounted for 2(8%) studies each. Only 4(16%) studies conducted a sensitivity analysis to evaluate the

robustness of the result.

Conclusion

All methods used for handling missing data in the included studies produced different estimates from the primary analysis, although in some studies the difference was not statistically significant. These differences highlight the need of considering using more advance methods when facing the problem of missing data in surveys to avoid producing biased results. As these estimates are a major source of information in planning interventions, treatment, allocation of resources and policy making.

Impact of improved drinking water source and sanitation, and urban residence on child diarrhoea in Sub Saharan Africa. A meta analysis of demographic and health surveys

[1]Samuel Manda, [2]Rejoice Msiska
[1]South African Medical Research Council, [2]Department of Mathematical Sciences, Chancellor College

Background
Several studies in Sub Saharan Africa have shown higher prevalence of diarrheal diseases in under five children living in households with unimproved drinking water sources, unimproved sanitation facility and located in rural areas. Randomised controlled trials comparing childhood diarrhoea across levels of drinking water sources and sanitation facility types are rare and subject to ethical problems. This study conducted a meta analysis of demographic and health surveys to examine effectiveness of household level source of drinking water and sanitation in combating childhood diarrhoea.

Methods
The summary data for 35 demographic health surveys (DHS) conducted between 2000 and 2016 in 11 sub Saharan countries were analysed. The pooled odds ratio of diarrhoea among children below 5 years of age was estimated using random effects meta analysis. The heterogeneity among the DHS data was tested by Q statistic and quantified by I2 statistic. Sensitivity analyses were conducted to examine the effect of omitting a particular DHS data in the analysis. Both meta regression and subgroup analyses were used to explain the source of observed heterogeneity.

Results
The pooled estimates across the studied countries showed that children living in households with improved source of drinking water were associated with lower odds of diarrhoea compared to those living in households with unimproved drinking water source (OR = 0.79, 95%CI = (0.77, 0.81)). Again, children from households with improved toilet facility had lower odds of diarrhoea compared to those from household with poor toilet facility (OR =0.81, 95%CI = (0.79, 0.84)). As regards to location type, children in urban areas of the studied countries were associated with lower odds of diarrhoea compared to those in rural areas (OR = 0.911, 95% CI =(0.89, 0.93)). The studies showed considerable amount of heterogeneity (I2 > 80%). When sub Saharan Africa was divided into East, West and Southern tip regions, the findings on effects of water sources and type of toilet facility were not different from the overall estimates. However, with place of residence, the results matched with overall estimates only in East and West sub African regions, and they were different in Southern tip of the region where the odds of diarrhoea were higher in children from urban than rural areas although the difference was not statistically significant (OR= 1.02, 95% CI= (0.98,1.09)). On effect of period of study, both meta regression and sub group analyses showed that the period (in years) in which the survey was conducted was a major source of heterogeneity in the effect sizes. Sensitivity analysis revealed that no single study had substantial influence on the pooled effect size.

Conclusion
The study has demonstrated that in the absence of randomised controlled trials, observational studies could be used to generate evidence on importance of notable risk factors to childhood diarrhoea across sub Saharan African region. The findings imply that ensuring good sanitation access at household level across all areas will greatly lower morbidity and mortality in children due to diarrhoea in the region.

Prediction models for newborn complications at birth: comparing multinomial logit models and supervised machine learning algorithm Naïve-Bayes classification methods

'Paul Mubiri [1]Joseph Akuze, [1]Frank Namugera, [1]Gertrude Namazzi, [1]Peter Waiswa, [1]Geraldine Agireambabazi
'Makerere University School of Public Health

BACKGROUND
Globally, preterm birth complications account for about 28% respectively of most neonatal deaths. Complications following preterm birth are a major cause of morbidity and mortality. Preemies are between 6 and 26 times more likely to die during the first four weeks of their lives than term newborns. However, the mechanism of understanding when these complications occur is lacking. Mathematical and statistical algorithms can be used to predict the risk infection, re-admission, development of complication and adverse outcomes including neonatal mortality among infants and newborns.

METHODS
This analysis aims to develop algorithms to predict newborn complications at birth. This study will test two modelling strategies for developing prediction algorithms for newborn complications using data from the preterm birth initiative and newborn complications study enrolling preemies in Eastern Uganda and following them for 28 days from the date of birth. The study is enrolling 600 neonates. Strategy one involves using multinomial logit models to classify the causation agents for complication among the preemies. In strategy two, Naives-Bayes classification methods using machine learning algorithms will be used with the assumption that there is no multicollinearity among covariates.

RESULTS
The study is still enrolling.

CONCLUSION
The results from the two strategies will be compared and conclude on the more robust method of predicting newborn complications.

Modeling proportion of infected tsetse flies over time in Kajiado and Narok districts, Kenya

'Caroline Wanja Mugo, [2]Samuel Mwalili, [3]Roel Braekers
'JKUAT [2]JKUAT, [3]Hasselt University

Background
Trypanosomiasis is a neglected vector-borne disease caused by trypanosomes that affects both human and animals in sub-saharan Africa. The vector of trypanosomes are tsetse flies which inhabit a large proportion of Africa and their distribution is greatly influenced by climate. Different species of these flies exist. The effect of the flies in the incidence of trypanosomiasis is modeled as the tsetse challenge which is the product of the tsetse density, the prevalence of the trypanosome in dissected tese flies and the proportion of infected blood meals from the host. With the prevalence of the disease in the flies being one of the major components, it is important to understand the population dynamics of the vector as it may aid in the development of effective and economic vector control strategies.In this study the prevalence of trypanosomiasis on tsetse flies over time was modelled to account for the spatial heterogeneity.

Methods
The data used for this analysis was sourced from a farm trial that was conducted to test the effect of using a repellent to control the flies between August 2004 and August 2005 in Narok and Kajiado districts of Kenya. To extract patterns of the proportion infected over time, the generalized additive mixed model (GAMM), a semi-parametric model was constructed by incorporating smoothing in the modelling process.

Results
The species of tsetse flies in the two regions were different.The proportion of infected tsetse flies was higher in Narok district compared to Kajiado district. The smoothed curves of the proportion infected over time were observed to have two peaks in the Narok region while for Kajiado it was constantly low over time. These variation in trypanosomiasis prevalence in the flies In Narok was observed between the dry season in March and the late wet season in October.

Conclusion
Since the proportion of infected tsetse flies was high in March and the late wet season in October for Narok then would be effective to have more tsetse control strategies at these periods to reduce the transmission of the trypanosome.

Developing Socio-Economic Status (SES) index for Basse region in The Gambia

'Abdul Khalie Muhammad [3]Miriam W. Wathuo, [4]Christian H. Hansen, [6]Momodou Jasseh, [7]Kathy Baisley, [8]John Bradley, [9]David Jeffries, '0Nuredin I. Mohammed
'Medical Research Council Unit The Gambia at London School of Hygiene & Tropical Medicine, [3]Medical Research Council Unit The Gambia at London School of Hygiene & Tropical Medicine, [4]'. Medical Research Council/ Uganda Virus Research Institute and London School of Hygiene & Tropical Medicine; [2]. London School of Hygiene & Tropical Medicine, [6]Medical Research Council Unit The Gambia at London School of Hygiene & Tropical Medicine, [7]London School of Hygiene & Tropical Medicine, [8]London School of Hygiene & Tropical Medicine, [9]Medical Research Council Unit The Gambia at London School of Hygiene & Tropical Medicine, '0Medical Research Council Unit The Gambia at London School of Hygiene & Tropical Medicine

Background

The Basse area has a population of approximately 137,000 with a semi-urban community and an important multilingual business center in the eastern part of The Gambia. The population comprises mainly farmers, herdsmen and businessmen. The Basse Demographic Surveillance System (DSS) collects and routinely updates data including asset ownership in this area. As part of a study assessing child mortality, data were extracted from 8000 households (HHs) with at least one infant born between 1st January 2014 and 28th February 2015.Although data on asset ownership are available, we do not have a standardised method of assessing the Socio-Economic Status (SES) of the DSS residents. Quantifying SES is important in itself and would also be of great statistical importance when adjustment for SES is required. Here we develop an SES index for the Basse town adapting the approach we used in S. Kapiga et al. (1). We also compare the similarity of results from this study in Basse, The Gambia and the previous study in our Mwanza, Tanzania study (1).

Methods

Item Response Theory (more specifically a Rasch Model, A. Tennant et al.(2)) was used to quantify the latent traits of each individual. The latent traits are based on the items they possess and their ability to possess such items. We identified 27 asset indicator variables that could be used to assess SES. These asset data were generally related to ownership and adequacy. Ownership assets includes items such as owning a car, bed, cart, bicycle, motorcycle, truck, tractor, DVD and TV, refrigerator, fan, computer, stove, solar lighting, generator, cattle, sheep, goats, donkeys, horses, chickens or duck. Adequacy assets includes having proper flooring, proper roofing, proper walling, water supply within same compound, proper toilet or an ability to earn above approximately $400 per year. Non-binary responses were dichotomised to facilitate computation for the Rasch model. The predicted latent variable SES was restricted to follow a standard normal distribution which could be turned to an ordinal scale to assess different quantiles of the population of the Basse DSS. The rationale is that the predicted probabilities of asset ownership or adequacy would vary with different levels of the SES latent variable in a manner representative of the study population. We used Stata for Unix (StataCorp. 2013. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP.) for the analyses.

Results

In general, the predicted probabilities associated with each of the assets used in predicting the SES variables showed relationships we expected in the Basse community. The results are summarised below. Probabilities of truck ownership, computer ownership and stove ownership was generally low in the

community. Ownership of these assets was strictly probable for the top 30th percentile of the SES latent variable and the probabilities of ownership was below 20%. Probability of bed ownership which was left in as a control variable was approximately 100% for any SES level. The probability of having a water supply within one's own compound and an adequate toilet (one which is either a flushed toilet or is a properly ventilated pit latrine) was generally less than 50% in the DSS. However, these respective probabilities started from zero and gradually increased with increasing level of the SES variable. The probabilities of having an adequate floor (a floor that has some sort of finishing such as cement, carpet, tiles or vinyl), an adequate roof (a roof made of asbestos, cement or iron-sheets) and an adequate wall (a wall made of cement or burnt bricks) started from zero for the lowest 5th percentile of the SES latent variable and quickly increased to approximately 100% around the 40th percentile of the SES latent variable. Probabilities of cow and sheep ownership increased with increasing SES level with probabilities ranging from 25% to approximately 75%. Probability of goat ownership, however, decreased with increasing SES level and ranged from 75% to about 60%. The resulting SES latent variable was normally distributed.

Conclusion

Assessing SES of study participants as a single variable could simplify analyses especially in prognostic and other models where adjustments for SES may be required. The relationships between SES and the asset indicator variables used in this study were generally as expected. We were unable to validate it internally in The Gambia but the trends in the relationships were as anticipated. We have showed that this methodology can provide an adequate tool for quantifying SES in a population but further validation studies are in order. ACKNOWLEDGMENT MAISHA team at MITU for the original data used to develop the method.The DSS team at MRCG for the data used to test the methodology. REFERENCE 1. Kapiga S, Harvey S, Muhammad AK, Stöckl H, Mshana G, Hashim R, Hansen C, Lees S, Watts C. Prevalence of intimate partner violence and abuse and associated factors among women enrolled into a cluster randomised trial in northwestern Tanzania. BMC Public Health. 2017; 17:190 2. Tennant A, McKenna SP, Hagell P, Application of Rasch Analysis in the Development and Application of Quality of Life Instruments. Value in Health. 2004; Volume 7 - Supplement 1

Time interval to modern contraceptive use following child birth among reproductive women in Tanzania: Evidence from Tanzania Demographic Health Survey2015/16

'Martin Mujuni Rwabilimbo [2]ReniBilikisu Elewonibi, [3]MichaelJ Mahande, [4]SiaE Msuya, [5]William Mauka

'Kilimanjaro Christian Medical University College (KCMUCo)' Tanzania [2]Associate at Harvard T.H. Chan School of Public Health' USA, [3]Kilimanjaro Christian Medical University College (KCMUCo)' Tanzania, [4]Kilimanjaro Christian Medical University College (KCMUCo)' Tanzania, [5]Kilimanjaro Christian Medical University College (KCMUCo)' Tanzania

Background

Postpartum contraception use is important to improve maternal and child survival. It helps to have desired family size, optimal child spacing,prevent unplanned pregnancies and associated adverse birth outcomes. However, postpartum contraceptive use in Tanzania is still low as 19%. Only 49% of Tanzanian women adhere to recommended optimal birth interval.Previous studies also have reported short median interval for resumption to sex after birth among African women may affect maternal and child health well being. This study aims to assess optimal time to contraceptive use and predictors of time to contraceptive use after birth among women of reproductive age in Tanzania

Methods

A cross-section study was designed using secondary data from TDHS 2015-16. A total of 13,664 women of reproductive age (15-49 years) were randomly selected from all 30 regions of Tanzania with national representativeness. Information on pregnancy, births and contraceptive use were recorded for women over the previous 5 years, using a contraceptive calendar. Data analysis was performed using Stata 15.0 accounting for complex cluster design and weighting. Both categorical and continuous variables were summarized using descriptive statistics.Time interval to contraceptive use after birth was computed using Kaplan Meier estimate.Adjusted hazard ratio with 95% CI were estimated using Cox Proportional Hazard regression. A p-value of <0.05 was considered significant

Results

A total of 3,964 women were studied. Majority (80%) were married. Nearly half (47%) were aged had 20-29 years. More than half had attended ANC (52.8%), with majority (68.5%) reported using short acting contraceptives. The median time to contraceptive use after birth was 7(range 2-13) months. It was longer among unmarried than married counterparts 9(2-17) vs. 5(2-12) months, respectively.On Crude analysis aged 15-19 and 40-49years had longer time to use (HR: 0.67 95%CI: .52-0.84)(HR:.68 95%CI: .56-0.84) respectively compared to aged 20-29years.Women being older than their partner had longer time to use(HR:0.65 95%CI:0.45-0.94)compared to man being older not more than 9yrs. Women never being married and ones separated or widowed had longer times(HR:0.26 95%CI:0.19-.35)(HR:0.67 95%CI:.56-.80)respectively versus married women. Women in southern zone took longer(HR.46 95%CI: .35-.61) compared to lakezone. Other predictors were home deliveries (HR: 0.88 95%CI: .78-1.00),exclusivebreastfeeding women (HR: 1.57 95%CI: 1.42-1.75)and long acting method type(HR: 1.40 95%CI: 1.23-1.61).Resumption to sex and menses were also significant predictors(HR:0.93 95%CI: 0.91-0.97)(HR: 1.28 95%CI: 1.11-1.49).With adjusted analysis women in the southern zone were more likely to have longer interval to start contraceptive use (AHR: 0.48 95%CI: 0.30-0.78). But it was shorter among users of long acting modern methods (AHR: 1.27 95%CI: 1.02-1.58) and longer for those who did

not resume sex (AHR .83 95%CI: .49-.99).

Conclusion

Optimal time to contraception was found to be 7 months whereby being a resident of southern zone, adoption of long acting method and nonresumption of sex after birth were significant predictors. The interceptive measures to facilitate postpartum contraception by timely availing methods of their choice and addressing on their perceptions can help to reduce delays in uptake.

Mixed method estimation of population HIV viral suppression in the Western Cape, South Africa

'Elton Mukonda [2]Maia Lesosky, [3]Landon Myer, [4]Nei-Yuan Hsiao
'Division of Epidemiology and Biostatistics' University of Cape Town [2]Division of Epidemiology and Biostatistics' University of Cape Town, [3]Division of Epidemiology & Biostatistics' School of Public Health & Family Medicine' University of Cape Town' Cape Town' South Africa, [4]Division of Medical Virology' Faculty of Health Sciences' University of Cape Town

Background

Viral load monitoring is the method of choice for evaluating treatment effectiveness and transmission risk in HIV-infected individuals. Despite its importance, few programmatic data from viral load (VL) monitoring in sub-Saharan Africa are available for analysis towards UNAIDS 2020 targets. It is also not well understood if routine laboratory data can be used to estimate progress towards population targets. We therefore aim to describe how routinely collected VL test data can be used to report on viral suppression rates among individuals, and to estimate population level viral suppression using a combination of empiric and model-based estimates.

Methods

We reviewed HIV VL test data for the Western Cape province for the period 1 January 2008 to 30 September 2018, obtained from the National Health Laboratory Service (NHLS). These data represent a catchment area of all public-sector ART clinics and hospitals in the Western Cape, South Africa for the period. A linkage procedure was applied to link repeat VL measures to individuals. This method was a combination of deterministic record linkage followed by probabilistic record linkage. After data cleaning, deterministic linkage followed by estimation of similarity scores using the Jaro-Winkler algorithm and hierarchical clustering were used to assign results to individuals. Sensitivity analysis was undertaken using different cut-off values for the Jaro-Winkler distance metric. Test level and individual level viral suppression rates were estimated using a definition of either any VL <1000 copies/ml (for tests) or the last VL test in a calendar year < 1000 copies/mL (for individuals). We calculated the population level viral suppression rate of all people living with HIV in the region by combining SA census estimates, Thembisa model estimates of the proportion living with HIV, and individual level viral suppression estimates from the laboratory data. suppression Estimates are disaggregated by age, year and gender where possible.

Results

Over 2,012,452 tests from 541 clinical sites were retained in the analysis. VL testing services increased in the area by nearly 500% between 2008 and 2018. Approximately 83% of the tests are for people aged between 25 and 54 years, with 68.6% on women. In all 1689525/2012452 (84.0%) of VL tests were <1000 cps/mL for the period, with women having higher rates of test viral suppression (84.7%) compared to men (82.3%). The number of people on ART estimated from this data source was between 1%- 9% of estimates published by the Western Cape Department of Health. The estimated percentage of individuals on ART and virally suppressed did not vary by linkage method and increased from 84% in 2008 to 90% in 2018. Men had lower rates of suppression (83.1% to 88.5%) compared to women (84.4% to 90.7%). Population viral suppression among all individuals living with HIV in the western cape was estimated to be 51% in 2017. This did not vary substantially by linkage method and is comparable to estimates from the Thembisa model (v4.1) and other published studies.

Conclusion

Significant progress in attaining population viral suppression rates towards the UNAIDS 2020 targets have been made, even with a massive scale up in service provision. Test, individual and population viral suppression rates have been consistently moving upwards over time. Population viral suppression can be estimated with reasonable accuracy providing record linkage is possible and the numbers of people living with HIV in the population can be estimated. Estimation of population estimates from routine 'in care' data should not be undertaken without record linkage as it overestimates the true population viral suppression.

Spatial distribution of HIV prevalence among young people in Mozambique

'Rachid Joel Guidion Muleia [2]Marc Aerts, [3]Christel Faes
'Eduardo Mondlane University/ Hasselt University [2]Hasselt University, [3]Hasselt University

Background

Southern and Eastern Africa are the most affected regions in Sub-Saharan Africa. Mozambique a country located in southern Africa has the eighth highest HIV prevalence among adults aged 15-49 years old, currently with an estimated prevalence of 13%. The HIV epidemic is present in all age groups, nevertheless, adolescent and youth are the one of hardest hit group by the epidemic with an estimated prevalence of 7.9% in 2009 and 6.9% in 2015. In Mozambique little is known about risk factors that are associated with HIV infection among adolescent and young people. This study will fill the gap that exists of studies about HIV infection in adolescents and young people. Additionally, there is also a huge gap of studies that seek to understand the geographical variation of HIV infection. We understand that a thorough understanding of the spatial distribution of HIV/AIDS will help health official and policy makers to formulate proper intervention measure to fight the epidemic and also know which areas and regions to prioritize for allocation of resources.

Methods

To understand the spatial distribution of HIV among adolescent and young people in Mozambique we analysis we a data set from INSIDA 2009, a first population-based nationally representative survey on HIV sero-prevalence. We use generalized geoadditive models, a merge of kriging and additive models. we use radial basis splines to assess the nonlinear spatial effect. The estimation is done using two-stage iterative penalized-quasi-likelihood, an extension of the estimation procedure by Vandendijck et al. (2017). This method enables to estimate the range (spatial decay) parameter associated with the basis spline of the spatial component when the geographically distributed outcome is binary. The proposed procedure also applies when the outcome is count.

Results

The results revealed a significant spatial pattern. After controlling for important covariates the epidemic was found to be more severe in the central and northern part of Mozambique. Furthermore, the results revealed several socio-demographic, biological and behavioral factors associated with HIV among young people, where education was found to have a protective effects. The analysis also showed that young girls are about six times more likely to be HIV positive than man. Additionally, results revealed that young individuals in female-headed-household are much more vulnerable than those in male-headed-household as their odds of being HIV positive were two times higher.

Conclusion

In summary, results revealed that intervention measure should place more emphasis in districts located in the central and northern part of the country. Additionally, policy maker should make education of easy access for adolescent and young, and also promote awareness campaign for this group. The government should also create laws to protect widows/ or female-headed-households, as the results showed that children under female-headed-household are more vulnerable to HIV infection.

Geospatial Computing of large scale spatial data: Faster computation of the dense spatial matrix

'Eustasius Musenge, [2]Gideon Nimako
'University of the Witwatersrand [2]AU-NEPAD

Background

Space and time related public health data have grown exponentially since the time that Dr John Snow used a map to resolve cholera in London in 1854. The importance of spatial modelling to inform interventions and policymakers on where to target has never been greater than our current times. Using simulated and real data from a remote area in South Africa, we explore the different approaches to reduce the computational time required to model the spatial correlated dense matrix.

Methods

The Bayesian approach requires iterations to produce posterior estimates from given data, priors and hyper-priors. Performing this procedure on a large spatial dense matrix, requires extensive computational resources and often results in slowing the whole modelling process.The datasets involved in such spatial modelling is primarily large. In our real datasets, for instance, there are approximately 15000 households and about 70000 at the individual level. The spatial random effects modelling using Bayesian iteration techniques requires creating square matrices from the households' geo-locations and individual level data and also involves matrix operations such as vector-matrix multiplication and matrix inversions at each iteration. Traditionally such operations on large datasets are computationally intensive and have the bottleneck of most scientific applications. The computational complexity of such matrix operations is of order three O(n^3). However, these operations are embarrassingly parallel and amenable to parallel and optimise algorithms.

Results

In this work, we present optimised matrix operation algorithms that reduce the time and resources taken by the spatial random effects modelling due to such computational complexity. As most of the square matrices in the modelling are often sparse, we first utilised Bit Encoded Sparse Storage (BESS) and extended Compressed Row Storage (xCRS) to compress the dataset before the start of the computation. These compression algorithms reduce the access time to an element in the space while still maintaining the spatial properties of the dataset. We then parallelized the spatial random effects models using MapReduce framework on both shared and distributed memory architectures using OpenMP and Message Passing Interface (MPI) programming models.

Conclusion

The algorithms employed were able to reduce the computational time required drastically and were yielding results that were accurate based on the simulated data. Models that would produce results after two weeks were now completing in days after employing these high computing procedures. Real-time modelling of spatial data is key to timeously controlling diseases and techniques such as these presented in this paper help give insight to making that a reality in the near future.

Inequalities in stunting among under five children in Tanzania: Decomposing the Concentration Indexes using Demographic Health Surveys from 2004/5 - 2015/6

'Edwin Musheiguza [2]Rune Philemon, [3]Melina Mgongo, [4]Sia Msuya, [5]Johnson Mahande, [6]Elia Malamala, [7]Festo Charles

'Kilimanjaro Christian Medical University College (KCMUCo) [2]KCMUCo, [3]KCMUCo, [4]KCMUCo, [5]KCMUCo, [6]College of Business Education, [7]Ifakara Health Institute

Background

Child stunting not only contributes to morbidity and mortality, but also to poor economic production in adulthood as it is associated with poor child growth and mental impairment. Different studies have reported that, a child's stunting was greatly associated with socioeconomic factors including household wealth, mother's education and area of residence. Furthermore, the absolute level of stunting has been decreasing in Tanzania although the prevalence is still high (34%). As we aim to reach 28% in 2020 and zero stunting in 2025 under Tanzania nutrition target and Sustainable Development Goals respectively; interventions should target the disadvantaged groups by prioritizing determinants with larger contribution on differentials in stunting. However, there is limited information on trends of stunting disaggregated by child's family socioeconomic characteristics useful for designing interventions basing on socioeconomic differentials and hence improve child's growth. This study aims to determine the trend, contributing factors and changes in inequality of stunting among children aged 3 – 59 months using Tanzania Demographic Health Surveys (TDHS) from 2004/5 to 2015/6

Methods

We analyzed TDHS data from 2004/5 to 2015/6 using STATA version 15. A total of 22, 450 children living with their mothers were studied whereby 6,780, 6,432 and 8,443 belonged to survey years 2004/5, 209/10 and 2015/6 respectively. Level of education was categorized as no education (Lo), Primary education (L1) and Secondary or higher education (L2). Household wealth quintiles were categorized as poorest wealth quintile (Qo), poorer wealth (Q1), middle wealth (Q2), rich wealth (Q3) and richest wealth Q4) while area of residence was categorized as urban (Ao) and rural (A1). Categorical data was presented using frequency and proportions while continuous data was summarized using measures of central tendency with their respective measures of dispersion. Inequality in stunting was measured using the concentration index (CIX). The multilevel Poisson regression model was used to determine the association between child's stunting status and socio-economic predictors; linear mixed models with Wagstaff and Wantanabe methods were used to decompose the CIX. The odds ratio (OR) and 95% CI were computed; the contribution of each determinant was reported in percentages.

Results

The prevalence of stunting was 45.47%, 42.82% and 35.56% survey years 2004/5, 209/10 and 2015/6 respectively. In the survey year 2004/5 stunting was higher among children born from mothers with Lo 48.6%, living in Lo 52.3% and 48.1% were living in A1; in the survey year 2009/10, majority of stunted children were born to mothers with Lo 46.5%, 49.3% were living in Qo while 45.5% lived in A1; in the 2015/6 survey, of stunted children 40.7% were born to mothers with Lo, 41.5% lived in Qo and 38.9 % lived in A1. Adjusted analysis revealed that stunting was associated with area of residence OR: 1.18 (0.82 - 1.71), mother's education OR: 0.98 (0.917 - 1.05) and household wealth quintiles OR: 0.91 (0.876 - 0.938) during the 2004/5 survey; in 2009/10 area of residence OR: 1.05 (0.901 - 1.228), mother's education OR: 0.847 (0.704 - 1.02) and household wealth quintiles OR: 0.94 (0. .85 - 1.04). Inequalities in stunting decreased

from -0.078 in 2004/5 to -0.74 in 2009/10 and increased to -0.083 in 2015/6.

Conclusion

In all three survey years, stunting was higher among the disadvantaged groups; children born to mothers with no education, living in rural areas as well as households with poorest wealth quintiles. As we aim to reach zero malnutrition, interventions should abide on the pro-rich inequality in stunting as they will be identified during further analysis during decomposing the concentration indexes. Priorities on combating child stunting should be on factors with larger contribution to the existing inequality.

