ISMS Bulletin

August 2022

Official Publication of
Indian Society for Medical Statistics (ISMS)

Editorial Office
Department of Biostatistics
National Institute of Mental Health and Neuro Sciences
Bengaluru - 560 029

Phone : +91-80-26995119, 5120

email:isms.editor@gmail.com

ISMS Bulletin
August 2022

Official Publication of
Indian Society for Medical Statistics (ISMS)

Regd. office: Dept. of Biostatistics, AIIMS
New Delhi-110029, India

Editorial Office
Department of Biostatistics
National Institute of Mental Health and Neuro Sciences
Bengaluru - 560 029

Phone : +91-80-26995119, 5120

email:isms.editor@gmail.com

Contents

Message from the President

Respected Members,

Greetings. As I complete my second year as President of ISMS by the end of this year, I wish to thank the members of Indian Society for Medical Statistics(ISMS) for electing me as the President. This opportunity provided me a chance to work closely with Dr. P. Venkatesan, former President with the capacity as President elect.

We together with the Secretary Dr. Sreekumaran Nair, were able to organize trend setting Pre-Conference workshops. I am sure that the trend will be followed in terms of quality of teaching in future as well.

Though the COVID-19 pandemic has toppled our face to face meeting, the great institutions NIMHANS, Bengaluru with Dr. Thennarasu as a leader and Ranchi Institute of Medical Sciences (RIMS) with Dr. S.B. Singh as a leader who have contributed immensely and upheld the qualities of the Pre-Conferenceworkshops. The quality of presentations of the Plenary session have improved significantly as well. These are the opportunities to build capacity in the society and we need to thrive to improve every year.

Capacity building (continuing education) should go on for the young members of the society. There are new developments in Bio-statistical methods and they need to be imparted. Today we have many institutions training students in Masters in Bio-statistics. At this juncture, beyond curriculum, we need to impart them the practical challenges and promoting the use of free source software etc.

The website is getting renovated with good objectives. This is going to serve asa good platform for disseminating scientific activities. I am sure the students and young members will be benefitted. I invite each one of you to participate in the activities and in the growth of ISMS.

Finally, I would like to appreciate and thank the Organizing Committee members of ISMSCON-2020 and 2021, Governing council, Special Committee Chairs and members for their constant support and encouragement.

Wishing you a good and successful conference in 2022, and looking forward to seeing you all in KIMS, Karad.

Regards,
Dr. L. Jeyaseelan
President, ISMS

Message from the General Secretary- ISMS

Dear Friends,

Happy to understand that the next issue of ISMS Bulleting is getting released and many members have contributed to the issue. Most likely this could be a step towards ISMS Journal. Congratulations to the Editor and Editorial team for the hard work.

The second announcement of the ISMS conference 2022 which is to be held at Krishna Institute of Medical Sciences, Karad has been circulated. We will be having physical meeting after a gap of two years. Please register for the conference at the earliest and have a good participation this year.

Dr. N. Sreekumaran Nair
General Secretary

Editor's Desk

Dear ISMS Members,

A warm Greetings to all ISMS members! The COVID pandemic has made the scientific community to a difficult situation with recurrent influx of new variants of the virus. We have relatively well managed the pandemic with the availability of vaccine from Government of India and dedication of our healthcare workers. This also provided environments for online meetings and newer learning platforms. This was evident from our ISMSCON2021 in hybrid mode, which was successfully conducted. On behalf of the editorial team, I sincerely appreciate all members for their effort in bringing this ISMS bulletin. The raised level of quality of contribution in the form of original research articles, technical notes are outstanding. We invite each one of you to contribute any significant content related to biostatistics and medical statistics, including opportunities, news and events to our members of the society.

This issue of the bulletin provides wide area of material required for medical statistics readers. We are happy to include unique global positions available in the form of fellowships for our members. I take this opportunity to thank our editorial board members for their support in bringing this August 2022 issue of the bulletin.

I appreciate the services of our senior members for their guidance and support.

Dr. K Thennarasu, *PhD*, *PDF*, *FSMS*
Editor, *ISMS Bulletin*

MED AI: A GRAPHICAL USER INTERFACE (GUI) FOR TRAINING MACHINE LEARNING MODELS FOR HEALTHCARE

Tejinder Singh[1], Dr Kamal Kishore[2]

1. JRF (Under ICMR project), Department of Biostatistics, PGIMER, Chandigarh
2. Assistant Professor, Department of Biostatistics, PGIMER, Chandigarh

Abstract

Machine learning has the potential to transform the healthcare sector; however, the lack of technical expertise hinders the development of advanced ML models for healthcare applications. Due to a shortage of technical expertise, healthcare is impervious to rapid technological advances. Coding an ML algorithm is a complex task for a healthcare researcher; it consumes alot of time and energy that can otherwise be utilized to find innovative solutions to complex health care problems. To facilitate researchers in developing ML models without writing the code, we developed GUI-based software: Med AI—a Graphical User Interface application that does not require coding knowledge to run ML models. The healthcare researcher can import data into the program; subsequently, researchers can train and test various ML algorithms through the GUI of the "Med AI" software. The health researchers can focus on researching AI's application in healthcare rather than the coding em—A major obstacle.

Background

Machine Learning (ML) and Artificial Intelligence (AI) are the engines of growth in the 21st century; the advances in the same have accelerated and expanded the development of the digital ecosystem. ML and AI have penetrated many domains of personal and professional life—personal assistance, self-driving cars, smart recommendations, and personal healthcare, to name a few. The academic community also benefits from the development of the digital eco—system it facilitates collaboration, communication, citation, and sharing. The digital ecosystem has also played a decisive role in combating the unprecedented challenge of COVID-19, wherein ML tools and techniques have helped monitor, manage, contain, and expedite drug development (1).

The role of ML and AI in healthcare research is invaluable; the integration of the same with medical data helps in the early diagnosis of severe diseases and the detection of people with a high risk of developing infections (2). Large tech companies have started focussing on the applications of ML and AI in healthcare. The progress, however, is slow to move into the mainstream due to technical requirements and the sensitive nature of data. Advanced computer knowledge and coding skills are the fundamental requirements to apply ML and AI; these requirements pose a significant challenge for healthcare researchers (3). Staying updated with the rapidly evolving landscape of AI is another challenge. Besides this, the ethical, legal, and administrative permissions also impede the growth and adoption of AI in healthcare.

Many commercial ML and AI tools and techniques such as BigML, Microsoft Azure ML studio, Google cloud AutoML, and RapidMiner are available to the researchers. The utilization of commercial tools and rapid adoption of ML and AI in healthcare is affected by the shortage of funds and is more so for developing and under-developed countries(4)—paradoxically, these countries require more funding for improving health. Therefore, we developed Med AI—free to use software. The researcher can use and circulate the Med AI under GNU General Public License. The non-commercial and coding-free attributes—twin advantages of Med AI are expected to facilitate the rapid adoption of ML and AI in patient care research.

Med AI—a Graphical User Interface (GUI) is an application that does not require researchers to have coding knowledge to run ML models. The healthcare researcher can import data into the program; subsequently, researchers can train and test various ML algorithms. The health researchers can focus on application to healthcare rather than coding—A major obstacle. The users can access and download the software from the Med AI web page; a brief technical manual is also available on the website. A detailed research article explaining the functioning of the software is in progress.

Software Link: *Med AI (https://sites.google.com/view/ml-for-healthcare/home)*

Programming Language– Python 3

Open-source Packages– Scikit-learn (5), Pandas (6), Matplotlib (7), Pandastable (8).

The software performs three broad tasks

1. Data preprocessing
2. Training machine learning models
3. Testing and evaluation of trained models

Data preprocessing

Data is loaded as a Pandas *"DataFrame"* object. Data preprocessing functions are implemented using Pandas and the scikit-learn library. Functions included are Feature Scaling, Feature Encoding, Missing Values Imputation, and Feature Selection. Details for data preprocessing are given below:

1. **Feature Scaling** –This includes functions for data normalization and standardization.
2. **Feature Encoding** – One-hot encoding and label encoding techniques are the available options to the user for encoding categorical features.
3. **Handling Missing Values** – The software provides two options

- Drop NaN values- The user can delete the rows or the columns containing the missing values.
- Fill NaN values. Missing values can be filled using mean, median, backward, and forward selection methods.

4. **Feature Selection** – Selecting the most informative features from the dataset boosts the performance of the ML algorithms. The software provides three feature selection methods

- SelectKBest – Select the k highest scoring features; based on the scoring function used.
- Variance Threshold – Remove features that do not meet the variance threshold value.
- Recursive Feature Elimination– Removes the lowest features one by one by recursively training the ML algorithms; users can select one out of 12 classifiers and 25 regression algorithms.

