Social Inequality Lab

Working Paper

WP-SIL-03

2023

Analyzing Infant Mortality Inequality: Identifying Risk Factors and Addressing Disparities

Author:
Khadija Mujahid

IBA SESS
School of Economics and Social Sciences

ANALYZING INFANT MORTALITY INEQUALITY: IDENTIFYING RISK FACTORS AND ADDRESSING DISPARITIES

Khadija Mujahid

khadijamujahid.07479@khi.iba.edu.pk

Abstract

The infant mortality rate (IMR) varies by area, with Pakistan and other developing nations seeing high prevalence. At 86 deaths per 1000 babies born, infant mortality in Pakistan was notably high until the beginning of the 1990s. According to the World Bank data, it decreased by 24 points in the past three decades to 62 fatalities per 1000 in 2015 though Pakistan's IMR discrepancy has not yet been adequately addressed. Pakistan continues to be among the nations with the highest infant mortality rates; thus, we looked at the IMR inequality and identified the potential risk factors using the most suitable methodologies, such as principal component analysis (PCA) for wealth index, to draw conclusions based on evidence. The analysis for this study included a total weighted sample of 12,708 mothers of childbearing age who reported experiencing infant mortality in the previous five years as part of the Pakistan Demographic Health Survey (PDHS). The count models, Poisson and Negative Binomial, assessed the risk indicators associated with infant mortality in Pakistan. A closer look at the number of infant deaths showed over-dispersion because the variance (0.389) is higher than the mean (0.250). AIC (14961.63), and BIC (15170.23) values of the multilevel negative binomial regression model are the lowest making it the best model for assessing infant mortality in Pakistan. Baluchistan had the highest number of infant deaths per mother against the babies born in the last five years (167, 6.19%). In FATA, 22 infants died out of 1883 babies born making the lowest proportion of infant deaths (1.17%). IMR was more prevalent in female babies from the most impoverished subpopulations, such as the uneducated, poorest, and rural residents. The poorest households, rural people, women without a formal education, and female infants must all be the focus of efforts to eradicate the observed inequities.

Keywords: Infant mortality, Inequality, DHS, Global Health

Introduction

Infant mortality (IMR) is essential to the community's physical and socioeconomic well-being. Globally, the infant mortality rate decreased from 8.8 million in 1990 to 4.1 million in 2017, i.e., 65 to 29 deaths in every 1000 live births, accounting for nearly three-fourths of all deaths of children under five. For the same period, infant mortality rates in Pakistan decreased from 90 to 65 for every 1000 live births (World Bank, 2019). About 330,479 infants in Pakistan died before observing their first birthday, making it one of the places with the highest IMR (64 on every 1000 live births), over eight times greater than the WHO European Region's average of 8 in every 1000 (World Health Organization, 2019).

In Pakistan, infant mortality significantly decreased in the 1950s and the beginning of the 1960s due to widespread government health initiatives such as vaccination campaigns. Between the early 1960s and the end of the 1970s, the decline in the rate of infant mortality slowed (Agha, 2000). According to two independent national sample surveys, the IMR was 101 on every 1000 live births from 1989 to 1992, indicating that by the end of the 1980s and the beginning of the 1990s, the infant mortality rate appeared to have reached a plateau (Agha, 2000). These infant mortality rates align with previous IMR estimates for this period (Hoadley, 1995). However, according to (Irfan, 1986), Pakistan's infant mortality appears to have plateaued when regional and urban-rural differences were notably large, and the IMR was comparatively high. Regional variations in IMR result from socioeconomic crises and distributional concerns.

Women's legal status advanced significantly in the early 1960s; however, this development was reversed by a government-endorsed "Islamization" movement that started at the end of the 1970s and continued through the course of the 1980s. The infant mortality rate in Pakistan was relatively high until the early 1990s; during the past three decades, it has fallen by 24 points, reaching 62 deaths per 1000; yet, Pakistan continues to be among the nations with the highest infant mortality rates (Patel, Rai, & Rai, 2021).

There is little doubt that the world's countries are reducing child mortality rates in diverse ways, and this difference in reduction seems to be a challenge that the global community is currently experiencing. Infants' overall health and mortality rates are essential proxies for a country's position socioeconomically and general quality of life (Puri, Gaye, Kurukulasuriya, & Scott, 2007). High infant mortality rates (IMR) typically point to unmet needs for essential healthcare, including hygienic conditions, nutrition, medical facilities, and education. According to the World Bank, as of 2021, the population of Pakistan is approximately 225.2 million. After countries such as India, China, the United States, and Indonesia, Pakistan is ranked as the fifth fastest-growing nation in terms of population in the world (World Bank, 2021). It scores inadequately on all human development metrics, and unequal access to finite economic resources compromises its political stability.

Literature Review

The life course perspective emphasizes the importance of considering health outcomes across the lifespan, from before birth through adulthood. A study carried out in Ghana discovered that, despite the disparity between children of mothers with no schooling at all and those with secondary or higher education appearing to have reduced in 2014, under-five mortality was still disproportionately high among children of mothers with no schooling at all (Agbadi, et al., 2021). The study concluded that interventions to improve early childhood development and maternal education could substantially impact reducing infant mortality rates.

The numerous and interconnected facets of social identities, such as race, gender, class, and sexual orientation, are highlighted by the intersectionality framework. For instance, a study undertaken in Yemen discovered that the incidence of IMR was higher in male babies from disadvantaged subpopulations, such as the poorest, uneducated, and residents of rural areas (Zegeye, Shibre, Haidar, & Lemma, 2021). To eliminate the observed inequities, interventions must focus on the poorest individuals, households in rural areas, women with no educational experience, and male infants.

The social determinants of health framework highlight the relevance of environmental, social, and economic factors in determining health consequences, notably the death rate of infants. A study in Brazil discovered that greater infant death rates were linked to higher poverty rates, inadequate sanitation, and lacking primary healthcare (Alves & Belluzzo, 2004). The study concluded that measures to reduce poverty and improve access to clean water and healthcare could help lower infant death rates. According to a study's findings, actions to eliminate poverty and increase access to healthcare and clean water could help reduce the number of infant death rates. Congenital heart disease (CHD) infant mortality in the USA varies significantly by region. This underscores the potential advantages of enhancing medical services for babies residing in disadvantaged areas and places away from quality pediatric cardiology institutions. (Udine, Evans, Burns, Pearson, & Kaltman, 2021). Researchers can develop a more nuanced understanding of the diverse and multifaceted mechanisms causing differences in infant mortality across regions and gender by considering these discussions and theories. Such understanding can help develop policies and actions to reduce these inequalities and advance health equity.