Feasibility of employing systematic random cluster selection with probability proportional to estimated size and without replacement, to obtain a nationally representative sample of 9204 clusters for estimating vaccination coverage in South Africa

'Portia Mutevedzi, [2]Shabir Madhi
'Respiratory and Meningeal Pathogens Research Unit RMPRU [2]Respiratory and Meningeal Pathogens Research Unit RMPRU

Background

Systematic random cluster selection without replacement and probability proportional to estimated size reduces potential for selection where both the probability of selecting population clusters from a sampling frame and the probability of selecting eligible respondents within the clusters is quantified and is non-zero. In 2015, World Health Organisation implemented new recommendations for assessing Expanded Programme on Immunization (EPI) coverage estimates at national and local health management levels. The most significant change from previous WHO recommended EPI survey methods was the implementation of true probability-based sampling to reduce bias, and improve the accuracy and precision of survey results. South Africa recently implemented these new guidelines to obtain a sample for the first ever national EPI coverage survey post-independence. The survey aims to produce robust and precise vaccination coverage estimates at the district and national level. The WHO recommended sampling methodology employed for this survey is expected to produce a representative sample of children requiring immunisation and the coverage thereof.

Methods

In line with the WHO cluster survey recommendations, clusters were randomly selected from the complete and most recent STATISTICS SOUTH AFRICA (STATS-SA) 2011 national census frame. The sampling frame listed all sampling units [province, districts, municipalities, wards and small area layers (SALs)] and covered the entire target population of children aged 24 to 25 months. Multi-stage probability proportional to size sampling was employed where the sampling frame was first split into 52 districts. For each district, stratification was done by dwelling type (formal residential, informal residential, traditional residential, farms, collective living quarters and small holdings) following which clusters were selected from each district guided by a calculated sampling interval based on total number of households and sample size required to calculate EPI coverage for that particular district. Probability of a cluster being selected was proportional to size of the target population in each district. Within each selected cluster, a random sample of 120 households was selected and included in the survey sample. Stata 13 and Microsoft excel were used for sampling procedures.

Results

The resultant sample consisted 9204 clusters from which 55 120 children aged 24 to 35 months will be interviewed to collect vaccination data. The map in the attached abstract document illustrates the resultant survey sample; the red areas represent selected clusters. Geographic distribution and density of the 9204 selected clusters mirrors population distribution and density within the 52 districts in South Africa.

Conclusion

The currently ongoing national EPI coverage survey in South Africa is the first such survey

post-independence and is required to produce robust and precise vaccination coverage estimates at district and national level. As such rigorous sampling methodology is required and must also be aligned with the WHO recommended methods for results to be acceptable to the international community. Our study demonstrates the feasibility of implementing the WHO recommended EPI cluster survey methodology to produce a large survey sample that is representative of the underlying target population. Results of this survey will be invaluable in guiding policy and planning for enhanced access to vaccines within each district and nationally.

Improving chest x ray classification using transfer learning

'Paul Mwaniki [2]Samuel Akech, [3]Timothy Kamanu, [4]Rene Eijkemans
'KEMRI Wellcome Trust [2]KEMRI Wellcome Trust, [3]University of Nairobi, [4]University Medical Centre Utrecht

Background
Chest x ray classification using machine learning models can aid in cheap and rapid diagnostics of common respiratory ailments. Developing machine learning models with high predictive performance however requires large labeled (with outcome measured) data sets. Labeling of chest x rays is time consuming and required highly trained specialists (clinicians/radiologist) and therefore does not scale well to large data sets hindering development of models with high predictive accuracy. Lack of large training data set can be overcome by using transfer learning. Transfer learning involves using machine learning models developed to solve one problem to improve a different but related problem.

Methods
We compared performance of machine learning models fitted with and without transfer learning in classifying paediatric chest x ray images for abnormalities. The chest x ray data set comprises of four thousand chest x rays collected as part of Pneumonia Etiology Research for Clinical Health (PERCH) study. The images are classified according to WHO standardized classification of paediatric chest radiographs for diagnosis of pneumonia: consolidation; other infiltrate; both consolidation and other infiltrate; normal or uninterpretable. Two transfer learning techniques were used: fine tuning publicly available machine learning models for image classification and unsupervised pretraining using large unlabeled chest x ray data set.

Results
Machine learning models involving transfer learning performed better compared to those without.

Conclusion
Transfer learning can overcome challenges posed by lack of large labeled data sets in developing machine learning models for regression and classification. Clinical data sets are often relatively small and even when large data sets exist, majority of observations are unlabeled. Labeling data sets is often a challenge whenever expensive lab test are needed or highly trained specialist are required.

Evaluating the Effect of Sampling Weights on the Predictors of Contraceptive Use in Uganda Using 2016 UDHS Data

'Edson Mwebesa [2]Emmanuel Obuya, [3]Lorna A Aol, [4]Nazarius Mbona Tumwesigye
'Makerere University [2]Makerere University, [3]Makerere University, [4]Makerere University

Background

Sampling weights are used in studies that involve complex sampling designs to cater for unequal probabilities of selection, nonresponse, and noncoverage. They are calculated as the inverse of the individual probability of participation. What is not known, however, is whether sampling weights affect surveys results despite its continued application in the analysis? This study investigated whether sampling weights have an impact on the predictors of contraceptive use using 2016 Uganda Demographic Health Survey (UDHS) data.

Methods

The study used secondary data collected by the Uganda Bureau of Statistics (UBOS) during the 2016 UDHS. All women of age 15 – 49 in selected households were eligible for the study. The weighted sample of 57,029 women and an unweighted sample of 57,906 were considered for analysis. The comparison was made of the results from weighted and unweighted samples to check if there was a difference in these results, both at descriptive and inferential levels, using the Confidence intervals and p-values

Results

The prevalence of contraceptive use for weighted sample was at 0.39 (95% CI: 0.39, 0.40) while for the unweighted sample was at 0.37 (95% CI: 0.37, 0.38). The proportions and their CIs for covariates of contraceptive use showed a wider CIs for weighted samples than for unweighted. The results from a modified Poisson regression model indicate that wider CIs from the weighted sample than the unweighted samples. For example, the adjusted prevalence rates for weighted sample for age group 25 – 29 years was 1.16 (95% CI: 1.08, 1.25) while that for unweighted was1.13 (95% CI: 1.07, 1.19) and that of 30 – 34 years was 1.17 (95% CI: 1.08, 1.27) for weighted and 1.17 (95% CI: 1.12, 1.24). A noticeable relationship between the point estimates is that both the weighted point estimates fall into the CI for unweighted and the reverse is true. The linktest of adjusted models for weighted and unweighted samples showed that both models were correctly specified.

Conclusion

Based on the CIs and point estimates, the researcher concluded that both weighted and unweighted samples do not give different results of the predictors of contraceptive use. The researcher, therefore, finds that whether samples are weighted or not, as long as the sample is large enough and representative, there is no need to weight the samples.

Modelling the Force of Infection for Hepatitis B among Heterogeneous Groups Reporting at Tertiary Hospital, Ghana

'Emmanuel Kweku Nakua [2]Emmanuel Kweku Nakua, [3]Nana Kena Frimpong
'Kwame Nkrumah University of Science and Technology [2]Kwame Nkrumah University of Science and Technology, [3]Kwame Nkrumah University of Science and Technology

Background

Globally, in 2015, an estimated 257 million people were living with chronic Hepatitis B virus (HBV) infection. The most common modes of HBV transmission are: from mother to baby at birth, child-to-child, and sexual transmission. The prevalence is high at infancy and childhood. The age at infection and severity of illness is important. Viral exposure at an early age usually elicits an asymptomatic or self-limiting liver infection, with resulting life-long protective immunity. In contrast, later viral exposure occurring in adolescents or adults which happen in more developed immune system may cause symptomatic acute cases with increased severity and higher disease burden. In any age group, the disease is fatal in 2% of cases. Acute hepatitis is rare in infants, only in 6% of infected children are acute and 30-50% of infected adults will develop cirrhosis and/or liver cancer. The main objective of this study is to model the prevalence of HBV infection as a function of age using Beta Regression technique and estimate force of infection from the estimated model.

Methods

A retrospective study was conducted using data from the virology/serology unit of the Komfo Anokye Teaching Hospital (KATH). The data was from January 2013 to June 2016 and 18528 records were involved. The facility serves about 80% of the population of Ghana and sees patients within the Ashanti, Northern, Upper East, Upper West, Central, Brong Ahafo and Western regions and neighbouring countries such as Cote d'Ivoire, Burkina Faso, and Togo. Ethical approval was obtained from Kwame Nkrumah University of Science and Technology and the Research and Development Unit of KATH The Beta regression model was used to estimate the mean response given age. The dependent variable in this study was the proportion of HBV+ at each age after infection. The response is constrained to the standard unit interval (0, 1) and the ages of the individual is the age diagnosed just after infection. The ages of the participants were categorised in accordance to the 2010 Population and Housing Census (PHC) by the Ghana Statistical Service, where 12-19 represents adolescents, 20-24 as young adults, 25-59 were considered as old adults and lastly 60+ are the aged. A linear and quadratic age-dependent model for the prevalence using single year age groups were fitted. The analysis was performed using R studio. Ethical approval was obtained from Kwame Nkrumah University of Science and Technology and the Research and Development Unit of KATH

Results

A total 18528 men and women were screened at the virology/serology unit. 2574 were positive for HBV, representing an overall prevalence of HBV+ of 13.9%. There was a relatively higher prevalence among age group 25-59 years old and lower for the above 60 years. HBV prevalence increases from adolescents and reach its peak at about age 40 and decreases for higher ages. From the quadratic model the prevalence of HBV depicts a concave shape while the linear model shows a negative slope for the prevalence with age. The Akaike's information criterion (AIC) was used to assess the fit of the two models. The quadratic model had the least AIC hence was chosen for further analysis. steady growth and a sharp decline of HBV infection over ages.The estimated quadratic model was used to estimate the force of infection (FOI). The

estimated force of infection increases exponentially from 0 to 12 and decreases from age 12 to age 38. The force of infection increases from the age 38 to 64 and decreases thereafter. An average force of infection (instantaneous per susceptible incidence) for HBV of 0.034% per annum, (3.4 per 1000 at birth per annum). Meanwhile, age-dependence in transmission is supported statistically indicating peak incidence rate in the total population of 0.04% per annum for 11 year olds and a lower incidence of 0.005% per annum in 38 year olds

Conclusion

In conclusion, HBV infections increases from lower ages and reach its peak at age 38 years old. The historical HBV infections decreases at a faster rate from age 38 years to age 80 years old. HBV prevalence is high at age 38 and the transmission rate is lower at age 38 years and higher at age 11 year and 64 year olds. While other studies have shown the shift towards intermediate infection to older age groups. This study findings revealed increasing infection at lower age groups. We believe that this findings of HBV can help to target pre-vaccination screnario for Ghana.

Non parametric techniques for multilevel discrete survival data

Thambeleni Nevhungoni, A Bere, D Chen, S Manda
University of Venda

In discrete survival modelling the parametric approach has been receiving a lot of attention due to its computational ease. However, recently attention has shifted to nonparametric methods due to lack of flexibility with parametric approach. A lot of work has been done on nonparametric modelling of the baseline hazard, link function and the random effect separately. This study seek to build a discrete survival model where the base line hazard, link function and the random effect are simultaneously modelled nonparametrically.

Age, Period and Cohort Analysis of young adult Mortality due to HIV and TB in South Africa: 1997-2015

'Tshifhiwa Nkwenika [2]Samuel Manda, [3]samuel Manda
'Biostatistics Unit' South African Medical Research Council' Pretoria' South Africa. [2]Biostatistics Unit' South African Medical Research Council' Pretoria' South Africa., [3]Department of Statistics' University of Pretoria' Pretoria' South Africa

Background
Young adult mortality is very important in South Africa with the impact of Human Immunodeficiency Virus /Acquired Immune deficiency Syndrome (HIV/AIDS), Tuberculosis (TB), injuries and emerging non-communicable diseases (NCDs). Investigation of temporal trends for adult mortality associated with TB and HIV is often based on age, gender and birth cohort separately. This has hindered an insightful interpretation of the mortality trends and effects by age, period, and birth cohort simultaneously. The overall aim of this study was to estimate age effect across period and birth cohort; period effect across age and birth cohort; and birth cohort effect across age and period.

Methods
Mortality data and mid population estimates were obtained from Statistics South Africa for the period 1997 to 2015. Three-year age, period and birth cohort intervals for 15-64 years, 1997-2015 and 1934-2000 respectively were used. Age-Period-Cohort (APC) analysis using the Poisson distribution was used to compute effects of age, period and cohort.

Results
The results showed a concave down association between age and TB mortality, with a peak at 36-38 years. There was a concave down relationship between TB cause-specific mortality in the studies mortality data period between 1997 and 2015. There was a downward trend between TB mortality and the effect of birth cohort from 1934 to 2000. There was an inverse flatter U-shaped association between age and HIV mortality, and was more pronounced at 36-38 years. The estimated relative risks showed approximately linear relationship between HIV mortality and effect of period from 1997 to 2015. An inverted U-shape relationship between birth cohort and HIV-related mortality with a slight increase at later birth cohort was observed.

Conclusion
Despite the limitations of official published mortality data, we still found a notable age and period effect according to APC analysis, which were similar to the results of previous studies. However, the pattern of period effect associated with HIV mortality differed from the one published by Statistics South Africa. Age-Period-Cohort Model of HIV and TB mortality offers a more robust assessment of effect of age, period and birth cohort, which would not be possible using traditional Poisson regression model on the death counts separately.

An exploratory analysis of multidimensional binary data using Correspondence Analysis, Non-metric multidimensional scaling and Cluster analysis as applied to Biolog EcoPlate data

'Sheroline Nombasa Ntushelo [2]Morné Lamont, [3]Marieta Van Der Rijst
'Agricultural Research Council [2]Stellenbosch University, [3]Agricultural Research Council

Background

This is an experiment that was conducted by researchers at Agricultural Research Council (ARC) in Stellenbosch. The experiment is about 12 differently treated soils being used to study the activities of micro-organisms in the soil. A completely randomized experimental design was used. Samples of the soil were collected at 2 depths to study the microbial activities at different depths. Samples were also collected for 3 months to study the microbial activity over time. Once the soil samples were collected, it was dissolved in water so that the soil would sink to the bottom and the micro-organism in the soil will rise to the top. A sample of this was then put in a Biolog EcoPlate to observe the microbial activity. A Biolog EcoPlate contains 96 wells, in which 31 are the most useful carbon sources and 32nd component is water and the 32 components are replicated 3 times. A microbial activity is when a carbon source is used by microbes and the well turns purple. The value 1 indicates the presence of microbial activity and 0 indicates the absence of microbial activity. The data is multidimensional binary data and is referred to Biolog data.

Methods

Multivariate statistical methods play a very important role in understanding data that is multidimensional in nature. Such data are often very complex to understand and very difficult to analyse. The challenge is that most multivariate methods have been designed to work on numerical data and there are not many multivariate methods which can be applied on other types of data such as categorical data or binary data. Categorical data or binary data are likely not to follow the assumptions required by most multivariate methods. Analysing multivariate numerical data is usually easier than analysing multivariate category or binary data sets. An exploratory analysis of Biolog data using multivariate statistical methods, which are applicable for binary data. These methods include Correspondence analysis (CA), Non-metric multidimensional scaling (NMDS) and Cluster analysis. CA analysis is applied to study associations between the treatments and the carbon sources. CA is a non-parametric method, which makes no distributional assumptions. CA makes use of a chi-square distance. The total variance explained in the data by the first two dimensions which make up the CA plot shows goodness of fit. NMDS is used to visualize the treatments or carbon sources as points in a 2 dimensional space. NMDS make use of rank order of entries in the data matrix not the actual dissimilarities to produce the plot. NMDS has no assumptions about the raw data. Cluster analysis is used to search data for groupings of both treatments and carbon sources. Input data for both Cluster analysis and NMDS is a proximity matrix based on the data type.

Results

Correspondence Analysis gives 2 CA plots as results where one is asymmetric row plot and the other is asymmetric column plot. These 2 CA plots give the results of the same analysis looking at both sides of the analysis. In asymmetric correspondence analysis, treatments and carbon sources are interpreted together. The percentage total variance explained by the first two dimensions of correspondence analysis is most of the time around 50 % which is quite good for a binary data. In correspondence analysis some treatments and carbon sources will cluster together near the origin. These treatments experienced microbial activity

with most carbon sources within each combination of Depth and Month. Some of the treatments and carbon sources will group farther away from the origin. These treatments had microbial activity with few carbon sources within each combination of Depth and Month. Non-metric multidimensional scaling gives a 2 dimensional plot of treatments and of carbon sources separately within the same combination of Depth and Month. Non-metric multidimensional scaling and cluster analysis groupings of treatments and of carbon sources within the same combination of Depth and Month are observed and interpreted separately. The cluster analysis gives a dendrogram showing the groups of treatments and groups of carbon sources for the same Depth and Month on separate plots. In non-metric multidimensional scaling (NMDS) and correspondence analysis, carbon sources and treatments which cluster together making it difficult to visualise relational patterns. Dendrograms always show clear groupings of treatments and carbon sources on the plots. Both Jaccard and Bray Curtis dissimilarities showed similar results in Non-metric multidimensional scaling and Cluster analysis.

Conclusion

All 3 multivariate statistical methods show similar results. Even though these methods use different types of distance measures, their results are in agreement. Even though the 3 methods analyse and present the data differently, they complement each other. It is best to use more than one multivariate method to thoroughly explore and verify the results for multidimensional binary data. In correspondence analysis, the treatments that tend to cluster around the origin experience microbial activity with most carbon sources. In addition, the treatments that tend to group farther away from the origin experience microbial activity with limited carbon sources. The shorter the distance between two points representing treatments or carbon sources in a 2 dimensional NMDS plot the more similar the treatments or carbon sources are. In a dendrogram the treatment or carbon source closer to dissimilarity measure =1, the treatment or the carbon source experience microbial activity with few treatments or carbon source. The treatment or carbon source closer to dissimilarity measure = 0 experiences microbial activity with most treatments or carbon sources.

Investigation of three instrumental variable methods in the presence of errors in explanatory variables: A comparative analysis

'Maureen Nwakuya [2]Esther Estherw Anana
'University of Port Harcourt [2]University of Port Harcourt

Background

One of the assumptions of linear regression analysis is that the error term and the predictor variable should not be correlated, and for this to hold it is also assumed that the predictor variables are measured without error. This problem of measurement errors has been described as one of the most fundamental problems in empirical economics. This is because the presence of measurement errors causes biased and inconsistent parameter estimates and therefore leads to erroneous conclusions to various degrees in statistical analysis. There are different methods of estimating regression parameters in the presence of error in variables. In this work we considered the instrumental variable method. The method of instrumental variables provides consistent estimates of regression coefficient in linear regression when the predictor variables and the error terms are correlated. The instrumental variable method consists of finding a set of variables which are correlated with the predictor variables in the model but uncorrelated with the composite disturbances. Under instrumental variable method we considered: Wald's method, Bartlett's method and Durbin's method. The limitation in their approach is that it was applied to regression with only predictor variable, so in this work we tried to extend their methods to accommodate three predictor variables, and also compared these methods using Akaike Information Criterion (AIC), Mean Square Error (MSE) and Root Mean Square Error (RMSE).

Methods

The methods make use of the discrete grouping variable either on the basis of median or rank. Wald (1940) suggested to use -1 and +1 for values less or greater than the median of the manifest variables; The following steps are involved in Wald method: Step 1: Find the median of given observations.Step 2: Then we classify the observations by defining an instrumental variable W such that W=1 if x is > median and w=-1 if x<median.Bartlett (1949) proposed to rank and divide the values of the explanatory variables into three groups, where the middle values take on 0, the upper values take on 1 and the down values take on -1 to define the instrumental variables. Also, Durbin (1954) proposed the use of ranks of the values of the predictor variable to define the instrumental variable. For our method the X variable was three giving us a four column matrix likewise the W (instrumental variable). Then Substituting the values of X, y and W in $(W'X)^{-1}(W'Y)= \beta$ to solve and estimate the regression coefficients. Also the variance of the estimates from the three methods were estimated and compared using the paired t test. We also computed the AIC=nlnSSE/n +2k (where k is the number parameters and n is the number of observations), the mean square error and the root mean square error, these we used for comparison purposes.

Results

The results of the analysis showed that the variances of the estimates were not significantly different from each other. But the Bartlet method showed the least AIC, MSE and RMSE,hence can be considered as the best method compared to Wald's and Durbin's method.

Conclusion

In conclusion, for linear regression analysis where there are errors in the predictor variables,the best approach is the use of instrumental variable method, but preferably the Bartlet's method.

Comparison of cox proportional hazard model and accelerated failure time model with application to data on tuberculosis/HIV patients in Nigeria

'Opeyemi Oyekola Ogungbola [1]Abayomi Ayodele Akomolafe
'Federal University of Technology Akure' Ondo State

Background

These were used for analyzing survival data on Tuberculosis/HIV co-infected patients in Nigeria. We apply the methods to a cohort of these patients managed in tertiary Directly Observed Treatment Short Course (DOTS) centre, Nigerian Institute of Medical Research (NIMR) for the period of six months, where we compare the effect of the accelerated failure time model with Cox proportional hazard model in determining the time to sputum conversion in TB patients who are co-infected with HIV.In this research, we adopted various techniques were adopted such as the survival function curve, the log-rank test, and the Cox proportional hazards (PH) model, omnibus test, survival function of the mean covariates, log minus log, the Accelerated failure time (AFT) model, the AFT model plot, the Log-likelihood test and Akaike Information Criteriion (AIC).

Methods

In this research, we adopted various techniques were adopted such as the survival function curve, the log-rank test, and the Cox proportional hazards (PH) model, omnibus test, survival function of the mean covariates, log minus log, the Accelerated failure time (AFT) model, the AFT model plot, the Log-likelihood test and Akaike Information Criteriion (AIC).

Results

The research established that AFT model provides a better description of the dataset as compared with Cox PH model because it allows prediction of Hazard function, survival functions as well as time ratio. Moreover, PH model does not fit appropriately when compared with AFT model; thereby provide less appropriate description of survival data.

Conclusion

The result revealed that the gamma model provided a better fit to the studied data than the Cox proportional hazards model. Hence, it is better for researchers of TB/HIV co-infection to consider AFT model even if the proportionality assumption of the Cox model is satisfied.

Gender differential and social determinant of tuberculosis/HIV co-infected patients

'Opeyemi OyekolaS Ogungbola [1]Sunday M. Abatan
'Federal University Of Technology Akure' Ondo State

Background
The study used the datasets that was obtained from the DOTS Clinic of the Nigerian Institute of Medical Research (NIMR). A parastatal under the Federal Ministry of Health that has treated over 5000 TB patients in the last 6 years (2011-2016). All the patients enrolled between 2011 and 2016 were study population for the study but the patients which were co-infected with TB and HIV was used as the focus of this study. The eligible patients for this study comprises of 190 male and 280 female patients making a total sum of 470 patients in all.

Methods
These patients were diagnosed, subjected to the same control treatment rate, pre-clinical evaluation, clinical observations and disease results outcome, these form the different levels of statistical analysis. Comparative Univariate Analysis (CUA) was employed while Separated cross-tabulation was used for bivariate analysis while control for the gender.

Results
The gender differential in the demographic factors among the TB/HIV co infected patients shows that females current pregnancy status, age and occupation were influential factors that caused TB with respective Chi-square (X2=95.85, p=0.000),X2=117.37,p=0.000 and X=207.68, p=0.000 while among the male, occupation, age and marital status gave a significant statistical result with respective chi-Square (X2=87.41, p=0.003), (X2=111.43,p=0.000) and (X2=35.57, p=0.017). There was no significant gender difference on treatment control rates, pre-clinical evaluations and clinical observations for both male and female patients. But there was a significant gender differential in the result treatment outcome in type of occupation of male patients with chi-square (X2) =25.00, p=0.02 whereas other factors has no statistical justification to support gender difference at 0.05 level of confidence

Conclusion
We discovered that there are gender differential in the social-determinants in the cases of TB/HIV co-infected patients in the study area. Thus it is very crucial to give special attention and consideration to that vulnerable groups; women in age (30-39years), married men, female petty traders and commercial drivers as well as bus stop hawkers/ vendors. These selected individuals should be given proper interventions and sensitization. Also, because of the close relationship between HIV and TB in notable sub-Sahara African in which Nigeria is inclusive, this call for structural and social determinants of TB risk. Another major area of concern is the wide spread of inequalities in opportunities, vulnerability and expectation for men and women which is being reinforced through cultural norms and socio-economic system (Hargreaves, et. al, 2011).

Mathematical Modeling of the Interruption of the Transmission of Soil Transmitted Helminths Infections in Kenya

'Collins Okoyo [2]Graham Medley, [3]Charles Mwandawiro, [4]Nelson Owuor
'University of Nairobi' and Kenya Medical Research Institute [2]Faculty of Public Health and Policy' London School of Hygiene and Tropical Medicine' London' United Kingdom, [3]Eastern and Southern Africa Centre of International Parasite Control' Kenya Medical Research Institute (KEMRI)' Nairobi' Kenya, [4]School of Mathematics' University of Nairobi' Kenya

Background

Kenya, just like other soil transmitted helminthes (STH) endemic countries, has been conducting regular treatment program for the last five years among school aged children as a way to reduce STH infections burden in the country. However, the point of interruption of transmission of these infections still remains unclear. In the current study, we analyzed an age-structured mathematical model to predict the point of interruption of these infections in Kenya. The main objective was to develop and analyze an age-structured model of the STH population dynamics under a regular STH treatment program to determine infection transmission rate, the point of infection interruption, and the optimal interpulse treatment interval sufficient to achieve STH infections elimination in Kenya.

Methods

The ongoing study utilized age-structured model of the STH population dynamics under a regular treatment program. The model was applied to two main age groups: school-age children (5-14 years) and adult populations (>15 years) and investigated the potential for STH elimination with finite rounds of treatment while allowing the STH distribution to change dynamically as a function of treatment frequency and treatment coverage. The model was verified using a five-year field data from the National School Based Deworming Program (NSBDP) for all the three main STHs; Ascaris lumbricoides, Trichuris trichiura and hookworms.

Results

The model behaviour demonstrated convincingly an accurate predictions of prevalence and mean intensities of infections during and after treatment rounds in each of the age groups. The model indicated that the benefit derived from the regular treatment increases non-linearly with the treatment rounds and coverage. Additionally, it depicted that for elimination to be achieved within a shorter time period in the general population and within each age group, higher treatment coverage and bi-annual treatment rounds are more effective.

Conclusion

The model captured the dynamics of the STH burdens in vulnerable populations under regular treatment program as elimination is approached. It aided in examining the role of age structure to the persistent STH infections in Kenya among these most vulnerable populations. As a result of these findings, we aim to advise the STH control programs on the right mix of strategies needed to achieve faster elimination of the STH infections in Kenya.