Training Machine learning models

When the data is ready for training, the user can select the algorithms of their choice to train. The software provides two types of machine learning algorithms–**Classifiers** (24 classifiers from the scikit-learn library) and **Regressors** (40 regression algorithms from the scikit-learn library). Hyperparameter optimization technique **"GridSearchCV"** can be used to select the best hyperparameter values automatically. Coding is not required to train and test the ML models—the data stays locally private to the researcher. All the algorithms are trained locally on the institute server, providing complete data privacy and security.

Testing and evaluation of trained models

For evaluation, the trained model is tested on the test dataset, and the performance output of the trained model is evaluated on various evaluation metrics. Metrics used for assessing the trained models are-

1. **Classification** - F1 Score, Log Loss, Accuracy, Average Precision, AUC.
2. **Regression** - Mean Squared Error, Mean Absolute Error, Explained Variance Score, Max Error, Mean Squared Log Error, R2 Score.
3. **Graphical Representation** - ROC-AUC Curve, Precision-Recall Curve, Bar Graph, Line Graph, Histogram, Scatter Plot.

The graphs and results can be exported and saved to the local drive. The software saves the trained models as *"pickle"* files to the local repository, which the user can use later for making predictions on future data.

Conclusion

ML and AI are indispensable in the 21st century—medical science is no exception. Medical researchers face many technical and regulatory hurdles besides the challenge of coding algorithms. However, ML and AI have become fundamental research components in the 21st century. Therefore, we developed Med AI GUI—A no-code ML for healthcare research. We hope that "Med AI" will help democratize the use of ML in the healthcare domain—providing the impetus for quick and effortless experiments with different machine learning algorithms.

References

1. Syeda HB, Syed M, Sexton KW, Syed S, Begum S, Syed F, et al. Role of Machine Learning Techniques to Tackle the COVID-19 Crisis: Systematic Review. JMIR Med Inform. 2021 Jan 11;9(1):e23811.
2. Butryn B, Chomiak-Orsa I, Hauke K, Pondel M, Siennicka A. Application of Machine Learning in medical data analysis illustrated with an example of association rules. Procedia Comput Sci. 2021;192:3134–43.
3. Faes L, Wagner SK, Fu DJ, Liu X, Korot E, Ledsam JR, et al. Automated deep learning design for medical image classification by healthcare professionals with no coding experience: a feasibility study. Lancet Digit Health. 2019 Sep;1(5):e232–42.
4. Kurian OC. Staff shortages, fund crunch/delays imperil India's healthcare delivery system [Internet]. 20. Available from:https:// policycommons.net /artifacts/1346989/staff-shortages-fund-crunchdelays-imperil-indias-healthcare-delivery-system/1959136/
5. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python [Internet]. arXiv; 2018 [cited 2022 Jun 2. Available from:http://arxiv.org/abs/1201.0490
6. Reback J, Jbrockmendel, McKinney W, Van Den Bossche J, Augspurger T, Roeschke M, et al. pandas-dev/pandas: Pandas 1.4.2 [Internet]. Zenodo; 2022 [cited 2022 Jun 20. Available from: https://zenodo.org/record/6408044
7. Hunter JD. Matplotlib: A 2D Graphics Environment. Comput Sci Eng. 2007;9(3):90–5.
8. Farrell D. DataExplore: An Application for General Data Analysis in Research and Education. J Open Res Softw. 2016 Mar 22;4(1):9.

LATENT GROWTH MODELING: AN APPLICATION FOR MODELING CD4 COUNTS OF HIV PATIENTS TREATED UNDER RANDOMISED CONTROLLED TRIAL

Mahalingam Vasantha[1]*, Chinnaiyan Ponnuraja[1], Adhin Bhaskar[1], Perumal Venkatesan[1]
[1]Department of Statistics, ICMR-National Institute for Research in Tuberculosis, Indian Council of Medical Research, Chetpet, Chennai 600031, India

*Corresponding author
Email: drvasantha.m.icmr@gmail.com
Source of financial support in the form of grants: We didn't receive any funding for this study

Abstract

Latent growth models (LGMs) have been employed in structural equation modeling (SEM) to evaluate the pattern of growth curve over time. SEM improves longitudinal data analysis by including growth of latent variable over time when individual and group changes modelled through slopes and intercepts. The aim of this study is to estimate growth trajectory of CD4 counts of HIV patients by applying LGM. A total of 343 HIV infected antiretroviral therapy (ART) naïve patients without active TB who were part of preventive therapy for TB under randomised controlled trail at ICMR-National Institute for Research in Tuberculosis, Chennai, Tamil Nadu were formed the database for this study. The patients with CD4 counts available at least any three time points from starting stage of treatment to 24th month were considered for analyses. The changes of CD4 counts during the treatment period were found to be nonlinear. The conditional Quadratic LGM by adding age and gender as covariates were fitted and identified as good fit. The Quadratic LGM model identified the changes in CD4 counts of the HIV patients during preventive therapy of TB over different time points.

1. INTRODUCTION

Repeated measurements are widely used in medical field to test temporal changes in individuals or groups. There are two statistical traditions used for studying these changes which are referred as growth curve analysis. One approach to study the growth curve is the statistical equation modeling of latent growth curves and the other one is longitudinal study of multilevel modeling. Latent growth models (LGMs) have been employed in structural equation modeling (SEM) to evaluate the pattern of growth curve over time. LGM represent repeated measures of dependent variables as a function of time and other measures[1]. SEM improves longitudinal data analysis by including growth of latent variable over time when individual and group changes modelled through slopes and intercepts[2,3,4]. The traditional repeated measures of analysis of variance are considered as a special case of general LGM which resemble confirmatory factor analysis[1].

Latent growth curve analysis explains two different components. The first step is to find whether there is a linear or non-linear trend in the repeated measures of each individual across time. The second step is to determine the difference in growth from a baseline using the individual's parameters (slope and intercept values)[5].The parameters of individual are employed as latent variables in LGM to model the changes over time.

A prior knowledge about the functional form of the growth is required for modeling the process of change which explains the changes for a sample. LGM may be also called as mixed models which are mixture of fixed and random effects. The fixed effects are the average slope and intercept for the group. In random effects, the variability around the mean slope and mean intercept from individuals' slopes and intercepts of individuals within the group are also estimated. These models are example of hierarchical linear models with two levels in the hierarchy. The level 1 represents the repeated values over time whereas the level 2 represents the individual within which the values are nested[6]. The purpose of the current study is to provide the concept of LGMs and its application for modeling CD4 counts of HIV individuals.

2. METHODS

2.1 Linear Latent Growth Model

The LGM is often comprised of two factors to represent the aspects of change. The intercept factor is the level of the outcome measure, where time variable equals zero and the slope factor is the linear rate at which the outcome measure changes. The LGM is written as

$$y = \tau_y + \Lambda_y \eta + \varepsilon \quad (1)$$

where,

y = p x 1 of vector of repeated measures,

τ_y = p x 1 vector of intercept,

η= m x 1 vector of dependent latent growth factor that contains scores on m factor for a given individual,

ε= px1 vector of residual errors,

Λ_y= p x m vector of factor loading representing the hypothesized a prior growth pattern of y.

The intercept term τ_y is fixed as zero for model identification reasons[7]. When m =2, the LGM for a single individual *i* on a variable of interest y_i can be written as linear function of time (t) and defined with two components namely the measurement model and latent model. The measurement model is

$$y_{it} = \eta_{0i} + \eta_{1i} t_{it} + \varepsilon_{it}, \text{ i = 1,2,3...n, and } t = 1,2,3...T \quad (2)$$

where,

y_{it} = the observation of individual *i* at the time point t,

η_{0i} = intercept of individual i which is the expected value of y_{it} at the time of origin, η_{1i} = regression coefficient for individual i

ε_{it} = measurement error for individual i at the time point t,

n = number of observations,

T = number of measurements.

Latent model is

$$\eta_{0i} = \alpha_0 + \zeta_{0i},$$
$$\eta_{1i} = \alpha_1 + \zeta_{1i},, i = 1,2,...n, \quad (3)$$

where, the latent variables η_{0i} and η_{1i} are referred as random coefficients, α_0 and α_1 are expected values (fixed part of model) of latent growth factors ζ_0 and ζ_1 respectively. The residuals ζ_0 and ζ_1 are random variables which form individual growth.

In general, the basic latent growth curve model can be written as

$$\left.\begin{aligned} y &= \Lambda\eta + \varepsilon \\ \eta &= \alpha + \zeta \end{aligned}\right\} \quad (4)$$

$$\Lambda = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ . & . \\ . & . \\ . & . \\ 1 & T-1 \end{bmatrix} \quad (5)$$

where,

y = T x 1 vector of measured variables,

η = 2x1 vector of the latent variables,

ε = T x 1 vector of measurements errors,

α = 2 x 1 vector of expected values of η.