The results of a systematic review indicate that socioeconomic traits and social circumstances are important risk factors for IMR (Kim & Saada, 2013). Differences in socioeconomic strata, access to healthcare facilities, and environmental factors directly impact IMR (Ruiz, et al., 2015). IMR discrepancies in middle- and low-income countries are typically caused by growing socioeconomic imbalance, such as differences in per capita national income, mother's level of education, and access to healthcare services (Emamgholipour Sefiddashti, Nakhae, Kazemi Karyani, & Ghazanfari, 2015). Gender differences and residencies are two other significant characteristics that affect IMR (Ely, Driscoll, & Mathews, 2017). At the same time, further research is needed to provide data regarding the scope of the issue, generally and explicitly concerning study settings.

Several countries have researched the risk factors contributing to infant mortality (Adulo & Zewudie, 2021). Numerous small-scale studies have been conducted to look into a certain collection of characteristics. Binary logistic and survival factors were examined in these analyses (Kanmiki, et al., 2014). Binary logistic regression gives enough data to examine multiple child fatalities, even though it undercounts the overall death toll since several deaths are aggregated into a unit or value to meet the requirements for the binary logistic model. In this work, multilevel count regression was used as the analysis technique. The number of newborn fatalities (count) is the response variable, and the objective is to estimate how this count evolves when more explanatory factors are added.

For count data analysis, Poisson regression modeling is a widely recognized approach. Its application is restricted in some real-life situations with uneven distributions (i.e., the mean is not equal to variance), despite its underlying premise that there is evenness in the distribution (i.e., the mean is equal to variance). Parameter estimates, standard errors, test scores, and confidence intervals may be skewed when there is considerable deviation. There are numerous reasons for over-dispersion. Examples include censorship and excessive zero counts. Several fields use over-dispersed count data, which stimulated the introduction of statistical modeling methods for these data types (Sellers & Shmueli, 2013). The Poisson and negative binomial distributions are similar, but the latter has a wider, fatter tail whenever the variance is greater than the mean. The multilevel negative binomial model can account for more zeros than the multilevel Poisson model, depending on the extent of over-dispersion.

There have been a few studies on newborns and child analyzing their survival in Pakistan, but they used a limited number of determinants and lacked a solid conceptual framework. Generally speaking, these research studies have not used nationally representing data. The determinants affecting infant survival in Pakistan are examined in greater detail in this study, using high-quality representative data from across the nation, covering a wide range of covariates that affect the health of babies, and applying a sound statistical approach.

Another fundamental issue that policymakers must be aware of to employ their limited resources wisely is how IMR varies even within a sovereign country (Sartorius & Sartorius, 2014). The initial prerequisite for policies based on evidence to address health inequities is obtaining evidence of health data stratified by the country's demographic groupings. With this preliminary knowledge, we used a variety of absolute and relative measures to analyze the size of IMR by sociodemographic status and economic position in Pakistan. The purpose of this study's policy implications is to provide scholars and policymakers with evidence.

Methodology

3.1 Data

There are four Pakistan Demographic and Health Surveys (PDHS): in 1991, 2007, 2013, and 2018. This is a routine cross-sectional survey that is carried out among Individuals and households that are typically nested within one another within clusters in the DHS data, which is typically hierarchical. The first step is to prepare the data by creating variables for the individual- and community-level factors likely to influence IMR.

3.2 Study Area

The DHS is carried out every five years, and Pakistan's 2018 survey, which covered all nine areas and two administrative cities, was the fourth to be conducted. The PDHS acquired data from key indicators and samples of all demographic groups that were nationally representative. The survey acquired information on sociodemographic, socioeconomic, and maternal variables.

To guarantee that the sample is representative of Pakistan's population, research participants were selected using the stratified two-stage cluster sampling procedure. The first stage involves selecting EAs, which are geographic areas defined by the Pakistan Bureau of Statistics, and the second stage involves selecting households within the selected EAs. Rural and urban areas, as well as the following provinces of Pakistan, Sindh, Punjab, Baluchistan, Khyber Pakhtunkhwa, Gilgit-Baltistan (GB), Islamabad Capital Territory (ICT), Azad & Jammu Kashmir (AJK), and Federally Administrative Tribal Area (FATA), define the strata in the PDHS.

A total of 16,240 households were selected for the survey's overall sample from 580 primary sampling units (PSUs) to form a nationally representative sample (Pakistan Demographic and Health Survey 2017-18, 2019). Random sampling was used to choose Pakistani households in two rounds. 580 enumeration areas (EAs) were chosen for the first stage, of which 561 of these EAs completed the survey successfully, using a population proportionating to the EAs recorded by the Census. 28 homes in each cluster (EAs) were picked for the second phase to produce statistically significant estimates of the nation's major demographic and health variables.

Women between the ages of 15 and 49 were the target demographic. These ladies were qualified to participate in the survey interviews since they were either regular home inhabitants or had spent the previous night there. 19,489 married women between the ages of 15 and 49 participated in interviews, and all survey data were adjusted for stigmatization at the national level. The analysis for this study included a total weighted sample of 12,708 mothers of childbearing age who reported experiencing infant mortality in the previous five years as part of the survey (Figure 1).