Assessing the Sensitivity and Robustness of Randomization Test in Analysis of Repeated Measures Design with Missing Observations

'Abimibola Victoria Oladugba [2]John Ajali Obasi
'University of Nigeria, Nsukka [2]University of Nigeria, Nsukka

Background

Missing observations are often faced while carrying out repeated measures experiment in which the need might arise to analyze them without necessarily replacing the missing observations with calculated values or making any parametric test assumption. There is need to examine the alternative options to the parametric methods that can be deployed in the light of inability of such data to meet normality and other parametric test conditions. This assessment can be done by evaluating sensitivity and robustness of such test in terms of statistical power and type-I-error rate respectively to ensure it outperforms other classical alternatives. In this study, a randomization test was used to analyze one factor repeated measures design (RMD) in the presence of missing observations under scenarios when the data are normal, non normal, had outliers and when sphericity condition is met or not for variant sample sizes and number of treatments.

Methods

In order to assess the sensitivity and robustness of this test, the behavior of the power and empirical type-I-error were observed in a simulation study of 10,000 iterations using Monte Carlo approach in R statistical computing package.

Results

In the complete case analysis with normal data and no missing observation, the randomization test compared favorably with the classical F-test in terms of sensitivity, while it was more sensitive than the F-test for data with skewed distributions. The two tests appeared to produce stable p-values around the nominal pre-assigned value in the complete case analysis. In the presence of missing observations, the randomization test was mildly more sensitive than the F-test in analyzing RMD data with normality while it was substantially more sensitive than the F-test when the data followed skewed distributions. In the presence of missing observations with outlier, the randomization test was more sensitive than the F-test in analyzing RMD data for both normality and non normality conditions. When sphericity condition was met, the randomization test and the F-test were approximately equally sensitive; whereas the randomization test was more sensitive than the F-test when sphericity condition was not met.

Conclusion

Summarily, the randomization test was more sensitive than the F-test when data were normally distributed or skewed. Also, the randomization test is as robust as the F-test in terms of maintaining type-I-error rate when data follow normal and more robust than the F-test in non normality case.

Modeling the Stability and Determinant Factors of Household Food Insecurity: A Pair Copula Construction Approach

'John Olaomi [2]Jemal Ayalew Yimam
'University of South Africa [2]Department of Statistics, College of Natural Science, Wollo University, Dessie, Ethiopia

Background

Assessment of household food security status as well as determinant predictors using cross sectional data does not reflect the entire situation of it, rather, the stability over time is the key component to assure household food security. The aim of this study was to assess the stability of household food security over time and its determinant factors using longitudinal data through pair copula construction (PCC) approach. This longitudinal study uses household food-security status, environmental and climate change related data collected three times at six months interval from June 2014 to 2015. The status of household food security was determined using both the quartile and composite food security index for each rounds. The study participants were 646 farmer households randomly selected from selected Weredas of the South Wollo Zone of Amhara regional state, Ethiopia. Food stability depends on the availability, access and utilization dimensions of food security over time. Moreover, the first three dimensions of food security levels were disaggregated into “chronically food in-secured”, “mildly food in-secured”, “moderately food in-secured” and “food secured”. These multiple dimensions and categories make it multifaceted to model the indicators as well as the determinant factors.

Methods

Pair copula construction (PCC) cumulative logit model was used to assess the stability of household food security status over time and the determinants of it simultaneously. For parameter estimation, we employ the full maximum likelihood estimation approach to compute jointly the dependence parameters using the selected Gumbel bivariate copula families and the marginal parameters of the cumulative logit model. R code was written using “arglag” optimization R package to estimate the parameters of the copula and the marginal distribution functions. In modeling the status and determinant factor of household food insecurity, the nice feature of the PCC approach in this setting is measuring the dependence of the consecutive food security status of the households using the copula parameter and the associated determinant factors of household food security using the parameters of the marginal distributions. As a result, the pair copula construction approach with D-vine is attractive since it allows pairwise positive dependence structures and has closed form cumulative distribution function (cdf), no other copula family has both these properties.

Results

The copula parameter revealed that individual household food security status is not stable over time. Moreover, the marginal parameter indicated that presence of crop disease; market price increase and medium weathering condition were the significant recurrent factors for households to be chronically to mildly food in-secured throughout the study period. One time cultivation per year was the temporal significant factor for household food insecurity.

Conclusion

The results of this model suggest that household food security is not stable over time. Great attention is thus require for granting households to be food secured taking valuable intervention for the identified recurrent determinant factors. Moreover, the model provides easily interpretable and understandable outputs, thus, we suggest the model for any longitudinal discrete data analysis.

Spatiotemporal patterns of successful TB treatment outcomes among HIV co-infected patients in Kenya

'Verrah Otiende [2]Thomas Achia, [3]Henry Mwambi
'PAUISTI [2]Epidemiology and Biostatistics Division' School of Public Health' University of the Witwatersrand' Johannesburg' South Africa, [3]School of Mathematics' Statistics & Computer Science' University of Kwa-Zulu Natal' Pietermaritzburg' South Africa

Background
The convergence of the Tuberculosis (TB) and HIV epidemics threatens the management of TB treatment. These have been evidenced by various studies describing how HIV co-infection propagates unsuccessful TB treatment outcomes. Information on the spatiotemporal patterns of successful TB treatment outcomes remains less understood despite the multiorganizational TB treatment efforts. This study uses case notification data to evaluate the spatiotemporal patterns of successful TB treatment outcomes for HIV co-infected patients in Kenya.

Methods
This study used the case notification data from the Kenya National TB control program to investigate successful TB treatment outcomes in forty-seven counties in the period 2012 - 2017. The population of study was HIV co-infected cases with known TB treatment outcome. A chi-square test was performed to determine the association between treatment outcomes and risk factors; TB-type, age, gender, ART therapy, and patient type. The study also assessed the geographic patterns and temporal trends by mapping the TB treatment success rate in each county for the six-year period. Using the Integrated Nested Laplace Approach (INLA), the TB treatment success of HIV co-infected patients was modeled. The spatial parameters assumed the Besag-York-Mollie (BYM) specification. The temporally structured effect was represented through a neighboring structure and the temporally unstructured effects using a Gaussian exchangeable prior.

Results
Among the 172233 HIV co-infected cases included in the analysis, 135973 (78.9%) achieved successful TB treatment outcomes. Female cases registered higher treatment success rates (80.1%) compared to the male cases (77.8%). The cases on Anti Retroviral Therapy (ART) recorded a success rate of 79.9% against 69.1% for their counterpart not on ART. The spatial trend depicted increased treatment success in some parts of the country with a relatively high level of associated certainty, characterized by a spatial relative success above 1 and posterior probabilities above 0.8. The temporal trend of treatment success showed an increase in the treatment success of TB in HIV coinfected cases. Overall, the success rate was still below 85% particularly for Homabay, Siaya, Kisumu, Migori and Busia counties in western Kenya

Conclusion
The successful TB treatment outcomes for HIV coinfected cases in Kenya were slightly below the 85% standard threshold set by the World Health Organization. Our study showed that even though coinfected cases have an increased risk of unsuccessful treatment outcomes, enhanced treatment monitoring improved the treatment outcome in most counties for the six-year period.

Robust estimation of single-index models with responses missing at random

'Masego Otlaadisa [2]Ash Abebe, [3]Huybrechts F. Bindele, [4]Boikanyo Makubate
'BIUST [2]Auburn University, [3]University of South Alabama, [4]BIUST

Background

In regression modelling, developing methodology for handling missing data has attracted significant attention within the statistical community. Data may be missing for a variety of common reasons including equipment malfunction, contamination of samples, manufacturing defects, drop out in clinical trials, inhospitable conditions for data collection, and incorrect data entry. In this talk, we are concerned with statistical inference regarding parameters of a single-index regression model (SIM) with missing responses. We assume that responses are missing at random (MAR) as discussed in Rubin (1976). Under this assumption, the probability that a response variable is observed can depend only on the values of those other variables that have been observed. The SIM on the other hand has gained popularity due to its ability to avoid the curse of dimensionality by searching for a single linear combination of covariates that captures the most information about the relationship between the response and the covariates. In addition to reducing the dimensionality in nonparametric estimation, the model under study has unspecified function g(.) which allows model flexibility yand thus has less risk of model misspecification. An elaborate discussion on the importance and use of single index models can be found in Wu et al. (2010). In the absence of missing data, papers that have dealt with this model include Ichimura (1993), Horowitz and Hle(1996), Carroll et al. (1997), Delecroix et al. (2006), Xia (2006), Wu et al. (2010), Kong and Xia (2012) and Bindele et al. (2018) among others. There are some existing results for the estimation of the regression coefficient β_0 in model (1.1) under the MAR assumption. These include those of Wang et al. (2010), Guo et al. (2015), Wang et al. (2016), and Niu and Zhu (2017) among others. Most of these are based on the least squares (LS) method.

Methods

Our interest is in the robust and efficient estimation of β_0 in single-index regression model (SIM) when some responses are MAR. When it comes to dealing with MAR responses, the standard approach involves imputing the missing responses and then analyzing the new data as if they were complete. Among the existing imputation methods, the commonly used in most regression problems include linear regression imputation (Healy and Westmacott, 1956), nonparametric kernel regression imputation (Cheng, 1994; Wang and Rao, 2002), multiple imputation(Schafer, 1997; Rubin, 2004), and semi-parametric regression imputation (Wang and Sun,2007). Another imputation approach that has proven to be an effective way of reducing the model parameter estimator bias that could be introduced when observations with missing responses are ignored, is the inverse probability weighting procedure introduced by Wanget al. (1997). This approach, however, usually depends on high dimensional smoothing for estimating the completely unknown propensity score function and therefore, suffers from the curse of dimensionality that may restrict the use of the resulting estimator. A more detailed discussion of this approach and related approaches can be found in Robins et al.(1994), Zhao et al. (1996), Wang et al. (1997), Wang et al. (2004) and references therein. Wang et al. (2004) proposed the inverse marginal probability weighting method as a way to avoid the curse of dimensionality problem that may occur when using the inverse probability weighting imputation approach. In an effort to derive a robust and efficient estimator of β_0 in SIM model with MAR responses, we propose a rank-based approach using the dispersion function proposed by Jaeckel(1972). The motivation comes from the fact that the minimization of this dispersion function results

in robust and efficient estimators (Hettmansperger and McKean, 2011; Bindele and Abebe, 2012; Bindele et al., 2018). Moreover, resulting estimators have simple geometric interpretations comparable of that of LS and furthermore, the proposed approach does not require specification of the model error distribution.

Results

When the errors follow a normal distribution, the least squares index direction estimator tends to be more efficient than the rank-based index direction estimator; however, the least squares link function estimator remains less efficient than the rank-based link function estimator. A real data example is analyzed and cross-validation studies show that the proposed procedure provides better prediction than the least squares method when the responses contain outliers and are missing at random.

Conclusion

A single-index regression model is considered, where some responses in the model are assumed to be missing at random. Local linear rank-based estimators of the single-index direction and the unknown link function are proposed. Asymptotic properties of the estimators are established under mild regularity conditions. Monte Carlo simulation experiments show that the proposed estimators are more efficient than their least squares counterparts especially when the data are derived from contaminated or heavy-tailed model error distributions.

Combining data from national surveys with facility-based HIV testing to obtain more accurate estimate of HIV prevalence in districts in Uganda

'Joseph Ouma [2]Joseph Valadez, [3]Caroline Jeffery, [4]Rhoda Wanyenze, [5]Jim Todd, [6]Jonathan Levin
'University of Witwatersrand' Johannesburg' South Africa [2]Liverpool School of Tropical Medicine,
[3]Liverpool School of Tropical Medicine, [4]Makerere University' School of Public Health, [5]' London School of Hygiene and Tropical Medicine' London' United Kingdom, [6]University of Witwatersrand' Johannesburg' South Africa

Background

National or regional population-based surveys are reduced to small sample sizes when HIV prevalence estimates are reduced to the district (or lower) level for program monitoring and decision making. Health facility program data collected as part of service delivery is widely available but being a convenience sample, has limitations for general population health indicator monitoring. We present a statistical annealing technique, Hybrid Prevalence Estimation (HPE) methodology, that combines a small population survey sample with a facility-based sample to generate more accurate population level health indicator estimates. We apply the methodology to combine information from non-facility testers captured in a population survey with health facility HIV testing data to obtain improved HIV prevalence estimates at district levels in Uganda.

Methods

We analyzed data from the 2011 Uganda AIDS indicator survey. Multilevel logistic regression was used to obtain the propensity to test for HIV in a health facility. The probability for testing for HIV in a health facility was then used to combine population survey data of the "non-facility testers" with facility testing data to obtain district level hybrid HIV prevalence estimates. We further adjusted regional prevalence estimates from health facility data to be similar to population survey prevalence estimates to minimize inherent bias in health facility data. Health facility data comprised of HIV testing data for the period January to December 2012.

Results

District level HIV prevalence estimates were obtained. The estimates obtained had narrower confidence intervals compared to survey-based estimates. The average difference between the hybrid prevalence estimates and population survey estimates was less than 1%, while the decrease in standard errors ranged from 5.4% to 95%. Average percentage decrease in standard errors was 28.7%. Prevalence estimates varied between districts, ranging from 1.2% to 17.8%.

Conclusion

Facility data can be combined with population survey data to obtain more accurate HIV prevalence estimates for smaller areas with small survey sample sizes. Applying the HPE technique while accounting for the unrepresentativeness of the HIS data, leads to more accurate indicator estimates required for effective area-specific program monitoring, decision making and efficient utilization of the already scarce resources.

Fertility Differential between Mombasa and Nairobi counties in Kenya

'Lilian Owino ²Elias Obudho
'none ²University of Nairobi

Background

In this current day and age policies are formulated based on data which sometimes are modeled and analyzed using mathematical models. This paper seeks to estimate fertility of Nairobi and Mombasa counties in Kenya using Gompertz mathematical model. These two counties are majorly urban centers and vibrant economic hubs in Kenya they are also the most populated counties in Kenya according to KDHS 2014 survey. Change in governance structure in Kenya in 2010 resulted in 8 provinces being subdivided into 47 counties. Demographic research and analysis shift from previous provincial administration to county level. Therefore this study using KDHS 2014 dataset, the first survey on health and demography to be conducted at county level in Kenya, will estimate fertility level at county level. Estimation of fertility rates aids in assessing impact of population intervention program that have been implemented while providing updated information on population growth. For adjustment of birth at differing age relational Gompertz model is used to estimate fertility rate. The model remedies errors associated with high fertility rate where fertility is concentrated within certain age brackets and very low outside this bracket. Relation model is suitable for estimating fertility rate of Mombasa and Nairobi counties as it will control variability and ensure reliability of estimates. Objectives The aim of the study was To estimate fertility rate of Mombasa and Nairobi using Gompertz model To estimate fertility rate at socioeconomic level And using past data to forecast fertility rate of the whole country Kenya in the near future

Methods

Method of approach Relational Gompertz model is a modification of Coale and Trussell model by Brass(1978) that sought to estimate age specific fertility and total fertility using recent data on births reported in census or surveys by determining shape of fertility schedule. Retrospective data on children ever born suffer from reporting errors that are adjusted when using relational Gompertz model. The function is F(x)/F = ABxWhere:F(x) is the cumulated age specific fertility rates up to age xF is the total fertilityA and B are constants between 0 and 1 A and B describe the pattern of fertility, A being the proportion of total fertility obtained by age x and B is the determinant for growth rate measuring spread of age specific fertility.The Gompertz function is converted into a linear form using logarithms for convenience purposes during analysis

Results

FindingsData variables used in this study are:• Five year age group of women (15-19,20-24,25-29,30-34,35-39,40-44,45-49) • Number of children born in the past one year by five-year age group of women • Marital status by five-year age group of women • Education level by five-year age group of women • Contraceptive use by five-year age group of women Results obtained show that women with higher level of education have low TFR (total fertility rate) in both counties (3.03 in Mombasa County and 2.25 in Nairobi County) while women with no education have high TFR (6.8 in Mombasa County and 5.11 in Nairobi county). Married women in Nairobi county fertility rate is 3.67 which is lower than those who have never been in a union (4.58). In Mombasa County married women had high fertility rate of 4.2 which was high compared to those who had never been in a union (3.52). TFR among those using contraceptive is low in both counties (2.9 in Nairobi and 3.4 in Mombasa) when contrasted against those not using contraceptives (3.9 in Nairobi and 4.1 in Mombasa). Teenagers (15-19 years) with no basic

primary education shy away from using contraceptive only 1% are users and 69% of female with primary education are the major contraceptive users.The total fertility rate in Nairobi is 2.85 and total fertility rate in Mombasa is 3.2, they are both less than TFR of previous years. TFR of Kenya in 2014 is 3.18 and is predicted to be 3 in the year 2019.

Conclusion

From the findings it seems interventions and policies that encourage female education beyond primary education will result in less births thereby resulting in lower TFR. Creating awareness on the various available contraceptive methods in both counties targeting all the females aged 15 years to 49 years will help them make informed decisions.

Probabilistic modeling for an integrated temporary acquired immunity with norovirus epidemiological data

'Emmanuel de-Graft Johnson Owusu-Ansah [2]Benedict Barnes, [3]Tine Hald, [4]Anders Dalsgaard, [5]Anders Permin, [6]Torben Wilde Schou, [7]Robert Clement Abaidoo

'Kwame Nkrumah University of Science and Technology [2]Kwame Nkrumah University of Science and Technology, [3]Technical University of Denmark, [4]University of Copenhagen, [5]Technical University of Denmark, [6]DHI, [7]Kwame Nkrumah University of Science and Technolgy

Background

The adoption of quantitative approach in modeling epidemiological risk over the years has gained interest in the field of disease modeling and epidemiology. This approach has led to the use of adopting techniques in characterizing uncertainty of input parameters with cumulative density functions, probability density functions and probability mass functions.Mathematical epidemiology of diseases with compartmental models have covered almost all forms of diseases emphasizing the different transmission approaches with intervention inclusion models. Yet, only a few studies have been concentrating on using the same approach for the pathogen of interest causing such diseases. The transition of infection to illness for pathogens has received less attention in quantitative risk modeling. This is as a result of lack of integration of epidemiological studies approach in quantitative risk assessment, such as a quantitative model of a susceptible individual suffering from pathogenic attack, transiting from partial recovery to become susceptible again.In the light of the above, this study attempts to apply the integration of mathematical epidemiological method into quantitative risk assessment for modeling probability of illness by incorporating acquired immune function to predict the expected probability of illness for exposure to Norovirus, a major gastrointestinal pathogenic virus.

Methods

Integration of acquired immunity into microbial risk assessment for illness incidence is of no doubt essential for the study of susceptibility to illness. In this study, a probabilistic model was set up as dose response for infection and a mathematical derivation was carried out by integrating immunity to obtain probability of illness models. Temporary acquire immunity from epidemiology studies which includes six different Norovirus transmission scenarios such as symptomatic individuals infectious, pre- and post-symptomatic infectiousness (low and high), innate genetic resistance, genogroup 2 type 4 and those with no immune boosting by asymptomatic infection were evaluated. Four different probabilistic models were considered, where compartmental model was integrated to account for events of immunity function.

Results

Simulated results on illness inflation factor as a function of dose and exposure indicated that high frequency exposures had immense immunity build up even at high dose levels; hence minimized the probability of illness. The transmission dynamics in all scenarios had probability of infection/illness incidence for dose-immunity models within 1 log of order, dose models and the naïve model did not see any change. However, A difference of 7 logs (median values) of magnitude was found between the dose-immunity and the naïve model for all epidemiological transmission dynamics. Five and three logs differences for dose-immunity as against the dose model and the immunity model, respectively. Hence, the probability of infection/illness decreases from the naïve, the dose model, the immunity model and the dose-immunity model. The individual infection/illness risk estimates for various immunity-incorporated models across the transmission scenarios gave a much lesser risk incidence as compared to the naïve and

the dose-model approachesTransmission dynamics.Across the different transmission dynamics scenarios with respect to their loss of partial and full temporary immunity protection levels, a comparison of models of the epidemiological scenarios using their median values did not show a difference in order of magnitude from transmission scenario to scenario with the exception of the immunity dose model which had 1 log less for pre-symptomatic and post-symptomatic low. Therefore, difference in the probability of infection/illness is not sensitive to the epidemiological transmission scenarios, thus, norovirus transmissions dynamics does not influence the probability of infection/illness predictions Furthermore, the findings observed no difference for all infection response models of the transmission dynamics for symptomatic infectiousness' and the no immune boosting after asymptomatic infectiousness. Considering the findings under models without temporarily acquired immune inclusion, both the median values and mean values as well as their deviations were found to be the same for pre-symptomatic and post-symptomatic (high) and symptomatic infectiousness as well as recording similar estimates for innate genetic resistance and geno-group 2 Type 4' transmissions dynamics.

Conclusion

Applying the models to the norovirus data, resulted with the same trend of movement on the various dose-response models, and individual level of illness incidence reduction was much better measured by immunity incorporated models.The immunity incorporated models tend to predict a lower illness/infection incidence, while the non-immunity incorporated models do not. Again, the immunity dependent models (immunity and dose-immunity models) meet the more stringent WHO infection/illness threshold of 1 : 10-6 per person per exposure in all transmission scenarios.The resulting magnitude of decrease in probability estimation of risk of illness is having a pronounced effect on the estimation of diseases as a result of incorporation of the temporary immune dose response, irrespective of the scenario of transmission of infected individuals.These modeling results throw more light on the overestimation of the probable infection/illness as a result of the use of naïve and dose models approach. It is important to note that the immunity model, which is a buildup of inclusion for immune system response in the first stage, is a better estimation in terms of predicting the reality of infection/illness of exposure than the naïve approach. The lower estimates for the immunity-incorporated models indicates the impacts of the temporary immune response to offer full protection and always results in lower estimates as compared to the naïve estimates. It is also important to note that the dynamics of the norovirus transfer from person to person does not influence the responses of the models, hence the prediction models superimpose on the type of the transmission of the virus

Some New Nonlinear Growth Models For Biological Processes based on Hyperbolic Sine Function

'Oluwafemi Samuel Oyamakin [1]Angela Unna Chukwu Chukwu, [1]Timothy Adebayo Bamiduro Bamiduro
'Department of Statistics' University of Ibadan

Background
Studies have shown that majority of the growth models emanated from the Malthusian Growth Equation (MGE), which is limited to growing without bounds. This study was designed to develop alternative growth models flexible to enhance internal prediction of biological processes based on hyperbolic sine function with bound.

Methods
The intrinsic rate of increase in the MGE and its variants were modified by considering a growth equation, which produces flexible asymmetric curves through nonlinear ordinary differential equations of the form; dH/dt=H[r+θ/√(1+t^2)].

Results
The developed hyperbolic growth models captured boundedness in Malthusian Growth Equation, improved general fitness and robustness over exponential, monomolecular, Gompertz, Richards and von Bertallanffy growth models.

Conclusion
The developed hyperbolic growth models improved general fitness over exponential, monomolecular, Gompertz, Richards and von Bertallanffy growth models in predicting tree heights, diameter at breast heights along with other published datasets used demonstrating the effectiveness of the developed models

Evaluations of error in variable regression methods for the analysis of ECG data

'Ring Arne [2]Genevieve Mensah, [3]Gifty Obeng
'University of the Free State, South Africa & medac Germany [2] African Institute of Mathematical Science (AIMS), Senegal, [3] African Institute of Mathematical Science (AIMS), Ghana

Background

Electrocardiograms (ECG) of patients can indicate health related issues. One example is the QT interval, which should not exceed specific thresholds, such as QT intervals of more than 450 ms duration indicate a prolonged repolarisation. However, the length of the QT interval is related to the heart rate (Heart rate = 60 / RR interval). Therefore it needs to be corrected for changes of the heart rate (HR=60/RR interval), and this relationship can characterised using a linear relationship after log-transformation of the RR and QT intervals. Quantitative data of ECG intervals often have high variability. Therefore, standard regression methods might not be suitable to analyse the relationship of these data, as they tend to lead to biased estimates of the slope, intercept and the residual variability. Various methods have been presented in the literature to correct for this bias. We want to evaluate the appropriateness of these methods with a particular focus on the structure of ECG data from clinical trials.

Methods

The relationship between RR and QT interval can be modelled using a parabolic model. This can be described by:QT = n [RR]^mwhere RR and QT are recorded in seconds After logarithmic transformation, this leads to ln(QT) = m ln(RR)+nThis leads to the regression model y = β_1 x+ β_0Hereby is x a representation of the RR interval, and y of the QT interval. Because of the natural heart-rate variability, both variables are recorded with error, denoted by z_i = y+ε_iw_i = x+δ_iwhere ε_i and δ_i are independent, normally distributed errors with mean 0 and standard deviations σ_y and σ_x. It had been shown [Ring 2010] that ignoring the variability in the independent RR interval leads to biased estimates of the regression parameters β_1 and β_0, hence to an incorrect heart rate correction. Quantitative ECG data are typically recorded in hierarchical levels, e.g. 4 wave forms are measured in each ECG, and 3 ECGs are recorded at each time point, which will extend the model to a mixed effects model. These repetitions are typically used to reduce the impact of the heart rate variability. The first step of this investigation was to simulate data using various well-defined scenarios using Monte-Carlo-Simulations. Simple scenarios use an equidistant schema of the independent variable (with addition of random error), while other scenarios model the hierarchical structure of ECG data. The data are analysed with standard regression as well as specific error-in-variable regression techniques, such as the Deming regression, Principal Components analysis and Passing-Bablok regression. For each method, the distribution of the estimates is determined. The results of the techniques are to be compared using key quality parameters such as bias and root-mean-squared error. In a second step, the methods are applied to a real data set of ECG data and the results are compared.

Results

The analysis is ongoing, results are to be presented at the conference. Preliminary results indicate that the Deming regression can partly correct for error in variables. However in scenarios with limited error in the independent variable are overcorrected by Deming regression.

Conclusion

Appropriate methods for the analysis of the relationship between log(QT) and log(RR) data require the application of error-in variable techniques to avoid bias of the parameter estimation.

Predictors of non-completion to Isoniazid Preventive Therapy among People Living with HIV attended Care and Treatment Clinics in Dar-es Salaam from 2013 to 2017

'Masanja Robert [2]Jim Tod, [3]Bernard Ngowi, [4]Sia Msuya, [5]Werner Maokola, [6]James Ngocho, [7]Michael Mahande

'Kilimanjaro Christian Medical University college (KCMUCo) [2]London School of Tropical Medicine (LSTM)' United Kingdom, [3]National Institute for Medical Research-Muhimbili Medical Research Centre' Tanzania, [4]Kilimanjaro Christian Medical University College (KCMUCo)' Tanzania, [5]Ministry of Health Community Development Gender' Elderly and Children (NACP)' Dodoma, [6]Kilimanjaro Christian Medical University College (KCMUCo)' Tanzania, [7]Kilimanjaro Christian Medical University College (KCMUCo)' Tanzania

Background

Introduction: People Living with HIV have higher risk of acquiring active tuberculosis compared to HIV negative counterparts. This leads to excess mortality in this group. Isoniazid preventive therapy has been introduced to reduce to progression to active tuberculosis. Conversely, tuberculosis incidences has remained to be high in Tanzania. Dose non-completion poses challenges towards tuberculosis control strategy, of which slow down attainment of sustainable development goals by 2035. This study aimed to determine predictors of non-completion to Isoniazid Preventive Therapy among People Living with HIV attended care and treatment in Dar es Salaam from 2013 to 2017.