2.2. Building Latent Growth Modeling

The stages of developing LGMs are 1) model specification, 2) model identification, 3) model estimation, 4) testing and evaluating overall model fit 5) doing diagnostics of parameters of the model. Repeat the steps 1-5, when the model fits poor, occurrence of non-significant parameter or model modification needed.

2.3. Model fit

The fit indices used in SEM other than χ^2 test are, the root mean square error of approximation (RMSEA), Tucker Lewis index (TLI), comparative fit index (CFI), and standardized root mean square residual (SRMR). The adequate fit is defined as CFI>0.9, TLI>0.9, RMSEA<0.08[10,11].

2.4. Quadratic Growth Curve Model

The quadratic growth model is

$$y_{it} = \eta_{0i} + \eta_{1i} t_{it} + \eta_{2i} t_{it}^2 + \varepsilon_{it} \quad (6)$$

where, the intercept and linear components, η_{0i} and η_{1i} are predicted values. The quadratic

component, η_{2i}, indicates acceleration in growth. t_{ij} is the time i corresponding to each measurement for person j. ε_{it} contains random measurement error with the assumption of normality and zero mean. The second derivate of the equation (6) with respect to time, is the rate of change in the linear component for a one unit change in time[12,6]. The three latent variables for the quadratic model are

$$\left.\begin{aligned} \eta_{0i} &= \mu_{\eta 0} + \gamma_0 x_i + \zeta_{0i}, \\ \eta_{1i} &= \mu_{\eta 1} + \gamma_1 x_i + \zeta_{1i}, \\ \eta_{2i} &= \mu_{\eta 2} + \gamma_2 x_i + \zeta_{2i}, \end{aligned}\right\} \quad (7)$$

The squares of loadings of the linear latent variable are the corresponding loadings for the square latent variable in quadratic LGM.

2.5. Application to HIV Clinical Trial Data

A total of 343 HIV infected[13], antiretroviral therapy (ART) naïve patients without active TB who were part of preventive therapy for TB under randomised controlled trail at ICMR-National Institute for Research in Tuberculosis, Chennai, Tamil Nadu were formed database for this study. For the 343 patients, CD4 counts were available at least any three time points and missing at random from the starting point of treatment to 24th month at 6 month time intervals. The aim of the current study is to evaluate growth pattern of CD4 counts of the HIV patients by applying LGM. For easy convergence of model, the CD4 counts at different time points were divided by 100. Conditional LGMs were fitted to define the pattern of growth curve of CD4counts of the HIV patients during the period of treatment, where age and gender added as time invariant covariates. The maximum likelihood method was used to fit LGMs in the M*plus* version 7.1[14]. The observed repeated measures CD40m, CD46m, CD412m, CD418m, and CD424m are CD4 counts of the patients during treatment period at every sixth month. For the first iteration, the slopes had been coded as 0, 2, 4 and 6 to create linear trend[5]. Assume the first factor loading as zero to find out average at starting time point[15]. The intercepts had been taken as 1 to indicate averages for various months. The factor loading to estimate the quadratic LGM[12] was given in the below matrix

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \\ 1 & 6 & 36 \end{bmatrix} \quad (8)$$

3. RESULTS

Among the 343 patients, 120 (35%) were males, the mean age was 29.17 years (range: 18-60 years). The scatter line diagram of CD4 counts of the HIV patients over different time points is given in Figure 1

Table 1 Parameter estimates of quadratic LGM for percentage of CD4 counts

Parameter	Estimate	S.E	p value
Regression weights			
Intercept (mean)	3.607	0.127	<0.001
Slope (mean)	-0.269	0.195	0.167
Quadratic (mean)	0.229	0.089	<0.05
Intercept on			
Age in years	-0.044	0.021	<0.05
Gender	-0.924	0.280	<0.005
Slope on			
Age	0.003	0.032	0.926
Gender	0.971	0.431	<0.05
Quadratic on			
Age	-0.006	0.015	0.698
Gender	-0.454	0.196	<0.05
Variances			
Intercept	2.638	0.344	<0.001
Slope	0.233	0.127	0.066
Covariance			
Slope with intercept	-0.043	0.173	0.803
R Square			
CD40m	0.564	0.053	<0.001
CD46m	0.696	0.037	<0.001
CD412m	0.670	0.032	<0.001
CD418m	0.735	0.032	<0.001
CD424m	0.777	0.042	<0.001
Intercept	0.116	0.047	<0.05
Slope	0.486	0.251	0.053

The changes of CD4 counts during the treatment period were found to be nonlinear.

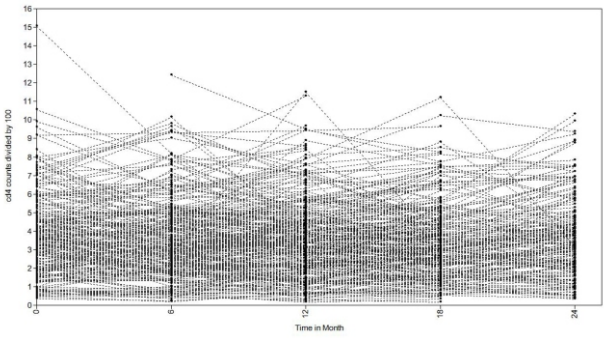

Figure 1 Scatter diagram of CD4 counts of HIV patients versus time in month

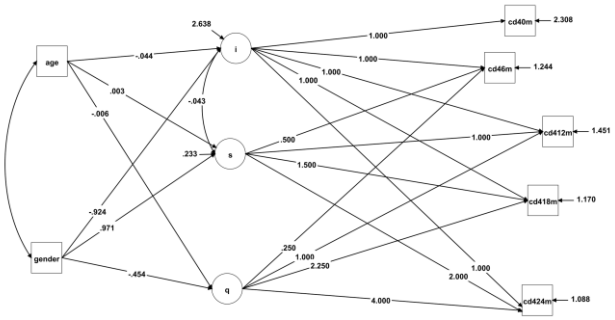

Figure 2 Quadratic growth model of percentage of CD4 counts in HIV patients over time

Legend: i - intercept of quadratic LGM of CD 4 count changes in HIV patients, s - slope of quadratic LGM of CD4 counts differences in HIV patients, q - quadratic factor of LGM of CD4 count differences in HIV patients, cd40 – CD4 counts of HIV patients at the starting stage of treatment, cd46m - CD4 counts of HIV patients at 6th month, cd412m - CD4 counts of HIV patients at 12th month, cd418m - CD4 counts of HIV patients at 18th month, cd424m - CD4 counts of HIV patients at 24th month

The conditional Quadratic LGM by adding age and gender as covariates were fitted and identified as good fit for identifying growth trajectories of Cd4 counts of the HIV patients (RMSEA=0.061, CFI = 0.976, TLI = 0.963) by fixing the variance of quadratic effect as zero. For this model, a significant quadratic effect (0.229, $p<0.05$) was identified as well as a significant intercept (3.607, $p<0.001$) where the slope effect was not found to be significant. The gender was found to be statistically significant in the difference of CD4 counts over time on the latent factors intercept, slope and quadratic effect (-0.924, $p<0.005$; 0.971, $p<0.05$; -0.454, $p<0.05$). The patients' age was negatively associated at the intercept level (-0.044, $p<0.05$) but not in slope and quadratic effect level (Table 1). The path diagram of the quadratic model of CD4 counts over different time points is given Figure 2.

4. SUMMARY

LGM is a power tool in SEM modeling to analyze dynamic changes. A LGM can be represented as a special case of SEM. The benefit of LGM is the usage of latent repeated measures that is not used in other growth curve models[15]. In the current study, LGM was found to be very helpful to identify nonlinear change of parameter estimations.

The Quadratic LGM model identified the changes in CD4 counts of the HIV patients during preventive therapy of TB over different time points. The gender was found to be statistically significant in the difference of CD4 counts over time on the latent factors intercept, slope and quadratic effect. The patients' age was negatively associated at the intercept level but not in slope and quadratic effect level. Further research is needed to study about the extension and application of growth curve models other than quadratic LGM.

References

1. Meredith, W. and Tisak, J. (1990). Latent curve analysis. Psychometrika 1990; 55: 107-122.
2. McArdle, J.J., and Epstein, D. (1987). Latent growth curves within developmental structural equation models. Child Development, 58: 110-133.
3. Stoolmiller, M. and Shyder, J. (2013).Embedding Multilevel Survival Analysis of Dyadic Social Interaction in Structural Equation Models: Hazard Rates as Both Outcomes and Predictors. Journal of Pediatric Psychology, 1–11.
4. Byrne, B.M. and Crombie, G. (2003). Modeling and testing change: An introduction to the latent growth curve model. Understanding Statistics, 2: 177-203.
5. Schumaker, R. E. and Lomax, R. G. (2004). A beginner's guide to structural equation modeling. Lawrence Erlbaum Associates, Inc., Publishers, New Jersy.
6. Llabre, M.M., Spitzer, S., Siegal, S., Sabb, P.G., and Schneiderman, N. (2004). Applying latent growth curve modeling to the investigation of individual differences in cardiovascular recovery from stress. Psychosomatic Medicine, 66: 29 – 41.