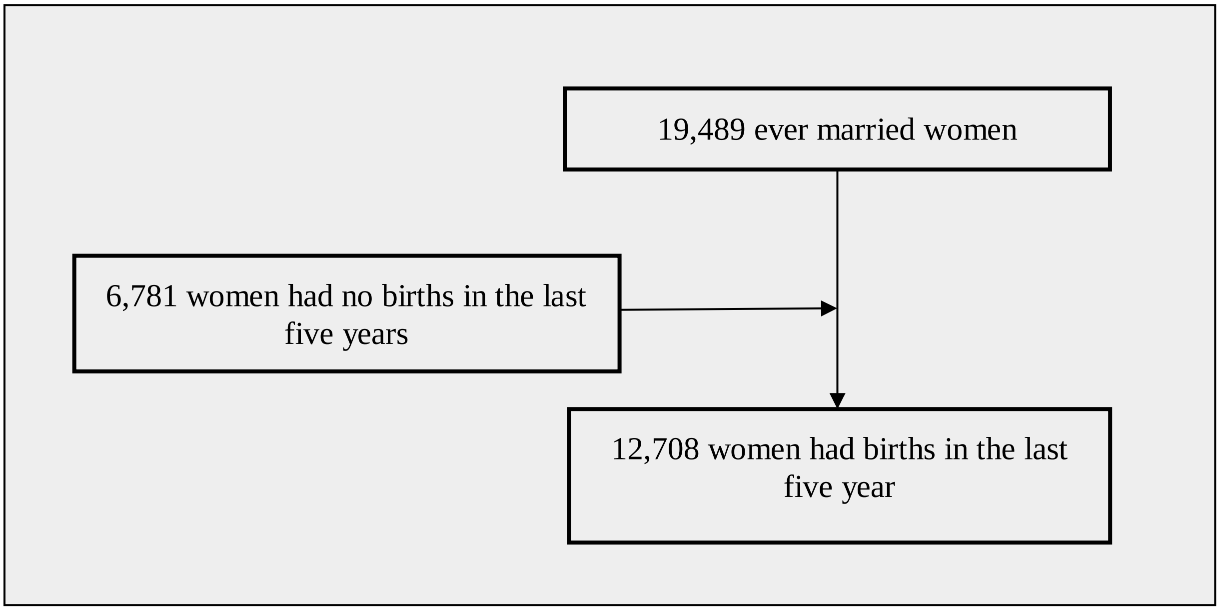

Figure 1: Sample selection flow

Note: 19,489 married women between the ages of 15 and 49 participated in interviews. The analysis for this study included a total weighted sample of 12,708 mothers of childbearing age who reported experiencing infant mortality in the previous five years as part of the survey. Due to confidentiality concerns, the titles of all enumeration areas selected in the PDHS are not publicly available.

3.3 Variables and Measurements

The Woman's Questionnaire's pregnancy history section, which includes questions on a woman's overall experience of having children, such as the number of children living with her or elsewhere, is the data source used to calculate child mortality estimates. Birth records also included all live births, including those of babies who later passed away but did not have stillbirths, miscarriages, or abortions. From first to last, birth histories are compiled chronologically. The analysis includes children who were born five years before the questionnaire.

Based on findings from the 1990–1991; 2006–07; 2012–13; and 2017–18 PDHS surveys, the 2017–18 PDHS illustrates a pattern of declining childhood mortality over the previous three decades. Infant deaths in every 1,000 live births went down from 86 to 62. For this research, however, the variable we are interested in measuring the inequality of is IMR.

A variance inflation factor (VIF) with a value below five implies that a variable correlates poorly with other predictors. A VIF value between 5 and 10 implies a moderate correlation, whereas VIF values greater than 10 imply a significant, unacceptably high association of model variables (James, Witten, Hastie, & Tibshirani, 2013). There is no multi-collinearity because the variance inflation factors for variables are smaller than 10.

The study categorized the variables into response and predictor variables.

3.3.1 Dependent Variable

The response variable is the number of infant deaths per woman, defined as the death of a baby before they turn one.

3.3.2 Independent Variables

Various factors could potentially predict infant mortality, including but not limited to the geographical region, the education levels of the mother and father, the occupation status of the mother, household size, the mother's age at her first delivery, access to an improved water source, wealth index, contraceptive use, place of residence, and female literacy.

3.3.3 Statistical Analysis

In a multilevel model, the dependent variable is modeled as a function of individual- and group-level predictors. In this case, the individual-level predictors are the mother's age at her first delivery, the education level of the father, etc., and the group-level predictor is the cluster. The number of infant deaths per woman is presumed to have a Poisson distribution, given that it is a count variable.

There are two levels in this multilevel model:

1. Level 1 (Individual level) - This level includes all the individual-level predictors, such as the mother's age at her first delivery, the education level of the father, the education level of the father mother, etc. The dependent variable, i.e., the number of infant deaths per woman, is also at this level.
2. Level 2 (Group level) - This level includes the group-level predictor, i.e., clusters. The random intercept for each cluster is also included in this level.

The multilevel model can be represented as:

Level 1:

$Y_{ij} = \beta_0 + \beta_1(father's\ education_{ij}) + \beta_2(mother's\ education_{ij}) +$

$\beta_3(father's\ occupation_{ij}) + \beta_4(mother's\ occupation_{ij}) + \beta_5(household\ size_{ij}) +$

$\beta_6(mother's\ age\ at\ first\ birth_{ij}) + \beta_7(improved\ water\ source_{ij}) + \beta_8(wealth_{ij}) +$

$\beta_9(contraceptive\ use_{ij}) + \varepsilon_{ij}$

(Equation 1)

Level 2: $\beta_0 = \gamma_{00} + \mu_{0j}$

(Equation 2)

where:

1. $\mathrm{E}(Y_{ij})$ is the expected number of infant deaths per woman in j region and i individual
2. β_0 is the intercept.
3. β_1 to β_9 are the coefficients for the individual-level predictors.
4. γ_{00} is the average intercept across all clusters
5. μ_{0j} is the random intercept for cluster
6. ε_{ij} is the random error term for individual i in cluster j

This model allows us to account for the variations in infant mortality rates among clusters in Pakistan and quantify the impact of each predictor on the number of infant deaths per woman. The random intercepts at the group level allow for variation in the intercept across clusters, i.e., estimate between cluster variability of the intercept, capturing the effect of unobserved factors that vary across regions.

The Pakistan Demographic and Health Survey (DHS) data can be used to estimate infant death risk variables using multilevel Poisson and negative binomial regression. The methodology for each model is as follows:

3.3.4 Multilevel Poisson Regression

The methodological steps for this model are similar to the ones described earlier. The main difference is that the outcome variable is assumed to follow a Poisson distribution, which is appropriate when the outcome is a count of events (such as the frequency of infant deaths per woman). In practical terms, there is a compromise on the assumption that issues such as dependency error and heteroscedasticity are present even while the quality of the test value, as assessed by R Square, demonstrates that the proposed model can explain the outcome variable. The following is the two-level variance-component Poisson regression model for a frequency Y_{ij} for unique women nested within jth region:

$y_{ij}|\mu_{ij} = poisson(m_{ij}, \mu_{ij})$

(Equation 3)

$\ln(\mu_{ij}) = \beta_0 + u_j$

(Equation 4)

$u_j \sim N(0, \sigma_u^2)$

(Equation 5)

where μ_{ij} denotes the predicted count, β_0 denotes the intercept, and u_j is the cluster random intercept effect, assumed normally distributed with a mean of zero and variance σ_u^2.