Methods

Methods: A retrospective cohort study was conducted using routinely collected data from 231 care and treatment clinics in Dar es Salaam region. A total of 41283 PLHIV were initiated IPT. Of these, 17356 of the participants who had an outcome status either completed or not completed at six months were analysed. Data analysis was performed using STATA 15.0. Bivariate and multi-level logistic regression analysis were used to identify predictors for non-completion to IPT.

Results

Results: The proportions non-completion to IPT was 17.21%. The mean age: 41.03[±SD:10.17] and in 2014 was 44.18% high. Predictors of non-completion of IPT were ARV status (stopped ARV versus on ARV) OR: 2.43, 95%CI [0.44, 13.26]; functional status (ambulatory versus work) OR: 12.73, 95%CI [4.93, 32.88], ICC: 8.3%; WHO stage (II versus III) OR: 1.11, 95%CI [1.008, 1.228], ICC: 8.1% and other medications(other versus cotrimoxazole) OR: 0.85, 95%CI [0.73, 0.98], ICC: 24.8%.

Conclusion

Conclusion: Non-completion of IPT among PLHIV is still high. This increases TB incidence in population and thereby slow down the effort towards achievement of sustainable development goals by 2035. Strategies to reduce predictors for non-completion to IPT is warranted.

Risk factors and spatial heterogeneity of childhood anaemia in four sub-Saharan African countries

'Danielle Roberts [2]Temesgen Zewotir
'University of KwaZulu-Natal [2]University of KwaZulu-Natal

Background

Anaemia in children is a significant health problem that receives little attention. It is responsible for an increased risk of morbidity and mortality in young children and pregnant women where the majority of the burden is experienced by these populations in sub-Saharan African countries. Anaemia contributes to a wide range of symptoms in children, from fatigue and weakness to impaired psychomotor, behavioural and physical development, and even death. Anaemia in children can be caused by a variety of factors, and due to the multifactorial and complex nature of these causes, the risk factors are multiple and vary not only from country to country, but also between the different regions of a country. A considerable number of studies assessing the risk factors and determinants of anaemia in children have been done, some of which have also assessed the spatial variation of anaemia. However, few studies have focused on a regional level within Africa. Identifying significant factors associated with an increased risk of anaemia in children is relevant to developing appropriate and effective interventions. Such studies can also aid in identifying subpopulations that are most at risk, which assists in creating a more efficient delivery system of limited national resources. However, these studies should account for heterogeneity due to spatial dependence in observations. Failure to do so may produce inaccurate estimates and thus misleading results and ineffective anaemia control programs. This study aimed to determine the significant risk factors of anaemia in children aged 6 to 59 months in Kenya, Malawi, Tanzania and Uganda (eastern sub-Saharan region countries) while accounting for the spatial heterogeneity within and between the countries.

Methods

The study made use of data collected from nationally represented Malaria Indicator Surveys (MIS) and Demographic and Health Surveys (DHS) conducted in all four countries between 2015 and 2017. The geographical coordinates of the clusters, which made up the primary sampling units, were recorded during all the surveys. Furthermore, all children under the age of five years old in the sampled households were tested for malaria and anaemia. A child's anaemia status was based on the World Health Organization's cut-off points where a child was considered anaemic if their altitude adjusted haemoglobin (Hb) level was less than 11 g/dL, or not anaemic otherwise. This child anaemia test outcome was used as the characteristic of interest in this study. The risk factors considered for investigation comprised of individual, household and cluster level factors, including the child's malaria Rapid Diagnostic test (RDT) result. Due to the survey as well as the data structure, a hierarchical structured additive regression (STAR) model with a logit link function was used. Furthermore, the effect of correlated unmeasured spatial effects at each cluster location was accounted for by the inclusion of a bivariate non-linear function of the longitude and latitude based on the geographical coordinates of the clusters. In addition, an independently and identically distributed random effect based on the district in which the child resided was included in order to further account for the heterogeneity in responses between the districts of the countries. Estimation was carried out using an empirical Bayesian approach where the variance components were estimated using restricted maximum likelihood (REML) estimation. Non-linear effects of continuous covariates were explored. Moreover, all the two-way interactions between the fixed effects were assessed in order to avoid possible confounding effects.

Results

The final data from the four countries consisted of 18027 children from 1562 clusters for which valid geographical coordinates were available. The overall observed prevalence of anaemia was 52.3% with a 95% confidence interval of 51.8% to 53.3%. The observed prevalence of anaemia was lowest in Kenya at 38.5%, with the other countries having a much higher observed prevalence ranging from 53.8% to 58.3%. The result of the spatial model depicted significant spatial heterogeneity between and within the districts of the countries. After accounting for such spatial heterogeneity, child level characteristics (gender, malaria RDT result, and mother's highest education level), household level characteristics (household size, household's wealth index Z-score, the type of toilet facility available, and the type of place of residence) as well as the country of residence were found to significantly associated with the child's anaemia status. Female children were associated with a lower risk of anaemia compared to males. The odds of anaemia was significantly higher for children who tested positive for malaria compared to those who tested negative. The odds of anaemia increased with an increase in household size, however the odds decreased with an increase in the mother's education level, household wealth index and improved toilet facilities. Furthermore, there was a significant non-linear effect of his/her age in months on the log-odds of anaemia, where the effect increased from 6 to 10 months of age, after which there was a decline in the effect. There was a significant interaction between the type of place of residence (rural/urban) and the country of residence (Kenya, Malawi, Tanzania or Uganda). This interaction displayed a considerable difference in the effect of the type of place of residence on the log-odds of anaemia between the four countries.

Conclusion

Programs that educate parents and create awareness about the causes and consequences of anaemia in children would be extremely beneficial, especially targeting the parents of children in the younger age group of 6 to 10 months as these children are more susceptible to anaemia due to the rapid growth during that stage of their lives. Due to the significant association between malaria and anaemia in children, highly malaria-endemic regions should ensure regular testing for anaemia and malaria in young children as it is important to treat the cause of anaemia and not just the symptom. Anaemia control measures in Kenya, Malawi, Tanzania and Uganda need to account for the spatial heterogeneity that is evident in these countries. Accordingly, efforts in creating an eradication or prevention plan for childhood anaemia within each country should be focused on the local district-specific situations instead of a one size fits all strategy.

Nonparametric cure rate estimation when cure is partially known

Wende Clarence Safari[1], Ignacio Lopez-de-Ullibarri[2], Mara Amalia Jacome[3]
[1] Universidade da Coruna, CITIC, MODES group, Department of Mathematics, Faculty of Computer Science, A Coruna, Spain, [2] Universidade da Coruna, MODES group, Department of Mathematics, Escuela Universitaria Politecnica, Ferrol, Spain, [3] Universidade da Coruna, CITIC, MODES group, Department of Mathematics, Faculty of Science, A Coruna, Spain

Background

A common assumption in standard survival modeling is that all individ- uals can experience the event if observed for enough time. Cure models have been developed because there might be situations where the standard survival model is not true. For example, in cancer studies, due to advances in cancer treatment there might be a proportion of patients who will get cured. A common aspect in traditional cure models is that cured and uncured subjects cannot be distinguished within the censored observations. Hence, the cure indicator is usually modeled as a latent variable. However, some-
times this assumption is not entirely valid, when some extra information allows to conclude that some individuals with censored lifetimes are cured or long-term survivors. One typical example is the case if individuals are assumed to be cured when their survival time is larger than a given thresh-old (e.g., 5 years when considering recurrence in some types of cancer). In this paper, a nonparametric estimator of the cure rate in the presence of a known cure fraction and conditional on a covariate is introduced.

Method

Suppose Y is a random variable representing time to event of interest, S(t) = P(Y > t) is the survival function, and C is the censoring time. Y and C are independent given a covariate X. It is assumed that the studied population is a mixture of individuals: those who will and those who will not Nonparametric cure rate estimation experience the event of interest. According to this assumption the survival
function can be written as where S0(tjx) is the survival function of the uncured or latency conditional on X = x, and 1p(x) is the probability of being cured. The estimation of the model is usually performed with parametric or semiparametric models. Xu and Peng (2014) and Lopez-Cheda et al. (2017) proposed a nonpara-metric mixture cure model which ignored the existence of known cures. In
the presence of known cures the observations are where X is a covariate, T = min (Y;C) is the observed time, = 1(YC) is an uncensoring indicator, is a binary variable which indicates the
cure status is known (= 1) or not (= 0), and is the cure indicator. Therefore, = 1 indicates that the individual is known to be cured. Given the observations, the proposed estimator of 1 p(x) = P(Y = 1jX = x)
is, where Bch(x) =Pnj=1 Bh[j] (x) 1[j][j] = 1is the sum of the weights of all the individuals known to be cured, are the Nadaraya-Watson weights with Kh (:) = 1=hK (:=h) a rescaled
kernel with bandwidth h. Finally, [i];X[i]; [i] and [i] are the concomitants of the ordered observed times T(1) < T(2) < : : : < T(n). The intuition behind this estimator is that the cure proportion corresponds to the limit of the survival function as t tends to in unity. It can be proved that 1 bph(x) is the nonparametric local maximum like-lihood estimator of the cure rate.

Sarcoma data results

While this estimator can be applied into dierent research areas, the moti-vation in this paper was from a data set related to patients with sarcomas. There were 233 sarcoma patients in the data set. The main outcome was death from sarcoma. A total of 59 (25:2%) patients died from sarcoma, and 174 (74:8%) patients were censored. Within the censored patients, a total of 18 patients were known to be long-term

survivors, as they were tumor free for more than five years. Our estimator was used to estimate the probability of cure conditional on patient's age (ranging from 20 to 90 years). The estimate was compared with that produced by the semiparametric estimator of Bernhardt (2016), which assumes a logistic model for the cure probability.

Conclusion

Although the later estimate suggests a uniformly decreasing effect of the age on the cure rate, the curves from the proposed estimator are more consistent with a pattern characterized by a sharp decrease of the cure rate at younger ages until reaching a plateau at older ages.

Superiority and non-inferiority hypothesis testing with functional data endpoints

'Sandie Arsene Brunelle [2]Tchatchueng-Mbougua Jule Brice, [3]Wanjoya Anthony
'Pan African University, Institute of basics Science, Technology and Innovation/ Jomo Kenyatta University of Agriculture and Technology (PAUSTI/JKUAT) [2]Centre Pasteur du Cameroun, [3]JKUAT

Background

The clinical trials with binary, time-to-event and continuous endpoints have been commonly used in practice. And, many statistical test procedures have been studied and proposed in the framework of non-inferiority, superiority and equivalence hypothesis tests. However the same cannot be said for clinical trials with functional data outcomes, although it could have some relevant applications in practice. That is the case for example of longitudinal trials with repeated continuous endpoint, where the non-inferiority, superiority or equivalence hypothesis testing is generally performed after a fixed period of time, commonly at the end of the follow-up period. However, those longitudinal data can be turned into functional data on a continuum domain, and then instead of performing the hypothesis testing at the end of the follow-up period, one can get the decision about non-inferiority, superiority or equivalence hypothesis testing on the whole continuum domain.

Methods

In this work, it is introduced superiority and non-inferiority hypothesis tests for functional data outcome. Then, the pointwise and global test approaches have been used in the purpose of statistical hypothesis testing. For the pointwise test, the optimal pointwise test proposed in Xu et al. (2018) has been adopted. Concerning the global test, simultaneous confidence bands have been used to construct a test procedure with a single decision. The proposed test procedures have been assessed through a Monte-Carlo simulation example. For the pointwise test, the marginal false discovery rate and modified power were evaluated according to the sample sizes. While, for the confidence bands based test procedure, the actual type I error rate and the power were estimated according to the sample sizes as well.

Results

both proposed methods have good performances for large sample sizes. For small sample sizes, the optimal pointwise test would be too conservative while the simultaneous confidence bands based test would be a bit liberal.

Conclusion

The introduced non-inferiority and superiority hypothesis tests with functional data outcomes are more flexible in the interpretation of the results since the decision is on the whole continuum domain instead of the test performed only at the end of the follow-up period. Moreover, the results of the study show that both proposed methods can be used in practice.

High Dimensional Surrogacy: Modeling and Computational Aspects

'Rudradev Sengupta, [2]Nolen Joy Perualila, [3]Ziv Shkedy, [4]Geert Molenberghs, [5]Ariel Alonso Abad, [6]Wim Van der Elst, [7]Luc Bijnens

'Quantitative Sciences' Janssen Pharmaceutical companies of Johnson and Johnson [2]HEMAR EMEA' Janssen Pharmaceutical companies of Johnson and Johnson, [3]Center for Statistics' Hasselt University, [4]Center for Statistics' Hasselt University, [5]Leuven Biostatistics and Statistical Bioinformatics Center' K U Leuven, [6]Quantitative Sciences' Janssen Pharmaceutical companies of Johnson and Johnson, [7]Quantitative Sciences' Janssen Pharmaceutical companies of Johnson and Johnson

Background

In drug discovery experiments, one of the main challenges is a very slow, but costly and inefficient development process. The choice of endpoint(s), to assess the drug efficacy, plays an important role and it influences the duration of the development process. However, measuring the endpoint(s) can become difficult, time consuming and expensive. Thus, detection of biomarkers has become an important task in the context of drug discovery. A large amount of research (Burzykowski et al., 2005; Alonso et al., 2016; Perualila Tan et al., 2016) has been devoted to the identification of biomarkers for a primary endpoint in a setting where the biomarker is a part of a high dimensional data (e.g., genetic biomarkers for transcriptomics studies). Moreover, the ability and the need to analyze routinely large scale datasets in early drug discovery has become a central issue in the last few years in data analysis workflow of early drug discovery experiments. In a typical drug discovery study, while exploring new compounds, the chemical structure of a specific compound is known, but not all biological processes related to the compound. The main aim of the analysis is to investigate the association between one (or more) transcriptomic biomarker(s) and a clinical endpoint of interest. Perualila Tan et al. (2016) proposed a joint model to achieve this goal, by using the information about transcriptomics data to understand the biological processes related to a new compound and to gain a complete insight about its mechanism of action (MoA).

Methods

We focus on two different extensions of the joint model proposed by PerualilaTan et al. (2016): (a) Modeling aspects (Sengupta et al., 2019) and (b) Computational aspects (Sengupta et al., 2019). So far, the methodology for biomarker detection was focused on detecting a single biomarker. In the current study, we proposed a new framework in which multiple biomarkers can be identified from a collection of potential biomarkers. Multiple surrogacy framework (Van der Elst et al., 2019) can be used to identify a subset of K biomarkers which can be used together as a biomarker for the primary endpoint of interest (joint surrogacy). Our model allows to estimate the surrogacy effect of the Kth biomarker, given the surrogacy effect of (K 1) biomarkers (partial surrogacy). Finally, orthogonal surrogacy is a special case of partial surrogacy, consisting of K independent biomarkers. When the number of potential biomarkers, K, increases, computational time becomes an important issue that needed to be taken into account. While existing R packages can be used for parallelization, the master slave framework, that was developed for the joint modeling setting, outperformed the existing R packages. The analysis conducted in this study requires an access to a computer cluster (VSC, in our case). Users without access to computer cluster can execute their analysis in publicly available cloud clusters such as Amazon EC2 Cluster, Microsoft Azure Cluster, etc. A possible implementation of the master slave framework using cloud computing platform is a subject of an ongoing research.

Results

The proposed method is illustrated using a case study with 3595 candidate biomarkers for a bioactivity variable which was measured for 35 compounds under development. We showed that upscaling an analysis from a setting of a single “for loop” to the setting, which become currently a standard in data analysis, of multiple “for loops” allows to analyze large scale datasets in short time but requires a careful planning of the data analysis configuration. The analysis presented in Buyse and Molenberghs (1998) was focused on the evaluation of a single surrogate endpoint. A similar analysis, upscaled for the evaluation of 496110 possible biomarkers, can be done in 97.64 seconds following the proposed master slave framework

Conclusion

This reduction in computation time led to further analysis and revealed some interesting results related to multiple, partial and orthogonal surrogacy when the gene FOSL1 was used as the known biomarker (Verbist et al., 2015).

Digital data capturing: Using REDCap for a multi-arm, multi-site randomised clinical trial

'Ishen Seocharan [2]Mandisa Singata, [3]Yusentha Balakrishna
'South African Medical Research Council [2]Effective Care Research Unit, [3]South African Medical Research Council

Background

Prevention of unintended pregnancy and prevention of HIV infection rank among the most important and cost-effective public health strategies to promote the health and social and emotional wellbeing of women, their families and society. Both are complex interventions delivered in community primary health care settings. Injectable contraceptives are the most popular method among South Africa women, including women from settings with high poverty levels in the Eastern Cape and KwaZulu-Natal.The Women's Health, Injectable Contraceptive and HIV (WHICH) study conducted a randomised clinical trial comparing HIV acquisition risk between users of NET-EN versus DMPA among women. However, without effective record keeping and follow-up with participants, related information is often difficult to manage. Additionally, laboratory tracking and results of samples taken by an independent lab is required to be recorded. The implementation of a database in a longitudinal, multi-arm study, spanning multiple geographic sites is described.

Methods

The two study sites selected were the Effective Care Research Unit (ECRU) in Eastern Cape and MatCH (Maternal, Adolescent and Child Health) Research Unit (MRU) in KwaZulu-Natal. Participants were recruited from nearby clinics by field workers. An electronic participant registration and follow-up database was implemented with the use of Research Electronic Data Capture (REDCap, Vanderbilt University), a free, secure web-application with mobile-application features hosted by the South African Medical Research Council.Custom data collection elements were created included areas for participant screening, demographics, medical history, social history, concomitant medication, vital signs, interview notes and check lists. The web-application was used by office based staff and the mobile-application was used on tablets by clinic based staff. During the registration process, the participant needed to be randomised to study-arm. The randomisation module in REDCap was utilized to allocate participants to arms according to a stratified random block design.

Results

A total of 520 woman registered into the database across the two sites. The database streamlined the management of the data for all staff involved by providing one central location to store and access information related to the project. All sites could upload patient information into REDCap, which greatly decreased the number of paper records managed. The electronic forms used during record follow-up eliminated the need for paper forms and subsequent data entry, reducing staff time and data entry errors. The built-in reporting feature also allowed for the quick generation of reports and basic statistics to identify missing or incomplete records, and for monitoring the overall status of the project. Additionally, the calendar feature in REDCap helped organise the schedule of participants and kept track of upcoming visits, effectively improving participant retention. The project team could remotely track participant registration and follow-up at the two study sites throughout the duration of the project.

Conclusion

The implementation of a central online electronic tracking database brought significant improvements in the handling, recording and follow-up of participants across both study sites. It improved efficiency of the data validation and significantly reduced staff time. The mobile capability reduced the need of additional resources being required to digitise data.

Bayesian Approach in Treating Missing Values with Metabolomics

'Jasmit Shah , [3]Guy Brock, [4]Jeremy Gaskins
'Aga Khan University , [3]The Ohio State University, [4]University of Louisville

Background

With the rise of metabolomics, the development of methods to address analytical challenges in the analysis of metabolomics data is of great importance. Missing values (MVs) are pervasive, yet the treatment of MVs can have a substantial impact on downstream statistical analyses. The MVs problem in metabolomics is quite challenging and can arise because the metabolite is not biologically present in the sample, or is present in the sample but at a concentration below the lower limit of detection (LOD), or is present in the sample but undetected due to technical issues related to sample pre-processing steps. The former is considered missing not at random (MNAR) while the latter is an example of missing at random (MAR). Typically such MVs are substituted by a minimum value, which may lead to severely biased results in downstream analyses.

Methods

We develop a Bayesian model that systematically accounts for missing values based on a Markov chain Monte Carlo (MCMC) algorithm that incorporates data augmentation by allowing MVs to be due to either truncation below the LOD or other technical reasons unrelated to its abundance.. A key piece of building the model that can accommodate MV imputation is the choice of the structure of the dependence. We adapt the sparse Bayesian infinite factor model for the consideration of a flexible, lower-dimensional choice for the covariance matrix. Statistical inference may be performed using either the posterior samples of the parameters or by using the data sets imputed during MCMC.

Results

Based on a variety of performance metrics (power for detecting differential abundance, area under the curve, bias and MSE for parameter estimates) our simulation results indicate that our Bayesian method outperformed other imputation algorithms when there is a mixture of missingness due to MAR and MNAR. Further, our approach was competitive with other methods tailored specifically to MNAR in situations where missing data were completely MNAR. Applying our approach to an analysis of metabolomics data from a mouse myocardial infarction revealed several statistical significant metabolites not previously identified that were of direct biological relevance to the study.

Conclusion

Our findings demonstrate that our Bayesian method has improved performance in imputing the missing values and performing statistical inference compared to other current methods when missing values are due to a mixture of MNAR and MAR. Analysis of real metabolomics data strongly suggests this mixture is likely to occur in practice and thus it is important to consider an imputation model that accounts for a mixture of missing data types.

Diagnostics for a two-stage joint survival model

'Isaac Singini [2]Freedom Gumedze, [3]Henry Mwambi
'University of Cape Town [2]University of Cape Town, [3]University of KwaZulu-Natal

Background

Two-stage joint survival models are a class of models that analyse time to event outcomes that could be associated with bio-makers that are repeatedly collected over time. These models are considered to be superior over standard modelling of survival outcomes with a time-varying covariate because they account for measurement error in a time-varying covariate and do not assume that the value of the covariate remains constant between any two measurements. These models have limited model checking tools and are usually assessed using standard diagnostic tools for survival models. These diagnostic tools can be improved. In the two-stage joint model the time-varying covariate might contain outlying observations or subjects. These outliers might influence estimates used to obtain fitted values required in the second stage of the joint model.

Methods

n this study we propose a two-level application of the variance shift outlier model (VSOM) to detect and down-weight outliers. This entails fitting a VSOM at the observation level and a VSOM at the subject level, and then fitting a combined VSOM for the identified outliers. We then extract the fitted values from the combined VSOM which we use as time-varying covariate in the extended Cox model. The results from the two-stage joint model, with a combined VSOM at the first stage, are contrasted with those from the standard extended Cox model using AIC and BIC. We performed model checking using the influence diagnostics: dfbetas and likelihood displacement under all the three scenarios, and compared the results. We conducted simulation studies and for the longitudinal data we the artificially inserted five outliers that eventually get down-weighted by implementing VSOM. We illustrate this methodology on a multi-center clinical trial with 1400 patients diagnosed with probable or confirmed tuberculosis (TB) pericarditis.

Results

In the simulation studies the combined VSOM down-weighted outliers at observation and subject levels hence filtering some level of influence exerted by outliers on parameter estimates compared to standard two-stage joint model. An illustration on a multi-center trial on TB pericaditis has similar results to the ones obtained in the simulation studies.

Conclusion

We note that implementing a combined VSOM, when desired, has a better fit based on the fact that outliers are down-weighted to be consistent with the rest of the data points

Order statistics approach to modeling and prediction of early mood swing

'Ajibola Taiwo Soyinka , ²Akin Adeseye Olosunde
'Department Of Statistics; Federal University Of Agriculture, Abeokuta, Nigeria , ²Obafemi Awolowo University Ile-ife Osun state. Nigeria

Background
Though several works has been done on order statistics of many distributions, as far as I know, NO work has been done in the area of order statistics from exponential power distribution (EPD); which is a probability density function (pdf) that will serve as a substitute for both normal and double exponential distribution as well as all elliptically contoured distributions in theory and application. However since bipolar disorder describe an extreme disease conditions that fluctuate between extreme happiness and sadness, the tail regulation of EPD will be useful in describing the behaviour of mood disorder which is a pre-disease condition to bipolar risk disorder and facilitate its early detection.

Methods
The method of conditional probability was used to derive the left, right and double truncated order statistics conditional survival function (probability of non-occurrence) of the EPD. The obtained theoretical results is then used in the modeling of mood disorder dynamics of individuals who are yet to report any psychiatric problem. 2.) In addition, the conditional model also accommodates for response bias which is common among suspected mood disorder patients; aimed at misleading the medical personnel. So, the model correct the problem of incoherency or falsehood in patients claim by recommending a status that can be adequately managed by the psychiatrist.

Results
The study established the nature of bipolar disorder risk using order statistics tool as the model to explain mood swing in different individuals within extreme happiness (Mania) and extreme sadness (Depression). The order statistics approach which developed a shape dependent survival model reveals an improved level of representation of the real life situation compared to that of cubic spline approach. The study provides templates to predict and detect early mood disorder and provides guides to obtaining adequate therapy.

Conclusion
Hence this study is important to humanity because it provides a survival function template that will facilitate early detection of abnormal mood swing which the medical doctors, parents, guardian and teachers (lecturers) only become aware of when the situation has gone bad. This template can be domesticated with minimal knowledge with the aim of monitoring mood swing and providing adequate counseling. The study also provides template that will guide the psychiatrist in providing adequate therapy to client who deliberately gave false information to avoid stigmatization or admission.

Causal Inference for Multiple Outcomes for Observational Studies

'Halima Twabi [2]Samuel Manda, [3]Samuel Manda, [4]Dylan Small
'Department of Mathematical Sciences' University of Malawi' Zomba' Malawi [2]Biostatistics Research Unit' South Africa Medical Research Council' Pretoria' SA , [3]Department of Statistics' University of Pretoria' Pretoria' SA, [4]Department of Statistics' University of Pennsylvania' Penn state' USA

Background

In public health, multiple outcomes are measured on subjects in a study. For example, in studying the effect of exclusive breastfeeding on child growth, child nutrition measurements height for age, weight for height and weight for age are obtained on children. In the context of HIV studies, the effect of HIV status knowledge on marital status, number of sexual partners and condom use is considered. In each case, the outcome is multivariate and purportedly to be correlated and causes analytical problems when estimating cause effects of hypothesized exposure.

Methods

Using two datasets from Malawi, 2015/16 Malawi Demographic and Health Survey and Malawi Longitudinal Study for Family Health (MLSFH), we derive the causal effect of on the multivariate outcomes based on the D-product outcome approach. We compare results to those using univariate causal effect methods.

Results

Results showed that exclusive breastfeeding has both marginal causal and joint effect on childhood height for age, weight for height and weight for age (($p < 0.05$). HIV status knowledge has a marginal effect on condom use and sexual partners and marital status ($p<0.05$). HIV status knowledge has a joint effect on condom use and sexual partners only excluding marital status.

Conclusion

Joint models are important in studying the causal effect on multivariate outcomes as they account for correlation between outcomes, unlike univariate causal effect estimation. The joint models are particularly important in addressing public health problems as most outcomes collected in public health studies are associated.