7. Preacher, K.J., Wichman, A.L., MacCallum, R.C. and Briggs, N.E. (2008). Latent Growth Curve Modeling. Thousand Oaks, CA: Sage Publications.
8. Bollen, K.A. (1989). Structural equations with latent variables.Wiley, NewYork
9. Tolvanen A. (2007). Latent Growth Mixture Modeling: A Simulation Study. University of Jyväskylä, Department of Mathematics and Statistics, Report 111.
10. Marsh, H.W., Muthen, B., Asparouhov, T., Ludtke, O., Robitzsch, A, et al. (2009). Exploratory structural equation modeling, Integrating CFA and EFA: Application to students' evaluation of university teaching. Structural Equation. Modeling, 16: 439-476.
11. Zudien, M.V, Heijenen, C.J, Van De Schoot, R., Amarouchi, K., Mass M, et al. (2011). Cytokine production by leukocytes of military personnel with depressive symptoms after deployment to a combat-zone: A prospective, longitudinal study. PlosONE 6: e291402.
12. Biesanz, J.C., Deeb-Sossa, N., Papadakis, A.A., Bollen, K.A., Curran, P.J. (2004). The role of coding time in estimating and interpreting growth curve models. Psychological Methods, 9: 30-52.
13. Swaminathan, S., Menon, P.A., Gopalan N., Perumal V., Santhanakrishnan., R.K.,Ramachandran, R., Chinnaiyan, P., Iliayas, S., Padmapriyadarsini C., P., Navaneethapandian, P.D., Elangovan, T., Pho, M.T., Wares, F., and Ramalyengar N.,P. (2012). Efficacy of a Six-Month versus a 36-Month Regimen forPrevention of Tuberculosis in HIV-Infected Persons in India: A Randomized Clinical Trial. PLosONE 7(12): e47400.
14. Muthèn, L.K., and Muthèn, B.O. (1998-2013). MPlus user's guide. Seventh edition. Los Angeles, CA: Muthèn and Muthèn.
15. Bollen, K.A. and Curran, P.J. (2006). Latent Curve Models: A structural equation perspective. John Wiley and Sons, Hoboken, NJ.

Technical Note

APPLICATIONS OF MIXTURE CURE MODELS IN HEALTH RESEARCH

Durai Murukan G, Binu.V.S, B. Binukumar
Department of Biostatistics
National Institute of Mental Health and Neuro Sciences, Bengaluru, Karnataka

Abstract

The commonly used survival models viz. Cox Proportional Hazard (PH) and Accelerated Failure Time (AFT) models assumes that the subjects who are censored will get the event of interest at some point of time as time tends to infinity. However, there are situations where this assumption is violated; other words the study participants won't get the event of interest no matter how long they have been followed. Cure models are preferred to deal with such survival data. This communication aims to provide an introduction to the mixture cure models, their types, and estimation procedures involved. Further, it also provides a demonstration of application of mixture cure models on colon cancer data available in *cuRe* package in R software.

Introduction

Mixture Cure Models refers to a class of survival models, conceptually different from the conventional survival models. Usually, conventional survival model assumes that the subjects who are censored will get the event of interest at some point of time as time tends to infinity [1]. However, there are situations where the study participants won't get the event of interest no matter how long they have been followed. In order to make it clear, let us consider the example of smoking cessation trial in which the study participants are individuals who currently quit smoking, and relapse of smoking behavior is one of the events of interest. In this case there is a possibility that some individuals can quit smoking permanently and for such individuals the event of interest, relapse, never occurs. Similarly, in the context of early detection and treatment of cancer in which the cancer cells have been removed from individuals at the very early stage. In this case there is a chance that some individuals may develop recurrence after a few years whereas some may not experience the recurrence of cancer in their life span. Therefore, in a real-world setting, there is a chance that some participants won't develop the event of interest no matter how long they are followed. This scenario indicates there are mixtures of two groups among the study population [2]. The first group consists of subjects who are at the risk of developing the event of interest at some point of time; whereas the second group consists of those who are not going to get the event of interest. The latter group is usually referred as cured subjects. Existing literature suggests that not only the subjects who are not going to get the event of interest, but also the subjects who survive longer without getting the event can also be considered as cured provided that necessary follow-up time has been considered for the study. Therefore, they can be considered either as cured subjects or long-term survivors [3,4]. When the study population consists of cured and uncured groups, it is ideal to consider only the uncured group in the risk set for the event. However, conventional survival models use the combination of both groups in risk set and estimates the parameters, subsequently, results in biased estimates. In this scenario, the mixture cure model is the most appropriate one. The mixture cure model help to identify variables that have an effect on the probability of cure, as well as helps to find those variables which have an effect on the failure time of occurrence of event under study among uncured subjects. Thus, mixture cure models allow the variables to have different influence on cured subjects and on subjects who are uncured.

Boag(1949) was the first researcher to introduce the mixture cure model [5], based on the proportion of cured patients among those who had received treatment for mouth cancer. Later, in 1986, Farewell introduced the nonparametric mixture model for cure rate estimation [6]. Kuk and Chen (1992) established the mixture model that combined the logistic regression with

proportional hazards regression [7]. Maller and Zhou in (1996) extensively worked on the methodologies in survival analysis with long-term survivors [8]. Moreover, Peng and Dear (2000) proposed a generalized F mixture model for cure rate estimation [9].

Model Formulization

In order to formulize the model, let the random variable T denotes the time to an event of interest and let Y be the cure indicator, where

$$Y = \begin{cases} 1 & \text{if the subject is not cured} \\ 0 & \text{otherwise} \end{cases}$$

Let $P(Y=1)= \pi$ be the proportion of uncured subjects in the population, or it can also defined as the probability of being uncured. Let $S_u(t)=P(T>t \mid Y=1)$ and $S_c(t)= P(T>t \mid Y=0)$ be the survival functions of the uncured and cured population respectively. The above two survival functions are conditional survival functions. By combining the above two survival functions, we can get unconditional survival function for survival time T as

$$P(T>t)=S(t)= \pi S_u(t)+(1-\pi) S_c(t) \qquad (1)$$

Since we have assumed that the subjects in the cured group won't get the event of interest no matter how long we follow them, its survival function $S_c(t)$ will be equal to 1. Therefore, the survival function described in (1) becomes

$$P(T>t)=S(t)= \pi S_u(t)+1-\pi$$

The mixture cure model is a combination of two models, namely the *incidence model* and the *latency model*. The *incidence model* models the probability of cure (or uncured), whereas the *latency model* models the event time for the subjects who are all uncured.

The incidence model is defined as

$$logit\,[\pi(z)\,]= z'\gamma \qquad (2)$$

Equation (2) models the effect of the set of covariates z on π using a logistic link function where γ be the vector of coefficients of the covariates on z, which includes an intercept term. Link functions like log link function can also be used. There are many options to model the effect of covariates on $S_u(t)$. If we consider x being the set of covariates, the effect of x on $S_u(t)$ can be modelled either parametrically or non-parametrically. It should be noted that the variable list in both x and z can be overlapping but not necessarily be the same. The most commonly used model for fitting $S_u(t)$ is the proportional hazard mixture cure model. Accelerated failure time models are another set of models for fitting $S_u(t)$.

Proportional Hazard Mixture Cure (PHMC) Models

The most commonly used model for the effects of x on $S_u(t)$ is based on the proportional hazards (PH) assumption [11]. The latency model based on proportional hazard assumption is given by

$$S_u(t)= S_u(t|x)= S_{u0}(t)^{exp(x')} \qquad (3)$$

where β is a set of coefficients of the x covariates. β either contains or does not contain an intercept term based on the fact that whether baseline survival distribution is specified or not. With conventional semi parametric Cox Proportional hazard (PH) model, we can fit a model without specifying the underlying survival distribution. Further, the model (3) is convenient to use and interpret because of its similarity with the popular Cox PH model. However, exponential or Weibull distribution can also be assumed as the underlying survival distribution since it satisfies the PH assumption.

Accelerated Failure Time Mixture Cure Models

Another popular set of mixture cure model is Accelerated Failure Time Mixture Cure (AFTMC) models. Just like PH models, Accelerated Failure Time (AFT) models can also be used to model the effect of χ on $S_u(t)$ and the survival function of the uncured population becomes

$$S_u(t)= S_u(t|x)= S_{u0}(te^{-x'\beta}) \qquad (4)$$

Unlike the PHMC, an AFTMC model allows crossing hazards and direct interpretation of the effects of x on the *logT* scale. Under AFT assumptions, one can use many distributions as an underlying survival distribution of uncured subjects like Weibull, log-normal, log-logistic, etc. There are other sets of mixture cure models like

Accelerated Hazard Mixture Cure (AHMC) models and Proportional Odds Mixture Cure (POMC) models. However, in terms of popularity and availability of software's PHMC and AFTMC are most commonly used.