3.3.5 Multilevel Negative Binomial Regression

The over-dispersion in the data, which happens when the mean is less than the variance, is considered in this model, comparable to the Poisson regression model. With the premise that the outcome variable follows a negative binomial distribution, the methodology is the same as for the Poisson regression model.

The model can be computationally explained as follows:

$y_{ij}|\mu_{ij} = poisson(m_{ij}, \mu_{ij})$

(Equation 6)

$\ln(\mu_{ij}) = \beta_0 + u_j + e_{ij}$

(Equation 7)

$u_j \sim N(0, \sigma_u^2)$

$exp\ (e_{ij}) \sim Gamma(\frac{1}{\alpha}, \alpha)$

Where the over-dispersion random effect is indicated by e_{ij}. Therefore, two units in this model with identical random intercept effect magnitude could possess different predicted counts ij. These disparities are explained by the two units having different values for the omitted unit-level variables. The exponentiated over-dispersion random effect $exp\ (e_{ij})$ is assumed to have a gamma distribution with shape and scale parameters $\frac{1}{\alpha}$ and α. It is consequently distributed with mean 1 and variance or over-dispersion parameter. The over-dispersion increases as α grows larger. When α becomes zero ($\alpha = 0$), the model converts into a Poisson regression model (Equation 4). When this occurs, we can compare the two models and perform a likelihood-ratio test to determine if the projected over-dispersion is statistically significant.

3.3.6 Variance partitioning component

Since multilevel modeling relies on the premise that clustering exists, reporting the degree of clustering is the first logical step in any multilevel analysis. When experts employ variance-component models, they note that a part of the variation in the dependent variable is attributable to systematically different behaviors (responses) among clusters. These statistics are also known as interclass correlation coefficients (ICCs) and variance partition coefficients (VPCs). These are more difficult to obtain when fitting multilevel models with count and categorical answers, such as binary, ordinal, or nominal.

A statistical method called variance partitioning analysis is used to decompose the total variance of a response variable into its components. These components represent the proportion of variance explained by different factors or variables.

Compared to other variance decomposition methods, such as ANOVA or regression analysis, variance partitioning analysis allows for the simultaneous consideration of multiple explanatory variables, including categorical and continuous predictors. Additionally, variance partitioning analysis can be applied to non-normally distributed data, such as count data, which is impossible with other methods. Regarding Poisson models and more adaptable negative binomial models that consider over-dispersion, we used derived VPC/ICC expressions (Leckie, Browne, Goldstein, Merlo, & Austin, 2022).

First, we tested these models by running single-level regression, with no level 1 and 2 predictors, to verify if any significant clustering is involved in our data. In the context of a single-level model, we are running the least square regression where one of the assumptions is that the residuals are independent. If there is a substantial clustering effect, the assumption of independent residual is violated, deflating the standard errors and increasing type 1 error in probability.

3.3.7 Parameter estimation and model comparison

The maximum likelihood estimation approach (MLE) is the most prevalent methodology used for finding out the parameters of a multilevel count regression model. To select the parameters, the MLE estimation method maximizes the likelihood of the sample data. Many models and forms of data can be used with MLE approaches, which are adaptable. We used Stata commands "mepoisson" and "menbreg" to fit all models using the MLE technique using adaptive quadrature (StataCorp., 2019). To examine alternative models, we employed Akaike's Information Criteria (AIC), Bayesian Information Criteria (BIC), and Deviance Information Criteria (DIC).

Result

The study involved 12,708 women in total, with 10,345 (81.41%) of the moms reporting no infant deaths and 4 (0.03%) reporting eight infant fatalities in five year indicating the possibility of twin, or triplet, etc. infant deaths (*Table 1*). A closer look at the number of infant deaths showed over-dispersion because the variance (0.389) is higher than the mean (0.250). Count data models considering surplus zeroes could provide a better fit for the data.

Table 1: Summary statistics of infant deaths per woman

Infant death per mother	Number of mothers[1]	Percent	Cumulative percentage
0	10,345	81.41	81.41
1	1,810	14.24	95.65
2	388	3.05	98.70
3	110	0.87	99.57
4	37	0.29	99.86
5	4	0.03	99.89
6	3	0.02	99.91
7	7	0.06	99.97
8	4	0.03	100.00
Total	12,708	100.00	
Mean	0 .2505		
Variance	0 .3898		
Skewness	3.85628		
Kurtosis	26.9880		

[1] A total weighted sample of mothers of childbearing age who reported experiencing infant mortality in the previous five years as part of the survey

Punjab had the share of mothers with infant deaths (273, 27.9%) in absolute terms; however, Baluchistan had the highest number of infant deaths per mother against the babies born in the last five years (167, 6.19%). In FATA, 22 infants died out of 1883 babies born making the lowest proportion of infant deaths (1.17%).

Table 2: Proportion of infant deaths in each region given the births in the last five years

Region	**Total number of mothers**[1] N (%)	**Births in the last five years** N (%)	**Infant deaths**[2] **n (%)**	**Infant deaths per mother**[3] **(%)**	**Deaths per 1000 babies born**[4]
Punjab	2759 (21.7%)	5275 (22.7%)	273 (27.9%)	5.18	52
Sindh	2278 (17.9%)	4162 (18%)	176 (18%)	4.23	42
KPK	2097 (16.5%)	3747 (16.2%)	151 (15.4%)	4.03	40
Baluchistan	1508 (11.9%)	2698 (11.7%)	167 (17.1%)	6.19	62
GB	915 (7.2%)	1607 (6.9%)	58 (5.9%)	3.61	36
ICT	810 (6.4%)	1430 (6.2%)	56 (5.7%)	3.92	39
AJK	1320 (10.4%)	2352 (10.2%)	75 (7.7%)	3.19	32
FATA	1021 (8.03%)	1883 (8.1%)	22 (2.2%)	1.17	12
Total	12708 (100%)	23154 (100%)	978 (100%)	4.22	42

[1] A total weighted sample of mothers of childbearing age who reported experiencing infant mortality in the previous five years as part of the survey. [2] Infant deaths represent babies per mother by province who died before reaching 12 months (less than 1 year). [3] The percentage of infant deaths pr mother is calculated by dividing infant deaths by births in the last five years. [4] Death per 1000 babies born is calculated by multiplying the percentage of infant deaths per mother by 10; for example, 0.0518 multiplied by 1000 or 5.18 multiplied by 10 gives 52 deaths per 1000 babies in Punjab.