Do parents' mental disorders affect child behaviour via home environments and/or primary caregiver's personal and social performance? A mediation analysis

'Md Jamal Uddin [2]Claus Thorn Ekstrøm, [3]Merete Nordentoft
'Section of Biostatistics University of Copenhagen [2]Section of Biostatistics, University of Copenhagen,
[3]Copenhagen University Hospital, Mental Health Centre Copenhagen, Denmark

Background

Children of parents with mental disorders, e.g. bipolar disorder or schizophrenia, are exposed to numerous environmental risk factors and may face different developments and behavioural problems. These problems may partly be linked to their home environments and/or primary caregiver's personal and social performance. Classical statistical methods (e.g. linear regression) are widely applied to assess the association between parents' mental disorders and child development problems in psychiatric epidemiology. However, research on this area using mediation analyses is not widely studied. Therefore, we aimed to investigate the association of parents' mental disorders and child behaviour when home environments and primary caregiver's personal and social performance act as mediators.

Methods

We used data from the Nationwide Danish High Risk and Resilience Study – VIA7 cohort. The VIA7 study was conducted in Denmark from January 1, 2013 to January 31, 2016. The study sample was 522 in which 202 7-year-old children living with their parents diagnosed with schizophrenia, 120 children living with their bipolar disorder parents and 200 control children living with their parents who did not have a diagnosis of schizophrenia or bipolar disorder. The outcome was considered as child behaviour checklist school-age version (CBCL) and the home environments and primary caregiver's personal and social performance were considered as mediators. A higher CBCL score indicates more behaviour problems of the children. We analysed data using different statistical models including parallel and serial mediation models and the analyses were also stratified by sex of the children.

Results

Our mediation analyses showed that the total effects of parents' mental disorder, bipolar disorder and schizophrenia, as compared to the control group on the CBCL was 6.50 [95% confidence interval (CI): 2.15, 10.85] and 9.96 [95% CI: 6.22, 13.69], respectively. When considering both mediators simultaneously (serial mediation model), there was a significant indirect effect of bipolar disorder, 4.15 [95% CI: 2.44, 5.97], and schizophrenia, 5.87 [95% CI: 4.02, 7.87], vs. control on the CBCL through home environments and primary caregiver's personal and social performance. Here, 66% total effects for bipolar disorder and 57% total effects for schizophrenia on the CBCL were explained by both mediators. For male children, the mediators could account for 58% and 48% of the total effect of the bipolar disorder and schizophrenia on the CBCL and for female children, they accounted for 79% and 73% of the total effect of the bipolar disorder and schizophrenia on the CBCL, respectively.

Conclusion

We conclude that there was a significant association between parents' mental disorders and child behaviour via home environments and primary caregiver's personal and social performance. Hence, adequate home environments and primary caregiver's personal and social performance are significant

mediating factors for child behaviour, which is more vital when either one of the parents is mentally sick. Our analyses show that mediation analysis is an essential tool to find out actual causal association in a complex study setting in psychiatric epidemiology.

A two parameter Rama distribution with applications

'Edith Umeh , [1]Ebele Umeokeke, [1]Amuche Ibenegbu
'Nnamdi Azikiwe University, Awka

Background
This paper proposed a two-parameter Rama distribution. This was coined from Lindley distribution, Akash distribution and Rama distribution (Exponential and Gamma). Its mathematical and statistical properties which include its shapes, moment, coefficient of variation, skewness, kurtosis, index of dispersion, hazard rate function, mean residual life function, stochastic ordering, mean deviation; Bonferroni and Lorenz curves were also discussed. The work shows the best model for estimating right censored data applied in hypertensive patients

Methods
The estimation of parameters was analysed using method of moments and maximum likelihood. AIC and BIC were used to test for goodness of the model.

Results
From the result of analysis for the estimation of parameters using method of Maximum likelihood, it was observed that the proposed method performs better than other distributions based on its f minimum value. Also, for testing of goodness of fit, the new model out performed other distributions based on the minimum AIC and BIC

Conclusion
The two parameter Rama distribution proposed were compared with other distributions which includes Lindley, Akash, Rama, Gamma and Exponential. This were applied on hypertensive patients. This is a biostatistics which involves life activities, it was observed that the new distribution proposed fits a better model than other distributions. And is preferred for modelling life time data sets.

Product profiling: Standard multivariate or multi-block statistical methods? An application to SA honeybush herbal tea

[1]Marieta van der Rijst [2]Neil le Roux, [3]Lizette Joubert
[1]Agricultural Research Council [2]Stellenbosch University, [3]Agricultural Research Council

Background

It is said that tea is the second most consumed drink in the world after water. People drink tea for a number of reasons. Although drinking is a basic need of the human body, in the developed world consumers are increasingly aware of the link between health and food. Consumers not only expect healthy, functional beverages, but also satisfactory flavour and taste. South African honeybush herbal tea has unique health benefits, yet critical factors in its success on global markets are sensory quality and consistency. Sensory evaluation of tea by expert tasters remains the most common practice used to evaluate the sensory quality and characteristics of tea. Descriptive sensory analysis (DSA) uses humans as instruments to generate quantitative data on the perceived sensory characteristics of a product. However, human responses are complex responses to a mixture of several visual, chemical or structural components of a product. The purpose of statistical analysis on DSA data is to reflect human perception of a product by revealing perceived dimensions latent in the product, composed of the product descriptors.Conventional statistical procedures employed in product profiling, based on DSA, combine all descriptors in one large data matrix and consider it simultaneously. The drawback of this approach is that interpretation might be obscured, resulting in unsatisfactory results, especially when products have very complex profiles. The intricate nature of DSA data requires statistical models that can extract core information from large data matrices. Recent literature propose multi-block methods to improve the interpretability of multivariate models. In this paper product characterization by means of methods that split data into conceptually meaningful blocks (multi-block), as opposed methods that use all variables simultaneously in one large data matrix (multivariate), will be discussed and compared, using honeybush herbal tea data as illustration.

Methods

A study was conducted to characterise the sensory profile of honeybush herbal tea and to determine which compounds contribute to its unique flavour and sensory qualities. A total of 132 samples, consisting of 6-8 batches each of three Cyclopia species, harvested over three seasons and “fermented” (oxidised) at two temperature x time regimes, were subjected to DSA, performed by 10 trained panel members, using 39 descriptors. To facilitate reliable correlation between sensory quality and instrumental parameters, instrumental and chemical analysis have been done on exactly the same infusions as used for DSA.Data generated from DSA typically consists of different groups or types of descriptors, for example appearance (visual), aroma, flavour (chemical) and texture (structural) descriptors. From the sensory lexicon and sensory wheel developed from the DSA data for honeybush herbal tea, five groups of key sensory components (data blocks) could be distinguished namely positive aromas (14 descriptors), aroma taints (6 descriptors), positive flavour (9 descriptors), flavour taints (6 descriptors), as well as taste-and-mouthfeel (4 descriptors). A sixth block of data consisted of the chemical data (20 phenolic compounds).The data were subjected to multivariate statistical methods, including principal component analysis (PCA) and discriminant analysis (DA), as well as multi-block methods including multiple factor analysis (MFA) and generalised Procrustes analysis (GPA).

Results

For the PCA including all 39 sensory descriptors and 20 phenolic compounds, the first principal component

(PC1) clearly separated samples of different species based on its phenolic compound composition, while the second principal component (PC2) mainly separated samples from different seasons based on its association with positive aromas and flavours or taints. DA reveals similar discrimination in terms of variables and samples. PCA biplots for each of the identified data blocks separately expose rather different patterns of association for the data blocks. This may emphasize the effect of specific variables within each data block, but may also complicate interpretation of overall results. MFA also reveals similar patterns of association to PCA. An advantage is that the data blocks are visually distinguished on the MFA variable loadings plot by making use of different colours. Furthermore, RV coefficients (correlation coefficients between two matrices) confirm suspected associations between data blocks. GPA again produce similar graphical results, now with additional information on configurations after transformation.

Conclusion

Popular multivariate methods like PCA and DA are designed to handle correlation between a series of measured variables and reveal the important latent structure in the data.Multi-block methods may be considered extensions of multivariate methods, to be introduced when the number of variables is large and additional information exists to distinguish meaningful blocks. Nevertheless, solutions of multi-block methods will depend heavily on how the different blocks are scaled relative to each other, as well as the dimensionality within each the block.The proposed simple alternative to multi-block methods to separately analyse individual data blocks using multivariate methods, should be used with caution, as no statistics regarding association between blocks will be available. In the case of the honeybush herbal tea case study multivariate and multi-block statistical methods generally led to the same conclusions, but graphical results for multi-block methods were more informative and extra information regarding association between blocks could support interpretation. Both multivariate and multi-block methods give clear graphical overviews of large amounts of information. However, multi-block methods have the added advantage that information on communalities and discrepancies between blocks are accessible, which may elucidate interpretation.

Statistical methods to link Demographic and Health Survey data with census data

'Ruth Vellemu ²James Chirombo
'University of Malawi ²Malawi Liverpool Wellcome Trust

Background

Demographic and health surveys (DHS) are large nationally representative surveys held every five years to provide estimates on health and demographic outcomes such as child health, malaria prevalence, HIV prevalence, and nutrition among others at the district level. These estimates are widely used by different programmes for monitoring and evaluation purpose. Typically, DHS surveys have sample sizes between 5,000-30,000 households. One of the general challenges facing national surveys is declining response rates which may affect the reliability of estimates. In a Malawian setting, it is also not possible to obtain estimates at sub-district level as the DHS is powered to provide district-estimates. To obtain estimates for subpopulations of interest, it may be necessary to combine the information from surveys with other data sources with wider coverage and higher spatial resolution such as population and housing census (PHC) data. A combined dataset can provide comprehensive estimates of useful health indicators at a very fine spatial scale for improved decision making at the local level.

Methods

Bayesian Hierarchical models for geostatistical and areal data were used to combine different data sources to provide small area estimates of key health indicators at sub-district level in Malawi. A final model, implemented in a Bayesian framework with local area effects, was then developed to generate estimates at sub-district level.

Results

Results show that combining survey with census data provides robust estimates at high spatial resolution compared to when using only one dataset.

Conclusion

The project improves our understanding of methods used in combining information from different sources and provides a means for performing small area estimation of population parameters of interest to different researchers. Estimates of health indicators at high spatial resolution is an important undertaking important for disease monitoring and surveillance purposes.

Random Forests application in missing data and predictive modelling for hierarchical routine clinical data: A case study of childhood pneumonia in Kenya

'Steven Wambua [2]Nelson Owuor, [3]Susan Gachau
'University of Nairobi/SSACAB [2]University of Nairobi, [3]KEMRI - Wellcome Trust/University of Nairobi

Background
Health stakeholders usually need complete, accurate and reliable estimates of various health outcomes to make decisions on improving health care delivery. Missing observations especially in clinical routine data is one of major setbacks in evaluating public health problems efficiently. One of the tools used to measure clinical quality is the Paediatric Admission Quality of Care (PAQC) score. We seek to identify factors that influence clinical quality in this study after dealing with missing values.The main objective of this study is to identify key determinants of Pediatric Admission Quality of Care (PAQC) score using Random Forests.

Methods
Data on a total of 2027 children between 2 and 59 months who were admitted in selected county hospitals in Kenya was used. The data contained clinical data from admission to treatment. Random forests missForest package was used to impute missing data. Cumulative logit mixed models were fit with PAQC score as an outcome and age, sex, comorbidity, weight, clinician sex and cadre, hospital workload, malaria prevalence, intervention arm and time of admission as predictors to determine the significant determinants of clinical quality. The models were nested within both hospital and clinician levels. Both Random forests and conditional random forests were used to determine variable importance.

Results
The cumulative logit mixed model nested within both clinician and hospital level was selected based on AIC. Weight of the child, clinician sex, cadre and the time of admission were significant determinants of PAQC score based on the P-values at 0.05 level of significance. A unit increase in weight increases the probability of a higher PAQC score by 0.06, while being attended by a medical officer relative to a clinical officer increases the probability by 0.27. The time of admission increases the probability by 0.11. On the other hand, PAQC scores would be lower if the clinician was male. The probability of a reduced PAQC score if a clinician is male is 0.49. Month, weight, intervention arm and hospital workload were the most important variables in predicting the quality of care while age and the number of comorbidities were the least important using Random forests models.

Conclusion
Based on the cumulative logit mixed models, the study concludes that hospital level, weight of the child, clinician sex, cadre and the time of admission are key determinants of PAQC score. On the other hand, age and the number of comorbidities for a given patient may not strongly influence the quality of care provided based on the random forests models. The mechanisms around these associations however need to be studied extensively. Pneumonia management strategies could incorporate these findings to improve quality of healthcare delivery and hence reduce burden of the disease.

ProtGear: A tool for protein microarray data pre-processing

'Kennedy Mwai Wambui [2]Nelson Kibinge, [3]Samson Kinyanjui, [4]Faith Osier, [5]Eustatsius Musenge
'University of the Witwatersrand [2]KEMRI-Wellcome Trust, [3]KEMRI-Wellcome Trust, [4]University of Heildelberg, [5]University of WItwatersrand

Background

Protein microarray technology is increasingly being used for antigen discovery in experiments where thousands of samples are analysed. However, data generated using this approach has been shown to contain systematic and non-systematic sources of bias. Careful pre-processing is essential to eliminate this technical bias while keeping the biological variation. This bias has been attributed to nonbiological sources, which is introduced by small variations in the experimental conditions during the experiment process. Unlike DNA microarray technology, however, protein microarrays have found more use just recently and present unique sets of data challenges.Guidelines for pre-processing the protein microarray data before analysis, therefore, continue to evolve. Pre-processing involves; background correction, quantifying within-sample variation, batch correction4, normalisation among other steps. Background correction is the statistical process of excluding the background noise that may attach to the array spots. Normalisation helps to adjust for any bias which arises from variation in the microarray technology rather than from biological differences. In this project, we develop a generic pre-processing tool for protein microarrays in R software with statistical functions for the above procedures.

Methods

The tool has been developed with different statistical functions to perform the pre-processing data procedures which are organized into an R package. Our package implements procedures for quantifying within-sample variation, background correction, batch-effect removal and normalization. A function implementing Coefficient of Variation (CV) and pooled estimate of variance (PEV) for the technical repeats is used to estimate the within-sample variation. Additionally, have included functions to examine within-sample variability based on the CV filtering for technical repeat spots. CV-based filtering provides an algorithm for the selection of less variable spots among technical repeat.Background fluorescence can be attributed to within-slide or spot-specific artefacts; thus, correction minimizes the resultant noise. protGear implements five different techniques for background correction. The techniques have been adopted from the Linear Models for Microarray Data (Limma). Data from these experiments have been shown to have a mean-variance dependence that may mask the true biological variability. To assist in minimising this problem, different normalization and batch correction techniques have been adopted. On our package, we have included but not limited to robust linear (RLM) regression normalisation-assumes the bias is linear dependent with the estimates; local regression normalisation -assumes a non-linear dependency; hyperbolic sine normalisation asinh and Variance Stabilization Normalisation (VSN) - adopted from a DNA microarray approach that utilises parametric transformations and maximum likelihood approaches and log2 normalisation. These transformations have been incorporated with batch correction techniques using (ComBat) empirical Bayesian framework from SVA package for removing batch effects.

Results

The package is on the final phases of development and most of the functions have been tested using data from KILchip v1.0: A Novel Plasmodium falciparum Merozoite Protein Microarray to Facilitate Malaria Vaccine Candidate Prioritization6. The methods were evaluated in their ability to decrease variation among technical repeats, reduce the mean-variance dependence of the measured intensities which is

caused by technical errors and still maintain the biological difference. The CV-based filtering of technical repeats significantly reduces the variation among the technical replicates. A CV of 20% was considered the cut off for agreeable technical repeats. A subtraction of the local background was selected as the best approach to handle the background noise for the KILchip v1.0 data after comparing with the other approaches. The subtraction of the local background approach was selected since it did not produce negative intensities as per the design of the experiment. The different normalisation techniques and batch correction were run on the KILchip v1.0 data. ComBat. The different techniques significantly reduced the mean-variance dependence on the data. The package implements the pooled median absolute deviation (PMAD) and the pooled estimate of variance (PEV) together with graphical representation to check on the performance of the different methods. On top of ComBat batch correction, we incorporated a within function batch correction by including the batch parameter in the normalisation model for VSN and RLM normalization. VSN+Combat normalisation approaches were selected for the KILchip v1.0 array data.

Conclusion

protGear provides a rational data pre-processing framework and platform for exploratory and visualization to facilitate separation of technical variation from biological variation. Most of the available open-source packages have been implemented for DNA but there are limited standard pre-processing guidelines for protein microarray data. Additionally, this tool allows the users to compare the performance of the different normalization techniques utilised in different protein micro-array projects and select the suitable for their data. Logarithmic normalisation using a base n was less effective with a dataset with outliers which was the case with the KILChip v1.0 data2. Additionally, subtraction of purification-tag intensities for lab experiments generated negative intensities which caused convergence problem for RLM, Loess and log transformations. We are working on approaches to help handle this statistically on our package. Finally, this tool is a reusable package developed in the R statistical programming language. It provides a unified platform for pre-processing protein and peptide array data. It has been developed in such a way that other users can contribute in development after it has been published

Analysis of student-lecturer interaction in online PHT 112 course at Maseno University

'Mary Dancilla Wanjiru Muthoni [2]Justa Wanja
'Maseno University [2]Maseno University

Background

Online learning has grown prolifically in school institution and corporate training during the past decade. To develop a successful online course, many scholars suggest that building and sustaining an online learning community is crucial and necessary .Research indicates that a learning community cannot only increase persistence of students in online programs, but also enhance information exchanges, learning support, group commitment, collaboration, learning satisfaction and more so good interaction skills among students and students, learners and their lecturers and lastly student-content variation in an online course. Interaction is very important in any kind of learning. Moore (1989) defined the three types of interaction levels as: learner-learner interaction, learner-teacher interaction and learner-content interaction. E-learning has become a very vital source of information and way of interaction viewing from the current world of technology and social networking in schools. It consists of classes that are fully delivered via the internet, or a combination of classes delivered via the internet and periodic meetings in a face to face classroom. E learning facilitates personal communication and provides access to all kinds of information, implying that there is a methodological change to teaching and a need for both lecturers and students to adapt to the use of such tools. Interaction is a basic element in any educational process. Such interaction, which is always explicit in classrooms due to the face-to-face relationship established between lecturers and students, and among students, is an element that various professionals have reconsidered in the light of technology in education.

Methods

Area of studyThe study was based in Maseno university main campus. Maseno University comprises a total of approximately 15,000 students of whom we have both the Self sponsored students and Government sponsored students. E-learning in the University is managed and supervised using a Learning Management System (LMS-Moodle) module in the E-campus. The E-campus offices are situated at Maseno University Varsity Plaza in Kisumu city, Kenya. Target PopulationThe population comprised all the Maseno university students in the main campus. It comprised of first years, second years, third years and fourth years for it is compulsory in Maseno university to undertake PHT-112 online course since the registration of the course is determined by the student furthermore it does not exceed the fourth year of study. Sample designSimple random sampling was used to draw 2 groups from the total 10 groups in LMS-module. 10 topics for both groups were compiled together. The topics were further divided into six event activities that took place in the discussion forum which included: discussions creation(by students), posts creation(by students), course module updating(by the lecturers), discussion views(by students) and contents in each discussion The lecturers involved were 5 who facilitated the two groups. The event activities were further ranged in months and the total in each month was calculated. Data collection method Our project involved secondary data which was sourced from the LMS-module of Maseno E-campus. The data was collected by means of a flash disc drive which we transferred in our PC's for further analysis. Data analysisData collected was analyzed using Descriptive statistics (bar graphs) and inferential statistics (Non-parametric tests, Q-Q plots and histograms). The above was achieved using SPSS and Excel statistical softwares.

Results

This section will be presented on the document below for my abstract. Kindly view the document below.

Conclusion

The aim of our study was to explore the three levels of interactions in the discussion forum which included: student-student, student-content and student-lecturer. The study was specifically based on PHT-112 course which is a common course in the university. In order to study student-student interaction, we were specific on the events to use which included discussions created and the discussions viewed. Every creation of a discussion had a view by a fellow student and at least each student created and viewed a discussion in the forum. From our hypothesis results, it was evident that there was negative interaction among the students which enabled us to answer the first research question. This basically was achieved by checking the normality of the data and then applying the non-parametric test to achieve our first objective. It was also evident that most students participated in the viewing of the discussions but did little on the discussions creation leading to the negative interactions.To study on the student-content interaction, we used the events done by post creating assuming that every student created a post. We analyzed using the post creations per month and also per topic but couldn't fully answer our objective. We further went on to use the scores that students attained in the overall events in the forum. In the figure 1.5 it shows the graphical representation of the scores that indicates topic six having the highest score which implied that most student had the content understandable to them. The hypothesis indicated that the students got the contents right and thus answering our second research question.For the lecturer-student interaction, we based our argument on how they carried out the updates in the module as students could not have continued in their forums if update was not done. It indicated that number of the lectures did update rarely but only one who indicated a peak of 15 updates.

Comparison of Random Survival Forests Split rules in selecting the determinants of under-five mortality using 2014 Kenya DHS data

[1]Kennedy Wanyonyi [2]Nelson Owour
[1]University of Nairobi [2]University of Nairobi

Background

Survival trees and Random Survival Forests are extensions of classification and regression trees and random forests in analyzing time to event data. These methods are alternatives to Cox Proportional hazards models, useful for handling cases such as when Proportional Hazard assumption is violated. Survival tree methods are flexible and can handle high dimensional covariate data as they are fully non-parametric. Random survival forests use the Brieman's approach, first, by employing a random selection of a bootstrap sample used for growing a tree then growing tree learners by splitting the nodes on the randomly selected predictors. The performance of the survival trees highly depended on the splitting rule that is applied while growing the tree.

Methods

In our analysis, we compare the performance of random survival forests in variable selecting based on the following split rules; Log-rank splitting, Log-rank score splitting, and R-Squared Splitting. This is ongoing work, and we intend to explore other split rules. Our outcome variable is the under-five child mortality in Kenya using 2014 DHS data.

Results

Preliminary findings from this study show that Log-rank score outperforms R-Squared and Log-rank splitting rules in analyzing time to event data based on the bootstrap cross-validated estimates for integrated Brier scores. This could be true for high dimensional datasets, such as the KDHS data that we are handling. Furthermore, R-squared performs comparatively better than the Log-rank split rule.

Conclusion

Despite the preliminary results, there is need to investigate other split rules and the nature of data that best suits each split rule, in order to be able to make comparisons about the split rules.

Comparing coverage and precision of Newcombe confidence intervals for proportion differences generated using different weighting systems

'Miriam Wathuo [1]Nuredin Mohammed, [1]Ed Clarke, [1]David Jeffries
'Medical Research Council Unit The Gambia at the London School of Hygiene and Tropical Medicine

Background

The Wilson score confidence interval for proportions and the Newcombe confidence interval for proportion differences are often used in clinical trials, particularly in non-inferiority trials. They provide better coverage than the Wald interval, which is unreliable for success probabilities close to 0 or 1. Both the Wilson and Newcombe methods do not provide stratified versions. Other methods have been used to estimate treatment differences after adjusting for covariates in trials with binary outcomes. E.g., the Taylor approximation involves fitting a logistic regression, estimating the log odds of response on each arm, and transforming these to the probability scale. Confidence intervals for the proportions and their differences are then calculated. However, the method does not take into account the covariance between proportion estimates. Yan and Su proposed stratified confidence intervals providing closed form solutions with weights constrained as ∑_i▒w_i = 1 and w_i >0. However, they did not suggest any specific weighting system. It is thus of interest to check whether different weighting systems affect the coverage and precision of confidence intervals.

Methods

We simulated 1000 datasets, each with the following parameters; n=1000, two treatment groups with a probability of 0.5 of participants being in either group, immune response with a probability of 0.94, and baseline seropositivity with a probability of 0.97. Immune response was defined as either seroconversion or a four-fold change in antibody titre. For each dataset, we calculated the stratified Newcombe confidence intervals for the difference in immune response rate between the two treatment groups, stratified by baseline seropositivity. We compared four different weights; frequency, INVAR (inverse-variance), CMH (Cochran-Mantel-Haenszel) weights, and equal weights. The coverage and width of all the confidence intervals was calculated. All analysis was performed using R version 3.5.3.

Results

The coverage and width for the stratified Newcombe intervals were very similar using frequency, INVAR and CMH weights. However, the width of the intervals using equal weights was much higher. The coverage and width estimates were: Frequency; coverage: 94.4%, width: 0.05873, INVAR; coverage: 94.9%, width: 0.05876, CMH; coverage: 94.4%, width: 0.05872, equal weights; coverage: 96.4%, width: 0.17553.

Conclusion

There seems to be no benefit in using equal weights (no weighting). In fact, the confidence intervals are less precise when equal weights are used. As for the different weighting schemes, their coverage and width looked quite similar, so any of these could be used to calculate the stratified Newcombe confidence intervals for proportion differences.

Inform Prior Elicitation for Bayesian Questionnaire Validation using Confirmatory Factor Analysis

'Hao Zhang ²Tibor Schuster
'McGill University ²McGill University

Background

The Bayesian statistical framework, incorporating expert knowledge or beliefs and/or available data from literature, is well-suited for updating information for evidence-based practise in medicine and health research. Recent literature and software developments propose that Bayesian methods can largely increase efficiency when conducting confirmatory factor analysis (CFA) for questionnaire validation. Bayesian CFA enables use of smaller sample sizes compared to the conventional frequentist approach when estimating item-domain correlations (factor loadings). However, CFA assumes that investigators can provide relevant prior information with respect to the latent variable model.

Methods

We propose a survey approach to collecting and decoding such prior information. Instead of model regression coefficients, prior distributions on the scale of item-domain correlation are visually presented and explained to experts. Based on these inputs, we derive a pooled prior distribution for each item that aggregates the beliefs of all experts. The corresponding Bayesian latent variable model is fitted using Hamiltonian Monte Carlo sampling in R-Stan. We developed a user-friendly web app with R Shiny for questionnaire validation under a single framework, where priors could be specified by domain experts on a continuum of zero to one indicating item-domain correlations. We apply the method to data of a diabetes empowerment questionnaire (28 questions in 4 domains) including prior input from 6 domain experts and 100 patient responses. Various graphs with informative Beta distribution density curves depicting different degrees of expected strength of item-domain correlation (low, medium, high) as well as different levels of subjective uncertainty around these presumptions (low, high). Using the aggregated prior distribution, we estimate factor loadings and their associated credible intervals.

Results

Inter-rater reliability among experts' prior input was low to moderate. We compared the estimated factor loadings and credible intervals with the results of a classic CFA using the same data. The comparison yielded substantial gain in precision of estimated item-domain correlations substantially using Bayesian CFA with prior distributions derived from the obtained expert ratings.

Conclusion

Feasibility and acceptability of the expert survey was high. We propose that survey approaches should be more routinely employed to inform priors for Bayesian confirmatory factor analysis, achieving acceptable estimate precision with smaller samples, shorter recruitment time and at lower costs. This further strengthens the potential impact of Bayesian statistical framework in the practice of evidence-based medicine, integrating clinical expertise with the best available evidence from systematic research and for a more time- and cost-efficient questionnaires validation.