Cautions while fitting the mixture cure models

The major problem with the mixture cure models happens because of uncertainty associated with the tail of the baseline survival distribution for the uncured population $S_{u0}(t)$. Since $\pi(z)$ is the parameter of the incidence model, it should ideally explain the cure rate at $t=\infty$. In any study, the follow-up time is never considered as infinite. So, if there is a plausibility of occurrence of more number of events after the longest follow-up then the estimate of $\pi(z)$ won't be reliable. Also, the different choices for $S_{u0}(t)$ can lead to different cure rates. In addition, the variance of the parameter estimates tends to be large for the different choice for $S_{u0}(t)$. Hence, the absolute value of the cure rate should be interpreted cautiously [11].

Estimation Procedure

The Expectation-Maximization (EM) algorithm can be used to obtain the maximum likelihood estimates of the parameters in mixture cure models. It is assumed that censoring which happened in the study is "independent" and "non-informative" [10]. Let $O=(t_i, \delta_i, x_i, z_i), i=1,\ldots,n$ denotes the observed data for the i^{th} individual where t_i is the observed survival time, δ_i is the censoring indicator with $\delta_i=1$ for the uncensored time and $\delta_i=0$ for the censored time, and x_i, z_i are the possible covariates in the latency and incidence part of the mixture cure models respectively.

Let $\Theta=(\gamma, \beta, S_{u0}(t))$ denotes the set of unknown parameters. Let y be the indicator variable that indicates that the i^{th} individual will eventually $(y_i=1)$ or never $(y_i=0)$ experience the event of interest, with the probability of $\pi(z_i)$. Given the information for $y=(y_1,y_2,\ldots,y_n)$ and observed data O, the complete likelihood function can be expressed as below (10)

$$\prod_{i=1}^{n}[1-\pi(z_i)]^{1-y_i}\ \pi(z_i)^{y_i}\ h(t_i \mid Y=1,x_i)^{\delta_i y_i}\ S(t_i \mid Y=1,x_i)^{y_i}$$

where h(•) is the hazard function corresponding to S(•). The logarithm of the complete likelihood function can be written as

$$l_c(\gamma,\beta,S_{u0}(t)|O,y) = l_1(\gamma|O,y) + l_2(\beta,S_{u0}(t)|O,y)$$

$$where,\ l_1(\gamma|\,O,\,y) = \sum_{i=1}^{n} y_i \log[\pi(z_i)] + (1-y_i)\log[1-\pi(z_i)]$$

$$l_2(\beta,\,S_{u0}(t)|\,O,\,y) = \sum_{i=1}^{n} \delta_i y_i \log[h(t_i \mid Y=1,x_i)] + y_i \log[S(t_i \mid Y=1,x_i)]$$

Estimation procedure consists of two major steps namely, E-step and M-step. E-step computes the conditional expectation of the complete log-likelihood with respect to y_i's, given the observed data and current estimates of parameters γ, β, $S_{u0}(t)$. Since the y_i values are not known, expectation of y_i given O and Θ can be calculated as follows

$$w_i = E(y_i|O,\Theta) = \delta_i + (1+\delta_i)\frac{\pi(z_i)S_u(t_i|x_i)}{1-\pi(z_i)+\pi(z_i)S_u(t_i|x_i)}$$

Initially, $w_i=1$ if $\delta_i=1$ and w_i is the probability of uncured patients if $\delta_i=0$.

Consequently, the M-step involves in maximizing both likelihoods $l1(\gamma|\,O,w)$ and $l2(\beta, S_{u0}(t)|\,O,w)$ with respect to the unknown parameters γ, β, $S_{u0}(t)$. The E-step and M-step iterate until a convergence is achieved. Newton-Raphson or Nelder-Mead algorithm can be used to maximize the $l_1(\gamma|\,O,y)$ while maximizing $l_2(\beta, S_{u0}(t)|\,O,y)$ is based on how $S_{u0}(t)$ is parametrized (11). As mentioned, in semi parametric mixture cure model the $S_{u0}(t)$ is non-parametrically specified. In that case $l_2(\beta, S_{u0}(t)|\,O,y)$ can be considered as the Cox PH model's log-likelihood function with $w_1,\ldots,w_n$ as offset values.

$$l_2(\beta,\,S_{u0}(t)|\,O,\,y) = \sum_{i=1}^{n}\delta_i \log[h_o(t_i)\exp(\beta x_i + \log(w_i))]$$
$$+ \log[S_0(t_i)^{\exp(\beta x_i+\log(w_i))}]$$

The above equation can be maximized using existing methods for Cox's PH model [12-16].

Software to fit Mixture Cure models

Although statistical software like R, Stata and SAS are supports to fit different mixture cure models, this communication primarily focuses on the R software and its packages. The R package *"smcure"* can be used to fit the semi-parametric PHMC models [17]. The package "cuRe" can be used to fit parametric mixture cure models [18]. These packages provide wide choice of distributions for the baseline survival distribution for the uncured population.

The dataset named "colon DC" from the R package "*cuRe*" has been used for demonstration purpose. The dataset contains the individual baseline and follow-up data on 15,564 colon cancer patients. It contains baseline information on gender, age, age at the time of diagnosis, clinical stage at diagnosis, etc., along with the follow-up time in years; and status (Dead/Alive) as the event of interest.

The following table gives the summary of the PHMC Cox model with age as a covariate.

Table 1: The Summary of the PHMC Cox model.

Incidence Model					
Variables	Estimate	SE	Odds ratio (OR)	95% confidence interval (CI) of OR	p-value
Age	-0.054	0.002	0.947	0.943, 0.951	<0.001
Latency Model					
Variables	Estimate	SE	Hazard ration (HR)	95% CI of HR	p-value
Age	0.024	0.001	1.024	1.022, 1.026	<0.001

Incidence model observed that, as age increases by one year, the odds of cure (free from death caused by colon cancer) decreased by 5.3%. On the other hand, as age increases by one year, the hazard of death from colon cancer increases by 1.024 times.

Codes for fitting PHMC Cox Model

```
library(smcure)

smcure_obj<- smcure(Surv(FUyear, status)~age, cureform=~age, data=colonDC, model="ph")
printsmcure(smcure_obj)
```

To fit the AFTMC model, the distributional assumption for the underlying baseline survival distribution of the uncured population is essential, and hence we assumed it to have Weibull distribution. The following table gives the summary of the AFTMC model with age as a covariate.

Table 2: The Summary of the AFTMC model

Incidence Model					
Variables	Estimate	SE	Odds ratio (OR)	95% CI of OR	p-value
Age	-0.057	0.004	0.945	0.938, 0.952	<0.001
Latency Model					
Variables	Estimate	SE	Survival time ratio (TR)	95% CI of TR	p-value
Age	-0.024	0.002	0.976	0.973, 0.979	<0.001

Incidence model showed that as age increases by one year, the odds of cure (free from death due to colon cancer) decreased by 5.5%. Latency model showed thatfor every one-year increase in age, the survival time decreases.

Codes for fitting AFTMC Model

```
library ("cuRe")

WeibullAFTcure_model<fit.cure.model(Surv(FUyear, status)~age,formula.surv=list(~age),data=colonDC,link="logit", dist="weibull",type="mixture") summary (WeibulAFTcure_model)
```

Summary

Mixture cure models are the appropriate set of models when a group of study participants has a possibility of being completely cured and shall not develop the event of interest. However, these models should be used with utmost caution as an insufficient follow-up period leads to the model failure in identifying the subjects who will completely cure from the event of interest, which results in a biased estimate for the covariate effect on the cure status and survival time.

References

1. Hosmer D, Lemeshow S, May S. Applied Survival Analysis. Hoboken: Wiley; 2011.
2. Lambert PC. Modeling of the cure fraction in survival studies. Stata J. 2007;7(3):351–75.
3. Arano I, Sugimoto T, Hamasaki T, Ohno Y. Practical application of cure mixture model for long-term censored survivor data from a withdrawal clinical trial of patients with major depressive disorder. BMC Med Res Methodol. 2010;10:1–13.
4. Hosseini SV, Rezaianzadeh A, Rahimikazerooni S,

Ghahramani L, Bananzadeh A. Prognostic Factors Affecting Short- and Long-Term Recurrence-Free Survival of Patients with Rectal Cancer using Cure Models: A Cohort Study. Iran J Med Sci. 2020;45(5):333–40.