4.1 Variance partitioning component identifying cluster effect

Equation 3 describes the two-level variance-components Poisson model, which integrates a cluster's random intercept to account for potential clustering effects. The calculated intercept, which represents the average over clusters, is -1.742. The model calculates the cluster variance to be 0.586, and the comparison between a multilevel Poisson model and a single-level Poisson model without covariates results in a p-value less than 0.001 and a chi-square value equal to 962.26, which is statistically significant. We answered whether the value of 0.586 is large for variance using VPC and the estimated marginal statistics.

Based on the parameter estimates, the marginal expectation is 0.235, and the marginal variance is 0.279. The marginal expectation is roughly the same as the sample mean deaths per woman (0.250). However, the marginal variance is below the sample variance of 0.3898; hence, this model is unsuitable considering our data. To understand this last point, it is revealing to decompose the estimated marginal variance into level-specific components. The variance component at level 2, a cluster in our case, equals 0.0439, while the level 1 component (individual) equals 0.235. The resulting level 2 VPC is equal to 0.158, indicating systematic group variations account for 15.8% of the marginal variance. This suggests that at least community-level variation/ factors in individual-level characteristics account for almost 16% of the modeled variation in infant deaths between individual mothers.

The negative binomial model eases the condition that each component of the marginal variance must match the marginal expectation, allowing the data to be more distributed than the mean would indicate. As a result, the model ought to provide a more precise estimate of the marginal variance and a better fit for the data. Equation 7 is a negative binomial model with two levels of variance components. The model incorporates an individual over-dispersion random effect to take into consideration any within-cluster variance brought on by the influences of omitted individuals; the model incorporates an individual over-dispersion random effect. The over-dispersion parameter calculated has a value of 0.9915, and a likelihood-ratio test reveals that there is significant over-dispersion (Multilevel negative binomial vs. Multilevel Poisson model: χ 2 = 513.98, p < 0.000).

The calculated marginal variance now goes up from 0.279 to 0.370, which is noteworthy. The cluster component continues to be stable (Poisson = 0.044 and negative binomial = 0.040), as expected. Therefore, the individual component, which grows from 0.235 to 0.3306 (no longer required to be equal to the marginal variance), causes an increase in the marginal variance. The model predicts some degree of intra-group variability driven by omitted characteristics of individuals, which shows that even within clusters, infant mortality is far from a random Poisson process. This increase in the individual-level component, in turn, dramatically impacts the estimated VPC. Given that the model's estimation of the cluster-level VPC is now 0.107, it appears that omitted mother-specific factors, rather than omitted cluster-specific factors, are the probable primary source of the disparity in infant mortality. The value of the calculated VPC is smaller than the anticipated VPC for a multilevel Poisson of 0.158. A noteworthy conclusion is that the Poisson model significantly exaggerates the relative relevance of clusters in this data by disregarding over-dispersion. When there is over-dispersion, the Poisson VPC is biased upward. More generally, the VPC for the Poisson model is biased upwards in the existence of over-dispersion.

4.2 Socio-demographic and economic characteristics of infant deaths

Punjab has the highest (0.297) mean infant death per woman, while ICT has the lowest (0.179). Infant deaths per woman had a greater mean number in rural than urban areas (mean = 0.197 vs. mean = 0.293). The infant death per woman with no education is (0.324), which is higher than for mothers with other levels of education. Similarly, the mean number of infant deaths for the poorest mothers (0.323) is higher than for the wealthiest mothers (0.154). As a result, a woman from the highest socio-economic stratum had the lowest average number of infant deaths than a woman in the middle and lowest. Male infant death per woman is higher (0.254) than female infant mortality.

The differences between conditional means and conditional variances suggest that over-dispersion exists and that a Negative Binomial model would be appropriate (Table 3).

Table 3: Summary statistics of infant deaths per woman and variable characteristics

	Infants deaths per woman			
Variables	**Mean**	**Variance**	**Number of women interviewed**[1]	**Infant deaths**
Mother Education Level				
No Education	0.3242498	0.533033	6532	2118
Primary Education	0.2574257	0.3474463	1717	442
Secondary Education	0.1691316	0.2247979	2637	446
Higher Education	0.097146	0.103133	1822	177
Total	0.2504721	0 .389844	12708	3183
Husband Education Level				
No Education	0.3629055	0.6422509	3607	1309
Primary Education	0.2652608	0.3582484	1802	478
Secondary Education	0.2289103	0.3126286	4469	1023
Higher Education	0.1318021	0.1795114	2830	373
Total	0.2504721	0.389844	12708	3183
Socio-economic strata				
Poorest	0.3234687	0.5014419	5812	1880
Middle	0.2498008	0.437774	2510	627
Richest	0.1541268	0.1983604	4386	676
Total	0.2504721	0.389844	12708	3183
Province				
Punjab	0.2968467	0.5024955	2759	819
Sindh	0.2537313	0.4072654	2278	578
KPK	0 .2470196	0.4065095	2097	518
Baluchistan	0.301061	0.4733366	1508	454
Gilgit Baltistan	0.2961749	0.3881145	915	271
ICT	0.1790123	0.2806467	810	145
AJK	0.194697	0.2342402	1320	257
FATA	0.1380999	0.1485568	1021	141
Total	0.2504721	0.389844	12708	3183
Residence				

Rural	0.1966482	0.2710587	5609	2080
Urban	0.292999	0.4796509	7099	1103
Total	0.2504721	0.389844	12708	3183
Sex of child				
Male	0.2540341	0.3921121	6507	1653
Female	0.2467344	0.3874994	6201	1530
Total	0.2504721	0.389844	12708	3183

[1] A total weighted sample of mothers of childbearing age who reported experiencing infant mortality in the previous five years as part of the survey

4.3 Model selection criteria

Using AIC, BIC, and Deviance statistics, we utilized the findings to select the appropriate count model from the six regularly used count models. Any of these characteristics with a lower value represents a better fit or a more frugal model. As a result, AIC, and BIC values of the multilevel negative binomial regression model are the lowest. We draw this conclusion since the multilevel negative binomial regression model fits the data more accurately than the other count regression models (Table 4).

