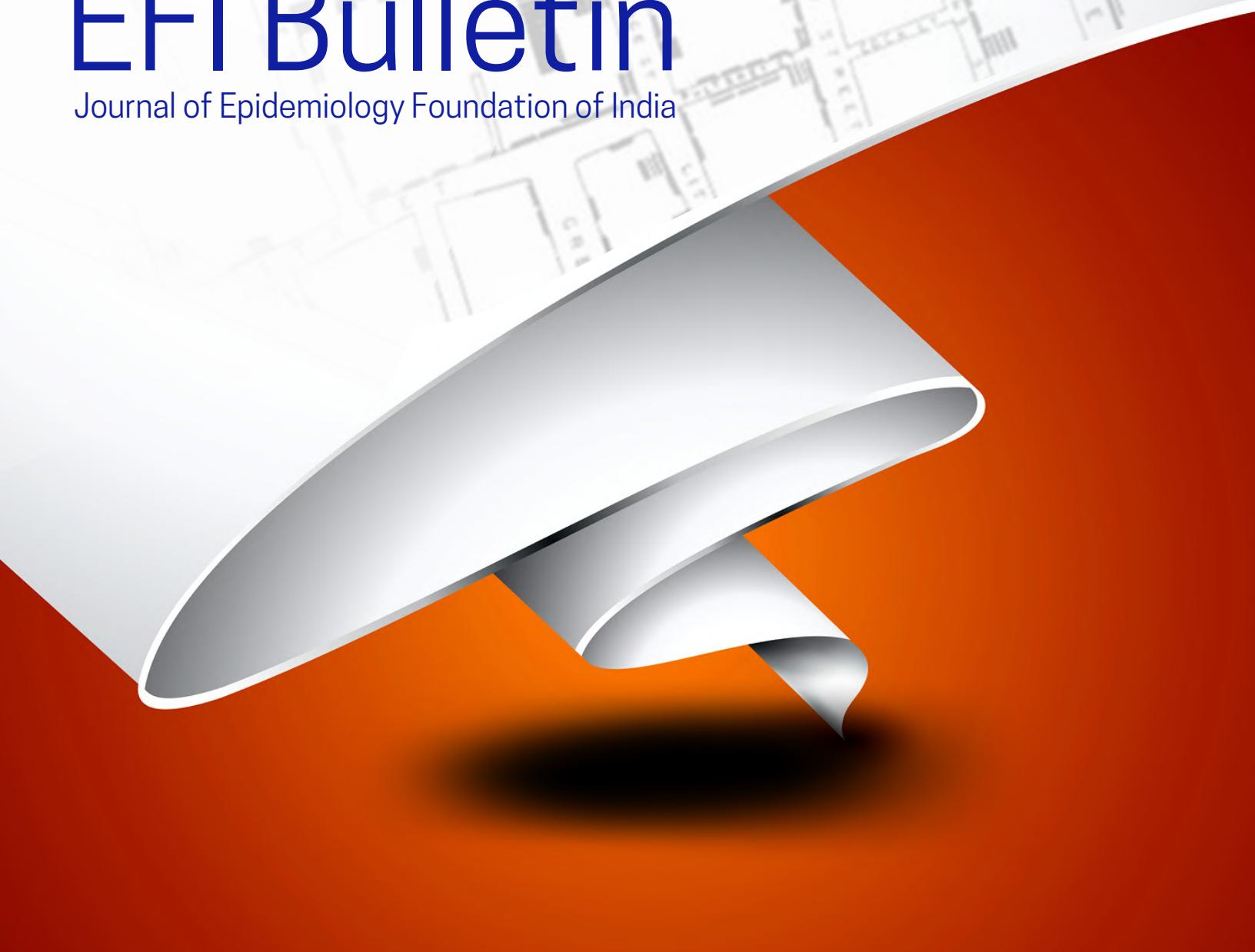




# **EFI Bulletin**

Journal of Epidemiology Foundation of India



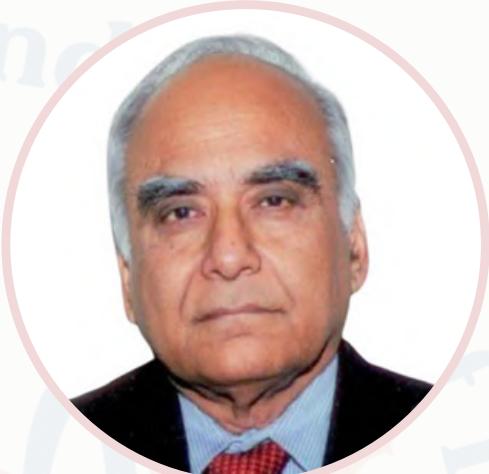


# Epidemiology Foundation of India

## PATRONS



**Dr. V M Katoch**  
Former Secretary DHR &  
Former Director General  
Indian Council of Medical Research (ICMR)



**Prof. Nirmal Kumar Ganguly**  
Former Director General  
Indian Council of Medical Research (ICMR)

## GOVERNING COUNCIL



**President**  
Dr. Umesh Kapil



**Vice-President**  
Dr. Ashok Kumar Bhardwaj



**Principal Scientific Adviser**  
Dr. Sanjay Zodpey



**Treasurer**  
Dr. Vinay Prakash Srivastava



**Secretary**  
Dr. Pradeep Aggarwal



**Member**  
Dr. R K Srivastava



**Member**  
Dr. Ajit Sahai



**Member**  
Dr. Arti Kapil



**Member**  
Dr. S K Singh



**Member**  
Dr. C M Singh



# EFI Bulletin

## Bulletin of Epidemiology Foundation of India

### EFI Governing Body

#### Patrons

Dr. V M Katoch

[vishwamohankatoch18@gmail.com](mailto:vishwamohankatoch18@gmail.com)

Dr. N K Ganguly

[ganguly1nk@gmail.com](mailto:ganguly1nk@gmail.com)

#### President

Prof. Umesh Kapil

[umeshkapil@gmail.com](mailto:umeshkapil@gmail.com)

#### Vice-President

Prof. Ashok K Bhardwaj

[ashoknonu@gmail.com](mailto:ashoknonu@gmail.com)

#### Principal Scientific Adviser

Dr. Sanjay Zodpey

[zodpeysp@gmail.com](mailto:zodpeysp@gmail.com)

#### Treasurer

Prof. V P Srivastava

[vinayp.1956@gmail.com](mailto:vinayp.1956@gmail.com)

#### Members

Dr R K Srivastava

[rakesh231149@gmail.com](mailto:rakesh231149@gmail.com)

Prof. Ajit Sahai

[ajit.sahai@gmail.com](mailto:ajit.sahai@gmail.com)

Prof. Arti Kapil

[akapilmicro@gmail.com](mailto:akapilmicro@gmail.com)

Prof. Shivendra K Singh

[shivmymail0522@gmail.com](mailto:shivmymail0522@gmail.com)

Prof. C M Singh

[drcmsingh@yahoo.co.in](mailto:drcmsingh@yahoo.co.in)

#### Secretary

Dr Pradeep Aggarwal

[drpradeep\\_aggarwal@hotmail.com](mailto:drpradeep_aggarwal@hotmail.com)

#### Chief Editor EFI Bulletin

Prof. Ajit Sahai

[editor2021.efibulletin@gmail.com](mailto:editor2021.efibulletin@gmail.com)

[ajit.sahai@gmail.com](mailto:ajit.sahai@gmail.com)

### Aims of EFI

To identify and promote areas of cooperation and understanding among researchers and like-minded organizations, individuals, scientific networks and other Governmental and Non-Governmental, National & International agencies which are contributing towards realizing the objectives of the Foundation.

#### Benefit of becoming a member of EFI

- Networking with renowned Epidemiology experts worldwide and partnership with Professional organizations in field of Epidemiology.
- Receiving announcements of EFI activities.
- Eligibility to receive travel scholarship / support for attending EFI sponsored courses / meeting.
- Reduced registration fees for attending EFI Training Courses, CME, Regional meeting and Annual Conference.
- Joint membership of International Epidemiological Association (IEA)

#### IEA joint memberships with EFI

Special rates for joint membership for regular member is 25 USD and for ECE is 10 USD do visit <http://www.ieaweb.org/> for more details.

#### Contact Us

Secretary,

Epidemiology Foundation of India

Website: [www.efi.org.in](http://www.efi.org.in)

E Mail Id: [epifindia@gmail.com](mailto:epifindia@gmail.com)



# EFI Bulletin

Bulletin of Epidemiology Foundation of India

## Editorial Board

### Chief Editor

Dr Ajit Sahai  
Former Director-Professor and Head,  
Department of Medical Biometrics, JIPMER, Pondicherry  
Email: [editor2021.efibulletin@gmail.com](mailto:editor2021.efibulletin@gmail.com); [ajit.sahai@gmail.com](mailto:ajit.sahai@gmail.com)  
Mobile Nos: 95804 25288; 94430 78433

### Executive-Editor

Dr (Col) Rajesh Kunwar  
Professor and Head,  
Department of Community Medicine,  
T S M Medical College, Lucknow.  
Email: [rkunwar14@yahoo.com](mailto:rkunwar14@yahoo.com)  
Mobile: 94452 10044

### Managing-Editor

Dr Pradeep Aggarwal  
Associate Professor,  
Department of Community and Family Medicine,  
All India Institute of Medical Science Rishikesh  
Email: [drpradeep\\_aggarwal@hotmail.com](mailto:drpradeep_aggarwal@hotmail.com)  
Mobile: 98372 15747

### Associate-Editors

Dr Sumanth M M  
Associate Professor,  
Department of Community Medicine,  
MMCRI, Mysore.  
Email:  
[drsumanthmmc@rediffmail.com](mailto:drsumanthmmc@rediffmail.com)  
Mobile: 74113 39709

Dr Subitha L  
Additional Professor,  
Department of Preventive and Social  
Medicine,  
JIPMER, Pondicherry.  
Email: [subitha.l@gmail.com](mailto:subitha.l@gmail.com)  
Mobile No: 84892 16058

Dr Abhishek Gupta  
Associate Professor,  
Department of Community Medicine,  
TSMMC, Lucknow.  
Email: [ab23kgmc@gmail.com](mailto:ab23kgmc@gmail.com)  
Mobile No: 98899 68009

### Board Members

Dr M Meghachandra Singh, New Delhi.  
Dr Manish Kumar Manar, Lucknow, Uttar Pradesh  
Dr Pankaj Bhardwaj, Jodhpur, Rajasthan  
Dr Sitanshu Kar, Pondicherry.  
Dr Sonu Goel, Chandigarh.

### Editorial Advisors

Dr A Indrayan, New Delhi.  
Dr F U Ahmed, Dibrugarh, Assam  
Dr J S Thakur, Chandigarh.  
Dr Rajiv K Jain, New Delhi.  
Dr Surya Kant, Lucknow, Uttar Pradesh



# EFI Bulletin

## Bulletin of Epidemiology Foundation of India

### Editorial

#### COVID-19 vaccination: the way forward

Sanjay Rai, Shreya Jha, Shashi Kant

1 - 2

### Review Article

#### Socioeconomic scales for rural versus urban settings in Health Research: A Review

Ravleen Kaur Bakshi, Neeta Kumar, Gagandeep Singh

3 - 11

Continuing Medical Education

#### Longitudinal data analysis methods – a primer

Nishaant Ramasamy, Ruben Raj, Namita Patel, Palanivel C

12 - 17

### Commentary

#### Adaptive platform trials present a new frontier for epidemiologists in the era of Covid-19

Manya Prasad, Umesh Kapil

18 - 19

#### Strengthening provisions for breast cancer screening in India

Ankan Mukherjee Das

20 - 22

### Original Article

#### Latent class analysis in a complex sampling design – an advanced modelling approach in Epidemiology

Ankita Dey, Diganta Mukherjee, Sugata Sen Roy

23 - 29

### Perspective

#### Anaesthesiologists as Community Physicians- Expanding Horizons

Naveen Malhotra, Divya Gupta

30 - 31

#### Snakebite Management in India: Challenges Remain

Kaushik Bharati

32 - 34

### Opinion

#### The future of Medical Writing: Artificial Intelligence as first author and not the ghost

Deepak Gupta

35 - 37

### Book Review

#### Book Review: Ergonomics in Healthcare: An Update

Rakesh Kakkar

38

### Report

#### Clinical epidemiology and biostatistics workshop 2023 report

Anil Mathew

39 - 40

## COVID-19 vaccination: the way forward

Sanjay K. Rai, Shreya Jha, Shashi Kant

Centre for Community Medicine, All India Institute of Medical Sciences, New Delhi



### CORRESPONDING AUTHOR

Dr. Shreya Jha, Consultant, Centre for Community Medicine, All India Institute of Medical Sciences, New Delhi

E Mail ID: [shreya.jh@gmail.com](mailto:shreya.jh@gmail.com)

### CITATION

Rai SK, Jha S, Kant S. COVID-19 vaccination: the way forward. *EFI Bulletin*. 2023;4(1):01-02.

<https://doi.org/10.56450/EFIB.2023.v3i01.001>

### ARTICLE CYCLE

Received: 16/03/2023; Accepted: 20/03/2023; Published: 31/03/2023

This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

©The Author(s). 2023 Open Access

It is nearly three years since the World Health Organization declared COVID-19 as a global public health emergency. COVID-19 continues to be a threat to global health system. WHO Director General Dr. Tedros Adhanom Ghebreyesus, during his press briefing on 22nd Feb 2023, advocated continuation of COVID-19 vaccination and booster dose.

More than 13 billion doses of different COVID-19 vaccines have been administered worldwide. The global scientific community has witnessed unprecedented research on COVID-19, including vaccine development. All future recommendations related to COVID-19 vaccination hence should be based on two factors; the existing evidence of community level immunity acquired against COVID-19, either through natural infection or through vaccination, and the comparison of risk benefit analysis of COVID-19 vaccination. If the potential benefit of COVID-19 vaccination outweighs the potential risk, then we continue with the COVID-19 vaccination.

The COVID-19 pandemic caused panic across the globe. Due to the urgency of the situation, the usual due diligence for vaccine development was not followed, and Emergency Use Authorization (EUA) was granted to the pharmaceutical companies. Now, that the situation has returned to near normal, it would be appropriate that the scientific rigor is brought back to the vaccine development. The

pharmaceutical industry must be asked to provide robust evidence that their COVID-19 vaccine is safe in long-term and effective. Due to the EUA of vaccines and universal vaccination policy for people recruited in the placebo group, global community has lost the opportunity to know the long-term impact of all these COVID-19 vaccines.

Currently 37 COVID-19 vaccines are in use. These vaccines have shown acceptable safety and efficacy against COVID-19 during clinical trials. During initial phase of the pandemic, COVID-19 vaccination played an important role in reducing the severity and death when majority of the population was susceptible. Now majority of the population worldwide has had exposure to SARS-CoV-2. Global evidence has demonstrated that exposure to COVID-19 provide good protection from subsequent infection, severity and death.<sup>(1)</sup> Like any other drugs, vaccination against COVID-19 may have some short term and long-term consequences. Many studies have reported serious adverse events following vaccination.<sup>(1)</sup> Women have reportedly experienced various menstrual abnormalities, including but not limited to menorrhagia, metrorrhagia, and polymenorrhea after COVID-19 vaccination.<sup>(2)</sup> Similarly, cardiac disorders following some COVID-19 vaccines have also been reported extensively from all over the world, especially among young male individuals.<sup>(3)</sup> A large study from New Zealand covering more than 4 million population, found

higher incidence rate ratio (IRR) of myo/pericarditis in individuals vaccinated with Pfizer vaccine. The IRR of myo/pericarditis following the first dose was 2.6 (2.2– 2.9) with a risk difference (95% CI) of 1.6 (1.1– 2.1) per 100,000 persons vaccinated and following the second dose was 4.1 (3.7– 4.5) with a risk difference of 3.2 (2.6– 3.9) per 100,000 persons vaccinated.(4) Similarly, another multinational study found a pooled 30% increased risk of thrombocytopenia after a first dose of the AstraZeneca vaccine compared to Pfizer vaccine.(5) After multiple reports of increased cardiac risk from AstraZeneca, vaccination with this vaccine was stopped in younger population in many European countries.

Recent studies have questioned the effectiveness of COVID-19 vaccines in preventing prevailing strains of the virus. Also, there are evidence now that COVID-19 vaccines do not prevent transmission of the virus, and the effectiveness of the vaccines declines over time.

Scientific and clinical data on the sub-acute and long-term effects of COVID-19 infection is continuously evolving, and evidences suggest that they can affect multiple organ systems. However, evidence from more robust study design is needed before one could arrive at a definitive conclusion.

In larger public interest, there is a need to establish an effective surveillance system to document the serious adverse events following immunization so that these events do not go unnoticed. Community based research is required to differentiate the role of naturally acquired immunity and vaccine acquired immunity. Risk-benefit analysis of different vaccines should be carried out. Decision whether to continue

vaccination with the booster doses, or halt vaccination altogether should be based on risk-benefit analysis. Continuation of vaccination should be considered only when the risk-benefit analysis is in favour of benefit, and clearly outweighs the risk. With a very high proportion of the population having been exposed to SARS-CoV-2 infection, unnecessary and irrational use of vaccination against COVID-19 will drive the resource and attention away from other public health issues.

## REFERENCES

- Stein C, Nassereldine H, Sorensen RJD, Amlag JO, Bisignano C, Byrne S, et al. Past SARS-CoV-2 infection protection against re-infection: a systematic review and meta-analysis. *The Lancet*. 2023 Feb 16. DOI: [https://doi.org/10.1016/S0140-6736\(22\)02465-5](https://doi.org/10.1016/S0140-6736(22)02465-5)
- Nazir M, Asghar S, Rathore MA, Shahzad A, Shahid A, Ashraf Khan A, et al. Menstrual abnormalities after COVID-19 vaccines: A systematic review. *Vacunas*. 2022 Sep-Dec;23:S77-S87. doi: <https://doi.org/10.1016/j.vacun.2022.07.001> ..
- Tuvali, O., Tshori, S., Derazne, E., Hannuna, R.R., Afek, A., Haberman, D, et sl. The Incidence of Myocarditis and Pericarditis in Post COVID-19 Unvaccinated Patients—A Large Population-Based Study. *J. Clin. Med.* 2022, 11, 2219. <https://doi.org/10.3390/jcm11082219>
- Walton M, Tomkies R, Teunissen T, Lumley T, Hanlon T. Thrombotic events following the BNT162b2 mRNA COVID-19 vaccine (Pfizer-BioNTech) in Aotearoa New Zealand: A self-controlled case series study. *Thromb Res.* 2023 Feb;222:102-108. doi: <https://doi.org/10.1016/j.thromres.2022.12.012> .
- Li X, Burn E, Duarte-Salles T, Yin C, Reich C, Delmestri A et al. Comparative risk of thrombosis with thrombocytopenia syndrome or thromboembolic events associated with different covid-19 vaccines: international network cohort study from five European countries and the US *BMJ* 2022; 379 :e071594 doi: <https://doi.org/10.1136/bmj-2022-071594>

## REVIEW ARTICLE

# Socioeconomic scales for rural versus urban settings in Health Research: A Review

Ravleen Kaur Bakshi<sup>1</sup>, Neeta Kumar<sup>2</sup>, Gagandeep Singh<sup>3</sup>

<sup>1</sup>Division of RCN, Indian Council of Medical Research-HQ, New Delhi;

<sup>2</sup>Division of SHI, Indian Council of Medical Research-HQ, New Delhi;

<sup>3</sup>Department of Microbiology, All India Institute of Medical Sciences, New Delhi

## CORRESPONDING AUTHOR

Dr Ravleen Kaur Bakshi, Scientist C, Division of RCN, Indian Council of Medical Research - ICMR-HQ, New Delhi

E Mail ID: [drravleen01@gmail.com](mailto:drravleen01@gmail.com)



## CITATION

Bakshi RK, Kumar N, Singh G. Socioeconomic scales for rural versus urban settings in Health Research: A Review. EFI Bulletin. 2023;4(1):03-11. <https://doi.org/10.56450/EFIB.2023.v3i01.002>

## ARTICLE CYCLE

Received: 23/02/2023; Accepted: 20/03/2023; Published: 31/03/2023

This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

©The Author(s). 2023 Open Access

Socioeconomic status (SES) is defined as a composite measure including economic, social, work status, and household assets.(1) Education and income are the most used measure of SES in epidemiological studies. Education helps in measuring social status, however occupations is used for measuring work status. Each status is considered as an indicator, and is related to each other but do not overlap. SES is an important social determinant of health and disease. Social status influences the accessibility, affordability, acceptability, and actual utilization of available health facilities.