The usefulness of prediction intervals in quantifying effect heterogeneity in randomized controlled trials

'Hao Zhang
'McGill University

Background

Primary findings of randomized controlled trials (RCTs) are typically reported as effect estimates (e.g. risk differences, risk- or hazard ratios) along with confidence intervals. In the presence of heterogenous subpopulations, the reported effect may, however, represent an average of a set of distinct underlying subgroup effects. This is a similar situation as it often occurs in meta-analyses that include heterogenous study populations. In meta-analyses, reporting of prediction intervals is recommended as these intervals reflect the variation in treatment or exposure effects across study populations. In contrast to confidence intervals, prediction intervals do not shrink to zero width if the sample size largely increases but cover a prespecified range of effects to be expected in future studies. The objective is to demonstrate the usefulness of prediction intervals in quantifying effect heterogeneity in randomized controlled trails.

Methods

We employ stratification factors used in the randomization to define distinct subpopulations within an RCT. Using standard meta-analysis software, we compute prediction intervals for the effect and compare these intervals with the respective confidence intervals. Changes in conclusions regarding the effectiveness and safety of the interventions under study are drawn. We apply the proposed method to trial data from the German Breast Cancer Study Group (GBSG) study. The stratification variables tumor grade (I, II, III) and menopausal status (yes / no) are used to define subpopulations within the GBSG study dataset.

Results

We demonstrate that prediction intervals are applicable in large randomized controlled trials that include subpopulations defined based on stratification factors. In the GBSG study, the prediction interval indicates potentially lower effectiveness than the confidence interval (upper limit 95% prediction interval: hazard ratio 0.99, upper limit 95% confidence interval: hazard ratio 0.90).

Conclusion

Randomized controlled trials should routinely provide prediction intervals when reporting treatment effect estimates, enabling better clinical applicability to subpopulations and allowing for better-informed patient and doctor decisions.

Poster List

* Indicates a SSACAB funded student

MONDAY, 9 SEPTEMBER 2019

NO	PRESENTING AUTHOR	ABSTRACT TITLE
1	Mary Adehi	Meta-analysis on the risk of mortality in patients with Ebola virus disease
2	Glorious Atukunda	Factors associated with modern family planning (MFP) use among women age 15-24 years in Uganda
3	Godfrey Biroma*	Determinants of blood pressure among pregnant women in Nebbi District-Uganda
4	Awino Diana*	Factors associated with modifying the relationship between district health performance using district league table score and the coverage of institutional deliveries in Uganda
5	Panzi Eric*	Recent trend in under-five mortality in the Democratic Republic of Congo (DRC): An application of Cox survival frailty regression modelling
6	Ogbonnaya Ezichi	Estimation of therapeutic window in crossover dose response design in phase II clinical trial of methylphenidate
7	Haile Mekonnen Fenta	Determinants of stunting among under-five years children in Ethiopia from the 2016 Ethiopia demographic and health survey: Application of ordinal logistic regression model using complex sampling designs
8	Paul Gatabazi*	Resampled marginal risk set model for infant mortality at the Kigali university teaching hospital
9	Lilian Giibwa	Comparative analysis of temporal trends in general fertility rate among four regions in Uganda
10	Tracy Glass	Impact of sampling bias in estimation of cumulative viremia
11	Nina Grundlingh	Modelling risk factors of diabetes and pre-diabetes in South Africa

TUESDAY, 10 SEPTEMBER 2019

NO	PRESENTING AUTHOR	ABSTRACT TITLE
12	Ndamona Haushona	A systematic review of spatial methods using health surveys in Sub Saharan Africa
13	Sesilia Rauha Ndeutalala Kapenda*	Suicide ideation and associated factors among school going adolescents in Namibia: A multilevel logistic regression
14	Emma Khaemba*	Comparing drug regimens for clearance of malaria parasites in asymptomatic adults in a clinical trial in Kilifi County, Kenya
15	Collen Makomane	Assessment of possible bias during sampling of health establishments for inspections
16	Jacqueline Materu	Psychometric properties of PHQ-4 anxiety depression screening scale among out of school adolescent girls and young women

17	Zvifadzo Matsena Zingoni*	Transitional Viral Suppression due to ART on a Zimbabwean Cohort from 2004 to 2017: A Bayesian Mixed effects Multistate Model
18	Mhlengi Mgaga	Statistical methods that detect publication bias in meta-analysis
19	Taurayi Mudzana	Harmonisation of longitudinal population data: Evidence from three rural Health and demographic surveillance system nodes in South Africa.
20	Edson Mwebesa*	Factors associated with prevalence of pregnancy termination in Uganda: A population-based study
21	Caroline Nakalema*	Factors associated with cigarette smoking in Uganda: Analysis of Uganda Demographic Health Survey (UDHS) 2016 Data

WEDNESDAY, 11 SEPTEMBER 2019

NO	PRESENTING AUTHOR	ABSTRACT TITLE
22	Suzan Nakasendwa*	Prevalence and factors associated with low birth weight in Uganda: Analysis of Uganda Demographic Health Survey (UDHS) 2016 Data
23	Justine Nassejje	A deep learning based survival prediction to identify trends in the predictiveness of socioeconomic factors in determining under-five mortality rates in sub-Saharan Africa
24	Qondeni Ndlangamandla*	Modeling Anemia in Children under five in Tanzania and Angola
25	Elias Obudho	Redefining agricultural statistics for the future
26	Emmanuel Obuya*	Spatial-temporal distribution of malaria prevalence among the under-fives in Tororo District, Uganda
27	Uchenna Petronilla Ogoke	A Bayesian statistical approach to investigate blood pressure cases
28	Morelearnings Sibanda*	Investigation of the effects of maternal related factors on infant and child mortality in Zimbabwe
29	Henry Sseguya	The role of climatic factors on malaria prevalence among children <5 years in Uganda in 2016: Analysis of uganda demographic health survey (UDHS) of 2016
30	Nazarius Mbona Tumwesigye	Alcohol abuse OPD cases in Uganda by region: A 2012-2017 A time series analysis
31	Anteneh Yalew	Multilevel Logistic Regression for Patients' Satisfaction in Ethiopian Public Hospitals
32	Ashenafi Yirga*	Factors affecting child malnutrition in Ethiopia

Meta-Analysis on the Risk of Mortality in Patients with Ebola Virus Disease

'Mary Unekwu Adehi, [2]Ekele Alih, [3]Angela Chukwu, [4]Halidu Haruna
'Nasarawa State University Keffi Nigera [2]Department of Mathematics and Statistics, Federal Polytechnic, PMB '037' Idah, Kogi State, Nigeria, [3]Department of Statistics, University of Ibadan, Ibadan, [4]Department of Statistics, Nasarawa State University, Keffi

Background
This paper illustrates the effective use of some well expanded DerSimonian and Laid methods in Meta-analysis, two values of the relative risk ratio were missing from the Ebola Virus disease data, rapid and UK studies.

Methods
Looking at the pattern of the missing values, the values were missing not at random. The strategy adopted to estimate the missing values is the single imputation method, sometimes referred to as person mean imputation.

Results
The relative risk ratios of the 17 complete studies were averaged and the resulting score allocated to the rapid study. The same procedure was carried out for the 18 studies and the allocation was made to the UK study. 19 studies altogether were meta-analyzed with summary effect, 1.833, confidence interval (1.553, 2.113), P-value = 0.000 and I2=97.4% indicating a high presence of heterogeneity. The meta-analysis favored mortality in patients with Ebola Virus disease. Due to the high percentage presence of heterogeneity, it was necessary to carry out the Begg's and Egger's tests for publication bias assessment. Both tests resulted in bias of the studies, sensitivity analysis indicated that the Sierra Leone 2 and Nigeria studies may be the source of bias, as such, the subsequent meta-analysis excluded both studies.

Conclusion
The results showed a summary effect of 1.37 in favor of mortality, confidence interval (1.26, 1.468), P-value = 0.000 and a lower presence of heterogeneity, I2=74.2%. In comparing both meta-analyses for Ebola Virus disease, the risk of mortality is less in the meta-analysis of the 17 studies than in the 19 studies, and the width of the confidence interval narrows as well, indicating higher level of precision.

Factors associated with modern family planning (MFP) use among women age 15-24 years in Uganda

'Glorious Atukunda ²Julius Ssempiira
'Makerere University ²Makerere University

Background

Uganda's total fertility, maternal mortality and teenage pregnancy rates remain among the highest globally. Uganda is committed to scaling up the uptake of MFP to ensure that every Ugandan woman can choose when and how many children to have. In this study, we analyzed the UDHS 2016 data to determine factors that are associated with the use of MFP among women of age 15 - 24 years in Uganda.

Methods

We analyzed the UDHS 2016 data. The outcome was a binary variable defined as use or non-use of MFP. The predictors consisted of socio-demographic factors including age, marital status, religion, education, socioeconomic status and region. Descriptive statistics were used to summarize study variables. Unadjusted and adjusted odds ratios (AOR) and 95% Confidence Intervals (95%CIs) from univariate and multi-variable logistic regression models, respectively were used to quantify the magnitude of association between predictors and the use of MFP.

Results

A total of 5256 women age 15-24 years were interviewed in the UDHS 2016. Of these, 1784 (33.9%) were using family planning methods, and of these 1664 (93.3%) were using a MFP. Marital status, Women from the central, eastern and western regions were 67%, 10% and 40% more likely to use MFP compared to northern Uganda women. (AOR=1.67, 1.10, 1.40 and 95%CI: 1.36-2.04, 0.93-1.31, 1.15-1.70), Education level (AOR=2.5,3.01,2.85, 95%CI: 1.76-3.53, 2.09-4.34, 1.75-4.64) respectively for primary, secondary and higher in reference to those without any education, socioeconomic status proxy of wealth index (AOR=1.23, 95%CI: 1.00-1.50 for the rich) and distance to health facilities (AOR=1.22, 95%CI=1.08-1.38 for those near facilities) were significantly associated with the use of MFP usage (P- value <0.05). Married women and those living with their partners were 56% and 66% more likely to use MFP compared to single women (AOR=1.56, 95%CI: 1.21-2.01) and (AOR=1.66, 95%CI: 1.31-2.11), respectively.

Conclusion

Although the uptake of family planning is still low in Uganda, but the use of MFP among family planning users is high. The low uptake of family planning should be improved by increasing the number of health facilities which provide MFP services, and increasing awareness of MFP methods through advertisements on radios and televisions. In addition, advocating for girl child formal education would also improve the usage of MFP.

Stunting in infancy, pubertal trajectories and adult body composition: Birth to Twenty-Plus cohort, South Africa

Chidumwa G[1,4]*, Said-Mohamed R[1,]*, Nyati LH[1], Mpondo F[1], Chikowore T[1,2,3], Prioreschi A[1], Kagura J[1], McGowan CJ[1], Ware LJ[1], Micklesfield LK[1], Norris SA[1]

[1]SAMRC/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, School of Clinical Medicine, Faculty of Health Sciences, University of Witwatersrand, [2]Division of Human Genetics, School of Pathology, Faculty of Health Sciences, National Health Laboratory Service & University of the Witwatersrand, Johannesburg, South Africa, [3]DST-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS), [4]Division of Biostatistics and Epidemiology, School of Public Health, Faculty of Health Sciences, University of Witwatersrand

Background

Rapid growth in childhood and puberty timing have been associated with adulthood obesity. However, the association between childhood stunting and obesity risk later in the life-course is less clear. We assessed whether the relationship between stunting at two years-old and body composition at 22 years-old is mediated by body mass index (BMI) from childhood, and by pubertal development, using the Birth-to-Twenty Plus cohort (BT20+) in South Africa.

Subjects

Anthropometric measurements were obtained for n=1036 male and female participants between birth and 22 years. Height-for-age z-scores and stunting at two years were generated using the WHO 2006 growth standard. BMI-for-age trajectories between 5 and 18 years were derived using latent class growth mixture modelling. The Tanner sexual maturation scale assessed pubic hair and genital/breast development, with pubertal trajectories generated using latent class growth analyses. Fat mass and fat free mass were determined by dual energy x-ray absorptiometry. Structural equation models unadjusted and adjusted for maternal factors were used to test our hypothesis.

Results

In males only, stunting at two years-old had an indirect positive effect on fat mass via adult height and contributed 66% of the total effect. Stunting, via adult height, accounted for 71% of the total negative effects on females' fat free mass while it contributed 50% of the total positive effects on males' fat free mass. Higher BMI trajectories, slowest and fastest pubertal trajectories, and being a female were associated with higher fat mass and fat free mass levels.

Conclusions

Shorter adult stature in previously stunted individuals is combined with more fat mass and fat free mass in males and lower fat free mass in females, independently of changes in BMI from childhood and pubertal development. Linear growth, weight gain and BMI should be monitored from infancy to promote healthy body composition profile in young adulthood.

Factors associated with modifying the relationship between district health performance using district league table score and the coverage of institutional deliveries in Uganda

'Diana Awino [2]Emmanuel Obuya, [3]Lorna Aol Akera, [4]Dr John Ssenkusu
'Makerere University School of Public Health [2]Makerere University School of Public Health , [3]Makerere University School of Public Health, [4]Makerere School of Public Health

Background

5% of Uganda's population is projected pregnancies, 4.85% carry the pregnancies to term and 4.3% are under one year. There is a 0.7% loss between conception and one year. Uganda has committed and prioritized efforts towards improving maternal and child health indicators; to achieve global targets in line with the sustainable development goals (SDGs). Implementation and scaling up of projects like results based financing (RBF) which remunerates the highest sums of money to institutional deliveries whose outcome is a live birth is underway. Recruitment of midwives, capacity building, and construction of maternity wards, theatres and stocking of emergency obstetric care commodities tops the list of both government and development partners' priorities in the sector. Over time, the parliament of Uganda has created new districts and municipalities to improve health service delivery among other reasons. 12 municipalities were approved in July 2015 and 6 new districts were opened in July 2018. The Ministry of health uses the district league table (DLT) approach to assess the performance of health systems at district level. The DLT was first prepared in 2003 and included in the Annual Health Sector Performance Report. The DLT has been prepared annually to date with the indicators aligned to the five core building blocks of health service delivery. The DLT provides feedback to the district health teams on their actions and how these compare to those of their peers. This paper examines the association between DLT score and coverage of institutional deliveries. It explores the effect of RBF, creation of new districts, region, urbanization and completeness of monthly HMIS reports on the coverage of deliveries. The paper further models factors that significantly modify the relationship and whether the interventions aimed at improving coverage of institutional deliveries leads to a better DLT score hence narrowing the gap of loss between conception and one year.

Methods

The study was conducted using secondary data from the annual health sector performance report FY2017/2018 prepared by the Ministry of health using routine health management information system (HMIS) records. The districts were geographically grouped into Eastern, Central, Western and Northern regions; RBF and non RBF districts according to RBF implementation status; new and old districts according to the minimum existence of 5 years, urban and non-urban according to City or Municipality status for a minimum of five years. District health performance was measured using the DLT score. The DLT score is calculated as a weighted composite score consisting of fourteen indicators including coverage of institutional deliveries. The outcome variable was District League Table score while the explanatory variable was institutional deliveries coverage. Size of the district population, geographical location, status of implementation of results based financing, status of district (Old vs New), type of settlement (Urban vs non-urban) and percentage completeness of monthly HMIS reports were investigated whether they modified the association between the outcome and predictor variable. Data analysis was done using Stata version 14. Descriptive analysis was conducted for all indicators in the data set. Pearson correlation coefficient was calculated to measure the strength of the association between DLT score and coverage of institutional deliveries while a scatter plot was used to illustrate the linear relationship. T test and ANOVA

were used to investigate the difference between group means, the Barlett's test was used to test for equal variance and then we applied the Bonferroni test to identify which group means were significantly different. We fitted a simple linear regression model for district league table score and institutional deliveries coverage. We added an interaction term to institutional deliveries coverage and district population, status of RBF district, status of district, type of settlement, completeness of HMIS monthly reports and geographical region one at a time to the model. We applied the testparm test to the interaction terms to identify which variables modified the association between DLT score and the coverage of institutional deliveries. The model was further adjusted for confounding. We tested for the assumptions of linear regression using the histogram and Kdensity plots and the Shapiro Wilk test to test for normality. For homogeneity of variance we plotted residuals against fitted values and then applied the Breusch pagan test. P values less than 0.05 were taken to be significant.

Results

We analyzed data for all the 122 districts in Uganda. 32 districts were from Northern region, 30 district from Central region, 30 from Western region and 30 from Eastern region. 18 were RBF districts while 104 were non RBF Districts. 79% of the districts had existed for more than 5 years while 21% were newly created. 14 of the districts were urban while 108 were non-urban. 2% of the districts had a total population of less than 100,000, 85% of the districts had a total population of between 100,001 -500,000, 11% of the districts had a total population of between 500,001 - 1,000,000 while 2% of the districts had a total population of above 1,000,001 people. 3 districts had reporting rates less than 80%, 12 of the districts had reporting rates between 81% - < 95% while 107 districts had reporting rates >95%. Table1 shows descriptive statistics analyzed. The district league table score was positively correlated with coverage of institutional deliveries with a Pearson correlation coefficient of 0.59 and p valve <0.0001 and the scatter plot showed a positive linear relationship between the variables too. The group mean for DLT score in RBF districts was significantly greater than the group mean in non RBF districts p value <0.0001. The group mean for DLT score in non-urban districts was significantly less than the group mean in urban districts p value =0.0013. The difference in group means between old and new districts was not significantly. There were regional differences between group means for the DLT score p value =0.0053 and the Barlett's test for equal variance was significant p value =0.0020. The Bonferroni test found that the group means for Eastern region was significantly different from Western region p value = 0.0100 and Central region was significantly different from Western region p value = 0.0170. The difference in group means for the categories of total district population and completeness of monthly HMIS reports were not significant. From the simple linear regression model fitted, the expected average DLT score is 61.1323 when the institutional deliveries coverage is 0, and the rate of change in the expected DLT score is 0.1814 for a unit change in the coverage of institutional deliveries. Instructional deliveries coverage accounts for 34.3% of the variations in the district league table score. The interaction term between coverage of institutional deliveries and geographical region was significant p value = 0.0368 and increased the percentage of variation explained by the model to 42.04%. The testparm test for effect modification for the following variables was statistically insignificant and none of the variables cofounded the association between DLT score and institutional deliveries coverage.

Conclusion

From the results of the study, geographical region significantly modified the association between DLT score and coverage of institutional deliveries. The group means for Western region were significantly different from the group means of Eastern and central Regions. From the model fitted, the expected average DLT score = 54.4019 + 0.2938(institutional deliveries coverage) + 8.6772 (Eastern) + 11.8788(Western) + 7.3595 (Central) – 0.1667(institutional deliveries coverage * Eastern) - 0.1585 (institutional deliveries coverage * Western) – 0.1489 (institutional deliveries coverage * Central). The expected average DLT score for districts

in the Northern Region is 54.4019 when institutional deliveries coverage is 0 and the interaction term is 0 too. The main limitation of the study was that it investigated one out of the fourteen indicators used in calculating the league table score which limits the power of the model. The study recommends that further research should be carried out to understand the regional variations in order to inform policy since the model explains 42.04% of the association between DLT score and coverage of institution deliveries modified by geographical location. One of the variables recommended for investigation is the portioning of geographical regions to specific implementing partners who aid in health service delivery on behalf of the ministry of health.

Recent trend in under-five mortality in the Democratic Republic of Congo (DRC): An application of Cox survival frailty regression modelling

[1,2]Eric Panzi [2]Antoine Dikoke, [3]Christine Engondo, [4]Theophile Tshimanga, [5]Martin Soda, [6]Nestor Ngoyi, [7]Catherine Nomopuane, [8]Bertin Tampwo, [9]Marie Claire Omanyondo, [10]Ngianga-Bakwin Kandala
[1-9]Kinshasa Institute of Medical Sciences [2]Kinshasa Institute of Medical Sciences, [2]Northumbria University

Background

Several years of war have created a humanitarian crisis in the Democratic Republic of Congo (DRC) with extensive disruption of civil society and the economy, which have devastated health services. The basis for political stabilization is not still in place and major efforts are required nationally and internationally for reconstruction. These challenges are faced against a background of heavy disease burden (AIDS, tuberculosis, malaria), malnutrition, maternal and under-five mortality. Recent crude estimates of under-five mortality rate of 205 per 1,000 live births for DRC are unacceptable high but this is likely to present a gross under-estimate because of displacement and conflicts. Mortality rates in the DRC are not only influenced by socio-economic, demographic and health variables; they also vary significantly across provinces. Despite some attention from the international community and the media, scientifically rigorous analysis of under-five mortality and economic consequences is scarce.

Methods

This paper analyses recent trends in under-five mortality in the DRC using flexible Cox survival frailty models, which enables the measurement of province-specific cluster effects simultaneously with possibly nonlinear or time-varying effects of other predictors using Cox regression modelling. Data for the study come from the 2007 and 2013 DRC Demographic and Health Sur-vey (DRC-DHS) and includes over 10000 children born between 2002 and 2008.

Results

Results indicate that despite the severity of conflicts in certain location, province-level socio-economic characteristics are important determinants of under-five mortality. More importantly, we find province clustering of under-five mortality, which indicates the importance of cluster effects.

Conclusion

The presentation of this clustering through maps facilitates visualisation and highlights differentials across geographical areas that would, otherwise, be overlooked in traditional data analytic methods.

Estimation of therapeutic window in crossover dose response design in phase ii clinical trial of methylphenidate

'Ogbonnaya Nzie Ezichi [1]Angela Chukwu
'University of Ibadan' Ibadan' Nigeria

Background
Attention-Deficit Hyperactivity Disorder (ADHD) is a chronic condition including attention difficulty, hyperactivity and impulsiveness. There are more than 1.5 million cases per year in Nigeria. ADHD often begins in childhood and can persist into adulthood. It may contribute to low self-esteem, troubled relationships and difficulty at school or work. Symptoms include limited attention and hyperactivity. Treatment can help, but this condition can't be cured. Treatments include medication and talk therapy. Methylphenidate is one of the medication and in this study we try to establish a therapeutic window for it under which ADHD can be controlled to a considerable extent.

Methods
Five different doses including a placebo were administered to the children in five different periods. To test for efficacy and toxicity of the test drug, data were analyzed using a Latin Square Design. A step-down procedure based on pairwise contrasts was also used in the analysis.

Results
The doses were clinically and statistically different and the therapeutic window [MinED, MaxSD] was established for Methylphenidate to control ADHD in children.

Conclusion
A Minimum Effective Dose and the Maximum Safe Dose have been established for Methylphenidate used to control ADHD in children. This would significantly reduce the abuse of the substance and also guard against toxicity due to overdose.

Determinants of Stunting among under-five Years Children in Ethiopia from the 2016 Ethiopia Demographic and Health Survey: Application of Ordinal Logistic Regression Model using Complex Sampling Designs

'Haile Mekonnen Fenta , [2]Demeke Lakew Workie, [3]Dereje Tesfaye Zikie, [4]Belaynew Wassie Taye, [5]Prafulla Kumar Swain 'Bahir Dar University , [2]Bahir Dar University, [3]Bahir Dar University, [4]University of Queensland, [5]India

Background

Stunting is a result of chronic under nutrition and a major public health issue in Ethiopia. This study aimed to calculate the prevalence of stunting, and identify associated factors among children younger than five years.

Methods

The national level cross-sectional EDHS 2016 survey data were accessed and used for the analysis. A total of 9,588 children from all the 11 regions of the country were included. A complex sampling technique has been implemented and based on HAZ score, stunting was classified in to three ordinal categories: normal, moderate and severe. Proportional Odds Model was used to identify determinants. The score test and plots were used to see the proportional odds model assumption and it was not satisfied. As a result, separate logistic regression models were used.

Results

The prevalence of stunting was 38% (with 21% moderately and 17% severely stunted). Children with illiterate mothers were 2 times more likely to be moderately and severely stunted (odds ratio= 1.8.0, 95% CI: 1.18-2.7) compared with their counterparts with secondary education. Children with multiple births were (0.438) times more likely to be stunted (odds ratio= 0.438, 95% CI: 0.522-0.762) than children with singleton birth.

Conclusion

The prevalence of stunting among under-five children in Ethiopia remained high though it decreased over time. Child age, child sex, birth interval, region, mother's educational status, wealth index, toilet facility and types of birth were the important determinants of stunting among children aged 0-59 months in Ethiopia. Addressing these factors will help to prevent future injury of physical and mental development in children and will assist in alleviating malnutrition and refining their quality of life. Moreover, in a DHS data set, complex sampling design should be incorporated in order to make valid statistical inference.

Resampled marginal risk set model for infant mortality at the Kigali university teaching hospital

'Paul Gatabazi [2]Sileshi Fanta Melesse, [3]Shaun Ramroop
'University of Kwazulu Natal [2]University of Kwazulu Natal, [3]University of Kwazulu Natal

Background
In Sub-Saharan Africa (SSA), the Infant Mortality Rate (IMR) remains the highest compared to the rest of the world. In the past decade, the policy on reducing infant mortality in SSA was reinforced and both infant mortality and parental death decreased critically for some countries of SSA. The analysis of risk to death or attracting chronic disease may be done for helping medical practitioners and decision makers for better preventing the infant mortality.

Methods
Dataset comprises 2117 newborns and socio-economic and clinical covariates for mother and children. The Bootstrap Marginal Risk Set Model (BMRSM) and Jackknife Marginal Risk Set Model (JMRSM) for the available covariates are conducted and then compared to the Marginal Risk Set Model (MRSM) for measuring stability of the MRSM. This study uses popular statistical methods of resampling and one selected method of multiple events analysis for measuring the risk of death or attracting chronic disease or critical complication per infant born in 2016 at the Kigali University Teaching Hospital in Rwanda, a country of SSA. Chronic disease or complications taken as alternative events include oliguria, severe prematurity, very low birth weight, macrosomia, severe respiratory distress, gastroparesis, hemolytic, trisomy, asphyxia and laparoschisis.

Results
Both BMRSM JMRSM and MRSM displayed the close results for significant covariates. The BMRSM displayed relatively higher standard error for some non-significant covariates and this emphasized their insignificance in MRSM.

Conclusion
The results suggested that babies from under 20 years old parents were at relatively higher risk and therefore, pregnancy of under 20 years old parents should then be avoided. Infant's abnormality in weight and head was also found to lead to increase the risk of infant mortality, clinically recommended ways of keeping pregnancy against any cause of infant abnormality was then recommended.

Comparative analysis of temporal trends in general fertility rate among four regions in Uganda

'Lilian Giibwa [2]Julius Ssempiira, [3]Godfrey Biroma
'Makerere University Kampala Uganda [2]Makerere University Kampala Uganda, [3]Makerere University Kampala Uganda

Background
Uganda's population keeps rising despite the unmatched slow economic growth. This strains the resources available. Reduction in total fertility rate and general fertility rate (GFR) is key for controlled population. In this study we analyzed the temporal trends in GFR by region using data from the Health Management Information System (HMIS) reported during 2014-2018.