5. Boag JW. Maximum Likelihood Estimates of the Proportion of Patients Cured by Cancer Therapy. Journal of the Royal Statistical Society: Series B (Methodological). 1949 Jan;11(1):15-44.
6. Farewell VT. The use of mixture models for the analysis of survival data with long-term survivors. Biometrics 1982;38:1041-6.
7. Kuk AYC, Chen C-H. A mixture model combining logistic regression with proportional hazards regression. Biometrika. 1992;79:531–41.
8. Maller RA, Zhou S. eds. Survival Analysis with Long-Term Survivors. New York: Wiley, 1996.
9. Peng Y, Dear KB, Denham JW. A generalized F mixture model for cure rate estimation. Stat Med 1998;17:813-30.
10. Beesley LJ, Taylor JMG. EM algorithms for fitting multistate cure models. Biostatistics. 2019;20(3):416–32.
11. Klein JP, Van Houwelingen HC, Ibrahim JG, Scheike TH. Handbook of Survival Analysis Handbooks of Modern Statistical Methods. New York: CRC , 2013.
12. Peng Y, Dear KBG. A nonparametric mixture model for cure rate estimation. Biometrics. 2000 Mar;56(1):237-43.
13. Peng Y. Estimating baseline distribution in proportional hazards cure models. Computational Statistics & Data Analysis. 2003 Feb;42(1-2):187-201.
14. Peng Y. Fitting semiparametric cure models. Computational Statistics & Data Analysis. 2003 Jan;41(3-4):481-90.
15. Sy JP, Taylor JMG. Estimation in a Cox Proportional Hazards Cure Model. Biometrics. 2000 Mar;56(1):227-36.
16. Sy J, Taylor J. Standard errors for the cox proportional hazards cure model. Mathematical and Computer Modelling. 2001 Jun;33(12-13):1237-51.
17. Cai AC, Zou Y, Peng Y, Zhang J. Package 'smcure': Fit Semiparametric Mixture Cure Models. Cran. 2015;
18. Jakobsen LH, Clements M, Jensen RK, Gjaerde LK. Package ' cuRe':Parametric Cure Model Estimation. Cran. 2018;

Programs and Events

ISMSCON-2021

39th Annual Conference of Indian Society for Medical Statistics (ISMS)

Organized by the

Dept. of PSM, RIMS, Ranchi

09th to 11th December 2021

A REPORT

The 4 Day Program was started with a Pre-Conference Workshop on 8thof December. Inauguration of the workshop was done in the presence of our honorable Guests, our illustrious Padma shri, Dr., Prof. Kameshwar Prasad, Medical Superintendent, HOD PSM Dr. Vivek Kashyap, Dean Academics Dr. Satish Chandra, Dean Research Akhouri, Dr. Vidya Sagar Prof. PSM, Organizing Secretary Dr. S.B Singh, and other head of departments and senior faculty members.

Pre-Conference workshop.

The hybrid mode pre conference started with a welcome speech from Padma Shri Prof. Dr. Kameshwar Prasad, who shared his research journey and the importance of systematic review and meta-analysis in his research career and inspired the young Indian researchers to take up quality research and publications in high-impact journals.

Prof. Dr. Kameshwar Prasad during his lecture

The workshop discussions began with the presentation by Padma Shri Prof. Dr. Kameshwar Prasad on "Introductory remarks including principles of systematic review and meta-analysis." The talk highlighted the importance of systematic review and meta-analysis over narrative reviews and the impact of such research on the clinical decision-making and guidelines generated by various associations, which are vital in saving millions of patient lives and the practice of medicine. The steps in the systematic review and 3 meta-analyses (6S) were discussed with the real-life examples and research experiences. The session ended with interaction from online and offline participants.

The next session started with another eminent speaker, Prof. Dr Sreekumaran Nair, Professor & Head Dept. of Medical Biometrics, JIPMER, Puducherry, India on the Method of Meta-analysis. The speaker highlighted when and when not to do meta-analysis and why to do meta-analysis. He also highlighted on fixed and random effect models in meta-analysis. The speaker also touched upon the heterogenicity and statistical methods, including forest plots, used in meta-analysis.

In the afternoon session Dr. B. Binukumar, Associate Professor, Dept. of Biostatistics, NIMHANS, Bengaluru, India on Methods of Bias Assessment. The speaker extensively covered biases in the research methodology and explained the risk of bias diagrams made with RevMan software. The final session of the day ended with Prof. Dr. Sada Nand Dwivedi, Former Professor of Biostatistics, AIIMS, New Delhi, India, on Network Meta-analysis. The speaker, with vast experience, explained the need for and various types of network meta-analysis. The speaker explained with a case study of doing network meta-analysis with his hardworking student, Dr. Mona Pathak, and shared all the difficulties and experiences with the audience.

At the end of the workshop Dr. S. B. Singh, proposed the vote of thanks to the dignitaries and participants for all the contributions made for the pre-conference.

ISMSCON-2021:

The next day i.e. 9th December started with the inauguration ceremony of ISMSCON-2021, Themed- Application of Bio statistical Techniques for summarizing Evidence: Systematic Review and Meta-Analysis. The inauguration started with the welcome speech by the cultural committee welcoming our chairman Prof. Dr. Vivek Kashyap, Co- Chairman Prof. Dr. Satish Chandra, Chairman Souvenir Prof. Dr. D.K. Mishra, Guest of honour Prof. N.C. Das, our Director cum Patron Padmashree Prof. Dr. Kameshwar Prasad, our Chief Guest Dr.(Mrs.) Kamini Kumar VC, Ranchi University and Organizing Secretary Dr. S. B. Singh followed by lamp lightning and Sarasvati vandana and was followed by honoring the guests by presenting sapling.

All the dignitaries present shared their views and enlightened the audience with their valuable words.

Prof. N.C. Das

Prof. Dr. Satish Chandra

Prof. Dr. D.K Mishra

Prof. Dr. Vidya Sagar honouring Chief Guest Dr. (Mrs.) Kamini Kumar

Dr. S. B. Singh

The inauguration was ended with the vote of thanks by the organizing secretary Dr. S.B. Singh. After this Prof. S.K. Bhattacharya oration by Prof. Dr N.K Tyagi, Prof. N. Sreekumaran Nair, General Secretary, ISMS Introduced the Orator & Prof. B. L. Verma, Founder General Secretary, ISMS gave a Brief about Prof. S.K. Bhattacharya. Then plenary session started and with the talk of Dr. Abhaya Indrayan, Professor of Biostatistics, And Head (Retd.) Department of Biostatistics and Medical Informatics, Delhi University College of Medical Sciences on the topic ***"Statistical Medicine"*** highlighting the role of statistics in the field of Medicine. Then Dr. Chandrakant Lahariya, Public Policy and Health System Specialist, New Delhi had a talk on the topic ***"Covid19 Vaccines Efficacy And Effectiveness: Opportunity To Bring Medical Statistics To Daily Life*** remembering late Prof Dr. V.K. Srivastava" highlighting the role of statistics in our daily life the session was followed by contributed oral talk and technical session where 42 presentations was given by the participants. The second day of the conference i.e. 10th December started with a brilliant talk by Dr. Debashree Ray all over Johns Hopkins, Maryland (U.S.A) on the topic "***New And Improved Meta-Analysis Approaches To Identify Risk Factors Of Diseases"***followed by plenary talk by Dr. Anil Mathew, Professor, PSG Institute of Medical Sciences and Research Coimbatore on the topic ***"Regression Modeling - Applications Of Ordinal Logistic Regression Analysis In Clinical Research "***. After these talks Professor R.N. Srivastava Award Competition was conducted & was given to After that a talk on the topic ***"Disease Informatics And Its Role In Disease Surveillance And Control"***was given by Dr. Prashant Mathur, Director, National Centre for Disease Informatics and Research Indian Council of Medical Research, Bengaluru. He highlighted that we somehow neglect air pollution & air pollution can be a cause of cancer therefore further studies should be done for more clarity. At the last of the plenary session for the day Dr. M. Vishnu Vardhana Rao, Scientist – G & Director, ICMR-National Institute of Medical Statistics, New Delhi talked on the topic ***"Application Of Meta-Analysis In Health Research And Pitfalls".***After the plenary session technical session was held with 38 presentations made by the participants.

Offline participants during Scientific Programme

On the 3rd day of the conference was started with a plenary talk by our director& CEO RIMS Ranchi Prof. Dr. Kameshwar Prasad on the topic ***"Biostatistical Aspects In Selecting And Summarizing Outcome Measure In Meta-Analysis".***Then Prof Dr. DebasisKundu, Dean of Faculty Affairs Indian Institute of Technology, Kanpur talked on the topic ***"Bayesian Inference Of A Dependent Competing Risk Data".*** Dr. Pankaj Bhardwaj, Vice-Dean Research, Nodal Officer, All India Institute of Medical Sciences (AIIMS), Jodhpur talked on the topic ***"Using Scopus Review And Economic Analysis For Assessment Of E Health Programs"***followed by Dr. Denny John, Adjunct Professor, Ramaiah Institute of Applied Sciences, Bangaluru on the topic ***"Evidence Synthesis: Past, Present And Future"***. At last Dr. Manya Prasad talked about "PUBLICATION BIAS". The conference was concluded with a valedictory function and obtained the feedback from the delegates.