Table 4: Model selection criteria for multilevel models

	DF	AIC	BIC
Poisson (Null)	2	15670.09	15684.99
Individual Level	15	15276.71	15388.46
Cluster Level	27	15246.07	15447.22
Negative Binomial (Null)	3	15315.14	15337.49
Individual Level	16	14989.53	15108.73
Cluster Level	28	14961.63	15170.23

DF Degrees of freedom, AIC Akaike's information criterion and BIC Bayesian information criterion

The model with AIC= 14989.53 is preferred over the model with AIC=15315.14.

Table 5: The results of random intercept MLNB model on infant deaths in Pakistan, 2018 PDHS

Infant deaths per mother	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
Mother's education level						
Primary education	-0.126	0.068	-1.840	0.066	-0.26	0.008
Secondary education	-0.402	0.072	-5.550	0.000***	-0.544	-0.26
Higher education	-0.745	0.105	-7.070	0.000***	-0.951	-0.538
Husband's education level						
Primary education	-0.165	0.066	-2.500	0.012*	-0.294	-0.036
Secondary education	-0.138	0.056	-2.440	0.015*	-0.248	-0.027
Higher education	-0.401	0.079	-5.050	0.000***	-0.556	-0.245
Mother's age at first birth						
above 18	-0.344	0.051	-6.680	0.000***	-0.445	-0.243

Infant deaths per mother	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
Mother's employment status						
not working	-0.266	0.064	-4.160	0.000***	-0.391	-0.141
Ever used contraceptive						
Yes	-0.081	0.047	1.720	0.086	-0.011	0.174
Individual wealth status						
Middle	-0.046	0.067	-0.680	0.494	-0.178	0.086
Rich	-0.204	0.083	-2.460	0.014*	-0.367	-0.041
Improved water source						
Yes	0.191	0.091	2.110	0.035*	0.013	0.369
Household size	-0.015	0.005	-2.970	0.003*	-0.024	-0.005
Region						
Sindh	-0.432	0.116	-3.710	0.000***	-0.660	-0.204
KPK	-0.231	0.118	-1.960	0.050	-0.462	0.000
Baluchistan	-0.279	0.137	-2.040	0.042*	-0.548	-0.010
Gilgit Baltistan	-0.076	0.150	-0.510	0.610	-0.371	0.218
ICT	-0.394	0.158	-2.500	0.013*	-0.703	-0.085
AJK	-0.236	0.130	-1.820	0.069	-0.490	0.018
FATA	-1.067	0.176	-6.070	0.000***	-1.412	-0.723

Residence						
Rural	0.06	0.087	0.690	0.493	-0.111	0.23
Cluster-level poverty						
Moderate	-0.033	0.108	-0.310	0.759	-0.245	0.179
High	0.164	0.122	1.350	0.177	-0.074	0.403
Cluster-level literacy						
Medium	-0.097	0.101	-0.960	0.338	-0.294	0.101
High	-0.108	0.135	-0.800	0.424	-0.373	0.157
Mean intercept	-0.551	0.198	-2.790	0.005**	-0.938	-0.163
Dispersion	-0.173	0.086			-0.343	-0.004
v001						
Cluster variance	0.334	0.040			0.263	0.423

LR test vs. nbinomial model: chibar2(01) = 282.32 Prob >= chibar2 = 0.0000

Std.Error = Standard error

*= significant at P-value < 0.05

**= significant at P-value < 0.01

***= significant at P-value < 0.001

The mothers with primary education experience 0.88 ($exp(-0.126)$) times less infant deaths compared to mothers with no education at all. We can also say infant deaths for mothers with primary education decreased by 12% ($0.88 - 1 = -0.12$). Similarly, the mothers with secondary education experience 0.67 ($exp(-0.402)$) times less infant deaths compared to mothers with no education at all and infant deaths decrease by 33% ($0.67 - 1 = -0.33$). The mothers with higher education experience 0.47 ($exp(-0.745)$) times less infant deaths compared to mothers with no education at all and infant deaths for such mothers decrease by 53% ($0.47 - 1 = -0.53$). In case of fathers, the infant deaths experienced will be 0.85 ($exp(-0.165)$) less for fathers with primary education, 0.87 ($exp(-0.138)$) less for fathers with secondary education, 0.67 ($exp(-0.401)$) less for fathers with higher education as compared to fathers with no education. It follows that the log count of newborn deaths will fall by 0.71 ($exp\ (-0.344)$) for every unit rise in the mother's age at the time of her first birth. The log count of infant deaths for not working mothers is 0.77 ($exp\ (-0.266)$) times the infant deaths for working mothers. The predicted log count of infant fatalities is

0.92 ($exp(-0.081)$) times fewer for mothers who do use contraception than those who don't. The predicted log count of infant fatalities is 0.96 ($exp(-0.046)$) and 0.82 ($exp(-0.204)$) times lesser for mothers who belong to middle-income and high-income households, respectively.

The expected infant mortality log count for women who reside in rural locations was 1.06 ($exp\ (0.06)$) times the anticipated infant death log count for mothers who reside in urban areas. The log count of infant deaths in Sindh is 0.65 ($exp\ (-0.432)$) times less as compared to Punjab. The coefficient associated with Sindh is −0.32, indicating that the average number of infant death per woman in Sindh is lower than the average infant death per woman in Punjab after controlling for all the other predictor variables. The log count of infant deaths in KPK is 0.79 ($exp\ (-0.231)$), Baluchistan is 0.76 ($exp\ (-0.279)$), Gilgit Baltistan is 0.93 ($exp\ (-0.076)$), ICT is 0.67 ($exp\ (-0.394)$), AJK is 0.79 ($exp\ (-0.236)$) and FATA is 0.34 ($exp\ (-1.067)$) times less as compared to Punjab (Table 5).

The comparison between a multilevel negative binomial model and a single-level Poisson model results in a p-value less than 0.005 and chi-square value equal to 282.32, which is statistically significant.