**SES Scales and health** – Status inequalities are common between individuals, which also reflects in their health status. Nutritional deficiencies and communicable diseases are presented by low SES population, whereas obesity, lifestyle diseases and noncommunicable diseases are more common in high SES population. High SES population can afford and access healthcare.(2)

**Definition of SES Scale** - The position that an individual or family occupies with reference to the prevailing average standards of cultural and material possessions, income, and participation in group activity of the community. (3)

The need Socio-economic status scale in health is to

1. It is used as a parameter to determine the affordability of the health systems by the community.
2. Understand the distribution of health and disease in population on the basis of SES.
3. To identify the most vulnerable population as priority group, and provide them health access in a low resource setting.
4. The inequities in access and utilization of health care with respect to the detailed. (4)

## Different parameters of Socio-Economic Scales available discussed in details –

### Socio-Economic Scales used in Urban Areas –

1. **Taussig Scale (1928):** Taussig scale was based on income. The Taussig Socio-Economic Scale, also known as the Taussig Index, was a system developed by economist and historian Frank William Taussig in 1928 to classify occupations according to their socio-economic status. The Taussig Scale consists of ten levels, each level representing a different socio-economic status. The levels are based on the income, education, and social status associated with different occupations. The levels range from Level I (highest) to Level X (lowest), with the level of an occupation

determined by its score on a set of criteria that include income, education, and social status. (5) Although the Taussig Scale has been criticized for being outdated and not accounting for the complexities of modern occupations and their changing socio-economic status, it is sometimes used as a reference point by researchers and historians interested in understanding historical patterns of occupational mobility and social class.

**2. Verma Scale (1962):** The Verma Socio-Economic Scale is a system developed by Indian sociologist and economist Yogendra Singh Verma in 1962 to classify households in India according to their socio-economic status.

The Verma Scale takes into account the education, occupation, and income of the head of the household, along with the size and composition of the household. The scale assigns a score to each household based on these factors and places the household into one of five categories, from the lowest socio-economic status to the highest. (6)

The five categories of the Verma Scale are: Category I: Lowest socio-economic status, Category II: Lower middle socio-economic status, Category III: Middle socio-economic status, Category IV: Upper middle socio-economic status, and Category V: Highest socio-economic status.

The Verma Scale has been used in India to study patterns of economic development and social inequality, and to inform policy decisions related to poverty alleviation and social welfare programs.

However, it has also been criticized for its narrow focus on the head of the household and its reliance on a single, fixed set of indicators that may not accurately capture the complexity of socio-economic status in modern India.

**3. Sims Scale (2011):** The Sims Socio-Economic Scale, also known as the Modified Kuppuswamy Scale, is a system developed by Indian epidemiologist and statistician P.R. Sreeramareddy and colleagues in 2011 to classify households in India according to their socio-economic status. (7)

The Sims Scale is a modification of the Kuppuswamy Scale, which was developed in the 1970s to classify households based on their occupation, education, and income in India. The Sims Scale takes into account additional factors, including the type of housing and the ownership of consumer durables.

In this scale a score is assigned to each household based on a set of criteria, including the education and occupation of the household head, the type of

housing, and ownership of consumer durables. The scores are then used to place households into one of five categories from the lowest to the highest socio-economic status.

The Sims Scale has been used in India to study patterns of health and disease, and to inform policy decisions related to public health and social welfare programs. However, like other socio-economic scales, it has been criticized for its reliance on a fixed set of indicators that may not accurately capture the complexity of socio-economic status in modern India.

**4. Kulshreshtha Scale (1972):** The Kulshreshtha Scale is a socioeconomic classification system developed by A.K. Kulshreshtha in 1972. It is based on the occupation and education level of the household head; and the total number of rooms in the dwelling unit. The six categories in which the households are categorised are - elite, upper middle, lower middle, upper lower, lower, and destitute; according to this scale. (8)

The Kulshreshtha Scale is commonly used in India for market research and social and economic research. It provides a useful tool for understanding the social and economic characteristics of different households and for identifying the needs of different segments of the population.

**5. Shrivastava scale (1978):** The Shrivastava Scale is a socioeconomic classification system developed by M.K. Shrivastava in 1978. It is based on the occupation and education level of the household head, and the total members earning in the family. (6)

The scale categorizes households into eight categories - Elite, Upper Middle, Lower Middle, Upper Lower, Lower, Marginal, Scheduled Castes, and Scheduled Tribes.

The Shrivastava Scale is commonly used in India for market research and social and economic research. It provides a useful tool for understanding the social and economic characteristics of different households and for identifying the needs of different segments of the population.

**6. Kuppuswamy's Scale (1976):** The Kuppuswamy's Scale is a classification system used to measure the socioeconomic status of a family or household in India. It was developed in 1976 by Dr. Kuppuswamy. He was a professor of medicine at the Christian Medical College in Vellore, India. The Kuppuswamy's Scale is a composite score taking into account three main factors to determine the socioeconomic status

of a family: education and occupation of the head of the family, total family income, and the size of the family. The scale provides a score based on these factors. (9) The scores obtained from these three components are added up to determine the final score of the household. The score is used to classify households into different socioeconomic categories ranging from a score of 3-29, and the study populations into high, middle, and low SES accordingly. (2)

Recently, Mishra et al have suggested an economic revision of Kuppuswami's scale in order to account for the devaluation of rupee. To get current income group, a conversion factor calculated based on current All India Consumer Price Index (AICPI) is applied. CPI in India comprises multiple series classified based on different economic groups. CPI is done to find out the difference in the prices and range of inflation.

Modification according to the income slabs - The CPI in 1976, was taken as 296 based on 1960. In 1982 new CPI base was set at 100 when 1960 based value became 460 (so multiplying factor 4.6). In 2001 CPI base was set at 100. In 1982, based value reached 490, so multiplying factor is 4.9. In January 2017, CPI value is 274 base year 200.

The Kuppuswamy classification is used in urban and peri urban areas. However, it has been criticized for not taking into account other important factors that may affect the socioeconomic status of a family, such as assets and liabilities, access to healthcare, and other social and cultural factors. (2)

**7. Modified Kuppuswamy scale (2018):** The Modified Kuppuswamy Scale is an updated version

of the original Kuppuswamy Scale that was developed in 2018 by Reddy and colleagues. It incorporates additional parameters to better capture the socioeconomic status of households in the urban areas of the contemporary India. In addition to the three main factors in the original Kuppuswamy Scale (education and occupation of the head of the household, and total family income); the Modified Kuppuswamy Scale includes the following additional parameters: type of house, sanitation facilities, social group and land ownership. Socioeconomic Class according to Kuppuswamy's scale are: 26–29 Upper (I); 16–25 Upper Middle (II); 11–15 Middle (III); 05–10 Lower Middle (IV); and < 4 Lower (V). The Modified Kuppuswamy Scale is used in a similar way to the original scale, with households being classified into different socioeconomic categories based on their scores.

However, the additional parameters provide a more comprehensive assessment of the socioeconomic status of households. In India, for research and policy purposes, such as in determining eligibility for government schemes and programs; the Kuppuswamy's Scale is used. (10)

Merits of this SES are that the three main domains considered. The limitations are that the steady inflation and consequent fall in the value of currency makes the economic criteria in the scale less relevant. Further, total family income part is considered without any consideration of the size of the family. There is an overemphasis on income. The material possessions not considered in this SES. (9–11)

#### Socio-Economic Scales available in different settings :

Urban SES	Rural SES	Both Rural and Urban SES	Student's SES	SES based on different economic theories*	Non-Indian/International SES	Community based study SES
Taussig Scale (1928)	Lewis and Dhillon (1955),	B G Prasad's classification (1961).	Bharadwaj scale (2001)	Multi-Dimensional Poverty Index scale (MPI),	Warner Scale,	Gour's classification
Verma Scale (1962)	Rahudkar scale (1960),	Modified B G Prasad's (1968 and 1970),		Unsatisfied basic needs (UBN)	Hollingshead Scale,	Wealth Index (WI)
Pandey (1966),	Freeman Scale (1961),	Udai Pareek Scale (1964)			Henderson Scale,	Standard of living index
Sims Scale (2011),	G Trivedi (1964),	Aggarwal Scale (2005),			Mac Arthur scale.	Mahajan-Gupta Scale
Jalota scale (1970),	Shirpurkar Scale (1967)	Tiwari Scale.				

Socio-Economic Scales available in different settings :						
Urban SES	Rural SES	Both Rural and Urban SES	Student's SES	SES based on different economic theories*	Non-Indian/International SES	Community based study SES
Scale (1972),						
Shrivastava scale (1978),						
Kuppuswamy's Scale (1976) and						
Modified Kuppuswamy's Scale (2018)						

\*capable of identifying non-income factors associated with social inequalities

### Socio-Economic Scales used in Rural Areas –

**1. Lewis and Dhillon Scale (1955):** The Lewis and Dhillon Socioeconomic Scale is a classification system developed by Michael Lewis and Harnam Dhillon in 1955 to know the socioeconomic status (SES) of individuals in India. The scale is based on the occupation of the head of the household. (6) The Lewis and Dhillon Scale divides the population into five groups or classes, ranging from the lowest to the highest SES - Class V; Class IV; Class III; Class II; and Class I.

The Lewis and Dhillon Scale is used in research to measure SES in India, as it provides a relatively simple and reliable way to categorize individuals based on their socioeconomic background. However, it is important to note that SES is a complex construct that is influenced by a wide range of factors beyond just occupation, such as income, wealth, and social status.

**2. Rahudkar Scale (1960):** The Rahudkar Socioeconomic Scale is a classification system developed by Dr. V. G. Rahudkar in 1960. The scale is based on the occupation and education of the head of the household, used to determine the socioeconomic status of the household.

The Rahudkar Scale divides the population into five groups or classes, ranging from the lowest to the highest SES: Class V, Class IV, Class III, Class II, and Class I.

The Rahudkar Scale is used in research to measure SES in India, as it provides a relatively simple and reliable way to categorize individuals based on their socioeconomic background. However, it is important to note that SES is a complex construct that is influenced by a wide range of factors beyond just occupation and education, such as income, wealth, and social status.

**3. G Trivedi (1964):** The G Trivedi Scale was developed in 1964 by G. Trivedi, an Indian sociologist. It is being used to evaluate the socioeconomic status of rural areas. The scale takes into account the education and occupation of the head of household, size of land owned, type of house, and type of fuel used for cooking. The scale classifies households into five categories: high, upper-middle, middle, lower-middle, and low. This scale is used in India for market research and social and economic research.

The merits are that it includes many domains. The limitations are that this scale does not emphasize the income aspect; and used for rural subjects only. Also there is variation in the caste system in each state.

### Socio – Economic Scales used in Both Rural and Urban Areas –

**1. Udai Pareek Scale:** The Udai Pareek Scale was developed in 1976 by Udai Pareek, an Indian industrial psychologist. It is used for both the urban and rural areas. The scale takes into account the education and occupation of the head of household, type of house, number of rooms in the house, and type of fuel used for cooking. The five categories used for classifying the household are - upper, upper-middle, lower-middle, lower, and bottom. It provides a useful tool for understanding the social and economic characteristics of different households and for identifying the needs of different segments of the population.

**2. B G Prasad's classification 1961, later modified in 1968 and 1970:** It was developed by B.G. Prasad in 1961. He was a sociologist from Lucknow University. B.G Prasad's classification is used both urban and rural areas. An update of Prasad's socio-economic classification (1961) of Indian population has been

proposed with relevance to the current consumer index price levels of January 2020. Per capita monthly income = Total monthly income of the family/Total members of family is used to calculate the socio-economic scores.(12,13). Accordingly, there are five categories – rich, upper middle class, middle class, lower middle class, and poor. It is now outdated, and its modification is being used.

**3. Modified BG Prasad socioeconomic scale:** The BG Prasad scale was formulated in 1961, and used as the base for Consumer Price Index (CPI) in 1960 as 100. This was revised in 1982, and a linking factor of 4.93 to convert CPI (1982) from the new base of 100 to the old base CPI (1960) was introduced. In 2001 again the base was revised done by introducing the linking factor of 4.63 during the price index year 1960. Further the linking factors for the year 2016 is 2.88. Classes under the Modified BG Prasad SES are I, II, III, IV, and V according to the per capita income per month of the family. The per capita family income includes all earning members.

The merits of this scale are that it can assess the socioeconomic status in both rural and urban areas as only income is considered. It can be updated. The limitations are that it is only income-based scale and no other domains addressed. Inflation is not considered in this SES.

**4. Aggarwal Scale (2005):** It can be used for urban areas, rural areas, resettlement colonies, slum areas, and mobile population. It consisted of 22 items giving suitable weightage to each item ranging from 3 to 9. The maximum total score is 100. Based on the final score, the socio-economic status of the family is divided into six categories - Upper high (combined score of more than 76), High (61-75), Upper Middle (46-60), Lower Middle (31-45), Poor (16-30) and Very Poor (combined score less than 15).

**5. Tiwari Scale:** Tiwari SES uses seven indicators or profiles for calculating the socio-economic status, i.e. housing, material possession, education, occupation, economic profile, cultivated land, and social profile. Initially the scale had seven profiles; and every profile contained five alternatives. The instrument of this scale has a 10-point scale. Weightage system of scoring (varying from 2 to 10) was followed from first to sixth profile; whereas additive pattern of scoring was followed in seventh profile.(7) The family could be placed in any of the five mutually exclusive sub-categories of each profile. It is a valid and reliable instrument for the assessment of socio-economic

status of the families/individuals from urban as well as rural areas in India.

### **Socio – Economic Scales for Student’s –**

There are also socio-economic scales that are specifically designed for students. These scales take into account the family background and education of the student and his/her parents, as well as their income and occupation. A SES on student’s scale is designed to provide a standardized way of assessing a student’s financial need and to ensure that financial aid and other forms of support are distributed fairly and equitably.

**1. Bharadwaj scale (2001):** The Bharadwaj Scale, which was developed by Dr. S. K. Bharadwaj. It could be used to measure the socioeconomic status of families with students. The Bharadwaj Scale takes into account the education level of father, education level of mother, father’s occupation and family income, to determine the socioeconomic status of the household. It categorizes individuals or families into five groups - upper, upper-middle, middle, lower-middle, and lower.

While the Bharadwaj Scale was not designed specifically for students, it could be used to provide information on the socioeconomic background of families with students. It is important in understanding the relationship between SES and academic performance. It should be noted that SES is a complex construct and is influenced by a wide range of factors beyond just occupation, education, and income, such as social status and access to resources. Therefore, it is important to use multiple measures to fully capture SES and its relationship with student outcomes. (9)

### **Socio – Economic Scales which are based on different economic theories (capable of identifying non-income factors associated with social inequalities)**

**1. Multi-Dimensional Poverty Index scale (MPI):** The United Nations Development Programme (UNDP) and the Oxford Poverty and Human Development Initiative (OPHI) in 2010, developed the Multi-Dimensional Poverty Index (MPI). It is a composite measure of poverty. This index takes three main dimensions of poverty into account, including health dimension, education dimension, and living standards dimension. Within these dimensions there are different indicators including nutrition, child mortality, years of schooling, school attendance,

cooking fuel, sanitation, drinking water, electricity, housing, and household assets. A household is considered to be deprived if the indicators falls below a certain threshold. It is considered to be multidimensionally poor if it is deprived in at least one third of the indicators. (5) It provides a more broad picture of poverty than traditional measures based on income alone.

It is calculated as the product of the percentage new multidimensionally poor households and the average proportion of deprivation experienced poor in various indicators. The resultant index ranges from 0 to 1. The higher values indicate greater levels of poverty.

The MPI has been used in many countries to track progress in poverty reduction and to inform policy decisions related to social welfare programs. By taking into account multiple dimensions of poverty, the MPI provides a more nuanced and accurate measure of poverty than traditional measures based on income alone. However, the MPI has also been criticized for its complex methodology and for the fact that it does not capture certain aspects of poverty, such as social exclusion and vulnerability.

**2. Unsatisfied basic needs (UBN):** The Unsatisfied Basic Needs (UBN) was developed in the 1970s as an alternative to income-based measures of poverty. UBN measures the extent to which people are able to meet their fundamental needs such as food/adequate nutrition, clean water, shelter, clothing, basic health care, and education. (14) The UBN approach identifies a set of basic needs that are considered essential for human development. It measures the extent to which people are able to satisfy these needs. The basic needs identified in the UBN approach may vary depending on the context.

UBN is measured by counting the number of basic needs that a person or household is unable to satisfy. The more unsatisfied basic needs a person or household has, the lower their socio-economic status is considered to be. This approach allows for a more nuanced understanding of poverty and socio-economic status, as it takes into account the ability of people to access basic necessities, rather than just their income level.

The UBN approach has been used in various contexts, including in developing countries where income-based measures of poverty may not accurately capture the level of deprivation experienced by some individuals and households.

However, UBN has also been criticized for its narrow focus on material needs and for the fact that it does not account for factors such as social exclusion and vulnerability.(6,14,15)

### Socio – Economic Scales for community based study :

**1. Gour's classification :** The Gour's socio-economic classification (SEC) is a system for classifying households in India based on their occupation and education level. It was developed by N.C. Gour, a professor of economics at Kurukshetra University in India, in 2013. The Gour SEC system divides households into eight categories. It is based on the occupation and education level of the household head. The categories are elite, upper middle, lower middle, semi-skilled, skilled manual, semi-skilled manual, unskilled, and destitute. The proposed SES scale includes the majority of the determinants of social class in a composite subjective manner. The applicability, reliability and validity of the proposed SES is high. This scale is able to categorize families into different socioeconomic statuses. So this proposed scale is fit to be used in community based study. The Gour SEC system is used in India for market research, social and economic research, and by government agencies for policy making. (16)

**2. Wealth Index (WI) :** The value of all the assets owned by a household minus the liabilities is the wealth. WI is a composite index, composed of the variables assessing the main assets owned. It is a substitute indicator of household level wealth. WI is constructed by all the indicators that allow us to comprehend the level of wealth of the household. It is a tool used to measure the socio-economic status of households or individuals, based on the principle that a person's or household's wealth is reflected in their access to assets and resources, such as housing, land, household goods, and savings. (17)

The WI is calculated by assigning a score to each asset or resource owned by a household or individual. The scores are based on the relative value of each asset or resource and are determined by comparing the distribution of the asset or resource in the population being studied. For example, if a household owns a car, which is a relatively rare asset in the population being studied, the car would be assigned a higher score than a more common asset, such as a bicycle. (18) The individual or household scores are then combined to create an overall WI score, which is used to rank individuals or

households according to their socio-economic status. In most cases, the WI is divided into quintiles, with the lowest quintile representing the poorest, and the highest quintile representing the wealthiest. The WI is widely used in public health research and is particularly useful in settings where income data is not available or unreliable. It is often used in surveys to help target interventions and programs to those who are most in need. However, the WI has been criticized for not capturing other important dimensions of poverty, such as social exclusion and vulnerability. (19)

In the National Family Health Survey 3 (NFHS-3) Government of India used wealth index variables. Wealth Index Variables is based on the 33 assets and housing characteristics, which include - material of exterior walls, type of roofing, cooking fuel, house ownership, number of household members per sleeping room, ownership of a bank or post-office account, ownership of mattress, household electrification, type of windows, drinking water source, type of toilet facility, type of flooring, pressure cooker; chair, cot/bed, table, an electric fan, radio/transistor; black and white television, color television; sewing machine; mobile telephone; any other telephone; computer; refrigerator; watch or clock; bicycle; motorcycle or scooter; an animal-drawn cart; car; water pump; thresher; and tractor. (20)

**3. Standard of living index (SLI):** SLI is a scale used to measure the overall well-being of population in a particular geographic area. It consists of factors which impact quality of life, including the type of house, food security, healthcare facility, education, and income. A scoring is given as per parameters and total score is calculated. It ranges from 0 to 100, a higher score indicating higher standard of living. (18) The Government of India in the National Family Health Survey (NFHS - II) had used the Standard of Living Index (SLI) scale which contains 11 items.(3,21) This scale is used to compare the standard of living across various geographic areas, such as countries. It helps policymakers to identify areas which require more support/investment.

**4. Mahajan Gupta Scale (1971):** The scale was developed by P.L. Mahajan and R.C. Gupta in 1971. The Mahajan-Gupta Scale is a socioeconomic classification system used in India to measure the social and economic status of households. The Mahajan-Gupta Scale classifies households into five categories based on a combination of factors

including occupation, education level, and income: upper class, upper middle class, middle class, lower middle class, and lower class. The Mahajan-Gupta Scale is commonly used in India for market and economic research. It is useful tool for understanding the social and economic characteristics of different households and for identifying the needs of different segments of the population.

### **Social security schemes and economic class**

Social security schemes in India are designed to provide financial assistance and social protection to individuals and families belonging to different economic classes. The government of India offers various social security schemes for people from different economic classes.

#### **Ayushman Bharat**

Pradhan Mantri Jan Arogya Yojana (PM-JAY) is a health insurance scheme giving coverage of INR 5 lakhs per family per year for hospitalization. This scheme is being targeted towards the deprived and weaker socio-economic sections. It has distinction between the rural and urban beneficiaries as well as the SES of the individual. Among the rural beneficiaries, there are total of seven deprivation criteria (D1 to D5 and D7) and automatic inclusion criteria of destitute/ living on donations, manual scavenger households, primitive tribal group, legally released bonded labour.

**For the “Urban Beneficiaries” there are 11 occupational categories of workers who are eligible for the scheme. These are namely**

Ragpicker, Beggar, Domestic worker, Street vendor/ Cobbler/hawker / other service provider working on streets; Construction worker/ Plumber/ Mason/ Labour/ Painter/ Welder/ Security guard/ Coolie and other head-load worker; Sweeper/ Sanitation worker/ Mali; Home-based worker/ Artisan/ Handicrafts worker/ Tailor; Transport worker/ Driver/ Conductor/ Helper to drivers and conductors/ Cart puller/ Rickshaw puller; Shop worker/ Assistant/ Peon in small establishment/ Helper/Delivery assistant / Attendant/ Waiter; Electrician/ Mechanic/ Assembler/ Repair worker; Washer-man/ Chowkidar.