Faculty Development Programme (FDP) on "Bayesian modeling using WinBUGS and R" during 15 – 16 July, 2022 at VIT, Vellore

On the behalf of Department of Statistics, ICMR-National Institute for Research in Tuberculosis, Chennai, Dr. M. Vasantha Conducted two days Faculty Development Programme (FDP) on "Bayesian modeling using WinBUGS and R" during 15 – 16 July, 2022 at School of Electrical Engineering, VIT, Vellore -632014, Tamil Nadu, India. A total of 25 Research Scholars who are pursuing their PhD and the Staff of VIT, Vellore attended the workshop and the hands on training on the software R and WinBUGS.

This workshop provided the concept of Bayesian Statistics which provides an elegant approach to many data science and decision making problems and hands on experience in conducting Bayesian analysis using WinBUGS and R.

The workshop covered:

Day 1:

- Concept of probability, Bayes theorem, Definition of Prior, Types of priors, Markov Chain Mante Carlo, Credible interval
- Types of Distribution Function, their assumptions and properties,
- Simple problems on Bayes theorem, Prior, Prior Precision, Posterior mean & variance
- Writing program on WinBUGS

Day 2:

- Bayesian data analysis using WinBUGS
- Writing program on R,
- Regression model using R, Writing codes for priors in R
- Fitting Bayesian model

FELLOWSHIPS & OPPORTUNITIES

Matsumae Short-Term Research Fellowship in Japan on Trialect

We have a posting soliciting applications for the Masumae Short-Term Research Fellowship in Japan on Trialect. Twenty Applications are accepted from **throughout the world.** Applicants should not have past or current experiences of staying in Japan (other than short-term stays such as for sightseeing or conferences). A monthly allowance is provided to cover expenses for research activities (including materials) and living expenses in Japan in addition to economy class airfare. The fellowship period is for three to six months between **April 2023 to March 2024.**

Click to know more: Matsumae Fellowship
Contact Details:
Sandra Miller
Trialect Support,
support@trialect.com
+1.805.852.1402 (7 AM-11 AM EST)

UJMT Fogarty Global Health Fellowship

The application process for Fogarty Fellowship 2023-2024 will be August – November, 2022 for focused countries including Indian nationals.
https://globalhealth.unc.edu/education/fellowships-and-training-programs/ujmt-fogarty-

Fogarty Global Health Program for Fellows and Scholars consortium with short term training for India and other LMIC for the year 2023. For a list of others, please visit:https://fogartyfellows.org/apply/

Short-term training in U.S. for LMIC trainees

LMIC Fellows are required to participate in a two to three-month training experience at the beginning of their fellowship (immediately following the July NIH Orientation) at their primary U.S. mentor's university. This experience should be tailored to each fellow's specific needs, incorporating structured learning, skills development, networking, and mentoring activities to enhance the trainee's fellowship year. Some of the training options potentially available at our US institutions can be found in the following table (please discuss with your mentor/program the logistics and affordability of participating in any of these opportunities). Please work with your mentor to develop your short-term training plan and identify opportunities including those that may not be mentioned below that are best suited for you and your fellowship timeline. The Support Center will coordinate coverage of fees associated with course attendance or laboratory experience, visas, lodging, and on-site orientation. Please contact the support center (qhfmqr@uw.edu) if you have any questions.

US University	Type of opportunity	Topic	Title	Duration
University of Michigan	Didactic	Epidemiology	Summer Epidemiology Institute	3-week or 1-week courses
		Big data / Data science	Big Data Summer Institute	3-week or 1-week courses
			Michigan Institute for Data Science (MIDAS) Introductory Data Science	5-day workshop
			ISR - Institute for Research on Innovation and Science	2-week long courses
		Survey design	Summer Institute in Survey Research Techniques	4-week Summer Sessions
		Varied (20 courses)	ICPSR Summer Sessions (20 courses)	4-week Summer Sessions
		Nursing	Global Reprod. and Sexual Health Summer Institute	3-day workshop
	Didactic & consultation	Mixed-methods	University of Michigan Mixed Methods Program	Varied dates, multi-day workshops
	Individual consultation	Statistics	Consulting for Statistics, Computing and Analytics Research (CSCAR) statistical workshops	Varied workshops in summer
	Experiential	Health disparities & clinical research	MICHR summer program	10-week program
University of Minnesota	Didactic	Public health	Summer Public Health Institute	3-week program
		Biostatistics	Graduate Courses in Biostatistics (11 courses)	Varied
		Global health	Minnesota In-person Global Health Course	4-week program
University of Washington	Didactic	Imp. science	UW Implementation Science Short Course	2-week course
		Infectious disease	UW Principles of HIV & STI course	1-2 week course
		Global health	eDGH courses (11 courses)	10-week course
University of Hawaii	Experiential	Laboratory Science	Training modules for working with A/BSL-3 and A/BSL-2 agents, with pathogen-specific standard operating procedures and maintenance of A/BSL-3 equipment	2-3 months
		Laboratory Science	Training in flow cytometry, enzyme immunoassay, and Luminex® technology	2-3 months
		Laboratory Science	Comprehensive training program in bioinformatics for Investigators.	2-3 months
Indiana University (associated partner)	Experiential	Infectious disease	Malaria Pathogenesis lab	2-3 months
		Data management	Data Management Course	2 months
		Infectious disease	Microbiome/immunology lab	3 months

40 th Annual National Conference of the Indian Society for Medical Statistics

ISMSCON 2022

24-26 November, 2022

Theme

Intersection of Medical Statistics, Epidemiology & Artificial Intelligence in Health Data Modeling

Organized by

Department of Community Medicine
Krishna Institute of Medical Sciences
Constituent Unit of
Krishna Institute of Medical Sciences "Deemed To Be University"
Karad, Dist: Satara, Maharashtra

Web page : http://ismscon2022.kimskarad.in
Email ID : ismscon2022@kimskarad.in

Organizing Committee

CHIEF PATRON

Hon'ble Dr. Suresh Bhosale
Chancellor, KIMSDU, Karad

PATRONS

Hon'ble Dr. Pravin H. Shingare
Pro-Chancellor, KIMSDU, Karad

Hon'ble Dr. Ved Prakash Mishra
Chief Advisor to Hon'ble Chancellor, KIMSDU, Karad

Hon'ble Dr. Neelam Mishra
Vice Chancellor, KIMSDU, Karad

Institutional Advisory Committee

Dr. Atul Bhosale
Dr. Rajani Gaonkar
Ms. Archana Kaulagekar
Dr. R. G. Naniwadekar
Dr. Rajendra Doijad
Shri. Vinayak Bhosale
Dr. A. R. Risbud
Dr. Asha Jadhav
Dr. Vaishali Mohite
Dr. G. R. Pathade
Dr. M. V. Ghorpade
Dr. D. K. Agarwal
Dr. S. T. Mohite
Dr. N. D. Shashikiran
Mr. S. A. Mashalkar
Mr. P. D. John
Dr. S. R. Patil
Dr. A. Y. Kshirsagar
Dr. G. Varadarajulu

ISMS Advisory Committee

Dr. L. Jayaseelan
Dr. K. Thennarasu
Dr. N. K. Tyagi
Dr. S. N. Dwivedi
Dr. P. Venkatesan
Dr. R. M. Pandey
Dr N. Sreekumaran Nair
Dr. B. L. Verma
Dr. Anil Mathew
Dr. B. Binukumar
Dr. C. M. Pandey

Organizing Chairman

Dr. P. M. Durgawale

Organizing Secretary

Dr. S. V. Kakade

Joint Organizing Secretaries

Dr. Mrs. Y. R. Kadam
Dr. Mrs. V. V. Raje
Dr. Mrs. S. S. Patil

Treasurer

Dr. Mrs. S. V. Patil

Conveners

Registration Committee	:	Dr. Sujata Patil (9960519000)
Scientific Committee	:	Dr. Rajsinh Mohite (9011571092)
Souvenir Committee	:	Dr. Shekhar Kumbhar, Dr. Tulsidas Patil
Sponsorship Committee	:	Mr. Rajendra Sande, Dr. Jayant Patil
Inauguration Committee	:	Dr. Uddhav Kumbhar, Dr. Ashutosh Jain
Cultural Committee	:	Dr. Arun Sansuddi, Dr. Manoj Ambali
Catering Committee	:	Dr. Sanjay Patil, Dr. Sharmishta Garud
Accommodation Committee	:	Dr. Anup Hendre (8446891638)
Transport Committee	:	Dr. Ganesh Ghanwat (9922643120)

Invitation

Dear Friends and Colleagues,

It gives us immense pleasure to invite you on behalf of the Organizing Committee to the XLth Annual National Conference of Indian Society for Medical Statistics (ISMS) to be held from 24 to 26 November, 2022 at Krishna Institute of Medical Sciences (KIMS), Karad, Maharashtra, India.