Discussion

Using a multilevel log-linear model, this work used a 2018 PDHS dataset to determine the risk factors linked to infant mortality in Pakistan. Policymakers can recognize advancement in people's lives and evaluate progress towards attaining the MDGs by determining these risk variables. The findings of this study provided insight into Pakistan's infant mortality factors. This finding makes it clear that women in urban areas experience lower infant mortality rates than mothers in rural areas. This outcome is consistent with the earlier studies (Muriithi & Muriithi, 2015). Even in rural India, infant mortality is a serious problem, and whether a newborn lives in a rural or urban area may impact their health. Infants could become highly unwell in remote locations due to a lack of basic infrastructure, including electricity, running water, toilets, and medical facilities. The issue is caused by inadequate healthcare infrastructure in rural areas. As a result, women cannot obtain better health care in urban regions since rural areas lack the same degree of infrastructure as metropolitan ones. To narrow the infant mortality rate disparity between urban and rural areas, targeted interventions should concentrate on enhancing healthcare infrastructure in rural areas.

According to the study, infant mortality rates decline as a mother's education level rises. Infant mortality rates are highest among mothers with no schooling at all, whereas they are lowest among mothers with a greater level of education. The husband's education level exhibits an inverse link with infant death rates, similar to the mother's. Despite the general trend of lowering rates as education level increases, the surge in infant death rates seen among fathers with secondary education may be caused by several factors. Here are several possible causes that might assist in explaining the observed result, even if further research is necessary to determine the precise variables underlying this phenomenon. Secondary-educated fathers may belong to a socioeconomic group with some knowledge but find it challenging to get the resources and services they need for the best possible maternal and newborn health. Due to financial limitations or differences in healthcare delivery systems, they may experience obstacles when acquiring adequate medical care, nutrition, or other vital services. Determining whether variations in the birth weight density associated with socioeconomic level result in changes in infant mortality (Gage, Fang, O'Neill, & DiRienzo, 2013) use CDDmlr and a statistical decision theory approach to separate direct from indirect effects of maternal education on infant mortality. They discovered that the indirect effects increase mortality with higher education.

Compared to couples with lower education levels, babies born to couples with higher education levels have lower infant death rates. This finding highlights the necessity of considering both parents' educational backgrounds when tackling infant mortality and that education considerably improves maternal awareness, healthcare-seeking behavior, and baby care practices.

Infant mortality rates vary across different socio-economic strata. The data indicate that infants from the poorest strata experience higher mortality rates than those from the middle and wealthiest strata. This association can be attributed to disparities in access to healthcare, nutrition, and living conditions. Addressing socio-economic inequalities should be a priority to reduce infant mortality.

Compared to working mothers, mothers not working had slightly higher infant mortality rates. Understanding the underlying causes of this pattern requires further research because the relationship between maternal employment and infant mortality is complex. Socioeconomic position and maternal employment status are linked. Mothers who are not working may have lower household earnings and less access to resources, such as food, healthcare, and other essentials for the well-being of their children. Higher infant mortality rates among mothers who are not working can be attributed to socioeconomic differences.

Maternal health and well-being may be impacted by maternal employment. Mothers who are not working may face increased stress levels, social isolation, or a lack of social support, which can harm their general health and capacity to give their children the best care possible. Poor maternal health can indirectly impact infant health outcomes and the risk of infant mortality. Mothers who work may have access to healthcare benefits or maternity support programs provided by their employers, which can help to improve maternal and newborn health outcomes. Our finding is supported by (Akinyemi, Solanke, & Odimegwu, 2018) study that revealed that the risk of infant and child mortality was higher amongst unemployed women.

Mothers who have taken contraceptives have slightly lower infant mortality rates than mothers who have not. This finding raises the possibility that access to and use of contraceptive methods may improve pregnancy outcomes and lower infant mortality. Utilizing contraception can prevent unwanted pregnancies, which are more frequently linked to poor prenatal care, maternal stress, and lifestyle choices that may harm the unborn child's health. Mothers who take contraceptives are more likely to have adequate time for preconception care, engage in healthy behaviors during pregnancy, and seek appropriate prenatal care because they are less likely to become pregnant unintentionally, which lowers the risk of infant death.

Contraceptives ensure that women have enough time to recover from prior pregnancies, refill their nutritional stores, and improve their general health before becoming pregnant again by enabling parents to plan and space their pregnancies. Contraceptive users are more likely to have access to healthcare providers, family planning advice, and details on good prenatal care. Lower infant mortality rates result from this access to reproductive health services, which improves mother and child health knowledge, encourages healthy habits, and increases the likelihood that women will seek the proper treatment during pregnancy and childbirth. (Bradshaw, et al., 2023) discovered that greater access to contraception decreased fertility, higher infant mortality, and household size increased fertility. According to their simulations, ensuring adequate housing to reduce the household size and expanding access to contraception will have the most impacts on lowering infant mortality and reducing global fertility.

Infant deaths are more likely to occur when the mother is 18 to 20 years at birth than when she becomes older, according to her age at delivery. When multiple births occur at a young age, both the mother's and the child's health may be at risk. This conclusion is backed up by a study conducted in low- and middle-income nations, Pakistan, Kenya, Nepal, and Kenya (Shukla VV, 2020), which indicated that infant mortality risks are raised for both young and elderly mothers, with the probability of risk for teenage mothers being bigger.

Conclusion

The risk factors associated with the death of infants in Pakistan were identified by this study using multilevel count models. As a result, we identified multiple sources of variability in the dataset. 12,708 mothers were enrolled in this study. NB was shown to be the sparsest way to quantify Pakistan's risk factors for infant mortality by multilevel Poisson model analysis. The findings show regional differences in the number of infant deaths per woman.

Infant mortality risk variables were the place of residence, number of family members in the household, mother’s age at her first delivery, wealth index, education and employment status of both father and mother, contraceptive use, and improved water supply. The results of the study point to the necessity for further efforts to broaden educational programs that inform parents, particularly mothers, about the benefits of contraception and intervals to decrease infant mortality per woman.