Even though PM-JAY uses the Socio-Economic Caste Census (SECC) as the basis of eligibility of households, many States are already implementing their own health insurance schemes with a set of beneficiaries already identified.

## Pradhan Mantri Jan Dhan Yojana (PMJDY)

This scheme aims to provide access to financial services like bank accounts, credit, insurance, and pensions to people from all economic classes. The scheme is targeted towards those who do not have access to such facilities, especially the poor and marginalized sections of society.

## National Social Assistance Programme (NSAP)

This scheme provides financial assistance to the elderly, widows, and disabled people who are living below the poverty line. It is aimed at providing social security to people who are economically vulnerable.

## Pradhan Mantri Jeevan Jyoti Bima Yojana (PMJJBY)

This scheme provides life insurance coverage to individuals from all economic classes. The premium for this scheme is very low, making it accessible to even the poorest sections of society.

## CONCLUSION

Kuppuswamy scale is the most preferred scale for urban, and B G Prasad SES for the rural population. SES help to develop socioeconomic classification of the population mostly based on the income with scientific basis. Udai Pareek and G. Trivedi scale, Kuppuswamy scale 1962, B G Prasad classification are some Indian SES scales most commonly used. (18) There is always a need to update SES scale based on the - income ranges in the scale as they lose their relevance following the depreciation in the value of the rupee.

There are many different factors that contribute to the link between socioeconomic status and health. Socioeconomic status is a predictor of health status of the community, as it determines the access and availability to healthcare. The access to healthcare tends to be limited for people with lower socioeconomic status. People with lower socioeconomic status may also have poorer living conditions, exposed to environmental toxins, and experience higher levels of stress, all of which can negatively impact health outcomes. Research has shown that people with lower socioeconomic status are more likely to suffer from a range of health problems, including infectious diseases, and nutritional deficiency disorders. They also tend to have higher rates of mental health problems, such as depression and anxiety. This relationship between SES and health has been observed in many different countries and across a range of health outcomes.

The relationship between SES and health is complex and multifaceted, and there is ongoing research aimed at better understanding this relationship and

identifying ways to address health disparities. Strategies to improve health outcomes for people with lower SES may include improving access to healthcare and other resources, reducing exposure to environmental toxins, and addressing the social determinants of health that contribute to health disparities.

## REFERENCES

1. Agarwal A. Social Classification: The Need to Update in the Present Scenario. *Indian J Community Med Off Publ Indian Assoc Prev Soc Med.* 2008;33(1):50–1.
2. Kumar N, Shekhar C, Kumar P, Kundu AS. Kuppuswamy's socioeconomic status scale-updating for 2007. *Indian J Pediatr.* 2007;74(12):1131–2.
3. World Health Organization. Handbook on health inequality monitoring with a special focus on low- and middle-income countries [Internet]. Geneva: World Health Organization; 2013 [cited 2023 Feb 21]. 105 p. Available from: <https://apps.who.int/iris/handle/10665/85345>
4. Cortinovis I, Vella V, Ndiku J. Construction of a socio-economic index to facilitate analysis of health data in developing countries. *Soc Sci Med* 1982. 1993;36(8):1087–97.
5. Joseph G, da Silva ICM, Fink G, Barros AJD, Victora CG. Absolute income is a better predictor of coverage by skilled birth attendance than relative wealth quintiles in a multicountry analysis: comparison of 100 low- and middle-income countries. *BMC Pregnancy Childbirth.* 2018;18(1):1–11.
6. Ankitha C, Srivatava D, Eshwar S, Jain V, Rekha D, Madhu N, et al. OVERVIEW OF SOCIO ECONOMIC STATUS SCALES IN INDIA. 2016;
7. Tiwari S, Kumar A, Kumar A. Development & standardization of a scale to measure socio-economic status in urban & rural communities in India. *Indian J Med Res.* 2005;122:309–14.
8. Ramesh Masthi NR, Gangaboraiah, Kulkarni P. An Exploratory Study on Socio Economic Status Scales in a Rural and Urban Setting. *J Fam Med Prim Care.* 2013;2(1):69–73.
9. Mishra D, Singh HP. Kuppuswamy Socioeconomic Status Scale - A Revision. *Indian J Pediatr.* 2003;70:273–4.
10. Bashar MdA. Modified BG Prasad Socioeconomic Status Scale: Updated for the Year 2022. *Indian Pediatr.* 2022 Oct;59(10):816–816.
11. Mohd Saleem S. Modified Kuppuswamy socioeconomic scale updated for the year 2019. *Indian J Forensic Community Med.* 2019;6(1):1–3.
12. Mangal A, Kumar V, Panesar S, Talwar R, Raut D, Singh S. Updated BG Prasad socioeconomic classification, 2014: A commentary. *Indian J Public Health.* 2015;59:42–4.
13. Prasad BG. Social classification of Indian families. *J Indian Medical Assoc.* 1961;37:250–1.
14. Rutstein S. The DHS Wealth Index: Approaches for Rural and Urban Areas. 2008;
15. Rutstein SO, Staveteig S. Making the Demographic and Health Surveys Wealth Index Comparable.
16. Gaur KL. Socio-Economic Status Measurement Scale: Thirst Area With Changing Concept For Socio-Economic Status. *Int J Innov Res Dev [Internet].* 2013 Sep 30 [cited 2023 Feb 21];

Available from:  
[https://internationaljournalcorner.com/index.php/ijird\\_ojs/article/view/133898](https://internationaljournalcorner.com/index.php/ijird_ojs/article/view/133898)

17. Howe LD. The wealth index as a measure of socio-economic position [Internet] [doctoral]. London School of Hygiene & Tropical Medicine; 2009 [cited 2023 Feb 21]. Available from:  
<https://researchonline.lshtm.ac.uk/id/eprint/768490/>

18. Deshmukh H, Dobriyal M, Tandel MB, Gunaga R, Sharma OP, Garde YA, et al. Development and Standardization of an Innovative Scale for Measuring the Socio-Economic Status of Agroforestry Farmers in South Gujarat, India. *Sustainability*. 2023;15(3):2691.

19. Georganos S, Gadiaga AN, Linard C, Grippa T, Vanhuyse S, Mboga N, et al. Modelling the Wealth Index of Demographic and Health Surveys within Cities Using Very High-Resolution Remotely Sensed Information. *Remote Sens*. 2019;11(21):2543.

20. Wealth index and maternal health care: Revisiting NFHS-3 [Internet]. ResearchGate. [cited 2023 Feb 21]. Available from:  
[https://www.researchgate.net/publication/281683338\\_Wealth\\_index\\_and\\_maternal\\_health\\_care\\_Revisiting\\_NFHS-3](https://www.researchgate.net/publication/281683338_Wealth_index_and_maternal_health_care_Revisiting_NFHS-3)

21. Howe LD. Handbook on Health Inequality Monitoring. *Int J Epidemiol*. 2014;43(4):1345–6.

# Longitudinal data analysis methods – a primer

Nishaant Ramasamy, Ruben Raj, Namita Patel, Palanivel C

Department of Preventive and Social Medicine,

Jawaharlal Institute of Postgraduate Medical Education and Research (JIPMER) Puducherry, India



## CORRESPONDING AUTHOR

Palanivel C, Additional Professor, Department of Preventive and Social Medicine, Jawaharlal Institute of Postgraduate Medical Education and Research (JIPMER) Puducherry, India

E Mail ID: [palaniccm@gmail.com](mailto:palaniccm@gmail.com)

## CITATION

Ramasamy N, Raj R, Patel N, C Palanivel. Longitudinal data analysis methods – a primer.

EFI Bulletin. 2023;4(1):12-17. <https://doi.org/10.56450/EFIB.2023.v3i01.003>

## ARTICLE CYCLE

Received: 14/03/2023; Accepted: 20/03/2023; Published: 31/03/2023

This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

©The Author(s). 2023 Open Access

Medical research frequently employs studies where the outcome/dependent variable is monitored or assessed repeatedly over different time points among the same study participants. Such repeatedly measured outcome over time are referred to as longitudinal data.

In a **longitudinal study**, participant's outcomes and possibly treatments or exposures are collected at multiple follow-up times (**on the same set of participants**). For e.g., in a group of HIV patients, every month their CD4 count or viral load is monitored.

### Properties of Longitudinal study design:

- Having repeated observations on individuals allows for studying the change in the outcome over time.
- Can be used to assess both the changes in the outcome over time within individuals and between the individuals.
- Use of sophisticated data statistical techniques because the repeated observations are usually correlated.
- Since certain types of correlation structures are likely to arise from longitudinal data, correlation must be accounted to obtain valid inference.
- Study participants act as their own control which economizes on participants and reduces unexplained variability in the response.

- Handles missing data and irregularly spaced measurement robustly (only if Mixed effect modeling was used)
- The dataset used for analyzing the effect of repeated observations on the outcome of interest should be in a **long format**.

In this article, we review the different statistical techniques used for analyzing longitudinal data with the help of research questions.

### There are different types of longitudinal studies:

- Panel study: Longitudinal follow up of subjects over a period of time, repeated measurements. Panel can be at the level of household or at individual level. E.g.: Longitudinal Ageing Study in India (LASI) – household panel study
- Cohort study: A group of people with similar characteristics will be followed up over a period of time. E.g.: Indian Migration study.
- Trials- RCT, Pre/ Post interventions

Different Modelling strategies used in the analysis of longitudinal data in which the outcome variable is a continuous variable are:

### Traditional Methods:

- Repeated-measures ANOVA

**Newer Methods:**

1. Generalized Estimating Equations (GEE) Models
2. Linear Mixed-effects Models

**Repeated Measures ANOVA:**

Repeated-measures ANOVA (RMANOVA) is a statistical method that has been extensively used for examining differences in the mean values of the dependent/outcome variable over a period of time. There are two types of repeated measures ANOVA – one-way and two-way repeated measures ANOVA. There are certain assumptions under which the repeated measures ANOVA can be used – they are as follows

**Assumptions:**

1. **Normality:** The outcome variable should follow the *normal distribution*.
2. **Sphericity:** The variances of the differences between all possible pairs of conditions or time points should be equal. Mauchly's test of sphericity is used for testing sphericity in RMANOVA.
3. **Homogeneity of variance:** The variances of the groups should be equal.
4. **Independence:** Observations should be independent of each other.
5. **Random sampling technique** should be employed.
6. **Time points of assessment should be constant.**
7. There should be **no missing values** in the outcome variable.

One-way repeated-measures ANOVA can be used to analyze differences in an outcome variable sequentially measured at different time points in the same individuals receiving the same intervention. The null hypothesis says that all related means measured at different time points are equal.

**Example 1**

A researcher wants to find the effect of a new topical anti-glaucoma medication on ***Intra-ocular pressure (IOP)*** in a group of individuals assessed at different time points.

**Timepoints for assessment of IOP** – prior to administration, 1<sup>st</sup> hour, 2<sup>nd</sup> hour, 3rd-hour post medication use.

The following figure shows the results of the one-way repeated measures ANOVA on the STATA

software. The descriptive statistics show there is a significant difference in the mean IOP values measured at different time points. In one-way RMANOVA, the model shows a significant decrease in the mean IOP values assessed over different time points with a statistically significant p-value of <0.001. Greenhouse-Geisser epsilon value is taken as the p-value even in the absence of Mauchly's test of sphericity.

. tabstat IOP, statistics(mean sd) by(Time)

Summary for variables: IOP

Group variable: Time

Time	Mean	SD
1	32.23529	6.3963
2	27.82353	5.534851
3	23.85294	5.700252
4	19.23529	6.3963
Total	25.78676	7.661745

anova IOP Sno Time, repeated(Time)

		Number of obs =	136	R-squared =	0.7987
		Root MSE =	4.01371	Adj R-squared =	0.7256
Source	Partial SS	df	MS	F	Prob>F
Model	6329.9412	36	175.8317	10.91	0.0000
Sno	3188.5662	33	96.623217	6.00	0.0000
Time	3141.375	3	1047.125	65.00	0.0000
Residual	1594.875	99	16.109848		
Total	7924.8162	135	58.702342		

Between-subjects error term: Sno  
Levels: 34 (33 df)  
Lowest b.s.e. variable: Sno

Repeated variable: Time

Source	df	F	Prob > F			
			Regular	H-F	G-G	Box
Time	3	65.00	0.0000	0.0000	0.0000	0.0000
Residual	99					

Typically, researchers seek to compare two or more groups (e.g., intervention arm with control arm) in which the outcome is serially and repeatedly evaluated over time. Thus, there are two variables of interest in this repeated-measures design (time and treatment). This statistical technique is known as two-way repeated-measures ANOVA.

The two-way repeated measures ANOVA model simultaneously tests several null hypotheses:

1. the main effect of time - all means at different time points are the same;

2. the main effect of treatment - all means in different treatment groups are the same;
3. there is no interaction between treatment and time

### Example 2

A researcher wants to find the effect of a palliative care intervention on Stage 4 breast cancer patients for a period of 2 months compared to usual care on the quality of life of patients assessed using Functional Assessment of Chronic Illness Therapy – Sp scale (FACIT-Sp)

**Timepoints for assessment of FACIT-Sp** – 1<sup>st</sup> month, 2<sup>nd</sup> month

The following figure shows the results of the two-way repeated measures ANOVA on the STATA software. The descriptive statistics show there is a significant increase in the mean FACIT-Sp scores in the intervention arm compared to the control arm measured at different time points. In two-way RMANOVA, the model shows a significant increase in the mean FACIT-Sp scores in the intervention arm compared to the control arm assessed over different time points with a statistically significant p-value of 0.0097.

```
. tabstat Sp if Group==0 , statistics(mean sd) by(time)
```

Summary for variables: Sp  
Group variable: time (time)

time	Mean	SD
1	93.46667	20.97913
2	89.4	21.8135
Total	91.43333	21.12962

```
. tabstat Sp if Group==1 , statistics(mean sd) by(time)
```

Summary for variables: Sp  
Group variable: time (time)

time	Mean	SD
1	87	18.10683
2	95.4	18.16905
Total	91.2	18.32729

```
anova Sp Group/ PATIENTDETAILS|Group time Group#time, repeated( time)
```

Source	Partial SS	df	MS	F	Prob>F
Model	20569.717	31	663.53925	8.77	0.0000
Group	.81666667	1	.81666667	0.00	0.9732
PATIENTDE-S Group	19915.667	28	711.27381		
time	78.4166667	1	78.4166667	0.93	0.3430
Group#time	582.81667	1	582.81667	7.70	0.0097
Residual	2119.2667	28	75.688095		
Total	22688.983	59	384.55904		

Between-subjects error term: PATIENTDE-S|Group  
Levels: 30 (28 df)  
Lowest b.s.e. variable: PATIENTDE-S  
Covariance pooled over: Group (for repeated variable)

Repeated variable: time

Huynh-Feldt epsilon = 1.0370  
\*Huynh-Feldt epsilon reset to 1.0000  
Greenhouse-Geisser epsilon = 1.0000  
Box's conservative epsilon = 1.0000

Source	df	F	Prob > F			
			Regular	H-F	G-G	Box
time	1	0.93	0.3430	0.3430	0.3430	0.3430
Group*time	1	7.70	0.0097	0.0097	0.0097	0.0097
Residual	28					

### Strengths:

1. Economical - requires relatively few participants
2. Model estimation is done on raw data

### Limitations:

1. RMANOVA cannot handle other predictor variables that vary over time.
2. Missing data – excludes the entire case from the analysis. Only complete case analysis (CCA) is performed in RMANOVA.
3. RMANOVA assumes each study participant to be measured at the same time points and does not allow for data collected at unequally spaced time intervals among different study participants.
4. Effect measure cannot be estimated.

The following reasons are proposed for the need for other methods of analysis:

#### 1. Irregularly timed data

Longitudinal studies typically advocate for frequent data collection trips. Yet study participation frequency, and total study visits fluctuate due to scheduling difficulties and drop out of the study participants.

#### 2. Missing data

Missing data can reduce the statistical power of a study and increase its bias. When the sample size of a study rises or when the variability of the study's outcome measure is accurate; statistical power increases. Sadly, missing data affects sample size and variability negatively. First, analyses that remove participants with missing values mistakenly limit the sample size of the study, so potentially diminishing its statistical power. Second, the variability of the study's outcome measure is wrongly understated when participants who would have had extreme data values opt-out.

#### 3. Effect measure estimation

Newer statistical techniques such as Generalized Estimating Equation (GEE) and Linear Mixed Effects Regression model (LMER) account for the

aforementioned reasons and provide the researcher with robust effect estimates.

There are two prevalent contemporary methods that utilize the adaptability of regression models while addressing for non-independent measures.

The first approach will specify the mean response of the outcome (also known as a marginal approach or population-average approach), while utilizing a technique that is valid in the presence of correlated observations to estimate the regression parameters. Often, a generalized estimating equation (GEE) is combined with so-called robust standard errors to assess the precision of an estimate when estimating the parameters of a marginal model.

Using random intercepts and/or slopes, the second method fully specifies the distribution of the outcome, going beyond the mean change in the outcome variable. These models have also been referred to as mixed-effect models, multilevel models, and hierarchical models.

	1	2	3	4	5
1	■				
2		■			
3			■		
4				■	
5					■

Independent

	1	2	3	4	5
1	■				
2		■			
3			■		
4				■	
5					■

Compound Symmetry

	1	2	3	4	5
1	■				
2		■			
3			■		
4				■	
5					■

Autoregressive

	1	2	3	4	5
1	■				
2		■			
3			■		
4				■	
5					■

Unstructured

Figure 1: Working correlation structure

- Independent:** Uncorrelated measurement
- Compound symmetry:** assumes all correlations are equal
- Autoregressive:** assumes correlations to decrease as the time interval between the measurements increases.
- Unstructured:** No assumptions, allows all correlations to differ
- Large Sample Size:** Assumes a large sample size - normally distributed, and standard errors are valid.
- Assumes missing values to be Missing completely at Random (MCAR)

### Example 3

A researcher wants to find the effect of the integration of early palliative care intervention with usual care on cervical patients undergoing surgery for a period of 4 months compared to usual care alone on the quality of life of patients assessed

### 1. Generalized Estimating Equation (GEE)

A generalized estimating equation can be applied to determine the regression parameters for the expected (mean) response of an outcome given a set of independent variables while taking repeated measurements over time into consideration. The following are the assumptions of GEE

- Independence:** The observations within a cluster - independent of each other.
- Model should be specified** such as the **family** of the distribution (eg., Gaussian, Binomial, Poisson, Gamma) and **identity** used (eg., identity, log, logit)
- Working correlation structure** appropriate for the research question under study should be used – independent, compound symmetry or exchangeable, autoregressive, unstructured.(Figure 1)

using Functional Assessment of Chronic Illness Therapy – Palliative (FACIT-Pal). Out of 44 patients, 43 patients completed follow-up in the intervention and 41 in the control arm.

**Timepoints for assessment of IOP** – prior to surgery, 1<sup>st</sup> month, 4<sup>th</sup> month

The following figure shows the results of the Generalized estimating equation (GEE) on the STATA software. The descriptive statistics show there is a significant increase in the mean FACIT-Pal scores in the intervention arm compared to the control arm measured at different time points.

Gaussian was chosen as the family, with the identity function. Autoregressive was the working correlation structure used in this model.

*In GEE, the model shows an increase of 15.2 points (95% CI 12.2 to 18.2) in the mean FACIT-Pal scores in the intervention arm compared to the control arm*

```

mixed FACITPal_Total _Randomization || _Period_of_Recruit, reml
Performing EM optimization ...

Performing gradient-based optimization:
Iteration 0: log restricted-likelihood = -975.54359
Iteration 1: log restricted-likelihood = -975.54359

Computing standard errors ...

```

```

Mixed-effects REML regression
Number of obs      =      260
Group variable: _Period_of_R~t
Number of groups  =        3
Obs per group:
    min =        84
    avg =      86.7
    max =        88
Wald ch2(1)      =     142.07
Prob > chi2       =     0.0000
Log restricted-likelihood = -975.54359

```

FACITPal_Total	Coefficient	Std. err.	z	P> z	[95% conf. interval]
_Randomization	15.05461	1.263051	11.92	0.000	12.57908 17.53015
_cons	112.1119	11.88593	9.43	0.000	88.81589 135.4079

over the baseline values with a statistically significant p-value of <0.001.

```

.tabstat FACITPal_Total if _Randomization ==0 , statistics(mean sd) by( _Period_of_Recruit )

Summary for variables: FACITPal_Total
Group variable: _Period_of_Recruit (40_Period_of_Recruit)

```

_Period_of_Recruit	Mean	SD
1	93.81818	5.226376
2	112.3864	7.530453
3	130.2195	11.98648
Total	111.7209	17.11906

```

.tabstat FACITPal_Total if _Randomization ==1 , statistics(mean sd) by( _Period_of_Recruit )

Summary for variables: FACITPal_Total
Group variable: _Period_of_Recruit (40_Period_of_Recruit)

```

_Period_of_Recruit	Mean	SD
1	99.65909	13.1555
2	138.9773	8.390162
3	142.8372	6.531718
Total	127.0382	21.88234

```
. xtset _Participant_No _Period_of_Recruit
```

Panel variable: \_Participant\_No (unbalanced)  
 Time variable: \_Period\_of\_Recruit, 1 to 3  
 Delta: 1 unit

```

.xtgee FACITPal_Total _Randomization, family(gaussian) link(identity) corr(ar 1) vce(robust)

Iteration 1: tolerance = .00560174
Iteration 2: tolerance = .00017111
Iteration 3: tolerance = 5.282e-06
Iteration 4: tolerance = 1.631e-07

```

```

GEE population-averaged model
Number of obs      =      260
Group and time vars: _Participant_No _Period_of_Recruit
Number of groups  =        88
Family: Gaussian
Link: Identity
Correlation: AR(1)
Scale parameter = 383.6999
Wald ch2(1)      =     97.13
Prob > chi2       =     0.0000
(Std. err. adjusted for clustering on _Participant_No)

```

FACITPal_Total	Coefficient	std. err.	z	P> z	[95% conf. interval]
_Randomization	15.22297	1.544592	9.86	0.000	12.19562 18.25031
_cons	111.7193	.931943	119.88	0.000	109.8927 113.5459

### Strengths:

1. GEE can account for other predictor variables in the model.
2. It can handle variation in time points at which the outcome is measured.
3. GEE assumed missing values as MCAR & provides an estimate

### Limitation:

1. Only Population average effects are estimated.
2. Model diagnostics cannot be compared.

### 3. Linear Mixed Effects Regression model (LMER)

Approach used by mixed-effect models is conceptually different than marginal models to account for an independence of repeated measurements over time. GEE focus on the mean outcome changes whereas mixed-effect models provide a fully specified model for the multivariate distribution of the repeatedly measured outcome.