Methods
We analyzed monthly data of live births for the four regions of Uganda in the HMIS during 2014-2018. GFR was calculated as number of live births reported per month in a given region divided by the population size of women aged 15-49 years. The population data was obtained from the census report of 2014.The population for each year was adjusted for population growth. Time series plots were used to describe the temporal trends and Analysis of variance (ANOVA) was used to test for differences among regional trends.

Results
Generally, GFR increased in Uganda during the study period from about 8 to 11 per 1000 women. However, regional trends varied during the study period. The Northern region had the highest GFR over the five years increasing from about 9 to 11 per 1000 women in 2018.The Eastern region had the lowest GFR which increased from 8 to 10 per 1000 women during the study period. The Western region had a static GFR over the period. The variations in the Central region were markedly different in 2015 and had a steep decline in 2016 followed by an increase thereafter. The ANOVA test indicated that the regional trends were significantly different (F-statistic=13.12, P-value <0.0001).

Conclusion
Although there was a general increase in GFR in Uganda during 2014-2018, this is masked by regional variations. This calls for targeted intervention implementation prioritizing the Northern region to address the problem of increasing GFR.

Impact of sampling bias in estimation of cumulative viremia

'Tracy Glass [2]Brian Rambau, [3]Landon Myer , [4]Maia Lesosky
'University of Cape Town [2]University of Cape town, [3]University of Cape town, [4]University of Cape Town

Background

Cumulative viral load (cVL), or viremia copy years , have been suggested as an estimate of an individual's cumulative exposure to raised HIV-RNA viral load over time. CVL is defined as the area under the plasma viral load (pVL) curve, which is the integral of repeated VL measures. This measure of cumulative exposure has typically been calculated using a trapezoidal approximation called "Simpson's rule". As with most measures of cumulative exposure, the underlying assumption is that larger cVL values place individuals at higher risk for poor health outcomes. The empirical evidence supporting cVL as a useful prognostic factor for long term health outcomes is mixed however. We hypothesise that these mixed findings are in part due to sampling bias introduced by viral load sampling frequency. We therefore sought to estimate the magnitude and direction of sampling frequency bias in cVL estimates.

Methods

We estimated cVL using three different source data sets containing repeat viral load measures: an intensively monitored cohort (MCH-ART) of pregnant women initiating antiretroviral therapy (ART), a simulated data set of longitudinal HIV viral load trajectories of pregnant women on ART (VLSIM) and a random sample of routine VL data collected by the National Health Laboratory Services (NHLS), Western Cape, South Africa. To estimate the bias in cVL due to sampling frequency, we calculated a reference cVL using all the available viral load data per individual and compared this estimate to cVL estimated on sub-sampled data sets. We used two different subsampling strategies: a) random subsample to a set number of viral load and b) sampling based on average interval between viral load measures. cVL measures were adjusted for individual follow-up times, resulting in standardised cVL estimate allowing for comparison across the different studies.

Results

The simulated dataset had the greatest number of viral load tests done (median 24, IQR: 19-28) with the least number of tests (33%) below the limited of detection (LOD) (50 copies/ml), while the MCH-ART dataset had the least number of viral load tests (median = 7: IQR: 6-8) with more than 70% of the proportion of VL tests below LOD (74%) . Over 30% and 20% of the viral load tests were above 1000 copies/ml in the WC and simulated dataset, respectively and fewer than 10% of the viral load tests in the MCH-ART study. Significant differences in estimated cVL resulted when using less than the full set of observations for all three data sets (cVLs = 2.1 vs 2.8 - simulated, 0.5 vs 2 - MCHART and 1.7 vs 1.8 - WCVL). There was a significant and meaningful bias towards underestimation, which worsened as the number of observations decreased or the duration between observations increased. The difference between reference measure and subsampled measure of cumulative viremia were more pronounced in the data sourced from empirical observation (MCH-ART; WCVL) than compared to the simulated data.

Conclusion

Sampling bias is significant in measures of cumulative viremia. Cumulative measures of viremia cannot be compared across studies with different frequencies of repeated viral load measurements. Cumulative measures of viremia, and all cumulative measures of exposure estimated with trapezoidal approximation should therefore be used with caution, and only in studies with sufficient frequency of sampling.

Modelling risk factors of diabetes and pre-diabetes in South Africa

'Nina Grundlingh [2]Temesgen Zewotir, [3]Danielle Roberts, [4]Samuel Manda
'University of KwaZulu-Natal [2]University of KwaZulu-Natal, [3]University of KwaZulu-Natal, [4]SAMRC

Background

Diabetes is a metabolic disorder associated with high blood sugar or glucose levels due to the absence or insufficient production of insulin. The number of people with diabetes globally has risen from 108 million in 1980 to 442 million in 2014. It was estimated that, of the 1.8 million people between 20- and 79 years old with diabetes in South Africa (SA) in 2017, 84.8% were undiagnosed. Diabetes was the 2nd leading underlying cause of death in SA in 2016. Furthermore, it was found to be the number one leading underlying cause of death for females in SA. If left untreated, diabetes can cause physical repercussions that could keep individuals from work and, in turn, can negatively affect the country's economy. In 2015 it was estimated that the economic cost due to diabetes in Sub-Saharan Africa was 1.2% of the gross domestic product (GDP) where these countries generally spend 5.5% of their GDP on health. This study aimed to assess the prevalence of diabetes and pre-diabetes and investigate the associated risk factors in the South African population.

Methods

This study utilised data from the South African Demographic and Health Survey (SADHS) done in 2016 where sampled individuals aged 15 years and older had their glycated haemoglobin level (HbA1c) tested. Individuals with test outcome levels of 6.5% or higher were classified as diabetic, those with levels between 5.7% and 6.4% were classified as pre-diabetic, otherwise they were classified as non-diabetic. The potential risk factors considered comprised of a range of demographic, socio-economic and anthropometric variables. Since the response outcome was non-diabetic, pre-diabetic or diabetic, the ordinal survey logistic regression model was fitted. This accounted for the complex survey design and the ordinal nature of the response.

Results

The observed prevalence of diabetes and pre-diabetes from the sampled and tested 6442 individuals was 21.9% and 66.6%, respectively. From which, 24.7% of females and 17.2% of males were found to be diabetic. Similarly, 64.9% of females and 69.5% of males were pre-diabetic. Of the individuals with a BMI classified as underweight to normal, 12.9% are diabetic and 71.8% are pre-diabetic. Similarly, of those with a BMI classified as overweight to severely obese, 30.2% are diabetic and 62.0% are pre-diabetic. Abnormal blood pressure was observed among 49.3% of the diabetics and 34.5% of the pre-diabetics. On the other hand, normal blood pressure was observed among 50.7% of diabetics and 65.5% of pre-diabetics. From the model fit, the variables significantly associated with diabetic status were gender, race, age, waist circumference, haemoglobin level, blood pressure and whether medication for high blood pressure was being taken, frequency of consuming fast food and having had smoked the previous 24 hours.Males had higher odds of being pre-diabetic or diabetic compared to females. The odds of a black/African individual being pre-diabetic or diabetic was higher than that of an individual from other race groups. Individuals with abnormal blood pressure had higher odds of being pre-diabetic or diabetic compared to individuals with normal blood pressure. Furthermore, it was less likely for an individual that was not taking high blood pressure medication to be pre-diabetic or diabetic compared to an individual that was taking high blood pressure medication. Individuals with a larger waist circumference had higher odds of being pre-diabetic or diabetic compared to individuals with a smaller waist circumference.

Conclusion

Diabetes is of major concern in South Africa however, it does not receive the attention it deserves where policy makers are generally unaware of its current prevalence. The results highlighted are from the data exploratory analysis of the early stage of the study. These preliminary analyses indicate the need for advanced modelling to contribute in identifying diabetes risk factors for public awareness, medical prognosis and preventative intervention.

A systematic review of spatial methods using Health Surveys in Sub Saharan Africa

Ndamona Haushona [1,3], Samuel Manda [1,2]
[1] Biostatistics Research Unit, South African Medical Research Council, South Africa
[2] Department of Statistics, University of Pretoria, South Africa, [3] Division of Epidemiology and Biostatistics, University of Stellenbosch, South Africa

Background
Statistical methods for mapping of disease are playing an important role to inform control efforts, identify and predict areas of highest risk of disease incidence and prevalence. Their implementation and application have increased exponentially in the sub-Saharan Africa (SSA) region as a result of availability and accessibility of Bayesian spatial methods and statistical software to fit them. There has also been a shift towards finding estimates of disease burden at lower resolution to support subnational health policy planning. Several of the applications have been based on using data from nationally representative household and population health surveys. The objective of this paper is to provide a review on the various spatial statistics methods applied and statistical software, type of pubic health themes and demographic groups studied, and to make recommendations for improved spatial methods and usage of health survey data in disease mapping in the SSA region.

Methods
An organized literature search for articles published from 1990 to 2018 using data from household and population health surveys related to spatial methods was done through PMC, PubMed/Medline, Scopus, NLM Catalog, and Science Direct electronic databases. We used the following search strategy to thoroughly retrieve necessary articles: (“Spatial Statistics” OR "Spatial modelling" OR "Spatial variation" OR "Small areas estimation" OR "Demographic Health Surveys" OR "AIDS indicator surveys" OR "Malaria Indicator survey" OR "Multiple Indicator Cluster Survey" OR "Health Surveys" OR "MIS" AND "sub-Saharan Africa"). Using pre-specified inclusion criteria, articles titles and abstracts were screened by two independent reviewers, and vital information answering the review’s objectives was abstracted.

Results and discussion
The search returned 4185 unique articles, of which 200 were included in the final review. The 200 articles were scrutinized and summarized in the following themes: (a) Spatial methods (b) Public health topics and (c) Special demographic groups of focus. Studies demonstrated a broad range of spatial analysis applications on characterizing geographic distribution of public health issues such as HIV, child and maternal mortality, and Hypertension. Nearly all of them did not account for survey design in the analyses and used health outcome in the analyses.

Conclusions
The review has found a rise in application of Bayesian geostatistical models, and the use computer packages such as integrated nested Laplace approximation (INLA), BayesX, and WINBUGs. Correspondingly, the literature shows that disease mapping methods had made substantial contribution to public health evidence base in SSA.

Suicide Ideation and Associated Factors among School Going Adolescents in Namibia: A Multilevel Logistic Regression

Kapenda S.R.N, Pazvakawambwa L.
University of Namibia, Department of Statistics and Population Studies

Suicide accounts for many deaths worldwide making it the second leading cause of death among 15-29 year olds globally. Adolescents who reported suicidal ideation at an early age are more likely to later attempt suicide. Suicidal ideation is the most commonly reported form of suicidality. Suicide ideation leads to mental health disorder and has been shown to "spread" suicidal thoughts to others. The objective of the study was to establish prevalence, patterns and risk factors for suicidal ideation among school-going adolescents in Namibia. The study analysed secondary data from the Namibia School-based Health Survey (GSHS) of 2013 (N= 3 531). Multilevel logistic regression, which allows for the incorporation of group-level effects (school, grade and region) and estimation of interactions was used to establish risk factors associated with suicidal ideation. Living arrangements, age, alcohol and drug abuse, sexual activity were significantly associated with suicide ideation. Preliminary results seem to suggest significant school level effects. Adolescents should be discouraged from taking alcohol and drugs and engaging in early sexual activity. Targeted interventions to reduce suicide ideation should be
school specific.

Comparing Drug Regimens for Clearance of Malaria Parasites in Asymptomatic Adults in a Clinical Trial in Kilifi County, Kenya

'Emma Nelima Khaemba [2]Caroline Ogwang, [3]Philip Bejon, [4]Sam Kinyanjui, [5]John Mutiso Muindi, [6]Joseph Koske, [7]Francis Ndungu, [8]Melissa Kapulu

'KEMRI-Wellcome Trust Research Programme' Kilifi' Kenya [2]KEMRI-Wellcome Trust Research Programme' Kilifi' Kenya, [3]KEMRI-Wellcome Trust Research Programme' Kilifi' Kenya, [4]KEMRI-Wellcome Trust Research Programme' Kilifi' Kenya, [5]Moi University, [6]Moi University, [7]KEMRI-Wellcome Trust Research Programme' Kilifi' Kenya, [8]KEMRI-Wellcome Trust Research Programme' Kilifi' Kenya

Background

In Phase IIb trials of candidate malaria vaccines where molecular methods are used to determine endpoints, anti-malarial drugs are used prior to commencing monitoring. This is done because vaccines only act against new infections and not old infections. However, if the anti-malarial used has a long half-life it may inhibit the acquisition of new infections. In this study the outcomes of malaria PCR monitoring were compared following randomization to three different drug regimens.

Methods

Ninety healthy adult participants aged between 18 to 50 years were randomized in 3 different groups to receive one of the three different anti-malarial drug regimens: Atovaquone/proguanil + Artesunate (AP+AS), Artesunate (AS) or Sulphadoxine pyrimethamine + Artesunate + Primaquine (SP+AS+PQ). Predictors of PCR positivity were examined by time and by drug regimen. Positivity was also assessed by gametocyte-specific and asexual-parasite-specific RT PCR. Classification of infections as either new or recrudescent was done using MSP2 genotyping data.

Results

There was an overall but incomplete reduction in the PCR positive rate for malaria parasites from 32.7% (147 out of 575 tests) between day 0 and 7 to 20.4% (55 out of 575 tests) between day 9 and day 14. MSP2 genotype data showed that PCR positive results after anti-malarial treatment were frequently due to new infections especially in the latter period of follow-up.

Conclusion

PCR positivity was persistent at low levels after all regimens were administered, hence none of the regimens were ideal in clearing malaria parasites.

Assessment of possible bias during sampling of health establishments for inspections

'Collen Makomane [2]Mantokeleng Matsaneng
'Office of Health Standards Compliance [2]Office of Health Standards Compliance

Background

The Office of Health Standards Compliance is responsible for the collection of information relating to prescribed norms and standards. Inspections are being conducted to collect data about the compliance of health establishments with the National Core Standards. The Office is expected to inspect more than 4000 public health establishment at least once every 4 years with the current limited resources. In each clinic, an inspector assesses 191 units of measurements, CHCs more than 200 and for a hospital more than 600 units of measurements. Other measures require inspectors to sample patients, records and wards. The current sampling process of health establishments is based on their proximity to limit travelling costs. The process does not give all the health establishments an equal chance to be selected and thus, the calculated compliance score process does not allow general statements to be made about the compliance. The review of the sampling process will assist the OHSC as a regulator to identify areas with compliance gaps. The aim of this study is to assess bias in the inspection sampling process that occurred due to resource limitations .

Methods

The published 2016/17 Annual Inspection Report methodology section was reviewed. The sampling technique used to select health establishments for inspections and the selection of the patients for interviews were assessed. The review considered that the same sampling method have been used from 2011 to 2019. The bias in the sampling method were assessed based on best statistical practices. Methods used to update the list of health establishments were reviewed to understand the completeness of the sampling frame. The method used to calculate the sample size was reviewed to allow recommendation of an optimal method. Grouping of health establishments for sampling were checked for cost-effectiveness and if the grouping allowed the adjustment of missing health establishments, patients, records, wards etc. Considering that not all the health establishments will be inspected the study also assessed how the missing health establishments are represented in the sample. The same process was followed to understand how the missing patients, records, wards are represented in the samples of the functional areas. Considering that some of the measures can be recorded as not applicable because of required information not available at that moment, the method to input the missing data was reviewed. The study assessed the feasibility of sampling the agreed number of patient and staff interviews to optimize the representation of the sample to the population. Analysis of the data with weights was explored to qualify general statement about the findings. The current method were assessed for gaps and recommendations were highlighted for improvement. Analysis of the precision and accuracy against the parameters were explored to support the recommendation; confirming that the results were not obtained by chance.

Results

The 2016/17 annual inspection report shows that health establishments were sampled using a convenient sampling method and the same method was used to sample patients and records at the facility level. Considering that convenient sampling method does not give members of the population an equal chance to be selected, the sample can be exposed to manipulation to favour the interests of individuals. Convenient sampling was selected based on the distance between the health establishment, budget allocated to conduct inspections, time required to complete targeted inspections and number of available

inspectors to complete inspections; the method resulted in bias which will lead the Office to criticism and possible litigations. The list of health establishments can be managed with a memorandum of agreement between The Department of Health and the OHSC to allow the recording and monitoring of newly opened, existing and closed down health establishments. The current list of health establishment does not indicate the date when the list was last updated and the number of newly opened and closed health establishments. The current sample size of 17% was sampled from the list of health establishments in proportion to the number of health establishments at provincial levels. To minimize the allocation of scares resources; the sampling was conducted at sub-district level and all health establishments within a sub-district were inspected. The sampling process did not show the percentage of error that were tolerated. Considering how the data is structured, random sampling can be performed at the sub-district and health establishment level. With an estimated target of 4000 public health establishments that need to be inspected every 4 years, a target of 1000 health establishment can be inspected annually. Considering that the total numbers of the health establishments are not equal in the sub-district; 5 facilities per sub- district will be enough.

Conclusion

Convenient sampling method does not give statisticians an opportunity to randomly sample the data to avoid selection bias; it is difficult to account for selection errors and random sampling can assist to provide enough statistics that can be used to account for selection errors. Probability sampling methods will be appropriate for any inspections sampling process. General statements needs to be made about the findings of the data, random sampling will be appropriate to allow missing health establishments to be represented in the sample. Regarding sampling of patients, staff, records and wards; inspectors can provide information that can be used to audit the findings; survey statisticians will apply the random sampling methods to calculate sampling errors to support the significance of the selected sample. With missing data random sampling can allow weights to be calculated during the sampling process to be used to adjust the findings. Random sampling method can be considered even when the resources are minimal while maintaining the accuracy and precision levels of the estimates.

Psychometric properties of PHQ-4 anxiety depression screening scale among out of school adolescent girls and young women

'Jacqueline Materu [2]Evodius Kuringe, [3]Daniel Nyato, [4]Gasper Mbita, [5]Amani Shao, [6]Baltazar Mtenga, [7]Soori Nnko, [8]Mwita Wambura

'Department of Sexual and Reproductive Health' National Institute for Medical Research' Mwanza' Tanzania [2]Department of Sexual and Reproductive Health' National Institute for Medical Research' Mwanza' Tanzania, [3]Department of Sexual and Reproductive Health' National Institute for Medical Research' Mwanza' Tanzania, [4]Jhpiego Tanzania - an affiliate of Johns Hopkins University' Dar es Salaam' Tanzania., [5]Department of Sexual and Reproductive Health' National Institute for Medical Research' Mwanza' Tanzania, [6] Department of Sexual and Reproductive Health' National Institute for Medical Research' Mwanza' Tanzania, [7]Department of Sexual and Reproductive Health' National Institute for Medical Research' Mwanza' Tanzania, [8]Department of Sexual and Reproductive Health' National Institute for Medical Research' Mwanza' Tanzania

Background

Literature suggests that most mental disorders go undiagnosed until adulthood, even though their onset is around childhood and adolescence. Several recommendations have been put forward in the literature on how to improve access and utilization of mental health services. Among the recommendations is decentralization and task shifting in the provision of mental health services, to maximize uptake of services both at lower lever facilities and in the community. To be able to realize this, screening scales must be made available to enable identification of probable mental health disorder patients to be referred to health facilities for diagnosis. Scales such as Patient Health Questionnaire-9 (PHQ-9 items) and Generalized Anxiety Disorders-7 (GAD-7 items) have been used for depression and anxiety respectively. However, in order to improve efficiency, shorter versions of the scales (PHQ-4) have been researched and successfully validated among clinic-based and general populations. This is a four-item screening tool for both depression (PHQ-2 items) and anxiety (GAD-2 items). This tool was validated in different population, such as in the general adult population in Germany and among college students in the United States. However, there is a paucity of research on testing and validation of the PHQ-4 on adolescent girls and young women (AGYW) who are out of school in the community. This study aims to explore the validity and reliability of the PHQ-4 in a large out-of-school AGYW population in Tanzania.

Methods

A cross-sectional analysis of first round data collected from June to July 2018 from an on-going cluster randomized controlled trial in North-West Tanzania was conducted. A total of 2426 out-of-school adolescent girls and young women (AGYW) aged 15 to 23 years were included. AGYW were given a questionnaire to respond to the items on the Patient Health Questionnaire (PHQ-4), a tool comprising of PHQ-2 and Generalized Anxiety Disorders (GAD-2) screener's. Data were collected using Audio Computer-Assisted Self-Interview (ACASI). The proportions of participants with the lowest (floor) and highest (ceiling) possible scores were assessed using frequency distributions. Cronbach's α was used to measure reliability of the PHQ-4 (i.e. the extent to which all items on a scale were inter-related or contribute positively towards measuring the same construct. Due to the ordinal nature of the PHQ-4 items, the estimation of the Cronbach's α was based on polychoric correlation matrix. Confirmatory factor analysis (CFA) and Principal components analysis (PCA) were used for construct validity assessment. In CFA, the analysis was based on the two stage estimation approach due to the ordinal nature of the items so as to ensure unbiased parameter estimates and standard error. In stage one, polychoric correlation was computed, then the model was analyzed with the appropriate weight matrix computed in stage one. In

addition, three criteria were used to assess how well the model fit the data: Standardized Root Mean Square Residual (SRMR), the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA) and 90% confidence interval for RMSEA. In PCA, a varmax rotation (orthogonal rotation) method was used with the assumption that there is no correlation between components. Furthermore, using chi-square test the association between the PHQ-2, GAD-2, and PHQ-4 scores for respondents who had reported suicidal thoughts or social dysfunction versus those who were not in the past two weeks was explored.

Results

Of the 2426 participants, positive screening for anxiety (GAD-2 ≥ 3) and depression (PHQ-2 ≥ 3) were 33.8% and 35.5% respectively. Cronbach's α of PHQ-4 was 0.81 and both items-correlation and corrected items-correlation have total correlations above 0.3 ($p< 0.01$). CFA showed that all items loaded significantly onto the single factor, and loadings were strong, ranging from 0.67 to 0.77 ($p<0.01$). CFA indicate that the PHQ-4 scale stand for a unidimensional construct with good model fit (CFI=0.995, SRMR=0.013 and RMSEA=0.054, 90% CI 0.031–0.079). PCA confirmed two distinct components (anxiety and depression) that explained by 72% of the total variance. Those who were reported having suicidal thoughts and social function problems had statistically significantly higher scores on PHQ-2, GAD-2, and PHQ-4 screening items ($p<0.01$). Floor effects were observed for all items.

Conclusion

The findings from this study, suggest that PHQ-4 item measure, can reliably and validly measure depression and anxiety among the out of school adolescent girls and young women. Overall, the PHQ-4 reliability and validity were reasonable and observed to be psychometrically sufficient and helpful for the measurement of anxiety and depression among adolescent girls and young women (AGYW). The PHQ-4 could be one of the useful tools for assessing depressive and anxiety disorders in a community as a mass screener or in hectic clinical settings within a short period of time. Then, those with positive scores on PHQ-4 questionnaire, could be administered a more meticulous and comprehensive screening tool for anxiety and depression as one among of other steps. The PHQ-4 results will have an important implication for screening, referral, and treatment of depression and anxiety disorders.

Transitional Viral Suppression due to ART on a Zimbabwean Cohort from 2004 to 2017: A Bayesian Mixed effects Multistate Model

'Zvifadzo Matsena Zingoni [2]Tobias Chirwa, [3]Jim Todd, [4]Eustasius Musenge
'Witwatersrand University [2]Witwatersrand University, [3]London School of Hygiene and Tropical Medicine, [4]Witwatersrand University

Background
Achieving viral suppression is a key milestone in monitoring HIV patients on antiretroviral therapy (ART). Multistate models can be used to simultaneous estimate the HIV disease progression rates among HIV patients in Zimbabwe between 2004 and 2017 during their ART uptake.

Methods
A Bayesian structural additive semiparametric mixed effects multistate Markov model was fitted to the data to estimate HIV disease progression trajectories. The model states were defined based on viral load undetectable threshold cutoff limits: ("suppressed-state" (<50 copies/mL), "unsuppressed-state" (≥50 copies/mL) and "death (absorbing) state"). Transition intensities were graphically modelled as smoothed time based penalised splines semiparametric Bayesian additive models.

Results
Amongst 18,150 participants, 66.36% were females and 54.43% were in WHO stage III/IV. The baseline transition rate from unsuppressed to the suppressed, increased with greater duration on ART at a rate of 31.4% per additional year (95%CI: 31.3-31.4%). Males were 68% more likely to die from unsuppressed state compared to females. WHO stage III/IV patients with suppressed viral load (reference: WHO stage I/II) were 26% more likely to have a viral load increase. Mortality from unsuppressed state increased with age. Compared to other provinces, patients in Matabeleland North and Manicaland were 7.39 and 4.48 times more likely to have viral load increase, respectively, whilst Midlands, Mashonaland West, Mashonaland Central and Matabeleland provinces were less likely (0.05 times) to have undetectable viral load.

Conclusion
Early and more sustained treatment interventions are essential to slow the progression of HIV patients to immune deterioration, as found that the longer the patients with unsuppressed viral load stayed on ART the more likely they became better. A unique aspect of this study was inclusion of spatial random effects on patient disease progression models, to help answer where ART is most effective. More focused interventions targeting defaulters on ART treatment which could prioritize high burdened areas, such as Matabeleland North, Manicaland and Masvingo provinces are required.

Statistical methods that detect publication bias in meta-analysis

Mhlengi Mgaga[1], Henry G. Mwambi[1], Ding-Geng Chen[2]

[1]School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa, [2]Department of Statistics, University of Pretoria, Pretoria, South Africa

Meta-analysis is a statistical analysis that combines results from different independent studies which investigate the same topic. In various fields of medicine, meta-analysis methods have become increasingly popular. For example interest may be to combine information on efficacy of a treatment that is available from a number of clinical studies with similar treatment protocols. However, meta-analysis results are subject to criticism for many reasons, one of those reasons is publication bias. Publication bias occurs when the results of published studies are systematically different from the results of unpublished studies. In particular, studies that are less likely to get published appear to be the less conclusive ones. The chance that studies with small sample size and low statistical precision to get published, is increased if they show stronger treatment effects. Publication bias may affect the conclusions of meta-analysis and systematic reviews and result in a biased overall estimate of the treatment effect. Therefore, it is important to evaluate the data for publication bias before conducting a meta-analysis. Graphical and formal statistical methods have been developed to investigate if the data might be affected by publication bias. In this presentation we discuss the statistical methods that detect publication bias in meta-analysis.