References

Adulo, L. A., & Zewudie, S. G. (2021). Identifying factors determining the survival time of under-five year children in rural parts of Ethiopia. *Research Square*. Retrieved from https://doi.org/10.21203/rs.3.rs-634955/v1

Agbadi, P., Agbaglo, E., Tetteh, J. K., Adu, C., Ameyaw, E. K., & Nutor, J. J. (2021). Trends in under-five mortality rate disaggregated across five inequality dimensions in Ghana between 1993 and 2014. *Public Health, 196*, 95-100. doi:https://doi.org/10.1016/j.puhe.2021.04.024

Agha, S. (2000). The determinants of infant mortality in Pakistan. *Social science & medicine, 51*(2), 199-208. doi:https://doi.org/10.1016/S0277-9536(99)00460-8

Akinyemi, J., Solanke, B., & Odimegwu, C. (2018). Maternal Employment and Child Survival During the Era of Sustainable Development Goals: Insights from Proportional Hazards Modelling of Nigeria Birth History Data. *Ann Glob Health, 84*(1), 15–30. doi:10.29024/aogh.11

Alves, D., & Belluzzo, W. (2004). Infant mortality and child health in Brazil. *Econ Hum Bio, 2*(3). doi:10.1016/j.ehb.2004.10.004

Bradshaw, C., Perry, C., Judge, M., Saraswati, C., Heyworth, J., & Le Souëf, P. (2023). Lower infant mortality, higher household size, and more access to contraception reduce fertility in low- and middle-income nations. *PLoS ONE, 18*(2), 1-16. doi:10.1371/journal.pone.0280260

Ely, D. M., Driscoll, A. K., & Mathews, T. (2017). *Infant Mortality Rates in Rural and Urban Areas in the United States, 2014.* US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. Retrieved from https://www.cdc.gov/nchs/products/databriefs/db285.htm

Emamgholipour Sefiddashti, S., Nakhae, M., Kazemi Karyani, A., & Ghazanfari, S. (2015). Decomposition socioeconomic inequality in infant mortality in EMRO countries. *International Journal of Pediatrics, 3*(4), 749-756. doi:10.22038/IJP.2015.4429

Gage, T. B., Fang, F., O'Neill, E., & DiRienzo, G. (2013). Maternal Education, Birth Weight, and Infant Mortality in the United States. *PubMed, 50*(2), 615-35. doi:10.1007/s13524-012-0148-2

Hoadley, W. E. (1995). Human Development Report 1994. . *Journal of Asian Studies, 54*(2), 523-525.

Irfan, M. (1986). *Mortality and health issues: mortality trends and patterns in Pakistan.* Asian Population Studies Series. Retrieved from https://mpra.ub.uni-muenchen.de/38619/

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: With Applications in R.* Springer New York, NY. doi:https://doi.org/10.1007/978-1-4614-7138-7

Junejo, P. A. (2019). *Pakistan Demographic and Health Survey 2017-18.* Islamabad: National Institute of Population Studies. Retrieved from https://dhsprogram.com/pubs/pdf/SR257/SR257.pdf

Kanmiki, E. W., Bawah, A. A., Agorinya, I., Achana, S., F., Awoonor-Williams, J. K., . . . Akazili, J. (2014). Socio-economic and demographic determinants of under-five mortality in rural northern Ghana. *BMC International Health and Human Rights, 14*, 1-10.

Kim, D., & Saada, A. (2013). The social determinants of infant mortality and birth outcomes in Western developed nations: a cross-country systematic review. *International journal of environmental research and public health, 10*(6), 2296-2335. doi:10.3390/ijerph10062296

Leckie, G., Browne, W., Goldstein, H., Merlo, J., & Austin, P. (2022). Variance partitioning in multilevel models for count data. *25*(6), 787-801. doi:10.1037/met0000265

Muriithi, D. M., & Muriithi, D. K. (2015). Determination of Infant and Child Mortality in Kenya Using Cox-Proportional Hazard Model. *American Journal of Theoretical and Applied Statistics, 4*(5), 404-413. doi:10.11648/j.ajtas.20150405.21

Patel, K., Rai, R., & Rai, A. (2021). Determinants of infant mortality in Pakistan: evidence from Pakistan Demographic and Health Survey 2017–18. *J Public Health (Berl.), 29*, 693–701. doi:https://doi.org/10.1007/s10389-019-01175-0

Puri, J., Gaye, A., Kurukulasuriya, S., & Scott, T. (2007). *Measuring Human Development: A Primer.* New York: Human Development Report Office, UNDP (HDRO). Retrieved from https://hdr.undp.org/content/measuring-human-development-primer

Ruiz, J. I., Nuhu, K., McDaniel, J. T., Popoff, F., Izcovich, A., & Criniti, J. M. (2015). Inequality as a Powerful Predictor of Infant and Maternal Mortality around the World. *PLoS ONE, 10*(10). doi:https://doi.org/10.1371/journal.pone.0140796

Sartorius, B. K., & Sartorius, K. (2014). Global infant mortality trends and attributable determinants–an ecological study using data from 192 countries for the period 1990–2011. *Population Health Metrics, 12*(1), 1-15. doi:https://doi.org/10.1186/s12963-014-0029-6

Sellers, K. F., & Shmueli, G. (2013). Data dispersion: now you see it… now you don't. *Communications in Statistics-Theory and Methods, 42*(17), 3134-3147.

doi:https://doi.org/10.1080/03610926.2011.621575

Shukla VV, E. B.-T. (2020). Predictive modeling for perinatal mortality in resource-limited settings. *JAMA Network Open, 3*(11). doi:10.1001/jamanetworkopen.2020.26750

StataCorp. (2019). *Stata statistical software: Release 16 [Computer software]. College Station, TX: StataCorp LLC.* . Retrieved from http://www.stata.com.

Udine, M. L., Evans, F., Burns, K. M., Pearson, G. D., & Kaltman, J. R. (2021). Geographical variation in infant mortality due to congenital heart disease in the USA: a population-based cohort study. *The Lancet Child & Adolescent Health, 5*(7), 483-490. doi:10.1016/S2352-4642(21)00105-X

World Bank. (2019). Estimates developed by the UN inter-agency Group for Child Mortality Estimation (UNICEF, WHO, World).

World Bank. (2021). *World Development Indicators.* The World Bank Group. Retrieved from https://data.worldbank.org/indicator/SP.POP.TOTL?locations=PK-IN-CN-ID-US

World Health Organization. (2019, July 27). *World Health Organization (2017) Infant mortality: Situation and trends*. Retrieved from https://www.who.int/gho/child_health/mortality/

Zegeye, B., Shibre, G., Haidar, J., & Lemma, G. (2021). ocioeconomic, urban-rural and sex-based inequality in infant mortality rate: evidence from 2013 Yemen demographic and health survey. *Archives of Public Health, 79*, 64. doi:https://doi.org/10.1186/s13690-021-00589-1