#### Linear mixed effect model assumptions:

1. Regression relationship between covariates and repeated response.
2. Correlation structure of the repeated response should be specified.
3. Fixed or Random effects model should be specified.
4. Estimation of cluster-specific effects – interindividual variability and cluster-specific predictions.
5. Gives both population-averaged and subject-specific estimates.
6. Assumes missing values to be Missing at Random (MAR)

Random intercept model was chosen with the period of recruitment as the time-varying predictor variable.

Using the same example 3 - In LMER, the model shows an increase of 15.0 points (95% CI 12.5 to 17.5) in the mean FACIT-Pal scores in the intervention arm compared to the control arm over the baseline values with a statistically significant p-value of <0.001.

#### Strengths:

1. LMER can account for other predictor variables in the model.
2. It can handle variation in time points at which the outcome is measured.
3. LMER assumed missing values as MAR & provides an estimate.
4. Model diagnostics can be compared.

### Limitation:

1. Complex, computationally intensive.

Table gives the difference between the Generalized estimating equation and Linear mixed effect model for estimating the change in mean value of the outcome variable measured over different time points.

**Table 1: Difference between GEE and MER**

Characteristics	GEE	MER
Assumptions	Working correlation structure	Random effect structure
Parameter estimation	Population-averaged effect	1. Population-averaged 2. Subject-specific effects
Interpretation	Similar to standard regression model	More complex
Handling of missing data	Robust; assumes MCAR	Assumes MAR
Computational complexity	1. Simple 2. Computationally efficient 3. Quasi-likelihood method	1. Complex 2. Computationally intensive 3. Maximum likelihood methods

**CONCLUSION**

We discussed the different statistical techniques employed in the analysis of longitudinal data. This article focussed mainly on the statistical techniques used for estimating the mean change in the outcome variable over a time period. Ranging from traditional methods such as repeated measures ANOVA to newer methods such as Generalized estimating equation (GEE) and Linear mixed effects regression model (LMER) discussing about the assumptions, strengths, and limitations of each model with the help of the research question. This article acts as a primer in providing the techniques used for the analysis of longitudinal continuous outcome variable. Further in-depth understanding of modern statistical methods is recommended to answer a particular research question.

**REFERENCES**

1. Mausner, J. S., & Kramer, S. (1985). Epidemiology: An introductory text
2. Schober P, Vetter TR. Repeated measures designs and analysis of longitudinal data: If at first you do not succeed—try, try again. *Anesthesia and analgesia*. 2018;127(2):569-575.
3. Garcia TP, Marder K. Statistical approaches to longitudinal data analysis in neurodegenerative diseases: Huntington's disease as a model. *Current neurology and neuroscience reports*. 2017;17(2):1-14.
4. Ballinger GA. Using generalized estimating equations for longitudinal data analysis. *Organizational research methods*. 2004;7(2):127-50.
5. Locascio JJ, Atri A. An overview of longitudinal data analysis methods for neurological research. *Dementia and geriatric cognitive disorders extra*. 2011;1(1):330-57.

## COMMENTARY

# Adaptive platform trials present a new frontier for epidemiologists in the era of Covid-19

Manya Prasad, Umesh Kapil

Department of Epidemiology, Institute of Liver and Biliary Sciences, New Delhi  
National Academy of Medical Sciences, New Delhi



## CORRESPONDING AUTHOR

Dr Manya Prasad, Department of Epidemiology, Institute of Liver and Biliary Sciences, New Delhi  
E Mail ID: [manya.2311@gmail.com](mailto:manya.2311@gmail.com)

## CITATION

Prasad M, Kapil U. Adaptive platform trials present a new frontier for epidemiologists in the era of Covid-19. *EFI Bulletin*. 2023;4(1):18-19. <https://doi.org/10.56450/EFIB.2023.v3i01.004>

## ARTICLE CYCLE

Received: 16/03/2023; Accepted: 21/03/2023; Published: 31/03/2023  
This work is licensed under a [Creative Commons Attribution 4.0 International License](#).  
©The Author(s). 2023 Open Access

The Covid-19 pandemic created an unprecedented opportunity for the conduct of a specific type of randomized controlled trials called platform trials. Indeed, the most valuable and practice-changing information on therapies for Covid-19 came from trials such as RECOVERY, SOLIDARITY, RECAP-MAP etc, all of which employ the platform trial design.(1,2,3)

A platform trial refers to a type of randomized clinical trial that allows for multiple interventions to be evaluated simultaneously.(4) They are designed with the flexibility of having new interventions added to the platform after the trial has begun, which constitutes a key difference when compared to conventional trials. Platform trials are typically intended to be perpetual, lasting a very long time, with different therapies added to the trial as evidence evolves. Multiple interventions can leave and enter the trial, a feature that was utilized for investigating therapies in Covid-19 since various interventions needed to be evaluated simultaneously. The following are some methodological issues that need to be accounted for while conducting platform trials.

### Master protocol

A platform trial adheres to a master protocol, which is a set of documents that contains the standard

operative procedures of the trial and guides how the platform trial is conducted. The IRB or ethics approval for this type of trial involves submission of the master protocol, and as new therapies are added, these are included as amendments rather than as a new protocol.(5)

### Response adaptive randomization

Most Covid-19 trials employed a response-adaptive trial design. Herein, the allocation ratio is adapted over time based on cumulative interim results in favour of the interventions that are performing better. Therefore, the allocation ratio changes in response to the evolving evidence.(4,6)

### Interim vs overall trial results

A key concern with platform trials is whether pre-specified plans were in place for interim analysis and whether the statistical tools were applied equally to all treatments. Interim analyses are planned to increase the efficiency of the trial but may be problematic if greater emphasis is placed on the interim rather than the overall trial results. The Data Safety and Monitoring Board must be responsible for avoidance of erroneous conclusions from the interim analyses.(4,7)

## Concurrently randomized vs non-concurrently randomized controlled groups

A concurrently randomized control group refers to patients that were randomized to the control arm at the same time as those that were randomized to an intervention. One of the issues with platform trials is whether to include non-concurrently randomized participants or limit the analysis to concurrently randomized patients since temporal variations between the two may bias the results. For instance, change in patient characteristics over time or change in standard of care over time may have to be accounted for.(8)

## Influence of other clinical trials

Evidence generated from other clinical trials may impact the platform trial in a couple of ways. Firstly, it may affect the decision of participants to take part in the platform trial. In addition, the standard of care may change due to the evidence generated by other trials. The changing standard of care poses a challenge in interpretation. However, there are statistical methods that can account for this after consideration of whether the control event rate changed according to differing standard of care. There are several sources of efficiency that the platform trial design possesses; these have been tapped for investigating therapy options for Covid-19. The pandemic has shown us the power of the platform trial approach to rapidly address pressing questions about the effectiveness of therapies. There is no doubt, however, that there is a high bar to entry and many nuances about the adaptive approach and the associated Bayesian statistics are unfamiliar even to experienced trialists.

Requirement of long-term funding and appropriate infrastructure presents further challenges. But as we forward and try to expand this approach, when appropriate, into other disease areas or in Covid-19, trialists around the world will hopefully begin to develop the infrastructure required for conducting this type of trial. Seeing that there is also a lot of opportunity for manipulation or 'spin' of trial results, the methods community should pay due time and attention in understanding this trial design.

## REFERENCES

1. RECOVERY Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, et al. Dexamethasone in Hospitalized Patients with Covid-19. *N Engl J Med.* 2021 Feb 25;384(8):693-704.
2. WHO Solidarity Trial Consortium, Pan H, Peto R, Henao-Restrepo AM, Preziosi MP, Sathiyanamoorthy V, et al. Repurposed Antiviral Drugs for Covid-19 - Interim WHO Solidarity Trial Results. *N Engl J Med.* 2021 Feb 11;384(6):497-511.
3. REMAP-CAP Investigators, Gordon AC, Mouncey PR, Al-Beidh F, Rowan KM, Nichol AD, et al. Interleukin-6 Receptor Antagonists in Critically Ill Patients with Covid-19. *N Engl J Med.* 2021 Apr 22;384(16):1491-1502.
4. Dimairo M, Pallmann P, Wason J, Todd S, Jaki T,; ACE Consensus Group. The Adaptive designs CONSORT Extension (ACE) statement: a checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design. *BMJ.* 2020 Jun 17;369:m115.
5. Chow SC, Chang M. Adaptive design methods in clinical trials - a review. *Orphanet J Rare Dis.* 2008 May 2;3:11.
6. Pallmann P, Bedding AW, Choodari-Oskooei B, Dimairo M, Flight L, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Med.* 2018 28;16(1):16-29.
7. Kairalla JA, Coffey CS, Thomann MA, Muller KE. Adaptive trial designs: a review of barriers and opportunities. *Trials.* 2012;13:145.
8. Park JJ, Thorlund K, Mills EJ. Critical concepts in adaptive clinical trials. *Clin Epidemiol.* 2018;10:343-351.

# Strengthening provisions for breast cancer screening in India

Ankan Mukherjee Das

Amity Institute of Public Health, Amity University Uttar Pradesh, Noida, India

## CORRESPONDING AUTHOR

Ankan Mukherjee Das, Senior Researcher, Amity Institute of Public Health, Amity University Uttar Pradesh, Noida, India



E Mail ID: [amdas@amity.edu](mailto:amdas@amity.edu)

## CITATION

Das AM. Strengthening provisions for breast cancer screening in India. *EFI Bulletin*. 2023;4(1):20-22.

<https://doi.org/10.56450/EFIB.2023.v3i01.005>

## ARTICLE CYCLE

Received: 20/02/2023; Accepted: 20/03/2023; Published: 31/03/2023

This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

©The Author(s). 2023 Open Access

## ABSTRACT

Although population-based clinical breast exam has been initiated in India utilizing the exiting human resource of frontline workers, there exists a tremendous lacuna in facilitating adequate training and developing necessary sustainable systems for screening, referrals and treatment of suspected and confirmed cases along with adequate follow up. Understanding and predicting the overall risk of breast cancer development in young women through identifying genetic signatures of susceptibility during screening is equally important and thus focus needs to be paid in implementing models incorporating genetic risk analysis and with digital structures involving provisions not just for increasing awareness and screening rates but also towards supporting active tracking and follow-up of cases during referral and treatment.

## KEYWORDS

Breast Neoplasm; Screening; Primary Health Care, Digital Health; Health Service Delivery; Epidemiology

## INTRODUCTION

Once thought to be the disease of the urban women, breast cancer (BC) has been equally affecting the rural populations with about 2.4 million cases and 685,000 deaths in 2020, making it the predominant cancer diagnosed among women world over.(1) In India, it is now the most common cancer affecting women, with an estimated 178,361 cases and 90408 deaths in 2020.(1) However, several cases continue to remain undetected and this number is an under-representation of the actual burden. Recent trends of breast cancer incidence suggest early-onset initiation of the disease, worryingly in early 20s to 40s, especially in India, characterized by advanced stage and aggressive disease, particularly the triple negative phenotype. Moreover, with the lacking awareness of breast cancer signs and symptoms among the general population, and almost non-existent population-wide active BC screening programmes with profound socio-economic

inequalities and cultural norms in India, majority women continue to present at an advanced metastatic stage, leading to increased disability and mortality.

Ergo the rapidly changing patterns of diet and lifestyle associated modulation of reproductive and developmental processes among the population along with the unforeseen early-life exposures to several environmental extremes including polluted air and water, the age-standardized rates and etiopathology of breast cancer have increasingly transformed. While, several established modifiable risk factors of this disease are known, these could not yet be completely attributable to Indian patients as compared to their western counterparts. This is due to the apparent dearth of region-specific population-wide causal studies and the evident variability of exposure to different risk factors in Indian women, thus rendering the existing models utilized for

clinically predicting risk(2) inadequate, and requiring integration, and evidence-based updating. Moreover, subsequent evaluation of risk among patients with BC signs and symptoms are not always performed at all Indian healthcare centres. Besides, the genetics, particularly carriers of breast cancer susceptibility; BRCA 1, BRCA 2 and other key high-penetrance gene mutations, independently stands as causal factors, but are not considered in majority of risk prediction models. Despite the fact that except for the BRCA genotype, other gene mutations are highly sporadic and potentially induced by the lifestyle and environment.(3)

The routine clinic procedure for BC diagnosis include an initial breast examination by a medical oncologist following or followed by USG and/or mammography and PET scans. And later, a biopsy to remove a portion of the affected breast tissue for pathological examination to confirm the type (invasive/metastatic or benign), BC phenotype based on the hormone-receptor status, stage and grade of the tumour, upon which the treatment decisions are also made. Breast self-examination (BSE), clinical breast examination (CBE), mammography and/or MRI are the exiting modalities through which early detection of the disease is possible, nevertheless, the understanding, and utilization of these screening options remains to be poor. Recent introduction of 3D breast tomosynthesis could potentially boost the overall detection rates, however, these are highly expensive, require additional logistics and are mostly inaccessible to women from low-resource settings.

The national programme for prevention and control of non-communicable diseases including cancer (NPCDCS),(4) has recently included provisions for education and training of front-line workers on competent practice of clinical breast examination of all women above 30 years for early identification of suspected cases with breast cancer signs and symptoms, and referral to the nearest health facility having access to cancer diagnostic and treatment services. The programme also encapsulates to elevate knowledge and awareness of breast cancer and importance of breast self-examination practice among both young and older women. However, the components of the program are currently unorganized and being piloted in a few specific regions of the country. Whilst the screening rates are found to have marginally improved according to the

MoHFW 2021 reports, no overall program evaluation have yet been carried out, and there still remains a humongous lacuna in CBE training of front-line workers along with detection, diagnosis and referral of suspected high-risk cases, from both the programme and individual end, including patient incompliance. There is also increasing concerns regarding access to diagnostic and comprehensive treatment options, which are typically long-term and mostly available at already overburdened tertiary care centres. Furthermore, the current programme with its existing activities alone have not helped in downstaging breast cancer cases and in alleviating the overall morbidity and mortality associated with the disease, as evident from the estimates of GLOBOCAN and India's national cancer registry programme.(1,5)

Digital technologies have rapidly evolved and are being developed across the globe for effective healthcare, and sustainable public health service delivery. For instance, the SMS appointment reminder system in the Philippines(6) and the Mobile-Alliance for Maternal Action which was implemented in India, Bangladesh, and South Africa(7) are models innovated for the same purpose. The iBreastExam is another highly predictive device, which works similar to a clinical breast exam.(8) This device is connected to a smart phone where the images are produced, can be saved and later compared, and used by a healthcare provider and/or front-line worker, for effective breast cancer screening in both rural and urban settings. Design and development of similar or incorporating existing devices and a mobile-health application in building a robust screening model involving multi-pronged strategies directed towards down staging breast cancer could accelerate awareness and active detection, including targeting vulnerable populations. The model could be integrated with the existing National and State health programme structure and help not only in detecting palpable tumors and predicting the overall risk of breast cancer development in women, segregating them as high and low-risk individuals based on genetic susceptibility and other parameters, but also provide reminder and access to clinical breast examinations by a trained ASHA/midwives along with facilitating adequate hands-on training and promoting self-examination through educational videos. This should alleviate the

breast cancer burden in low-income and middle-income countries such as India through adequately linked active referral systems, possibly leading to early treatment and improved prognosis. Digital health models for improving maternal and child health service provisions have been successfully implemented in several resource-limited areas of the country. For instance, the mobile-tablet based Maternal and Child Tracking System (MCTS), which integrates community-based electronic records of care instead of paper forms with SMS or pre-recorded voice messages to pregnant women, and their families reminding them that they are due for or are nonadherent to the scheduled care.(9) Another example is of BEMPU, a temperature monitoring watch implemented for early detection of hypothermic conditions in neonates,(10) which allows the opportunity for quick interventions including family oriented kangaroo-hold care. In the case of breast cancer, involving and training female family members and/or male partners in helping women carry out self-examinations could be beneficial in the long run. This has the potential to facilitate overall uptake of screening and healthcare practices to ameliorate early complications, and could also lead to better understanding, support and care of the patients, particularly focusing on rural populations.

## REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71(3):209-249.
2. Kim G, Bahl M. Assessing Risk of Breast Cancer: A Review of Risk Prediction Models. *J Breast Imaging.* 2021;3(2):144-155.
3. Nickels S, Truong T, Hein R, et al. Evidence of gene-environment interactions between common breast cancer susceptibility loci and established environmental risk factors. *PLoS Genet.* 2013;9(3):e1003284.
4. Directorate General of Health Services. National Program for Prevention and Control of Cancer, Diabetes, Cardiovascular Diseases and Stroke: Training module for medical officers for prevention, control and population level screening of hypertension, diabetes and common cancers (oral, breast and cervical). New Delhi: Ministry of Health & Family Welfare, 2017.
5. Mathur P, Sathishkumar K, Chaturvedi M, et al. Cancer Statistics, 2020: Report From National Cancer Registry Programme, India. *JCO Glob Oncol.* 2020;6:1063-1075.
6. Coleman J, Eriksen J, Black V, et al. The Mobile Alliance for Maternal Action Text Message-Based mHealth Intervention for Maternal Care in South Africa: Qualitative User Study. *JMIR Hum Factors.* 2020;7(2):e14078. Published 2020 Jun 29.
7. Sarmiento, Raymond Francis et al. "Design and implementation of SMS-based enhancements to a community-based health information management system: final technical report." (2011). Available from:[https://www.researchgate.net/publication/276837374\\_Design\\_and\\_Implementation\\_of\\_SMS-based\\_Enhancements\\_to\\_a\\_Community-based\\_Health\\_Information\\_Management\\_System](https://www.researchgate.net/publication/276837374_Design_and_Implementation_of_SMS-based_Enhancements_to_a_Community-based_Health_Information_Management_System). Accessed on 20/03/2023
8. Mango VL, Olasehinde O, Omisore AD, et al. The iBreastExam versus clinical breast examination for breast evaluation in high risk and symptomatic Nigerian women: a prospective study. *The Lancet Global Health.* 2022; 10(4), e555–e563.
9. Negandhi P, Chauhan M, Das AM, et al. Computer tablet-based health technology for strengthening maternal and child tracking in Bihar. *Indian J Public Health.* 2016;60(4):329-333.
10. Lei D, Tan K, Malhotra A. Decreasing Hypothermia-Related Escalation of Care in Newborn Infants Using the BEMPU TempWatch: A Randomised Controlled Trial. *Children.* 2021; 8(11):1068.

## ORIGINAL ARTICLE

# Latent class analysis in a complex sampling design – an advanced modelling approach in Epidemiology

Ankita Dey<sup>1</sup>, Diganta Mukherjee<sup>2</sup>, Sugata Sen Roy<sup>3</sup>

<sup>1</sup>National Institute of Tuberculosis and Respiratory Diseases, New Delhi;

<sup>2</sup>Indian Statistical Institute, Kolkata, West Bengal;

<sup>3</sup>Department of Statistics, University of Calcutta, Kolkata, West Bengal

## CORRESPONDING AUTHOR

Ankita Dey, National Institute of Tuberculosis and Respiratory Diseases (NITRD), An Autonomous Institute under Ministry of Health & Family Welfare, Sri Aurobindo Marg, New Delhi -110030

E Mail ID: [ankitadey14@gmail.com](mailto:ankitadey14@gmail.com)



## CITATION

Dey A, Mukherjee D, Roy SS. Latent class analysis in a complex sampling design – an advanced modelling approach in Epidemiology. *EFI Bulletin*. 2023;4(1):23-29. <https://doi.org/10.56450/EFIB.2023.v3i01.006>

## ARTICLE CYCLE

Received: 16/02/2023; Accepted: 21/03/2023; Published: 31/03/2023

This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

©The Author(s). 2023 Open Access

## ABSTRACT

In an Epidemiological study, researchers are often interested to identify population subgroups according to variables relevant to the context of the study. The subgroups may be identified according to some observed characteristics of the underlying disease or subgroups may be defined according to an unobserved or latent profile of patients. Latent Class Analysis (LCA), identifies hidden subgroups among the respondents, based upon the response pattern of a respondent to various surveys. There are certain assumptions for a traditional latent class model to work in a particular scenario. Use of traditional models in case of violation of one of these assumptions would lead to increased misclassification rate and biased estimates of the parameters. The present study provides a modified approach of latent class modelling, when one of the underlying assumptions of traditional latent class models is relaxed. Traditional models of LCA are compared with the newly developed models under more complex sampling designs.