The Conference will consist of keynote address, symposia, plenary sessions, paper presentations on variety of topics. The programme will focus on issues related to medical statistics, epidemiology, artificial intelligence, modeling etc. There will also be two pre-conference workshops for benefit of young scientists.

We look forward for your active participation and contribution to make this Conference a grand success. We will try our best to make this event a memorable one and eagerly look forward to see you all at Karad in November 2022.

Dr. S. V. Kakade
Organizing Secretary
Associate Professor (Statistics)
Dept. of Community Medicine
KIMS, Karad
Mobile: 9423272351

Dr. P. M. Durgawale
Organizing Chairman
Professor and Head
Dept. of Community Medicine
KIMS, Karad

Dr. S. T. Mohite
Dean
Krishna Institute
of Medical Sciences,
Karad

About Institute

KIMS, a constituent college of Krishna Institute of Medical Sciences "Deemed To Be University" (KIMSDU), Karad, is located on Pune-Banglore National High Way in Western Maharashtra which is better known as 'Sugar Bowl'. KIMS offers UG, PG, Ph.D., MCH and DM programmes. KIMSDU has five more constituent faculties; Dental, Nursing, Physiotherapy, Pharmacy and Allied Sciences (Microbiology and Biotechnology); all offering UG, PG and Ph.D. programmes.

KIMS was founded by the great visionary Hon'ble Late Shree Jayawantraoji Bhosale in 1984. It was started as a 200 bedded hospital in 1982 and currently it is well known as one of the best medical colleges in Maharashtra with 1125 bedded multi-specialty hospital and 250 UG intake capacity and is accredited by NAAC A+ grade (CGPA 3.39) under leadership of Hon'ble Dr. Suresh Bhosale. Today the institution is renowned for its research, teaching and dedication of services to humanity. KIMS's uniqueness lies in its commitment to provide the equal care to all with marginalized charges by utilizing the establishment of the hi-tech medical care.

Research being the immense part of the medical education, Statistical expertise is provided to all kinds of researches conducted in all constituent colleges of KIMSDU by statisticians under guidance of the Statistician from Department of Community Medicine, KIMS.

Scientific Focus

The scientific programme will consist of symposia, plenary sessions, and invited talks focusing on the theme

Intersection of Medical Statistics, Epidemiology and Artificial Intelligence in Health Data Modeling.

However, including the theme, the scientific programme will be accepting contributory papers; for oral as well as poster presentation; covering a wide range of topics listed below:

Epidemiological studies, Non-communicable diseases, Disease burden, Nutritional Epidemiology, Vector Borne diseases, Statistical models, Sampling and analytical methods, Clinical trials, Multi-centric studies, Gender dimensions, Challenges in urban and rural health, Qualitative and Quantitative research methods, Survival analysis, Medical Demography, Meta-Analysis, Health Economics, Population and Public Health Policy, Bio Informatics, etc.

Pre-Conference Workshops :

Two workshops are planned for young statisticians as well as for clinical and Bio-medical researchers. As a part of capacity building of enthusiastic researchers following Pre-Conference Workshops will be held on 23rd November 2022:

1. Statistical Methods for Medical Research.
2. Applications of Survival analysis in Medical Research.

General Information

Venue -

Krishna Institute of Medical Sciences, Karad, District Satara, Maharashtra.

Important Dates :

Last date for Conference Registration	:	31-10-2022
Last date for Submission of abstracts	:	31-10-2022
Date of Pre-Conference workshop:		23-11-2022
Dates of Conference	:	24-11-2022 to 26-11-2022

Abstract Submission :

- Abstract should be submitted as per the following pattern :
  Title of the study, Authors, Affiliation, Background, Objectives, Methodology, Results & Conclusion.
- It should be structured and should not exceed 300 words.
- The abstract should be submitted by an electronic mode on the following link along with the completed registration form and registration fees on or before 31st October 2022.
- Acceptance of the abstract will be communicated after its approval by the Scientific Committee.

Registration Fees

Category	Up to 10th October 2022	11th – 31st October 2022	Spot Registration **(Cash Payment Only)**
ISMS Member	₹ 5000	₹ 5500	₹ 6000
Non Member	₹ 6000	₹ 6500	₹ 7000
Student (PG, PhD)	₹ 3000	₹ 3500	₹ 4000
Foreign Delegate	US $ 150	US $ 175	US $ 200
Accompanying person	₹ 3000	₹ 3000	₹ 3500
Pre-Conference Workshop	₹ 2000	₹ 2500	₹ 3000

Please Note :

- For PG and Ph.D. students, Recommendation Letter from HOD is mandatory.
- Delegate Kit will not be given to Accompanying Person & for spot registration.
- For cancellation up to 31st October 2022, 35% of the amount after deduction of the taxes will be refunded one month after the Conference.

Link for registration of Pre-Conference workshop/Conference, Abstract submission & payment :

http://ismscon2022.kimskarad.in/register

For Registration please contact :

Dr. (Mrs.) Sujata Patil
Associate Professor, Dept. of Community Medicine
K.I.M.S., Karad – 415539

Ph. No. : 02164-241555 Ext. 303
Mobile : 09960519000
E-mail ID : ismscon2022@kimskarad.in

Hotel tariffs (Participants must contact Hotel to book accommodation)

Name of the Hotel & E-mail ID/Contact No.	Distance from KIMS (in kms)	Direction from KIMS	AC/ Deluxe AC (₹)	Non AC/ Non Deluxe AC (₹)	Executive/ Super Deluxe AC (₹)	Suite (₹)	Executive Suite (₹)
Hotel Gandharv Palace (02164) 241380	800 m	South	D: 1500 EB : 200	D: 1200 EB : 200			
Hotel Sapphire *# hsapphire2040@gmail.com 9175428640/ (02164) 243776, 243777	500m	North	S: 2100/ 2700 D: 2400/ 3000 T: 2800/ 3500			S:4000 D:5000 T:6000	
Hotel Shahi Executive hotelshahi313@gmail.com 9673723111	1.3	North	D: 2000 EB: 400	D:1600 EB:400			
Hotel Sangam*# contact@sangamhotel.com 9922514996	2	North	S: 2016 D: 2352 EB: 560	S:1624 D:1904 EB:392	S: 2464 D: 2799 EB: 560	D:4130 EB:590	
Hotel Pankaj*# htpankaj@gmail.com (02164) 222570, 222579	2.5	North	D: 2700 EB: 450	S:1900 EB:450		D:4480 EB:450	
Hotel Alankar* hotelalankarkarad@gmail.com 9172048398/99	3	North-East	S: 1814 D: 2117 EB: 560	S:1210 D:1613 EB:448		S:3931 D:4547 EB:616	
Hotel Krishna Palace palace_krishna@yahoo.com (02164) 227399, 225699	3.4	North-East	S: 1490 D: 1690 EB: 400	S: 990 D:1390 EB:300	S: 1690 D: 1900 EB: 400		D: 2200 EB: 400
Hotel Annapurna Prasad djoshi88@gmail.com (02164) 220364	3.5	North		D: 800 EB:150			
Hotel Royal Palace hotelroyalpalacekarad@gmail.com 8605224599	3.5	North	S: 1800/ 2100 EB: 500	S:1600 EB:500	T: 2500 EB: 500		D:2700 EB:500
Hotel Bhagyalaxmi motelbhagyalaxmi@gmail.com 8432038432	5	South	D: 1600/ 1800 EB: 300	D:1300 EB:300			
The Fern Residency*# fom.fr.karad@fernhotels.com 7796615249/ (02164) 228166, 229166	5	North	S: 3360 D: 4480 T: 5040			S:6720 D:7280 EB:1120	
Hotel Mahendra# hotelmahendra.netkarad@gmail.com (02164) 226199, 224199, 225399	5	North	D: 1800 EB: 350		D: 2100 EB: 350	D:3000¢ EB: 350	D:3500¢¢ EB: 350

*: Rates including GST
#: Rates including Complementary Breakfast
¢: Classic Suite
¢¢: Royal Suite

S: Single Bed
D: Double Bed
T: Triple Bed
EB: Extra Bed

For students (girls) few free hostel accommodation will be made available on first-come-first-serve basis.

ISMS Bulletin

August 2022

Department of Biostatistics

Dr. M. V. Govindaswamy Centre

National Institute of Mental Health and Neuro Sciences (NIMHANS)

Bengaluru - 560029, Karnataka

Email: isms.editor@gmail.com Ph: 080-26995119 / 080-26995118