Harmonisation of Longitudinal Population Data: evidence from three rural Health and Demographic Surveillance System Nodes in South Africa

'Taurayi Mudzana [1]Mark Collinson, [1]Kobus Herbst
'South African Medical Research Council

Background

Harmonised longitudinal population data can be a national asset for a developing country such as South Africa, due to its ability to support research on population dynamics, to understand the determinants and outcomes of population processes, and to calibrate national datasets. Data employed by the national government to measure progress towards specific developmental goals, such as the Sustainable Development Goals and National Development Plan, can be complemented and evaluated using harmonised longitudinal population data. Additionally, harmonised longitudinal population data increases the study power through increasing sample size thus allowing for more sophisticated statistical analyses.The South African Population Research Infrastructure Network (SAPRIN) represents a network of three Health and Demographic Surveillance System (HDSS) nodes located in rural South Africa, namely: MRC/Wits University Agincourt HDSS in Bushbuckridge District, Mpumalanga, which has collected data since 1993; the University of Limpopo DIMAMO HDSS in the Capricorn District of Limpopo, which has collected data since 1996; and the Africa Health Research Institute (AHRI) HDSS in uMkhanyakude District, KwaZulu-Natal, which has collected data since 2000. Individual and household indicators that are routinely collected and assessed include: vital events, such as births and deaths, residence and migration, socio-economic status, and measures of wellbeing represented by labour status, education and social protection. This network will shortly be expanded to include urban HDSS nodes. The purpose of this abstract is to describe the harmonisation process used to align the existing health and demographic surveillance data. As a case study we show how the harmonised dataset has been used to examine trends in child mortality across the three nodes.

Methods

Data harmonisation is the consolidation of multiple sources of data to present a singular perspective on objective reality. The data collected by SAPRIN nodes since their inception from their geographically-defined research areas have been harmonised to form this SAPRIN dataset. The dataset is derived from a core database structure developed by the SAPRIN team in collaboration with data managers from the three HDSS nodes. The data were passed through a series of quality measures and checked for accuracy and completeness using various demographic indicators and entity counts. If data failed the quality standards, they were returned to the nodal data managers to resolve the issues. When the data quality achieved an acceptable threshold, the data was integrated into the core database structure. The final harmonised dataset is shared with the scientific community via the SAPRIN Data Repository (http://saprindata.samrc.ac.za). The FAIR guiding principles were used in the development of the Data Repository to ensure that the data are Findable, Accessible, Interoperable and Reusable.To demonstrate the use of the SAPRIN harmonised data a case study is shown in which child mortality trends are examined. The analytic approach used was continuous-time event history analysis using survival-time, 'stset' commands, in Stata 15. Time event history analysis is a collection of statistical methods that focus on questions relating to timing and duration until the occurrence of an event.

Results

The results indicate that child mortality trends in the three HDSS nodes show a rise and subsequent fall in each node, but the scale and timing of the mortality peaks are different. Child mortality trend for Agincourt

shows mortality rates stable at around 6/1000 child-person years at risk (PYAR) from 1993 to 1997, a steep growth in mortality occurred from 1998 to 2002, rising from roughly 7 to 15/1000, to a peak of between 11/1000 to 23/1000 over an eight-year period 2003 to 2009, followed by a steady reduction between 2009 and 2016, dropping from 9/1000 to 3/1000. Whereas for the children at the Africa Health Research Institute (AHRI) study-site in uMkhanyakude, at the onset of population surveillance, in 2000, there was already a high child mortality rate, with estimates of around 20/1000 PYAR. The mortality rate remained high until 2003. From then, there is a rapid reduction. The first part of the reduction, between 2004 and 2007, is steeper with the mortality rate dropping from about 20/1000 to 10/1000 in only 4 years. In the case of DIMAMO child population numbers show that a large population size is needed to adequately analyse child mortality. From 1996 to 2010 the study population had a total of 8000 people. A decision to substantially enlarge the HDSS surveillance population was taken in 2009, with initial enumeration in adjacent communities conducted in 2010. This added a further population of 28 000, bringing the total to around 36 000 individuals. Of these, under-five children made up around 3 300.

Conclusion

One of the benefits of using the SAPRIN harmonised longitudinal dataset is that the data from the three nodes can be presented side-by-side and directly compared. As a result, this has a potential to improve informational support for decision-making and strengthening of policy on health and socio-economic matters.

Factors Associated with Prevalence of Pregnancy Termination in Uganda: A Population-Based Study

'Edson Mwebesa [2]Henry Sseguya, [3]Emmanuel Obuya, [4]Mary Nakafeero, [5]Julius Ssempiira, [6]Nazarius M. Tumwesigye

'Makerere University School of Public Health [2]Makerere University School of Public Health, [3]Makerere University School of Public Health, [4]Makerere University School of Public Health, [5]Makerere University School of Public Health, [6]Makerere University School of Public Health

Background

Abortion in Uganda is illegal, only permitted when it places the pregnant mother at risk. This lack of choice is associated with secrecy abortion; in most cases, unsafe increasing the probability of maternal mortality. This study aimed to find out the prevalence of pregnancy termination and its associated factors based on the 2016 Uganda Demographic and Health Survey (UDHS).

Methods

The 2016 UDHS data were used in this study. More than 57,000 women of the age of 15 – 49 years participated in this study. Data were cleaned and analyzed using Stata v14. A modified Poisson regression model was used to investigate factors associated with pregnancy termination.

Results

Of the women who participants in this study, 15384 (27%) had ever had a pregnancy terminated. The woman's age (APR = 1.35, 95% CI: 1.17, 1.56) and increased with age, marital status (APR = 1.63, 95% CI: 1.13, 2.33), contraceptive use and intention (APR = 1.34, 95% CI: 1.11, 1.60), frequency of listening to radio (APR = 1.22, 95% CI: 1.10, 1.34) were all positively and significantly associated with pregnancy termination. On the other hand, doesn't use cigarettes & tobacco (APR = 0.81, 95% CI: 0.66, 1.00) and whether visiting the health facility is a not a big problem (APR = 0.90, 95% CI: 0.84, 0.98) were negatively and significantly associated with pregnancy termination.

Conclusion

There exists a significant proportion of women who have had their pregnancies terminated in Uganda. Mother's age, marital status, contraceptive use and intention, frequency of listening to radio are the main predictors. Based on these results, researchers concluded that emphasis should be put on improving access to post-abortion care, family planning services, contraceptive use, and media exposure. Perhaps, abortion law should be revised in light of these outcomes. This is the first population-based study in Uganda.

Factors associated with cigarette smoking in Uganda: Analysis of Uganda Demographic Health Survey (UDHS) 2016 Data

'Caroline Nakalema ²Suzan Nakasendwa
'Makerere University School of Public Health ²Makerere University School of Public Health

Background

Cigarette smoking poses a growing public health problem to the country and is risk factor for Cardiovascular Diseases (CVDs), Lung Cancer, Diabetes and Chronic Respiratory Diseases which result into death. Statistics from the Uganda Cancer Institute also indicate that 25 % of lung cancer patients were cigarette smokers and 16 %, 13.7 % and 12.6 % of oral, stomach and throat cancer patients were former smokers. However, the number of cigarette smokers is rising and if left unchecked the risk of mortality is likely to increase especially in Africa (WHO 2015).The prevalence of cigarette smoking in Uganda was 4.8% (WHO 2017). The study was aimed at assessing the factors associated with cigarette smoking in Uganda which include age, highest level of education attained, type of residence, marital status, wealth index, region and the sex of the household head.

Methods

Secondary data analysis of Uganda Demographic Health Survey 2016 was conducted. An analytical sample of 15,522 individuals aged between 15-49 years participated in the study, the data was weighted and descriptive statistics were used to get the mean and standard deviation. Multiple Logistic regression was used for statistical analysis to determine the variables that were significantly associated with the cigarette smoking using Stata version 13 with 95% confidence level.

Results

Out of 15,522 respondents, the mean age of the respondents was 28.5 years, (SD=6.83). Individuals aged 35-44yrs were more likely to smoke cigarette compared to those <25 years OR=2.94[1.3147,6.5682]. Compared to participants with no education at all, those with a higher level of education were less likely to smoke cigarette OR=0.11[0.0137,0.9456] as were those with secondary level education OR=0.22[0.0660,0.7021]. Compared to the poorest participants, the richest participants were less likely to smoke cigarette OR=0. 31[0.1021,0.9555] as were the poorer participants OR=0. 37[0.1792,0.7579]. Compared to participants who lived in Kampala region, those who lived in Karamoja region were less likely to smoke cigarette OR=0.06 [0.0092,0.4039] as were those who lived in Acholi region OR=0.08[0.1070,0.5467] and those who lived in Lango region OR=0.12[0.0182,0.8277].

Conclusion

The highest number of cigarette smokers was among the poorest people aged 35-44 years in Kampala region with no education. These are the ones who need more attention through implementation of educational programs for awareness on media platforms as well as translating the information into local languages to reach people with no education, improved attitudes and practices as well as cigarette smoking cessation programs.

Prevalence and Factors Associated with Low Birth Weight in Uganda: Analysis of Uganda Demographic Health Survey (UDHS) 2016 Data

'Suzan Nakasendwa [2]Caroline Nakalema
'Makerere University School of Public Health [2]Makerere University School of Pulic Health

Background
The World Health Organization (WHO) defines low birth weight (LBW) as a new born having a weight of less than 2,500g at birth. Low birth weight is one of the major determinants of prenatal survival, infant morbidity and mortality as well as the risk of developmental disabilities and illnesses in future lives. WHO estimates that about 30 million low birth weight babies are born annually (23.4% of all births) and they often face short and long term health consequences. Whereas the global prevalence of LBW has slightly declined, the rate in many developing countries is still quite high. In Uganda, low birth weight is a problem with a prevalence rate of 13.5% reported in the recent Uganda Demographic and Health Survey (2011). The study aimed to estimate the prevalence and identify factors associated with low birth weight of babies in Uganda.

Methods
Using the Uganda Demographic Healthy Survey (UDHS) 2016 secondary data with an analytical sample of 10,429 mothers aged 15-44 years participated in the study. Data was weighted where descriptive statistics and multivariable logistic regression was fitted with using a backward elimination method were used to determine the association between birth weight of babies and prenatal care, type of cooking fuel, wealth status, sex of the baby, smoke status and socio-demographic factors using odds ratios and 95% confidence interval. Data was analyzed using STATA version 13.

Results
The mean birth weight was 3342g and mean mothers' age was 29 years with 6.2% aged 15–19 yrs. The prevalence of LBW was 9.8%. Compared with poorest mothers, those with middle wealth status were less likely to give birth to low birth weight babies OR = 0.83 [0.64-1.08], as were those with richer wealth status OR = 0.67 [0.52-0.87] and those with the richest wealth status OR 0.65 = [0.48-0.89]. Mothers who did not attend prenatal care were more likely to give birth to low birth weight babies compared to mothers who attended prenatal care OR = 2.81 [1.35-5.87]. Pregnant mothers who use biogas for cooking are more likely to give birth to low birth weight babies as compared to mothers to use liquid petroleum gas (LPG) OR = 10.13 [1.62-63.30].

Conclusion
The prevalence of LBW was 9.8%. Mothers with a poor wealth index, who don't attend prenatal care, and those who use biogas are more likely to give birth to low birth weight babies. Health professionals need to address maternal health. Health workers should encourage pregnant mothers to attend focused prenatal care as recommended by the Uganda ministry of Health. A specialized maternal facility centre that is friendly to pregnant mothers is advisable so as to improve on completion rates and capture high risk teenage mothers early.

Modeling Anemia in Children under five in Tanzania and Angola

'Qondeni Ndlangamandla
'University of KwaZulu-Natal

Background

Anemia is a common disease in Sub-Saharan Africa. It is a disorder whereby the body has insufficient hemoglobin (Hb) level to provide enough oxygen into the body tissues. In young children, it is characterized by the fragility and distortion of bones. As such, the lack of iron supplementation remains the main cause of anemia. This study aims to investigate the socio-economic and socio-geographic factors that are statistically associated with childhood anemia in children from Tanzania and Angola.

Methods

The 2015/2016 TDHS (Tanzania Demographic and Health Surveys) and 2015/2016 ADHS (Angola Demographic and Health Surveys) data sets were used to study the variables that are significantly associated with childhood anemia. The IBM Statistical Package for Social Science(SPSS) was used to give a summary of descriptive statistics information(cross-tabulation). Further, Survey Logistic regression and Generalized Additive Mixed Models would be fitted to data in helping in the identification of the factors that are significantly associated with childhood anemia.

Results

At 5\% level of significance, the survey logistic regression model reveals that factors that said to be common causes of anemia in both countries, includes: wealth index, sex of the child, region, mother literacy and child nutritional status (wasting and stunting). In Tanzania, children who were from poor families were 37.7\% more likely to suffer from childhood anemia than children born from rich families (reference category) OR = 1.347. In Angola, Children from rich families were 13.45\% less likely to suffer from childhood anemia disorder in comparison to children from poor families, OR=0.1345. In both countries, the model further reveals that the factors named: child nutrition(specifically, stunting), region, the current age of the child and the sex of the child, was significantly associated with childhood anemia.

Conclusion

According to the results from the survey logistic regression model, the factors that were early identified as the main causes of anemia remain significant. Thus, government and health departments have to do interventions on fighting with such factors.

Redefining agricultural statistics for the future

'Elias Otieno Obudho [2]Ratemo W. Michieka
'University of Nairobi and IBS Kenya Chapter [2]University of Nairobi and IBS Kenya Chapter

Background

During the 20th century Agricultural research played a key role in the development of statistical methods. Wide heterogeneity in the experimental materials led to the development, refinement and application of various statistical tools. Increasingly high demand of agricultural produce, fluctuations in productivity, dwindling arable land, rapid population growth, per capita income and climate change are transforming agricultural research. Agriculture has undergone yield enhancing transformations through mechanization, introduction of new crop varieties/breeds and agricultural chemicals to precision agriculture; subsequently broadening research scope. Precision agriculture is an integrated crop/livestock management system that combines information technologies with rational agricultural industries to provide timely and variable amounts and type of inputs based on actual localized needs. This study explored a statistical toolbox that will redefine research methodology, specifying the direction quantitative information will change agricultural landscape.

Methods

Qualitative and quantitative research approaches were used to explore the synergy between statistical methodology and agricultural research. Key informant interviews (KII) and literature review was carried out on the current and emerging issues to gain an understanding of underlying statistical methods, challenges and motivations behind their use. Sixty articles were selected from a sampling frame of seven major African agricultural related journals published between 2012 and 2018 using stratified method. The study focused on the research type, objectives, conceptual relationships and statistical contents, considering emerging and alternative methods.

Results

The studies sampled were of cross cutting disciplines; 38% livestock/animals, 38% plants, 36% disease/pests and 28% on food nutrition, human health and Socio economic. They were low on research impact, agrometeorology and climate change. The use of common designs and statistical methods were noted in traditional research. Since precision agriculture is built on big data, multivariate nature with spatiotemporal variability, evaluation of treatments goes beyond classical ANOVA. From the research objectives and conceptual relationships, spatiotemporal analyses such as repeated measures, survival, longitudinal/panel, forecasting/time series, geospatial modeling/simulation were possibilities. For causality and interactive relationships, path analysis was a possibility.

Conclusion

The data showed that the use of modern technology to increase yields has transformed farming management concept both at micro and macro level, opening a new horizon in agricultural research. Smart farming builds on precision agriculture, big data, digital mapping and data driven crop models for optimal farming based on local climate and soil characteristics is redefining statistics. Transformation in agricultural research and production remains an important contributor in the redevelopment of statistical methods and learning through a feedback process, redefining the role of statisticians.

Spatial-Temporal Distribution of Malaria Prevalence among the Under-Fives in Tororo District, Uganda

'Emmanuel Obuya [2]Simon Kasasa, [3]Edson Mwebesa, [4]Henry Sseguya, [5]Nazarius Tumwesigye
'Makerere University [2]Makerere University , [3]Makerere University, [4]Makerere University, [5]Makerere University

Background
Malaria transmission varies in space and time, and it is being driven by many factors namely; the mosquito vectors, climate, vegetation, elevation and land use. Uganda is among the countries that have registered a decline in malaria prevalence. However, the disease endemicity is not uniformly distributed. Tororo district is one of the high burdened districts in Uganda with a community Plasmodium prevalence of 48% in 2012. Interventions, namely; Universal distribution of Long Lasting Insecticide Treated Mosquito Nets (LLITNS) and Indoor Residual Spraying (IRS) have been applied to control malaria in the district. However, the current distribution of the disease burden in space and time has not been fully documented. We analyzed data from the district health information system (DHIS2) to identify the spatial-temporal patterns of malaria among the under-fives in Tororo District.

Methods
We extracted data from DHIS2 between 2015 and 2018 into Ms Excel on confirmed malaria cases and Out Patients Department (OPD) per month by sub-County. Tororo district is located in Eastern Uganda. Malaria prevalence was computed per 1000 OPD 1000 cases. The QGIS software was used to generate prevalence maps over time.

Results
The highest number of cases (36541) was observed in 2017 and lowest (14745) in 2018. The highest prevalence of 232/1000 OPD cases occurred in 2017. June registered the highest numbers of cases in 2016 (6332) and 2018 (2477). Highest malaria cases occurred between April and July every year. Osukuru sub-county recorded the highest malaria prevalence of 113 and 92 per 1000 OPD in 2016 and 2017 respectively. The 21 sub-counties registered varying prevalence over the 3 years. Urban divisions in Tororo Municipality registered less than 20 cases per 1000 OPD.

Conclusion
The results from this research can be used to inform the health sectors of when say the next highest prevalence of the malaria is most likely to happen and they can use this information to prepare for the increase in the number of cases by stocking enough drugs for treatment of people who will get infected.These results can also be used to reach the people who need the most immediate attention because the results show which particular say sub counties are more affected. Services therefore will be extended to those who need it mostThe results also can enable policy makers to identify which particular district needs extension of control strategies such as supply of Insecticide treated mosquito nets, Residual Indoor spaying and many others as this will ensure effective control of the disease.In terms of performance Vs time, the results can also enable government check the performance of implemented programs and how to better them in future and which place to consider or give more support because these performances can easily be detected from the mapping results in terms of reduction of disease over time.

A Bayesian Statistical Approach to Investigate Blood Pressure Cases

Uchenna Ogoke

University of PortHarcourt, Nigeria

This research is aimed at applying the Bayesian approach in the health sciences. The data used was obtained from the health clinic of one of the Nigerian universities on their freshly admitted students. The blood pressure (BP) of each student was checked and the results produced. We wish to check the sensitivity of the diagnostic test. This is because some of the students claim to be wrongly diagnosed with high BP. We are therefore interested in finding the probability that a positive high BP result reflects the true health status of the students. The Baye's Rule is used to analyse the data which is arranged in a contingency table. Thereafter, a Receiver Operating Characteristic (ROC) Curve is produced to identify the true positive rate and the false positive rate of the considered cases using SPSS. The result shows that the probability that a student tests positive given that he/she has high BP is 0.65.

Investigation of the effects of maternal related factors on infant and child mortality in Zimbabwe

'Morelearnings Sibanda [2]Eustatius Musenge
'University of Witwatersrand [2]University of Witwatersrand

Background
Infant and child mortality is a great area of concern in Sub Saharan Africa and poverty strictness in most of the countries in this region isn't helping the case. Zimbabwe is one such country and most of its population resides in a rural setup were health service delivery is not as efficient. A study was done to investigate the effect of maternal related factors on infant and child mortality in Zimbabwe to draw results which might help in making informed recommendations on interventions.

Methods
Demographic and Health Survey data on infant and child mortality captured from January 2011 to October 2015 was used for this study. The outcome variable was child death named Child-alive, with responses yes if the child is alive and no if the child died. The study targeted babies (dead or alive) from 0 months to 60 months born during the specified period. A sample of 6132 babies from Zimbabwe was use and both rural and urban communities from the country's 10 provinces were represented. Stata version 15 was used to extract the variables of interest from the full dataset. Cleaning of the data was done, and data analysis was then done, and the results were summarised in a table. The Pearson Chi-square test was used for comparisons between categorical factors and the outcome variable, then for continuous variables t-test was used to check if the explanatory variables were related to infant and child mortality.

Results
A total of 325 babies, a constitution of 5.50 % of the total population of babies died during the mentioned period. Factors such as birth-type ($p < 0.001$), postnatal check ($p = 0.003$), mother's education ($p < 0.001$), mother's first birth ($p = 0.050$), mother's residence ($p = 0.007$), baby birth-weight (0.0054) and mother's number of antenatal care visits ($p = 0.0013$) were significantly related to infant and child mortality.

Conclusion
Health service delivery might prove crucial in reducing infant mortality because the factors which seemed related to child death are linked to failure to get adequate and efficient health checkups and other related services. Maternal education is also crucial in reducing infant and child mortality, so mother's education on maternal health might be necessary to help reduce the number of children lost due to lack of knowledge on how to better take care of themselves and the baby

Alcohol abuse OPD cases in Uganda by region: A 2012 2017 A time series analysis

[1]Nazarius Mbona Tumwesigye, [2]Agaba Katureebe
[1]Department of epidemiology and Biostatistics [2]Department of epidemiology and Biostatistics

Background
Uganda is among countries with highest per capita alcohol consumption in Africa. Alcohol abuse is frequently reported in the national health information system. In 2009 2010 over 120 people died due to alcohol abuse. This work aims at analyzing a trend of alcohol abuse cases reporting at outpatient department of all facilities in Uganda from 2012 to 2017. Beside establishing the trend by region this work determines the seasonality of alcohol related OPD cases in the country.

Methods
The data are extracted from the national DHIS II data base to which all health facilities transmit monthly reports by electronic transfer system. STATA's arima function was used to fit linear models with autoregressive moving average (ARMA) models. There is a worldwide concern over non use of accumulated data in health management information systems in different countries. This paper attempts to address this concern.

Results
Preliminary results show that there is an increasing trend in number of alcohol abuse OPD cases in facilities in all regions except the eastern region. Northern region has the highest rise in OPD cases. Major high points are around July October and end of year while major low points are January to April 2019. There is strong correlation between each monthly OPD cases and the previous month cases (>0.6).

Conclusion
Seasonality terms were not significant (P>0.05). Possible causes of the spike in July October could be the harvest season when there is plenty of food to sell and get money and cereals to make alcohol End of year is also a festive season that leads many to alcohol abuse. This supports requests to stock relevant drugs readiness for the spike in alcohol abuse cases during the identified periods. The results are supported by previous studies and this adds weight to a campaign to increase use of accumulated data for planning of health services. Further work on models will add more variables and identify regional and gender differences.

Multilevel Logistic Regression for Patients' Satisfaction in Ethiopian Public Hospitals

'Anteneh Yalew [2]Hiwot Hailu, [3]Adinew Desale, [4]Muluken Azage, [5]Habtamu Asrat, [6]Hiwot Abebe, [7]Sisay Kebede, [8]Daniel Dejene, [9]Adisu Kebede, '0Ebba Abate

'Department of Statistics' College of Natural and Computational Sciences' Addis Ababa University' Addis Ababa' Ethiopia' and Division of Epidemiology and Biostatistics' Department of Global Health' Faculty of Medicine and Health Sciences' Stellenbosch University' Cape Town' South Africa [2]Ethiopian Public Health Institute, [3]ILEX Biotech Ltd' CRO Ethiopia, [4]School of Public Health' College of Medicine and Health Sciences' Bahir Dar University' Bahir Dar' Ethiopia, [5]Ethiopian Public Health Institute' Addis Ababa' Ethiopia , [6]Ethiopian Public Health Institute' Addis Ababa' Ethiopia , [7]Ethiopian Public Health Institute' Addis Ababa' Ethiopia , [8]Ethiopian Public Health Institute' Addis Ababa' Ethiopia, [9]Ethiopian Public Health Institute' Addis Ababa' Ethiopia, '0Ethiopian Public Health Institute' Addis Ababa' Ethiopia

Background

Knowing customers' level of satisfaction is relevant to improve and provide quality health care services. In a clinical laboratory, monitoring customers' satisfaction is an important indicator of the quality management system and required by international laboratory standards. However, in Ethiopia, there has not been baseline data related to the satisfaction level of patients with laboratory services at the national level. The aim of the study was to assess patients' satisfaction level with laboratory services at public hospitals in Ethiopia.

Methods

A national survey was conducted using an institutional based cross-sectional study. The study was employed from 01 to 30 November 2017. A total of 2,399 patients were selected randomly from 60 public hospitals in Ethiopia using stratified random sampling. A multilevel binary logistic regression model was fitted to identify predictors of patient's satisfaction with the services.

Results

Overall, 78.6% of the patients were satisfied with the clinical laboratory services. They were dissatisfied with cleanness of latrine (47%), long waiting time (30%), clear and understandable advisory service during specimen collection (26%), adequacy of waiting area (25%), latrine location (20%), the easy accessibility of laboratory (19%), availability of requested service (18%), unfair payment of service (17%) and missing of result (12%). The frequency of visit for more than two visits with reference of one visit to the hospital) (95% CI AOR: 1.53(1.15, 2.02), distance to a hospital (95% CI AOR: 1.003 (1.002,1.004)), needle stick attempts (P=0.00) were significantly associated with patients' dissatisfaction.

Conclusion

patients were satisfied with the service provided by public hospital laboratories in public hospitals in Ethiopia. However, they were dissatisfied with the accessibility of sites, adequacy of waiting area, cleanness of latrine, long TAT, communication, missing of results, availability of requested service and cost of service. Therefore, responsible bodies in each level should act on the identified gaps and improve the need of clients in each hospital laboratory. In addition, all hospital laboratories may conduct a satisfaction survey and meet the needs of patients.

Factors affecting child malnutrition in Ethiopia

'Ashenafi Yirga [2]Henry Mwambi, [3]Sileshi Melesse, [4]Dawit Ayele
'University of KwaZulu-Natal [2]University of KwaZulu-Natal, [3]University of KwaZulu-Natal, [4]The Johns Hopkins University

Background
One of the public health problems in developing countries is child malnutrition. A principal factor for children's well-being is good nutrition. Therefore, the malnutrition status of children under the age of five is an important outcome measure for children's health. This study uses the proportional odds model to identify risk factors associated with child malnutrition in Ethiopia using the 2016 Ethiopian Demographic and Health Survey data.

Methods
Based on weight-for-height anthropometric index (Z-score) child nutrition status is categorized into four levels namely - underweight, normal, overweight and obese. Since this leads to an ordinal variable for nutrition status, an ordinal logistic regression (OLR) proportional odds model (POM) is an obvious choice for analysis.

Results
The findings and comparison of results using the cumulative logit model with and without complex survey design are presented. The study results revealed that models fitted by considering the complex nature of the design produce more appropriate parameter estimates and standard errors. It has also been found that for children under the age of five, weight of a child at birth, mother's age, mother's BMI, marital status of mother and region (Affar, Dire Dawa, Gambela, Harari and Somali) were influential variables significantly associated with under five children's nutritional status in Ethiopia.

Conclusion
This research revealed that age of a child, sex, weight of child at birth, mother's BMI and region of residence were significant determinants of malnutrition of children under five years in Ethiopia. The effect of these determinants can be used to develop strategies for reducing child malnutrition in Ethiopia. Moreover, these findings clearly justify that OLR proportional odds model is appropriate in assessing the determinants of malnutrition for ordinal nutritional status of under five children in Ethiopia.