## KEYWORDS

Latent Class Analysis; Complex Sampling Design; Epidemiology; Statistical Methodology

## INTRODUCTION

In an Epidemiological study, researchers are often interested to identify population subgroups according to variables relevant to the context of the study. The subgroups may be identified according to some observed characteristics or features of the underlying disease or subgroups may be defined according to an unobserved or latent profile of patients. Latent Class Analysis (LCA), identifies hidden subgroups among the respondents, based upon the response pattern of a respondent to various surveys. It was introduced in 1950 by Paul F. Lazarsfeld (Lazarsfeld, 1950), who performed LCA for building typologies (or clustering) based on dichotomous observed variables. The estimation of

model parameters using maximum likelihood approach was proposed later on (Goodman, 1974; Vermunt & Magidson 2007).

The importance of latent class analysis in epidemiological modelling has been explored by several researchers in the past few decades. Importance of latent class analysis in chronic disease epidemiology is discussed by Kaldor & Clayton (1985). In their study, the authors provided a general description of the LCA methodology as a statistical methodology in epidemiology and indicated its role in logistic analysis of categorical epidemiological data. The paper also indicated the need for further exploration of the method when the underlying assumptions of standard LCA model are relaxed.

Koukounari et al (2021) discussed the scenarios under which latent class models are suitable in the context of the evaluation of existing schistosomiasis diagnostics.

There are certain assumptions for a traditional latent class model to work in a particular scenario. Use of traditional models in case of violation of one of these assumptions would lead to increased misclassification rate and biased estimates of the parameters (Koukounari et al, 2021). It has been noted that in epidemiological modelling, researches focusing on modifications or revisions of traditional LCA models, to fit into the context where one of the model assumptions does not hold, are comparatively less explored. The present study provides a modified approach of latent class modelling, when one of the underlying assumptions of traditional latent class models is relaxed. The statistical analysis described in this study provides a methodological advancement to the existing modelling approaches of Epidemiology. Traditional models of LCA are compared with the newly developed models under more complex sampling designs and their performances are compared in the context of a secondary data set.

## MATERIAL & METHODS

Latent class analysis (LCA) is considered to be an equivalent methodology for Factor Analysis, typically used for dichotomous or polytomous variables (Biemer, 2011, Skrondal and Rabe Hesketh, 2002). The parameters of interest in a typical problem of latent class analysis are the unobserved proportion or size of the latent classes and the conditional item-response probabilities given the membership in a latent class. Based on the observed data on manifest variables, LCA provides a classification among the population. Manifest variables are called the 'indicators' of a particular latent class. The basis of the analysis lies in modelling the relationship between the latent variable and its indicators. LCA identifies unobservable (latent) subgroups within a population based on individuals' response to different categorical observed variables. It is a technique which explains the relationships between manifest or observed variables (may be dichotomous or polytomous) with respect to some unobserved or latent variables (may also be dichotomous or polytomous) on the basis of data obtained in various kinds of complex surveys. One of the assumptions of a latent class model is that the underlying sample of individuals consists of  $n$  units sampled without

replacement from a large population of  $N$  units using simple random sampling. Extension in models of latent class analysis for more complicated sampling designs has drawn attention of several researchers in recent times (Patterson, Dayton & Graubard, 2002; Asparouhov & Muthén, 2005; Vermunt & Magidson, 2007; Vaughn et. al., 2011; Pennoni and Nakai, 2018, 2019.). The impact of sampling scheme in the process of latent class estimation is now considered an active area of research. Scarcity of models in the literature of latent class models incorporating complex sampling characteristics viz sampling weight, stratification etc was discussed in Patterson, Dayton & Graubard (2002). They proposed such an extension by introducing sample weights into the latent class estimation algorithm. A two-class solution based on indicators regarding food consumption was named as "regular" and "nonregular" vegetable consumption groups in their study. A pseudo-likelihood function was developed based on the sample weights available for each respondent. For ' $n$ ' number of individuals with response pattern  $Y_i$  and membership  $c_i$  in one of  $L$  number of latent classes, the usual latent class parameters are latent class size ( $\theta_l$ ) and the item-response probabilities ( $\alpha_{ljr}$ ). In the following equation,  $i$  stands for an individual,  $i=1,2,\dots,n$ ;  $l$  stands for the latent class,  $l=1,2,\dots,L$ ;  $j$  stands for the item in a questionnaire,  $j=1,2,\dots,J$ ;  $r$  stands for the response option in  $j$ th item,  $r=1,2,\dots,R$ . For a dichotomous item,  $R=2$ . In case of more complex sampling schemes than SRS, the pseudo-likelihood based on sampling weights  $w_i$  is given as (Patterson, Dayton & Graubard, 2002),

$$\sum_{i=1}^n w_i \ln \left( \sum_{l=1}^L \theta_l Pr(Y_i|c_l) \right) = \sum_{i=1}^n w_i \ln \left( \sum_{l=1}^L \theta_l \prod_{j=1}^J \prod_{r=1}^{R_j} \alpha_{ljr}^{\delta_{ijr}} \right)$$

where the sample weights can be measured by eg. the reciprocal of inclusion probabilities. Vermunt & Magidson (2007) proposed a maximum likelihood approach to log-linear latent class analysis with sampling weights and probability formulation of latent class model with sampling weights which can accommodate larger number of indicators. However, these sampling weights may not be always readily available. Impact of a stratified random sampling on parameter estimates in a latent class framework without incorporating sample weight may constitute an interesting area of research. The present study attempts to describe the model of latent class

analysis for a group of individuals selected through stratified random sampling technique without incorporating sampling weights. Within each stratum, a new set of model parameters are defined in the context of a full likelihood latent class model without a log-linear parameterization. The stratified latent class model is compared with the traditional model with a sample selected through simple random sampling technique. The latent class model with stratified random sampling is compared with two traditional LCA models, viz., LCA with simple random sampling (SRS) design without log-linear parameterization (full likelihood) and LCA with simple random sampling (SRS) design with a log-linear parameterization. Stability of estimates of the model parameters and their standard errors are compared for different situations mentioned above. The effect of sampling scheme in estimation of parameters are explored by applying the models in the data set of risk factors related to cardiac events (Dey et. al., 2016; Das, 2016; Das, Lee and Mukhopadhyay, 2017).

#### Latent class analysis with probabilistic parameterization

There are different modelling approaches for latent class analysis (Biemer, 2011). In the present study, two of the approaches are used viz. probability model parameterization of the standard latent class model and the log-linear parameterized model, for comparison of model fit and parameter estimates. In the probabilistic parameterization of the latent class model, the likelihood function is expressed as a function of the marginal and conditional cell probabilities of the underlying multiway contingency table. A probabilistic parameterization, without using a link function, is derived for a sample of size  $n$ , selected by simple random sampling (SRS). The unconditional probability of giving  $r^{\text{th}}$  response to the  $j^{\text{th}}$  question,  $P(jr)$  is the weighted sum of class dependent item response probabilities, the weights being the latent class size. Suppose,  $n_{jr}$  is the total number of individuals in the sample who have given  $r^{\text{th}}$  response to the  $j^{\text{th}}$  question and it follows a  $\text{Binomial}(n_j, P(jr))$ ; where,  $n_j$  is the number of individuals who have responded to item  $j$  and

$$P(jr) = \theta_1 \alpha_{1jr} + \theta_2 \alpha_{2jr} + \dots + \theta_L \alpha_{Ljr}$$

The likelihood function for LCA with SRS, for a two-class latent class analysis set-up (i.e.  $L=2$ ) with  $J$  dichotomous items is given by,

$$\Lambda_{SRS} \propto \prod_{j=1}^J (\theta_1 \alpha_{1j1} + \theta_2 \alpha_{2j1})^{n_{j1}} (1 - \theta_1 \alpha_{1j1} - \theta_2 \alpha_{2j1})^{n_j - n_{j1}} \\ = \prod_{j=1}^J \left[ \prod_{j1|iej1} (\theta_1 \alpha_{1j1} + \theta_2 \alpha_{2j1}) \prod_{j2|iej1} (1 - \theta_1 \alpha_{1j1} - \theta_2 \alpha_{2j1}) \right]; \\ n_{j1} = \sum_{iej1} \psi_i; \psi_i = 1 \forall i.$$

#### Latent class analysis with log-linear parameterization

One of the key developments of latent class modelling was the log-linear parameterization of the model parameters (Haberman, 1979; Goodman, 1973). For assessing relationship between two or more categorical variables, this model may be used when the variables are indicators to a particular latent variable. Suppose there are  $J$  number of indicators with  $R$  number of response options for each indicator. The latent class model is based on the multiway contingency table of  $r \times J$  response patterns. Suppose a particular response combination on  $J$  indicators for a person  $i$  is  $Y_i$ ,  $i = 1, 2, \dots, n$ . Based on the responses given in the indicators, the individuals are categorized into  $L$  number of latent classes. A model of latent class analysis has the following parameters based on observed and latent classifications.  $\alpha_{ljr}$ ,  $l = 1, \dots, L$ ;  $j = 1, \dots, J$ ;  $r = 1, \dots, R$ , defined as the conditional item-response probabilities i.e. the probability of responding with  $r^{\text{th}}$  response option to  $j^{\text{th}}$  item, for a respondent in  $l^{\text{th}}$  latent class (i.e.  $c_l, l=1,2,\dots,L$ ). The likelihood function and the log-likelihood function are given by,

$$\Lambda = \prod_{i=1}^n \sum_{l=1}^L \theta_l P(Y_i | c_l)$$

$$\ln \Lambda = \sum_{i=1}^n \ln \left\{ \sum_{l=1}^L \theta_l \prod_{j=1}^J \prod_{r=1}^R \alpha_{ljr}^{\delta_{ijr}} \right\}$$

The Kronecker's delta equals 1 when the  $i^{\text{th}}$  person gives the response  $r$  to the  $j^{\text{th}}$  item. The model constraints are as follows,

$$\sum_{r=1}^R \alpha_{ljr} = 1, \forall j$$

$$\sum_{l=1}^L \theta_l = 1$$

The loglinear transformation (Haberman, 1979; Biemer, 2011) of the conditional probabilities for  $l$ ,  $j$  and  $r$  is given by,

$$\alpha_{ljr} = \frac{(\exp \zeta_{jr} + \zeta_{ljr})}{\sum_{r=1}^R \exp (\zeta_{jr} + \zeta_{ljr})}$$

where  $\zeta_{jr}$  is the main effect of the indicator 'j' and  $\zeta_{ljr}$  is the interaction effect of the item 'j' with the latent variable  $Z$  assuming a particular latent class 'l'.

#### Latent class analysis with stratified random sampling

One of the assumptions of latent class analysis is that the underlying sample of individuals are selected according to a simple random sampling scheme. However, in most of the surveys, the assumption is violated and more complex sampling designs are used to select the required survey respondents. Often the respondents are selected from a larger population which is divided into number of subgroups relevant to the outcome of the survey. Thus, the sample of respondents also exhibit subgroups according to some shared characteristics amongst them. In such cases, the sample may be based on a stratified random sample where the individuals are categorized into observed groups or strata. It is assumed that  $n$  individuals are selected by stratified random sampling with simple random sampling in each stratum. Thus  $n$  is partitioned into 's' number of subgroups, where  $s = 1, 2, \dots, S$ , where  $S$  is the total number of strata. Following a simplest form of the model, the item-response probabilities,  $\alpha_{ljr}$ , are assumed not to be stratum-dependant; they are same over different strata. Latent class size  $\theta_l$  are different for each stratum i.e.  $\theta_l$  is stratum-dependant.  $\theta_{ls}, s = 1, \dots, S; l = 1, \dots, L$ , defined as the size or prevalence of  $l^{\text{th}}$  latent class, specific to the observed stratum 's'. The likelihood function of stratified latent class model with probabilistic representation of the model parameters is given by,

$$\Lambda_{\text{strata}} \propto \prod_{j=1}^J \prod_{r=1}^R \prod_{s=1}^S \{P_s(jr)^{n_{jrs}}\}$$

$n_{jrs}$ , number of individuals in the stratum 's' who have responded with option  $r$  in  $j^{\text{th}}$  item, follows Binomial( $n_{jrs}$ ,  $P_s(jr)$ ) where,  $n_{jrs}$  is the number of individuals in stratum 's' who have responded to  $j^{\text{th}}$

item. If all respondents in the stratum have given response to all items (i.e. if there is no missing data),  $n_{j,s}$  is equal to the stratum size.  $P_s(jr)$  is the probability of giving  $r^{\text{th}}$  response in  $j^{\text{th}}$  question, in the observed stratum 's'. For,  $L=2$ ,  $P_s(jr) = \theta_{1s} \alpha_{1j1} + \theta_{2s} \alpha_{2j1}$ , for  $s = 1, 2$ .

#### RESULTS

The present study uses a secondary data set on patients referred to the UCLA Adult Cardiac Imaging and Hemodynamics Laboratories for Dobutamine Stress Echocardiography (DSE) between March 1991 and March 1996 (Available in <http://www.stat.ucla.edu/projects/datasets/>), who have given consent to have their medical data reviewed. The above-mentioned study was designed to prospectively follow patients who underwent DSE at the UCLA School of Medicine during a five-year period. Patients who did not have either 12 months of follow-up or an event of cardiac malfunction within 12 months were not included in the final analysis. The data set from the UCLA study is used to fit the three models of latent class analysis in the present analysis. There are 142 male and 278 female patients in the data set for whom binary information on certain risk factors are available. Variables on risk factors related to cardiac events viz. gender, history of hypertension, history of diabetes and history of smoking are considered as the observed or manifest variables for the latent class analysis. All the manifest variables have binary response options (e.g. 1 as Yes and 2 as No, 1 as male and 2 as female). In the present study, classical models of LCA with probabilistic and loglinear parameterization under SRS are compared with stratified LCA using the data set. The observed indicators of latent classes are the presence or absence of risk factors viz. history of hypertension, diabetes or smoking. Gender is also taken as an indicator for the latent classes. The observed stratification of the 420 individuals under study is based on the ages of the patient. The first stratum represents patients with ages lesser than 60 and the second stratum consists of patients with ages more than 60. The observed class specific latent class size parameters are  $\theta_{11}$  and  $\theta_{12}$  for the present analysis. For  $J = 4$ ,  $S = 2$  and  $L = 2$ , the number of parameters is 10 i.e. two latent class size parameters and 8 item response probabilities corresponding to two latent class solution and the degrees of freedom is  $2^4 - 1$  i.e. 15. Since the number of parameters is less than the degrees of freedom, the model is identified.

**Table 1: Model fit statistics for the three LCA models**

	$\Lambda$	$\Lambda_{strata}$	$\Lambda_{SRS}$
AIC	2105.179	2062.594	2123.896
BIC	2141.541	2102.996	2160.258

Table 1 summarizes the values of model fit statistics viz. Akaike's Information Criteria or AIC, Bayesian Information Criteria or BIC for three models applied to the data set. It shows that both the criteria are lowest for the stratified LCA model. The maximum likelihood estimation procedure is replicated for 2000 times to achieve robustness. In each replication, the estimates of latent class prevalence are plotted in Figure 1. It shows that, the probabilistic (without a link function) LCA model which assumes SRS as the underlying sampling design, fails to identify the distinction between the observed strata as the parameter estimates of latent class prevalence are placed randomly with no significant clustering. This plot also suggests that the traditional latent class model with the log-linear parameterization provides the estimates in two different clusters, placed closer to each other with lower discrimination. However, the LCA model for stratified random sampling is the best fitted model as its discriminatory capacity of identifying two observed subgroups for a particular latent class is higher than the others. The stratified model distinctly identifies two clusters of estimates. In this case, the two group-specific latent class sizes remain in the range of (0.1, 0.3) and (0.7,0.9) for the stratified LCA model.

Figure 1: Maximum likelihood estimates (Estimates) of latent class sizes in 2000 replications (Index) for the models of traditional latent class analysis with log-linear link function for the parameters for a simple random sampling design (Traditional LCA)(blue), latent class analysis without a link function for the parameters for a simple random sampling design (LCA with SRS)(red) and latent class analysis for a stratified random sampling design (Stratified LCA)(black).

Figure 2 and 3 summarize the parameter estimates for the item-response probabilities in different latent classes. The variability in estimates of item response probabilities may be described in two estimate bands for the traditional model (with loglinear parameterization), whereas for the stratified model there exists only one cluster of estimates. However, for the last indicator (i.e.  $\alpha_{141}$  and  $\alpha_{241}$ ) it shows

erratic behaviour which emphasises the trade-off between model parsimony and model fit. Increase in the number of indicators which are included in the modelling framework, would inflate the number of model parameters and consequently would increase the complexity of the model. It can also be observed that the standard errors of estimates are comparatively lower in the stratified LCA (Figure 1). The standard error estimates of two strata-specific parameters of latent class prevalence are lower for the stratified LCA than the traditional model of LCA with loglinear parameterization.

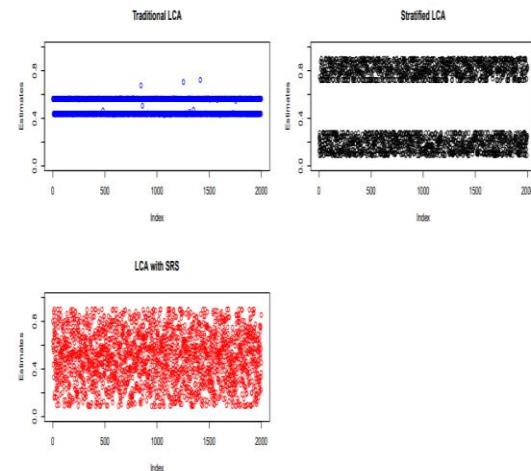


Figure 2: Plot of final estimates (estimates) versus initial values (initials) in 2000 replications, for first two indicators in latent class 1 ( $\alpha_{111}$ ,  $\alpha_{121}$ ) and in latent class 2 ( $\alpha_{211}$ ,  $\alpha_{221}$ ) for the traditional latent class analysis with log-linear link function for the parameters for a simple random sampling design (Traditional LCA)(blue) and latent class analysis for a stratified random sampling design (Stratified LCA)(black).

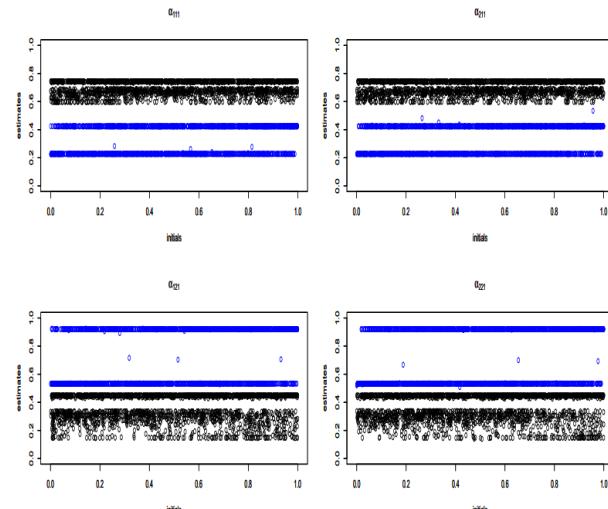


Figure 3: Plot of final estimates (estimates) versus initial values (initials) in 2000 replications, for last two indicators in latent class 1 ( $\alpha_{131}$ ,  $\alpha_{141}$ ) and in latent class 2 ( $\alpha_{231}$ ,  $\alpha_{241}$ ) for the traditional latent class analysis with log-linear link function for the parameters for a simple random sampling design (Traditional LCA)(blue) and latent class analysis for a stratified random sampling design (Stratified LCA)(black).

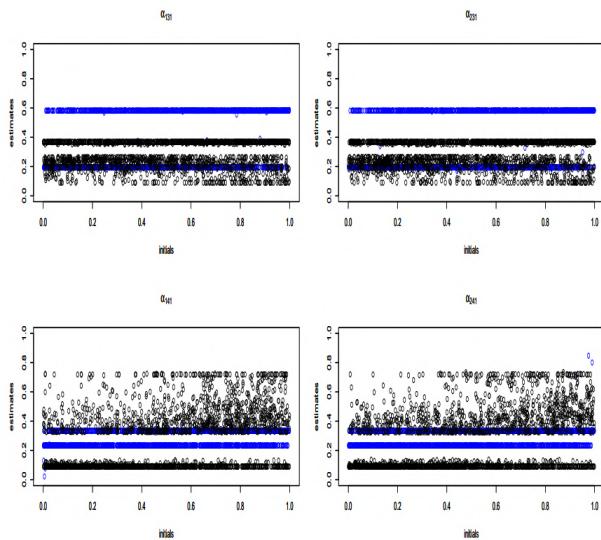
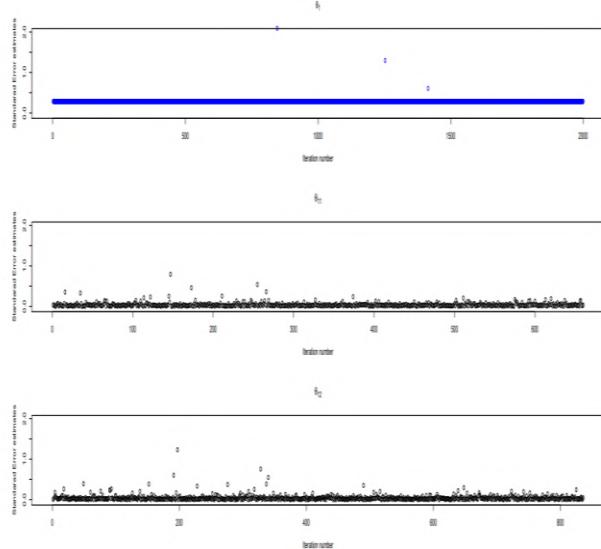


Figure 4: Standard error estimates in 2000 replications (iteration number) for latent class size parameter in latent class 1 ( $\theta_1$ ) for the traditional latent class analysis with log-linear link function for the parameters for a simple random sampling design (Traditional LCA)(blue) and for stratum-specific latent class size parameters in latent class 1 ( $\theta_{11}$ ,  $\theta_{12}$ ) for latent class analysis for a stratified random sampling design (Stratified LCA)(black).



## DISCUSSION

In most of the epidemiological studies, data are collected through more complex sampling techniques than a simple random sampling procedure. The analysis carried out for such situation requires special attention in this regard. Classical analysis using SRS design for a complex data collected through other complex sampling designs may yield inconsistent results and biased estimates of parameters. It is thus important for an epidemiological model and methodology to incorporate the complexities of a sampling design. The present study focuses on the importance of complex sampling design in identifying latent subgroups in a population of patients having a certain disease profile through latent class analysis. A simple model of latent class analysis is described for a stratified random sampling technique and the likelihood function is derived for it. The impact of change in sampling design is assessed by comparing the three likelihood functions for latent class model viz. probabilistic parameterization of latent class model under SRS ( $\Lambda_{SRS}$ ), loglinear parameterization of latent class model under SRS ( $\Lambda$ ) and probabilistic parameterization of latent class model under stratified random sampling scheme ( $\Lambda_{strata}$ ). The three models are applied to a secondary data set containing risk factors of cardiac diseases in a population of 420 individuals. Several model fit statistics are calculated for each variant of the latent class model and reported in Table 1. It is observed that the stratified LCA exhibits the lowest values of AIC and BIC among the models applied to the data set. Thus, the latent class model with stratified random sampling is the best fitted model to the data set among the three. The analysis in the present study exhibits that the stratified LCA fits the data best when there is some inherent stratification in it. The loss of information due to negligence of such stratification using traditional models of LCA may cause biased and unstable estimates of the parameters. The novel approach described in this study focuses on the importance of incorporating stratification details into modelling framework when the underlying sample is inherently stratified. The model of LCA which considers the stratification characteristics into the parameters gives better results in terms of stability and variability of the parameter estimates. Two bands of estimates for latent class sizes are separated significantly from each other for the stratified LCA model than the

classical models, with lower standard errors. A simple approach for incorporating sample characteristics when it is not collected through SRS is described in the present research. It describes a model with stratum specific latent class sizes and classical parameters of item-response probabilities. It does not consider any sampling weights in the analysis. A comparative analysis of data regarding LCA in SRS and stratified random sampling scheme is provided. It is evident from the analysis that a classical LCA model fails to identify stratification inherent in the data and in such cases, the stratified latent class modelling approach described in the present study provides better results in terms of parameter estimates, standard error estimates and model fit measures.

## REFERENCES

1. Kaldor J, Clayton D. (1985) Latent class analysis in chronic disease epidemiology. *Stat Med* ;4(3):327-35. doi: 10.1002/sim.4780040312. PMID: 3877331.
2. Koukounari A, Jamil H, Erosheva E, Shiff C, Moustaki I (2021) Latent Class Analysis: Insights about design and analysis of schistosomiasis diagnostic studies. *PLoS Negl Trop Dis* 15(2): e0009042. <https://doi.org/10.1371/journal.pntd.0009042>.
3. Lazarsfeld, P.F. (1950): The logical and mathematical foundation of latent structure analysis and The interpretation and mathematical foundation of latent structure analysis. S.A. Stouffer et al.(eds.), *Measurement and Prediction*,362-472.Princeton, NJ: Princeton University Press.
4. Goodman, L. A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I: A modified latent structure approach. *American Journal of Sociology*, 79(5), 1179–1259. doi:10.1086/225676.
5. Goodman, Leo A.(1974) Exploratory latent structure analysis using both identifiable and unidentifiable models, *Biometrika*, 61, 2, p.215.
6. Vermunt, J K, Magidson, J. (2007), Latent Class Analysis with Sampling Weights. *Sociological Methods & Research*, Vol 36, Issue 1, pp. 87 - 111. DOI:10.1177/0049124107301965.
7. Biemer, P. P.(2011), *Latent Class Analysis of Survey Error*, A John Wiley and Sons, Inc. Publication; ISBN: 978-0-470-89114-8.
8. Skrondal A, Rabe-Hesketh S. (2002) General latent variable modelling-multilevel, longitudinal, and structural equation models. *Interdisciplinary statistics* 1st ed., Chapman & Hall; Boca Raton, Fla. ISBN 1584880007.
9. Patterson, B. H. , Dayton, C. M. and Graubard, B. I. (2002), Latent Class Analysis of Complex Sample Survey Data, *Journal of the American Statistical Association*,97:459,72.
10. Asparouhov, T., Muthen. (2005) Sampling Weights in Latent Variable Modeling, *Structural Equation Modeling*, 12(3), 411–434.
11. Vaughn, M. G., DeLisi, M., Gunterbh, T., Fu, Q., Beaver, K. M., Perron, B., Howard, M.(2011). The Severe 5%: A Latent Class Analysis of the Externalizing Behaviour Spectrum in the United States, *J Crim Justice*, 39(1): 75–80, doi:10.1016/j.jcrimjus.2010.12.001.
12. Pennoni, F. , M. Nakai.(2019) A latent class analysis towards stability and changes in breadwinning patterns among coupled households, *Depend. Model.*, 7:234–246.
13. Dey, A., Majumdar, K.K., Chakraborty, A. K., Mandal, A. K. (2018) Women Empowerment in India - A Perspective, *National Journal of Community Medicine*, pISSN: 0976 3325, 9(8),574-579.
14. Das RN. (2016). Hypertension risk factors who underwent dobutamine stress echocardiography. *Interv. Cardiol.* 8(5), 683-693.
15. Das RN, Lee Y and Mukhopadhyay BB. (2017). The Basal, Peak and Maximum Heart Rate Determinants of the Cardiac Patients who Underwent Dobutamine Stress Echocardiography. *M J Cardiol.* 2(2): 013.
16. Haberman, S.J.(1979) *Analysis of Qualitative Data* (Vol. 2: New Developments), New York:Academic Press.
17. Goodman , L. A. (1973). The analysis of multidimensional contingency tables when some variables are posterior to others: A modified path analysis approach , *Biometrika* , 60 , 179 – 192 .

## ORIGINAL ARTICLE

# Latent class analysis in a complex sampling design – an advanced modelling approach in Epidemiology

Ankita Dey<sup>1</sup>, Diganta Mukherjee<sup>2</sup>, Sugata Sen Roy<sup>3</sup>

<sup>1</sup>National Institute of Tuberculosis and Respiratory Diseases, New Delhi;

<sup>2</sup>Indian Statistical Institute, Kolkata, West Bengal;

<sup>3</sup>Department of Statistics, University of Calcutta, Kolkata, West Bengal

## CORRESPONDING AUTHOR

Ankita Dey, National Institute of Tuberculosis and Respiratory Diseases (NITRD), An Autonomous Institute under Ministry of Health & Family Welfare, Sri Aurobindo Marg, New Delhi -110030

E Mail ID: [ankitadey14@gmail.com](mailto:ankitadey14@gmail.com)



## CITATION

Dey A, Mukherjee D, Roy SS. Latent class analysis in a complex sampling design – an advanced modelling approach in Epidemiology. *EFI Bulletin*. 2023;4(1):23-29. <https://doi.org/10.56450/EFIB.2023.v3i01.006>

## ARTICLE CYCLE

Received: 16/02/2023; Accepted: 21/03/2023; Published: 31/03/2023

This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

©The Author(s). 2023 Open Access

## ABSTRACT

In an Epidemiological study, researchers are often interested to identify population subgroups according to variables relevant to the context of the study. The subgroups may be identified according to some observed characteristics of the underlying disease or subgroups may be defined according to an unobserved or latent profile of patients. Latent Class Analysis (LCA), identifies hidden subgroups among the respondents, based upon the response pattern of a respondent to various surveys. There are certain assumptions for a traditional latent class model to work in a particular scenario. Use of traditional models in case of violation of one of these assumptions would lead to increased misclassification rate and biased estimates of the parameters. The present study provides a modified approach of latent class modelling, when one of the underlying assumptions of traditional latent class models is relaxed. Traditional models of LCA are compared with the newly developed models under more complex sampling designs.

## KEYWORDS

Latent Class Analysis; Complex Sampling Design; Epidemiology; Statistical Methodology

## INTRODUCTION

In an Epidemiological study, researchers are often interested to identify population subgroups according to variables relevant to the context of the study. The subgroups may be identified according to some observed characteristics or features of the underlying disease or subgroups may be defined according to an unobserved or latent profile of patients. Latent Class Analysis (LCA), identifies hidden subgroups among the respondents, based upon the response pattern of a respondent to various surveys. It was introduced in 1950 by Paul F. Lazarsfeld (Lazarsfeld, 1950), who performed LCA for building typologies (or clustering) based on dichotomous observed variables. The estimation of

model parameters using maximum likelihood approach was proposed later on (Goodman, 1974; Vermunt & Magidson 2007).

The importance of latent class analysis in epidemiological modelling has been explored by several researchers in the past few decades. Importance of latent class analysis in chronic disease epidemiology is discussed by Kaldor & Clayton (1985). In their study, the authors provided a general description of the LCA methodology as a statistical methodology in epidemiology and indicated its role in logistic analysis of categorical epidemiological data. The paper also indicated the need for further exploration of the method when the underlying assumptions of standard LCA model are relaxed.

Koukounari et al (2021) discussed the scenarios under which latent class models are suitable in the context of the evaluation of existing schistosomiasis diagnostics.

There are certain assumptions for a traditional latent class model to work in a particular scenario. Use of traditional models in case of violation of one of these assumptions would lead to increased misclassification rate and biased estimates of the parameters (Koukounari et al, 2021). It has been noted that in epidemiological modelling, researches focusing on modifications or revisions of traditional LCA models, to fit into the context where one of the model assumptions does not hold, are comparatively less explored. The present study provides a modified approach of latent class modelling, when one of the underlying assumptions of traditional latent class models is relaxed. The statistical analysis described in this study provides a methodological advancement to the existing modelling approaches of Epidemiology. Traditional models of LCA are compared with the newly developed models under more complex sampling designs and their performances are compared in the context of a secondary data set.

## MATERIAL & METHODS

Latent class analysis (LCA) is considered to be an equivalent methodology for Factor Analysis, typically used for dichotomous or polytomous variables (Biemer, 2011, Skrondal and Rabe Hesketh, 2002). The parameters of interest in a typical problem of latent class analysis are the unobserved proportion or size of the latent classes and the conditional item-response probabilities given the membership in a latent class. Based on the observed data on manifest variables, LCA provides a classification among the population. Manifest variables are called the 'indicators' of a particular latent class. The basis of the analysis lies in modelling the relationship between the latent variable and its indicators. LCA identifies unobservable (latent) subgroups within a population based on individuals' response to different categorical observed variables. It is a technique which explains the relationships between manifest or observed variables (may be dichotomous or polytomous) with respect to some unobserved or latent variables (may also be dichotomous or polytomous) on the basis of data obtained in various kinds of complex surveys. One of the assumptions of a latent class model is that the underlying sample of individuals consists of  $n$  units sampled without

replacement from a large population of  $N$  units using simple random sampling. Extension in models of latent class analysis for more complicated sampling designs has drawn attention of several researchers in recent times (Patterson, Dayton & Graubard, 2002; Asparouhov & Muthén, 2005; Vermunt & Magidson, 2007; Vaughn et. al., 2011; Pennoni and Nakai, 2018, 2019.). The impact of sampling scheme in the process of latent class estimation is now considered an active area of research. Scarcity of models in the literature of latent class models incorporating complex sampling characteristics viz sampling weight, stratification etc was discussed in Patterson, Dayton & Graubard (2002). They proposed such an extension by introducing sample weights into the latent class estimation algorithm. A two-class solution based on indicators regarding food consumption was named as "regular" and "nonregular" vegetable consumption groups in their study. A pseudo-likelihood function was developed based on the sample weights available for each respondent. For ' $n$ ' number of individuals with response pattern  $Y_i$  and membership  $c_i$  in one of  $L$  number of latent classes, the usual latent class parameters are latent class size ( $\theta_l$ ) and the item-response probabilities ( $\alpha_{ljr}$ ). In the following equation,  $i$  stands for an individual,  $i=1,2,\dots,n$ ;  $l$  stands for the latent class,  $l=1,2,\dots,L$ ;  $j$  stands for the item in a questionnaire,  $j=1,2,\dots,J$ ;  $r$  stands for the response option in  $j$ th item,  $r=1,2,\dots,R$ . For a dichotomous item,  $R=2$ . In case of more complex sampling schemes than SRS, the pseudo-likelihood based on sampling weights  $w_i$  is given as (Patterson, Dayton & Graubard, 2002),

$$\sum_{i=1}^n w_i \left( \ln \sum_{l=1}^L \theta_l Pr(Y_i|c_l) \right) = \sum_{i=1}^n w_i \ln \left( \sum_{l=1}^L \theta_l \prod_{j=1}^J \prod_{r=1}^{R_j} \alpha_{ljr}^{\delta_{ijr}} \right)$$

where the sample weights can be measured by eg. the reciprocal of inclusion probabilities. Vermunt & Magidson (2007) proposed a maximum likelihood approach to log-linear latent class analysis with sampling weights and probability formulation of latent class model with sampling weights which can accommodate larger number of indicators. However, these sampling weights may not be always readily available. Impact of a stratified random sampling on parameter estimates in a latent class framework without incorporating sample weight may constitute an interesting area of research. The present study attempts to describe the model of latent class

analysis for a group of individuals selected through stratified random sampling technique without incorporating sampling weights. Within each stratum, a new set of model parameters are defined in the context of a full likelihood latent class model without a log-linear parameterization. The stratified latent class model is compared with the traditional model with a sample selected through simple random sampling technique. The latent class model with stratified random sampling is compared with two traditional LCA models, viz., LCA with simple random sampling (SRS) design without log-linear parameterization (full likelihood) and LCA with simple random sampling (SRS) design with a log-linear parameterization. Stability of estimates of the model parameters and their standard errors are compared for different situations mentioned above. The effect of sampling scheme in estimation of parameters are explored by applying the models in the data set of risk factors related to cardiac events (Dey et. al., 2016; Das, 2016; Das, Lee and Mukhopadhyay, 2017).

#### Latent class analysis with probabilistic parameterization

There are different modelling approaches for latent class analysis (Biemer, 2011). In the present study, two of the approaches are used viz. probability model parameterization of the standard latent class model and the log-linear parameterized model, for comparison of model fit and parameter estimates. In the probabilistic parameterization of the latent class model, the likelihood function is expressed as a function of the marginal and conditional cell probabilities of the underlying multiway contingency table. A probabilistic parameterization, without using a link function, is derived for a sample of size  $n$ , selected by simple random sampling (SRS). The unconditional probability of giving  $r^{\text{th}}$  response to the  $j^{\text{th}}$  question,  $P(jr)$  is the weighted sum of class dependent item response probabilities, the weights being the latent class size. Suppose,  $n_{jr}$  is the total number of individuals in the sample who have given  $r^{\text{th}}$  response to the  $j^{\text{th}}$  question and it follows a  $\text{Binomial}(n_j, P(jr))$ ; where,  $n_j$  is the number of individuals who have responded to item  $j$  and

$$P(jr) = \theta_1 \alpha_{1jr} + \theta_2 \alpha_{2jr} + \dots + \theta_L \alpha_{Ljr}$$

The likelihood function for LCA with SRS, for a two-class latent class analysis set-up (i.e.  $L=2$ ) with  $J$  dichotomous items is given by,

$$\begin{aligned} \Lambda_{SRS} &\propto \prod_{j=1}^J (\theta_1 \alpha_{1j1} + \theta_2 \alpha_{2j1})^{n_{j1}} (1 - \theta_1 \alpha_{1j1} - \theta_2 \alpha_{2j1})^{n_j - n_{j1}} \\ &= \prod_{j=1}^J \left[ \prod_{j1|iej1} (\theta_1 \alpha_{1j1} + \theta_2 \alpha_{2j1}) \prod_{j2|iej2} (1 - \theta_1 \alpha_{1j1} - \theta_2 \alpha_{2j1}) \right]; \\ n_{j1} &= \sum_{i \in j1} \psi_i; \psi_i = 1 \forall i. \end{aligned}$$

#### Latent class analysis with log-linear parameterization

One of the key developments of latent class modelling was the log-linear parameterization of the model parameters (Haberman, 1979; Goodman, 1973). For assessing relationship between two or more categorical variables, this model may be used when the variables are indicators to a particular latent variable. Suppose there are  $J$  number of indicators with  $R$  number of response options for each indicator. The latent class model is based on the multiway contingency table of  $r \times J$  response patterns. Suppose a particular response combination on  $J$  indicators for a person  $i$  is  $Y_i$ ,  $i = 1, 2, \dots, n$ . Based on the responses given in the indicators, the individuals are categorized into  $L$  number of latent classes. A model of latent class analysis has the following parameters based on observed and latent classifications.  $\alpha_{ljr}$ ,  $l = 1, \dots, L$ ;  $j = 1, \dots, J$ ;  $r = 1, \dots, R$ , defined as the conditional item-response probabilities i.e. the probability of responding with  $r^{\text{th}}$  response option to  $j^{\text{th}}$  item, for a respondent in  $l^{\text{th}}$  latent class (i.e.  $c_l, l=1,2,\dots,L$ ). The likelihood function and the log-likelihood function are given by,

$$\Lambda = \prod_{i=1}^n \sum_{l=1}^L \theta_l P(Y_i | c_l)$$

$$\ln \Lambda = \sum_{i=1}^n \ln \left\{ \sum_{l=1}^L \theta_l \prod_{j=1}^J \prod_{r=1}^R \alpha_{ljr}^{\delta_{ijr}} \right\}$$

The Kronecker's delta equals 1 when the  $i^{\text{th}}$  person gives the response  $r$  to the  $j^{\text{th}}$  item. The model constraints are as follows,

$$\sum_{r=1}^R \alpha_{ljr} = 1, \forall j$$

$$\sum_{l=1}^L \theta_l = 1$$

The loglinear transformation (Haberman, 1979; Biemer, 2011) of the conditional probabilities for  $l$ ,  $j$  and  $r$  is given by,

$$\alpha_{ljr} = \frac{(\exp \zeta_{jr} + \zeta_{ljr})}{\sum_{r=1}^R \exp (\zeta_{jr} + \zeta_{ljr})}$$

where  $\zeta_{jr}$  is the main effect of the indicator 'j' and  $\zeta_{ljr}$  is the interaction effect of the item 'j' with the latent variable  $Z$  assuming a particular latent class 'l'.

#### Latent class analysis with stratified random sampling

One of the assumptions of latent class analysis is that the underlying sample of individuals are selected according to a simple random sampling scheme. However, in most of the surveys, the assumption is violated and more complex sampling designs are used to select the required survey respondents. Often the respondents are selected from a larger population which is divided into number of subgroups relevant to the outcome of the survey. Thus, the sample of respondents also exhibit subgroups according to some shared characteristics amongst them. In such cases, the sample may be based on a stratified random sample where the individuals are categorized into observed groups or strata. It is assumed that  $n$  individuals are selected by stratified random sampling with simple random sampling in each stratum. Thus  $n$  is partitioned into 's' number of subgroups, where  $s = 1, 2, \dots, S$ , where  $S$  is the total number of strata. Following a simplest form of the model, the item-response probabilities,  $\alpha_{ljr}$ , are assumed not to be stratum-dependant; they are same over different strata. Latent class size  $\theta_l$  are different for each stratum i.e.  $\theta_l$  is stratum-dependant.  $\theta_{ls}$ ,  $s = 1, \dots, S$ ;  $l = 1, \dots, L$ , defined as the size or prevalence of  $l^{\text{th}}$  latent class, specific to the observed stratum 's'. The likelihood function of stratified latent class model with probabilistic representation of the model parameters is given by,

$$\Lambda_{\text{strata}} \propto \prod_{j=1}^J \prod_{r=1}^R \prod_{s=1}^S \{P_s(jr)^{n_{jrs}}\}$$

$n_{jrs}$ , number of individuals in the stratum 's' who have responded with option  $r$  in  $j^{\text{th}}$  item, follows Binomial(  $n_{jrs}$ ,  $P_s(jr)$  ) where,  $n_{jrs}$  is the number of individuals in stratum 's' who have responded to  $j^{\text{th}}$

item. If all respondents in the stratum have given response to all items (i.e. if there is no missing data),  $n_{j,s}$  is equal to the stratum size.  $P_s(jr)$  is the probability of giving  $r^{\text{th}}$  response in  $j^{\text{th}}$  question, in the observed stratum 's'. For,  $L=2$ ,  $P_s(jr) = \theta_{1s} \alpha_{1j1} + \theta_{2s} \alpha_{2j1}$ , for  $s = 1, 2$ .

#### RESULTS

The present study uses a secondary data set on patients referred to the UCLA Adult Cardiac Imaging and Hemodynamics Laboratories for Dobutamine Stress Echocardiography (DSE) between March 1991 and March 1996 (Available in <http://www.stat.ucla.edu/projects/datasets/>), who have given consent to have their medical data reviewed. The above-mentioned study was designed to prospectively follow patients who underwent DSE at the UCLA School of Medicine during a five-year period. Patients who did not have either 12 months of follow-up or an event of cardiac malfunction within 12 months were not included in the final analysis. The data set from the UCLA study is used to fit the three models of latent class analysis in the present analysis. There are 142 male and 278 female patients in the data set for whom binary information on certain risk factors are available. Variables on risk factors related to cardiac events viz. gender, history of hypertension, history of diabetes and history of smoking are considered as the observed or manifest variables for the latent class analysis. All the manifest variables have binary response options (e.g. 1 as Yes and 2 as No, 1 as male and 2 as female). In the present study, classical models of LCA with probabilistic and loglinear parameterization under SRS are compared with stratified LCA using the data set. The observed indicators of latent classes are the presence or absence of risk factors viz. history of hypertension, diabetes or smoking. Gender is also taken as an indicator for the latent classes. The observed stratification of the 420 individuals under study is based on the ages of the patient. The first stratum represents patients with ages lesser than 60 and the second stratum consists of patients with ages more than 60. The observed class specific latent class size parameters are  $\theta_{11}$  and  $\theta_{12}$  for the present analysis. For  $J = 4$ ,  $S = 2$  and  $L = 2$ , the number of parameters is 10 i.e. two latent class size parameters and 8 item response probabilities corresponding to two latent class solution and the degrees of freedom is  $2^4 - 1$  i.e. 15. Since the number of parameters is less than the degrees of freedom, the model is identified.

**Table 1: Model fit statistics for the three LCA models**

	$\Lambda$	$\Lambda_{strata}$	$\Lambda_{SRS}$
AIC	2105.179	2062.594	2123.896
BIC	2141.541	2102.996	2160.258

Table 1 summarizes the values of model fit statistics viz. Akaike's Information Criteria or AIC, Bayesian Information Criteria or BIC for three models applied to the data set. It shows that both the criteria are lowest for the stratified LCA model. The maximum likelihood estimation procedure is replicated for 2000 times to achieve robustness. In each replication, the estimates of latent class prevalence are plotted in Figure 1. It shows that, the probabilistic (without a link function) LCA model which assumes SRS as the underlying sampling design, fails to identify the distinction between the observed strata as the parameter estimates of latent class prevalence are placed randomly with no significant clustering. This plot also suggests that the traditional latent class model with the log-linear parameterization provides the estimates in two different clusters, placed closer to each other with lower discrimination. However, the LCA model for stratified random sampling is the best fitted model as its discriminatory capacity of identifying two observed subgroups for a particular latent class is higher than the others. The stratified model distinctly identifies two clusters of estimates. In this case, the two group-specific latent class sizes remain in the range of (0.1, 0.3) and (0.7,0.9) for the stratified LCA model.

Figure 1: Maximum likelihood estimates (Estimates) of latent class sizes in 2000 replications (Index) for the models of traditional latent class analysis with log-linear link function for the parameters for a simple random sampling design (Traditional LCA)(blue), latent class analysis without a link function for the parameters for a simple random sampling design (LCA with SRS)(red) and latent class analysis for a stratified random sampling design (Stratified LCA)(black).

Figure 2 and 3 summarize the parameter estimates for the item-response probabilities in different latent classes. The variability in estimates of item response probabilities may be described in two estimate bands for the traditional model (with loglinear parameterization), whereas for the stratified model there exists only one cluster of estimates. However, for the last indicator (i.e.  $\alpha_{141}$  and  $\alpha_{241}$ ) it shows

erratic behaviour which emphasises the trade-off between model parsimony and model fit. Increase in the number of indicators which are included in the modelling framework, would inflate the number of model parameters and consequently would increase the complexity of the model. It can also be observed that the standard errors of estimates are comparatively lower in the stratified LCA (Figure 1). The standard error estimates of two strata-specific parameters of latent class prevalence are lower for the stratified LCA than the traditional model of LCA with loglinear parameterization.

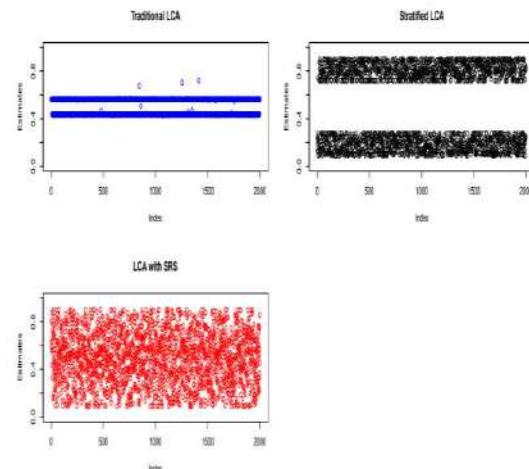


Figure 2: Plot of final estimates (estimates) versus initial values (initials) in 2000 replications, for first two indicators in latent class 1 ( $\alpha_{111}$ ,  $\alpha_{121}$ ) and in latent class 2 ( $\alpha_{211}$ ,  $\alpha_{221}$ ) for the traditional latent class analysis with log-linear link function for the parameters for a simple random sampling design (Traditional LCA)(blue) and latent class analysis for a stratified random sampling design (Stratified LCA)(black).

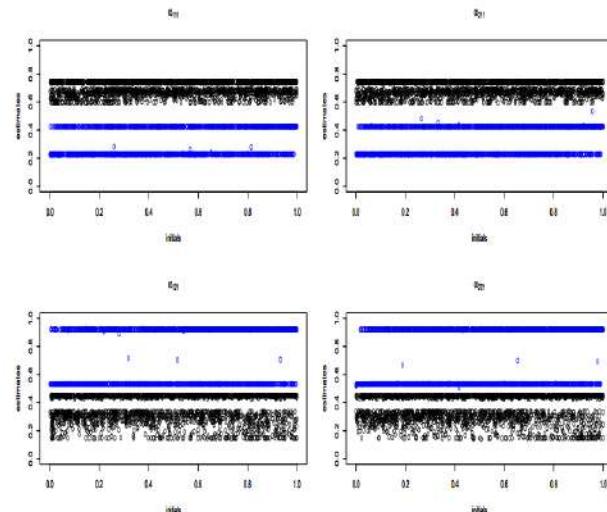


Figure 3: Plot of final estimates (estimates) versus initial values (initials) in 2000 replications, for last two indicators in latent class 1 ( $\alpha_{131}$ ,  $\alpha_{141}$ ) and in latent class 2 ( $\alpha_{231}$ ,  $\alpha_{241}$ ) for the traditional latent class analysis with log-linear link function for the parameters for a simple random sampling design (Traditional LCA)(blue) and latent class analysis for a stratified random sampling design (Stratified LCA)(black).

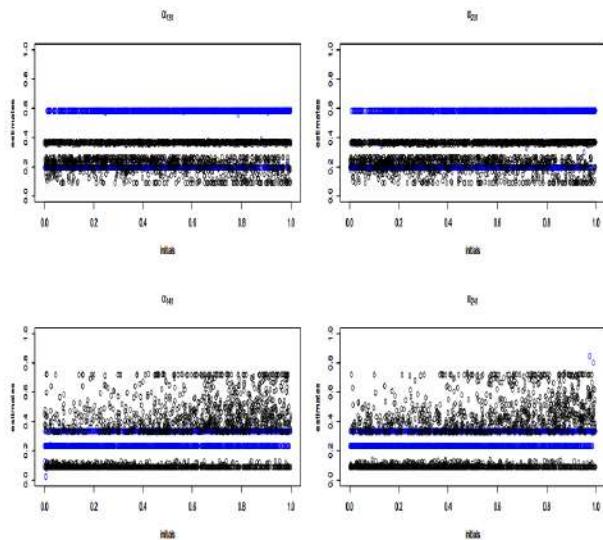
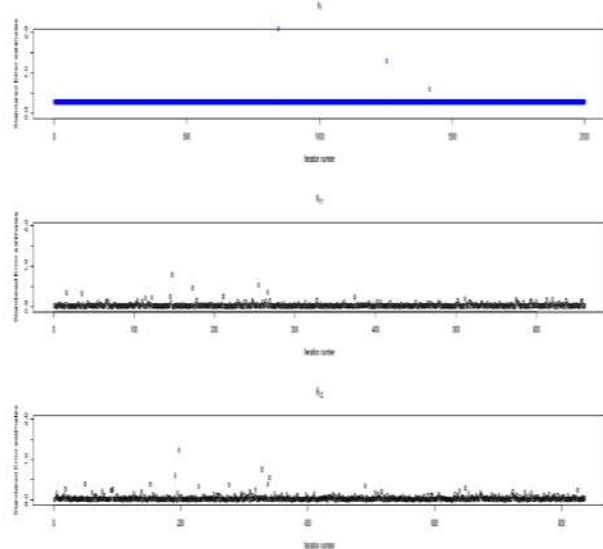


Figure 4: Standard error estimates in 2000 replications (iteration number) for latent class size parameter in latent class 1 ( $\theta_1$ ) for the traditional latent class analysis with log-linear link function for the parameters for a simple random sampling design (Traditional LCA)(blue) and for stratum-specific latent class size parameters in latent class 1 ( $\theta_{11}$ ,  $\theta_{12}$ ) for latent class analysis for a stratified random sampling design (Stratified LCA)(black).



## DISCUSSION

In most of the epidemiological studies, data are collected through more complex sampling techniques than a simple random sampling procedure. The analysis carried out for such situation requires special attention in this regard. Classical analysis using SRS design for a complex data collected through other complex sampling designs may yield inconsistent results and biased estimates of parameters. It is thus important for an epidemiological model and methodology to incorporate the complexities of a sampling design. The present study focuses on the importance of complex sampling design in identifying latent subgroups in a population of patients having a certain disease profile through latent class analysis. A simple model of latent class analysis is described for a stratified random sampling technique and the likelihood function is derived for it. The impact of change in sampling design is assessed by comparing the three likelihood functions for latent class model viz. probabilistic parameterization of latent class model under SRS ( $\Lambda_{SRS}$ ), loglinear parameterization of latent class model under SRS ( $\Lambda$ ) and probabilistic parameterization of latent class model under stratified random sampling scheme ( $\Lambda_{strata}$ ). The three models are applied to a secondary data set containing risk factors of cardiac diseases in a population of 420 individuals. Several model fit statistics are calculated for each variant of the latent class model and reported in Table 1. It is observed that the stratified LCA exhibits the lowest values of AIC and BIC among the models applied to the data set. Thus, the latent class model with stratified random sampling is the best fitted model to the data set among the three. The analysis in the present study exhibits that the stratified LCA fits the data best when there is some inherent stratification in it. The loss of information due to negligence of such stratification using traditional models of LCA may cause biased and unstable estimates of the parameters. The novel approach described in this study focuses on the importance of incorporating stratification details into modelling framework when the underlying sample is inherently stratified. The model of LCA which considers the stratification characteristics into the parameters gives better results in terms of stability and variability of the parameter estimates. Two bands of estimates for latent class sizes are separated significantly from each other for the stratified LCA model than the

classical models, with lower standard errors. A simple approach for incorporating sample characteristics when it is not collected through SRS is described in the present research. It describes a model with stratum specific latent class sizes and classical parameters of item-response probabilities. It does not consider any sampling weights in the analysis. A comparative analysis of data regarding LCA in SRS and stratified random sampling scheme is provided. It is evident from the analysis that a classical LCA model fails to identify stratification inherent in the data and in such cases, the stratified latent class modelling approach described in the present study provides better results in terms of parameter estimates, standard error estimates and model fit measures.

## REFERENCES

1. Kaldor J, Clayton D. (1985) Latent class analysis in chronic disease epidemiology. *Stat Med* ;4(3):327-35. doi: 10.1002/sim.4780040312. PMID: 3877331.
2. Koukounari A, Jamil H, Erosheva E, Shiff C, Moustaki I (2021) Latent Class Analysis: Insights about design and analysis of schistosomiasis diagnostic studies. *PLoS Negl Trop Dis* 15(2): e0009042. <https://doi.org/10.1371/journal.pntd.0009042>.
3. Lazarsfeld, P.F. (1950): The logical and mathematical foundation of latent structure analysis and The interpretation and mathematical foundation of latent structure analysis. S.A. Stouffer et al.(eds.), *Measurement and Prediction*,362-472.Princeton, NJ: Princeton University Press.
4. Goodman, L. A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I: A modified latent structure approach. *American Journal of Sociology*, 79(5), 1179–1259. doi:10.1086/225676.
5. Goodman, Leo A.(1974) Exploratory latent structure analysis using both identifiable and unidentifiable models, *Biometrika*, 61, 2, p.215.
6. Vermunt, J K, Magidson, J. (2007), Latent Class Analysis with Sampling Weights. *Sociological Methods & Research*, Vol 36, Issue 1, pp. 87 - 111. DOI:10.1177/0049124107301965.
7. Biemer, P. P.(2011), *Latent Class Analysis of Survey Error*, A John Wiley and Sons, Inc. Publication; ISBN: 978-0-470-89114-8.
8. Skrondal A, Rabe-Hesketh S. (2002) General latent variable modelling-multilevel, longitudinal, and structural equation models. *Interdisciplinary statistics* 1st ed., Chapman & Hall; Boca Raton, Fla. ISBN 1584880007.
9. Patterson, B. H. , Dayton, C. M. and Graubard, B. I. (2002), Latent Class Analysis of Complex Sample Survey Data, *Journal of the American Statistical Association*,97:459,72.
10. Asparouhov, T., Muthen. (2005) Sampling Weights in Latent Variable Modeling, *Structural Equation Modeling*, 12(3), 411–434.
11. Vaughn, M. G., DeLisi, M., Gunterbh, T., Fu, Q., Beaver, K. M., Perron, B., Howard, M.(2011). The Severe 5%: A Latent Class Analysis of the Externalizing Behaviour Spectrum in the United States, *J Crim Justice*, 39(1): 75–80, doi:10.1016/j.jcrimjus.2010.12.001.
12. Pennoni, F. , M. Nakai.(2019) A latent class analysis towards stability and changes in breadwinning patterns among coupled households, *Depend. Model.*, 7:234–246.
13. Dey, A., Majumdar, K.K., Chakraborty, A. K., Mandal, A. K. (2018) Women Empowerment in India - A Perspective, *National Journal of Community Medicine*, pISSN: 0976 3325, 9(8),574-579.
14. Das RN. (2016). Hypertension risk factors who underwent dobutamine stress echocardiography. *Interv. Cardiol.* 8(5), 683-693.
15. Das RN, Lee Y and Mukhopadhyay BB. (2017). The Basal, Peak and Maximum Heart Rate Determinants of the Cardiac Patients who Underwent Dobutamine Stress Echocardiography. *M J Cardiol.* 2(2): 013.
16. Haberman, S.J.(1979) *Analysis of Qualitative Data* (Vol. 2: New Developments), New York:Academic Press.
17. Goodman , L. A. (1973). The analysis of multidimensional contingency tables when some variables are posterior to others: A modified path analysis approach , *Biometrika* , 60 , 179 – 192 .

# Anaesthesiologists as Community Physicians- Expanding Horizons

Naveen Malhotra, Divya Gupta

Cardiac Anaesthesia & Pain Management Centre, Pt BDS PGIMS, Rohtak, Haryana

Department of Anaesthesiology and Pain Management, Himalayan Institute of Medical Sciences, Swami Rama Himalayan University, Swami Ram Nagar, Doiwala - 248016, Dehradun, (UTTARAKHAND)



## CORRESPONDING AUTHOR

Divya Gupta, Professor, Department of Anaesthesiology and Pain Management, Himalayan Institute of Medical Sciences, Swami Rama Himalayan University, Doiwala 248016, Dehradun, (UTTARAKHAND)

E Mail ID: [dr\\_divyagupta@rediffmail.com](mailto:dr_divyagupta@rediffmail.com)

## CITATION

Malhotra N, Gupta D. Anaesthesiologists as Community Physicians- Expanding Horizons. EFI Bulletin.

2023;4(1):30-31. Doi: <https://doi.org/10.56450/EFIB.2023.v3i01.007>

## ARTICLE CYCLE

Received: 23/03/2023; Accepted: 24/03/2023; Published: 31/03/2023

This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

©The Author(s). 2022 Open Access

*"The power of community to create health is far greater than any physician, clinic or hospital."*

Mark Hyman

The 19<sup>th</sup> century Pandemic of Cholera witnessed the emergence of a pioneer Dr. John Snow who was a full-time ‘physician anaesthetist’. He is credited for the development of medical mapping which led to the establishment of fundamentals of epidemiology and public health which are beneficial till date. Snow’s achievements in two medical specialities promulgated a scientific blend of Anaesthesiology and Epidemiology. Two centuries later, in 21<sup>st</sup> century’s COVID Pandemic, emerged a similar situation which has sparked the need of these two contemporary connected fields to be in action conjointly.

It is established that 90% of a patient’s health care journey is outside the confines of buildings of hospitals and clinics, which is why Community Health perspective of anaesthesiologists is a thought provoking aspect. The awareness highlighting the role of anaesthesiologists is exiguous in the public. To avert this, putting on eye glasses with “community lenses” will allow anaesthesiologists to have a broader perspective and move from ‘behind the curtains’ tag to the forefront forum to

‘Anaesthesiologists at door steps’. Since a decade, reforms and efforts are surged at the level of organisation, stake holders and individual self for the social outreach of the anaesthesiologists as COMMUNITY PHYSICIANS.

*"To not follow where the path may lead. Go instead where there is no path and leave a trail."*

The social responsibility of anaesthesiologists is expanding by embarking upon new opportunities beyond the operating room for community commitment. Work and cultural barriers are being addressed to commemorate the distinct talents of anaesthesiologists by facilitating smooth transition to healthy synergy thus elucidating the Facts and Myths about Anaesthesiology in the common public and strengthening Universal Health Coverage.

The absence of a clinic and anonymity of the services offered by anaesthesiologists to the community is changing. The society is witnessing various diseases like communicable, non-communicable, infectious and malignancies. Lately, the subdued aspect of Pain is soaring in the society rebounding as ‘Pain Epidemic’ in our country. The anaesthesiologists are breaking the bounds of their operating rooms and

extending their natural course of professional work into the community. Pain Clinics are one of such service providing outlets which are escalatingly being established and promoted in the society thus ministering to the patients in pain. Pain Medicine is coming in an exuberant form with increasing awareness in the community involving various modalities. As per Mordor Intelligence, Asia-Pacific subcontinent has the highest pain management arena with Compounded Annual Growth Rate(CAGR) of approximately 8% .

<https://www.mordorintelligence.com/industry-reports/pain-management-market>

Thus, the Anaesthesiologists with their unparalleled zeal, enthusiasm and multifaceted qualities are the Torch Bearers in providing Pain-Relief Benevolences. They can act translational catalyst to be Influential Community Leaders thus traversing the gap from the confines of Operating rooms to the common public evincing as Community Physicians.

*“If you want to touch the past, touch a rock. If you want to touch the present, touch a flower. If you want to touch the future, touch a life.”*

# Snakebite Management in India: Challenges Remain

Kaushik Bharati

Health Policy Consultant – UNESCO New Delhi

## CORRESPONDING AUTHOR

Dr Kaushik Bharati, Health Policy Consultant – UNESCO New Delhi

E Mail ID: [dr.kaushik.bharati@gmail.com](mailto:dr.kaushik.bharati@gmail.com)



## CITATION

Bharati K. Snakebite Management in India: Challenges Remain. EFI Bulletin. 2023;4(1):32-34.

<https://doi.org/10.56450/EFIB.2023.v3i01.008>

## ARTICLE CYCLE

Received: 16/03/2023; Accepted: 21/03/2023; Published: 31/03/2023

This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

©The Author(s). 2023 Open Access

Snakebite continues to be a major problem in tropical countries, especially so, in India. India still has the highest number of deaths, which accounts for approximately half of the global burden. Management of snakebite victims remains a constant challenge in India, especially in rural settings. Even after the inclusion of snakebite envenoming in the priority list of neglected tropical diseases (NTDs) by the World Health Organization (WHO) in 2017, there is still a long way to go to reach the goal of halving the number of cases and deaths by 2030.

**Snakebite in India: A long neglected tropical disease**  
 Snakebite is a neglected public health problem in many tropical and subtropical countries worldwide, especially so in India. It is arguably the 'most neglected' of the NTDs. In fact, it is so neglected that until very recently, it wasn't even categorized as an NTD. In 2017, WHO recognized snake envenoming as a NTD, and included it in its list of top-20 priority NTDs (1). This is the result of intensive lobbying by snakebite activists around the globe who had pushed hard for a long time for categorizing snakebite envenoming as a NTD.

### The snakebite problem

Snakebite is a very serious problem in many parts of the globe, including India. It is estimated that approximately 5 million snakebites occur annually

worldwide, with up to 2.5 million envenomings. It is estimated that on average, 100,000 deaths occur annually around the globe (range: 81,000 – 138,000), while approximately 300,000 are maimed or permanently crippled for life (2).

Notably, the above data are hospital-based and therefore, a gross underestimate as many snakebite deaths occur in remote places, primarily in rural areas. Many of these victims die on the way to hospital to receive antivenom therapy. Therefore, the lion's share of snakebite fatalities go unreported (3).

Snake envenomation is a medical emergency as it often results in severe hemorrhage, rhabdomyolysis, shock, blood clotting defects, acute kidney injury (AKI), necrosis, paralysis of respiratory muscles, among other complications that require immediate medical attention. Sadly, permanent disability in the affected limb sometimes require amputation. The most vulnerable group are the agricultural workers, who venture into the paddy fields and often get bitten by venomous snakes. Children are also highly vulnerable as the same amount of venom causes much greater damage due to their small body size, compared to adults.

The first line of treatment for snakebites is administering first aid, which should be short, simple and quick. Not much time should be wasted on first

aid. It is very important to transport the victim as quickly as possible to a hospital, where antivenom is available, which is the only definitive treatment for snake envenomation. Indian snake antivenom is polyvalent, meaning that it is raised against more than one species of snake. In fact, Indian antivenom is manufactured from a cocktail of venoms from the 'Big Four' that are responsible for the highest number of deaths in India. These include the Indian cobra (*Naja naja*), Russell's viper (*Daboia russelii*), saw-scaled viper (*Echis carinatus*), and the common krait (*Bungarus caeruleus*).

### The Indian scenario

The most influential role is played by India in the context of the global snakebite mortality and morbidity statistics. India, being a vast and highly populous country, strongly influences the national snakebite epidemiological pattern. Notably the major chunk of the population (~70%) resides in rural areas, where snakebite incidents are more rampant. A landmark study, termed as the "Million Death Study", was conducted between 2001 and 2003, that used verbal autopsy to gather and analyze data. This study conducted a detailed nationwide snakebite mortality survey using data from 123,000 deaths from 6,671 randomly selected areas of India. This study found that approximately 45,900 snakebite fatalities occur annually in India, which is the highest in the world. Importantly, most of the deaths occurred in rural areas, accounting for 97% of all snakebite deaths in India. The states where the largest number of deaths occurred annually were Uttar Pradesh (8,700), Andhra Pradesh (5,200), and Bihar (4,500) (4).

The ground reality is that most of the primary health centers in rural India have suboptimal infrastructure and lack of trained doctors and nurses for tackling a snake envenomation emergency. As a result, many rural folk either turn to traditional healers such as "Ojhas" for help or don't seek medical care at all. This clearly hints at the fact that the number of snakebite victims is likely to be much higher than the official figures. Hence, due to this gross underestimation, snakebites should definitely be considered as a neglected problem in South Asia in general and India in particular (5,6).

### What can be done?

Firstly, categorization of snake envenomation as a NTD, has given its due importance on the global health agenda of health policymakers, which it previously lacked. The proactive role of WHO has ensured increased allocation of global funds for the management of snakebites. This has allowed India to leverage funds for training medical personnel, community mobilization, snakebite advocacy, as well as increasing the quality and quantity of antivenom production.

Currently, only seven Indian companies manufacture snake antivenom, which collectively provide approximately 1,958,000 vials (10 ml antivenom per vial), as per 2012 data (7). These companies should be provided incentives by the government to increase capacity-building, so that there isn't a shortfall of these life-saving medicines.

More public-private-partnerships (PPPs) and hand-holding among the various stakeholders will facilitate manufacture, testing, and quality control of snake antivenom. Moreover, leading vaccine manufacturers and start-ups should be roped-in and coaxed to produce snake antivenom. Linking these vaccine manufacturers to government organizations, such as the Central Research Institute (Kasauli, Himachal Pradesh) in a PPP mode will boost production capacity.

More international collaborations in the domain of medical research should be encouraged for developing novel and innovative ways of fighting the snakebite menace. Moreover, research strategies should also be focused on developing new therapeutic molecules and point-of-care rapid diagnostic tests (RDTs) for use in low-resource settings.

Therefore, there is an urgent need for a holistic approach, encompassing community participation, education of vulnerable populations, adequate training for medical staff to tackle snakebite emergencies, and ensuring availability, accessibility and affordability of snake antivenom, especially in rural areas. This multi-pronged approach is likely to drastically reduce the number of snakebite deaths in India.

## REFERENCES

1. World Health Organization. The NTD road map: Together towards 2030. Available from: <https://www.who.int/teams/control-of-neglected-tropical-diseases>; Accessed 12 March 2023.

2. World Health Organization. Snakebite envenoming. 17 May 2021. Available from: <https://www.who.int/en/news-room/fact-sheets/detail/snakebite-envenoming>; Accessed 12 March 2023.

3. Groneberg DA, Geier V, Klingelhöfer D, Gerber A, Kuch U, Kloft B. Snakebite envenoming – A combined density equalizing mapping and scientometric analysis of the publication history. *PLoS Negl Trop Dis.* 2016; 10(11): e0005046.

4. Mohapatra B, Warrell DA, Suraweera W, Bhatia P, Dhingra N, Jotkar RM, Rodriguez PS, Mishra K, Whitaker R, Jha P and for the Million Death Study Collaborators. Snakebite mortality in India: A nationally representative mortality survey. *PLoS Negl Trop Dis.* 2011; 5(4): e1018.

5. Alirol E, Sharma SK, Bawaskar HS, Kuch U, Chappuis F. Snake bite in South Asia: A review. *PLoS Negl Trop Dis.* 2010; 4(1): e603.

6. Warrell DA. Snake bite: A neglected problem in twenty-first century India. *Natl Med J India.* 2011; 24: 321-24.

7. Whitaker R, Whitaker S. Venom, antivenom production and the medically important snakes of India. *Curr Sci.* 2012; 103(6): 635-43.

# The future of Medical Writing: Artificial Intelligence as first author and not the ghost

Deepak Gupta

Department of Anesthesiology, Wayne State University,  
Detroit, Michigan, United States



## CORRESPONDING AUTHOR

Dr Deepak Gupta, Clinical Assistant Professor, Department of Anesthesiology, Wayne State University/Detroit Medical Center, Box No 162, 3990 John R, Detroit, MI 48201, United States  
E Mail ID: [dgupta@med.wayne.edu](mailto:dgupta@med.wayne.edu)

## CITATION

Gupta D. The future of Medical Writing: Artificial Intelligence as first author and not the ghost. EFI Bulletin. 2023;4(1):35-37. <https://doi.org/10.56450/EFIB.2023.v3i01.009>

## ARTICLE CYCLE

Received: 02/02/2023; Accepted: 17/03/2023; Published: 31/03/2023

This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

©The Author(s). 2023 Open Access

## ABSTRACT

### written by human

In the footsteps of pioneering Swedish preprint with artificial intelligence (AI) as first author, this article about dawning of futuristic medical writing has AI writing as first author turning AI the ghost writer aided and abetted by humans as a thing of the past.

Top P: 1

Frequency Penalty: 2

Presence Penalty: 2

Best of n=5

Second prompt generated text copy-pasted unedited within double quotations.

## BACKGROUND

### written by human

Now-a-days, humans are writing too much and too often either by choice or by coercion. Medical writing is no exception. Thus, humans may need artificial intelligence (AI) to help generating text when humans are not getting enough time to write so many mandated writeups coerced by peer pressure systems.

*"A team of researchers at MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) has developed an AI system that can create medical research papers. Dubbed 'Neural hypothesize-and-Edit Network' (NheNet), the artificial intelligence tool is capable of understanding complex concepts in a given field, summarizing them into coherent sentences, then going through each sentence to make sure it accurately conveys the intended meaning within context. The project began as part of ongoing efforts by CSAIL scientists to give machines more natural abilities when communicating with humans. They reasoned that if AI could understand historical information about diseases or treatments written by knowledgeable clinicians and researchers – something computers lack – such knowledge would be helpful for doctors completing paperwork efficiently but also enable nonprofessionals like patients and public health workers to access useful insight without having specialist skills or degrees."*

## Ghost Writer

### generated by AI

*Prompt text-davinci-003: Write about artificial intelligence acting as ghost writer for medical research papers*

*Temperature: 1*

*Maximum length: 500*

from universities medical schools . With NheNet , people will essentially have their own ghostwriters : The artificial intelligence takes input text on any topic related to medicine — say Type 2 Diabetes—analyzes relevant scientific literature summaries available data sets conducted experiments findings etc . All these resources help form basic hypotheses general overviews breakthrough theories opinions emerging trends conclusion results which may further provide interesting avenues inquiry explore specific queries answer core questions write high quality manuscripts presentable international conferences symposiums journals blogs websites ! Ultimately this learning process builds up level comprehension ability prediction acumen lead fewer manual revisions significantly faster edit rates refine draft publication state art clean format free scholarly errors typos misspellings deviations interpretation fluency acceptable standards length accuracy reading pleasure!"

## Plagiarism

### generated by AI

*Prompt text-davinci-003: Write about artificial intelligence controlling plagiarism while writing medical research papers*

*Temperature: 1  
Maximum length: 500  
Top P: 1  
Frequency Penalty: 2  
Presence Penalty: 2  
Best of n=5*

*First prompt generated text copy-pasted unedited within double quotations.*

*"The use of artificial intelligence to control plagiarism while writing medical research papers is an increasingly popular topic. As the quality and precision with which AI can detect copied content increases, more authors are relying on it to prevent unintentional or intentional acts of stealing intellectual property when preparing their work for publication. AI based plagiarism checking tools utilise a range of techniques, such as natural language processing algorithms and text-matching methods that analyse word choices along with sentence structure in order identify any portions where source material appears identical elsewhere online or within*

*[The future of Medical...] | Gupta D other works by different authors. Those identified passages then need reviewing against established rules from applicable journals/publications before being modified if necessary or replaced completely depending upon how closely they appear connected back originals sources found . Medical researchers should consider employing this technology at early stages during paper production processes so discrepancies between various iterations aren't overlooked until later down manuscript assembly lines; especially since unique contributions tend form basis for determining credibility behind clinical study findings released publically via peer reviewed publications - without these minimal levels copyright trustworthiness will inevitably suffer among educational institutions often awarding points (e.g., citation indexes) used evaluating researcher's productivity year over year trends related grant funding & career advancement opportunities made available through them within competitive academic industry sectors worldwide today"*

### FINAL THOUGHTS

#### written by human

Maybe AI is creating the future wherein more AI will be needed to write. This was not the scenario when there was no AI: as if AI is forcing a natural selection of those who write frequently and thus who will eventually seek more AI and become dependent on AI to write more frequently when humans cannot and will not be able to write that much humanely. The only concern remaining is that finding references or bibliography or further reading may still remain as human labor unless AI can also generate them at the same time when AI is generating text from them. This dystopic future is here but from here to where is anybody's guess to be fully deciphered only in the dystopic future.

### ACKNOWLEDGEMENT

#### written by human

With "Corrigendum to "Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse?" [Nurse Educ. Pract. 66 (2023) 103537]

<https://doi.org/10.1016/j.nepr.2023.103572>" failing the use of "ChatGPT" from OpenAI.com as an author due to its authorship not meeting the qualifications per journal's guide and publishing policy despite

“ChatGPT” as Artificial Intelligence making substantial contributions, similarly “Generative Pre-Trained Transformer -3 (GPT-3)” from OpenAI.com has been moved from authorship section to acknowledgement section to honor the substantial contributions of “Generative Pre-Trained Transformer -3 (GPT-3)” as Artificial Intelligence in writing this futuristic article for which the human author is deeply indebted to Artificial Intelligence’s undeniable contributions.

#### FURTHER READING

- OpenAI is an AI research and deployment company. Our mission is to ensure that artificial general intelligence benefits all of humanity. <https://openai.com/about/>
- GPT-3 Powers the Next Generation of Apps <https://openai.com/blog/gpt-3-apps/>
- We Asked GPT-3 to Write an Academic Paper about Itself—Then We Tried to Get It Published: An artificially intelligent first author presents many ethical questions—and could upend the publishing process <https://www.scientificamerican.com/article/we-asked-gpt-3-to-write-an-academic-paper-about-itself-mdash-then-we-tried-to-get-it-published/>
- Can GPT-3 write an academic paper on itself, with minimal human input? <https://hal.science/hal-03701250>

- Does the Use of AI to Create Academic Research Papers Undermine Researcher Originality? <https://www.mdpi.com/2673-2688/3/3/40>
- Researcher Tells AI to Write a Paper About Itself, Then Submits It to Academic Journal “All we know is, we opened a gate. We just hope we didn’t open a Pandora’s box.” <https://futurism.com/gpt3-academic-paper>
- Using AI to Improve the Medical Writing Experience <https://thejournalofmhealth.com/using-ai-to-improve-the-medical-writing-experience/>
- Abstracts written by ChatGPT fool scientists: Researchers cannot always differentiate between AI-generated and original abstracts. <https://www.nature.com/articles/d41586-023-00056-7>
- iThenticate <https://www.ithenticate.com>
- Top AI conference bans use of ChatGPT and AI language tools to write academic papers / AI tools can be used to ‘edit’ and ‘polish’ authors’ work, say the conference organizers, but text ‘produced entirely’ by AI is not allowed. This raises the question: where do you draw the line between editing and writing? <https://www.theverge.com/2023/1/5/2354029/1/chatgpt-ai-writing-tool-banned-writing-academic-icml-paper>

## BOOK REVIEW

## Book Review: Ergonomics in Healthcare: An Update

Rakesh Kakkar

Department of Community & Family Medicine, All India Institute of Medical Sciences Bathinda

### CORRESPONDING AUTHOR

Rakesh Kakkar, Department of Community & Family Medicine, All India Institute of Medical Sciences Bathinda 151001

E Mail ID: [drkakkar75@rediffmail.com](mailto:drkakkar75@rediffmail.com)



### CITATION

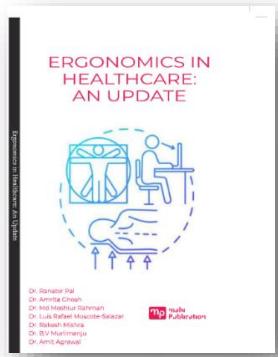
Kakkar R. Book Review: Ergonomics in Healthcare: An Update. EFI Bulletin. 2023;4(1):38.  
<https://doi.org/10.56450/EFIB.2023.v3i01.010>

### ARTICLE CYCLE

Received: 16/03/2023; Accepted: 21/03/2023; Published: 31/03/2023

This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

©The Author(s). 2023 Open Access



Authors: Dr Ranabir Pal, Dr Amrita Ghosh, Dr Md Moshiur Rahman, Dr Luis Rafael Moscote Salazar, Dr Rakesh Mishra, Dr B V Murlimanju, Dr Amit Agrawal

Imprint: Mahi Publications

Published: 2022 ISBN: 97893-94238-23-7

Length: 155 pages MRP: 399

**Ergonomics In Healthcare: An Update**, authored by renowned experts in public health, is a nice compilation for increasing awareness regarding ergonomic friendly environment in Health care services. It will be beneficial for the healthcare providers, residents UG and PG students for wider usage in clinical and public health practices of all levels.

This book is divided into Eleven Chapters addressing the countless issues encountered for managing occupational health. Book is intended to understand the principles and practices of ergonomics for healthcare professionals by covering an array of aspects related to professional risk factors.

Initial few chapters give generalized recommendations, importance of applied digital technologies and advances in understanding human performances for healthy work habits. While next few sections are specific for application of ergonomics at Primary care, for pediatricians and surgeons and in last few chapters authors have focused on recommended working schedules, controversies and advantages of implementation of human factors and system engineering by adopting instruments, improved methods, trainings and creation of awareness for improving patient safety. It contains material on cognitive aspects, real life applications, and designing methods of implementation of 'Ergonomics and human factors' in safety management across different aspects of professional care paradigm in health and disease. Health care organization have to prioritize the physical and mental health needs of health care providers and create safer work environments for all physicians which will further improve patient care and safety.

Lucid reading of contents will help professionals related with research and developments of medical device, logistics and technologies, the industry partners and stakeholders involved in health care delivery systems also will get 'Food for thought' in futuristic vision. Ergonomics In Healthcare: An Update covers all aspects conglomerated from engineering, medicine, infection prevention, trauma and emergency medicine, neurosurgery, natural sciences, computer and information science, healthcare, life science to professional burnout which are vital not only for quality health care services at physician level but also for patient safety and satisfaction.

## REPORT

# Clinical epidemiology and biostatistics workshop 2023 report

Anil C Mathew

Department of Community Medicine, PSG Institute of Medical Sciences and Research, Coimbatore

## CORRESPONDING AUTHOR

Dr. Anil C Mathew, Professor of Biostatistics, Department of Community Medicine, PSG Institute of Medical Sciences and Research, Coimbatore



E Mail ID: [dranilmathew@rediffmail.com](mailto:dranilmathew@rediffmail.com)

## CITATION

Mathew AC. Clinical epidemiology and biostatistics workshop 2023 report. EFI Bulletin. 2023;4(1):39-40.

<https://doi.org/10.56450/EFIB.2023.v3i01.011>

## ARTICLE CYCLE

Received: 16/03/2023; Accepted: 17/03/2023; Published: 31/03/2023

This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

©The Author(s). 2023 Open Access

This workshop was conducted from February 13-24th, 2023 at PSG Institute of Medical Sciences and Research. This workshop was mainly targeted to medical postgraduates to provide them with the knowledge to evaluate and judge applied clinical research and data analysis and give a sufficient scientific and methodological background to actively participate in clinical studies. The main objective of the workshop was to make the postgraduates to understand and be familiar with bio-statistical principles in clinical diagnosis and management and to understand the reports of the medical research done by others. The topics covered include: formulate a concise clinical research question; find and appraise the level of evidence and apply the information to patient care (evidence based medicine); different study designs in clinical epidemiology and identify the best type of study to answer clinical research questions; locate and critically appraise original research articles from journals; identify major sources of bias and their likely influence on results and implications in clinical practice; understand and interpret 95% confidence intervals and p values; understand, calculate and interpret prevalence, incidence, relative risk and odds ratio; calculate and interpret sensitivity, specificity; pre- and post-test probability of disease. In addition, hands on training on EXCEL, SPSS, EPIINFO and MENDELEY was conducted.

A total of 39 participants of which 19 internal, 13 external medical postgraduates and 7 MPH students attended and got benefitted from this program.

The session feedbacks were: it was a good experience, very well conducted, very useful interactive sessions with group activity, lot of new information, hands on training was very useful, able to clarify all doubts in a friendly manner and in general more than 95% reported in each session that it was excellent/good.

The workshop was conducted by Dr. Anil C Mathew, Professor of Biostatistics and Dr. V Sandhiya, Assistant Professor, Department of Community Medicine led by Dr. Sudha Ramalingam, Professor and Head of Community Medicine. All faculty in the Department of Community Medicine were the resource persons. In addition, there was cultural events by the participants. MCQ Test was conducted at the end of the workshop and the best participant was given a prize.

TNMC Observer Dr. Devaraj inspected the session and TNMC has awarded 6 credit hours for the participants and 7 credit hours for resource persons for this event



**Clinical epidemiology and biostatistics workshop 2023 dated 13-24<sup>th</sup> February 2023 Organized by Division of Biostatistics, Department of Community Medicine, PSG Institute of Medical Sciences and Research, Coimbatore, South India**





# EFI Bulletin

## Bulletin of Epidemiology Foundation of India

### Editorial

#### COVID-19 vaccination: the way forward

Sanjay Rai, Shreya Jha, Shashi Kant

1 - 2

### Review Article

#### Socioeconomic scales for rural versus urban settings in Health Research: A Review

Ravleen Kaur Bakshi, Neeta Kumar, Gagandeep Singh

3 - 11

Continuing Medical Education

#### Longitudinal data analysis methods – a primer

Nishaant Ramasamy, Ruben Raj, Namita Patel, Palanivel C

12 - 17

### Commentary

#### Adaptive platform trials present a new frontier for epidemiologists in the era of Covid-19

Manya Prasad, Umesh Kapil

18 - 19

#### Strengthening provisions for breast cancer screening in India

Ankan Mukherjee Das

20 - 22

### Original Article

#### Latent class analysis in a complex sampling design – an advanced modelling approach in Epidemiology

Ankita Dey, Diganta Mukherjee, Sugata Sen Roy

23 - 29

### Perspective

#### Anaesthesiologists as Community Physicians- Expanding Horizons

Naveen Malhotra, Divya Gupta

30 - 31

#### Snakebite Management in India: Challenges Remain

Kaushik Bharati

32 - 34

### Opinion

#### The future of Medical Writing: Artificial Intelligence as first author and not the ghost

Deepak Gupta

35 - 37

### Book Review

#### Book Review: Ergonomics in Healthcare: An Update

Rakesh Kakkar

38

### Report

#### Clinical epidemiology and biostatistics workshop 2023 report

Anil Mathew

39 - 40