ISSN 2564-0127 (Print)
ISSN 2564-0135 (Online)

Canadian Journal of Medicine 3(2021) 153-161

Canadian Journal of Medicine

WWW.CIKD.CA

journal homepage: https://www.cjm.cikd.ca

Original Article

Accuracy Measure of Separate and Joint Modelling for a Correlated Binary Outcome: The Case Study of Mother Education and Immunization in Bangladesh

Md Asadullah[1*], Nahid Hosen[2], Mamunar Rashid[3], Priyanka Basu[4], Farjana Boby Mukta[5], Emon Ahmed[6]

[1, 2, 3, 4]Department of Statistics, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj-8100, Dhaka, Bangladesh
[5]Shahjalal University of Science and Technology, Sylhet, Bangladesh
[6]ASA University Bangladesh, Shyamoli, Dhaka, Bangladesh

Keywords:
Joint model, Separate model, Immunization, Underestimates, Correlated data

Received
03 June 2021
Received in revised form
14 July 2021
Accepted
24 July 2021

Correspondence:
asadullahstat@gmail.com

ABSTRACT

Joint modelling is a statistical approach that is used to analyze correlated data when two or more outcome variables are correlated. By joint modeling, we refer to the simultaneous analysis of two or more different response variables from the same individual. But in a separate model, it is unable to measure the effect of covariate simultaneously. This article focuses on separate and joint modelling for correlated discrete data, including logistic regression models for binary outcomes. Since most of the women are illiterate in Bangladesh and most of the people are living in urban areas, as a result, most of the women are not aware of immunization. But an educated mother is always aware of her child's health which is dependent on immunization. Therefore, mother education and immunization are interdependent. We jointly address the effect of maternal education and immunization. Joint modeling of these two outcomes is appropriate because mother education helps raise awareness of the child's health and immunization is the prevention of various diseases for the child's health. We also identified factors influencing maternal education and immunization among women in Bangladesh. By jointly modelling we found the correlation between maternal education and immunization and the most important contributing factor. The joint model removes a less significant impact of covariates as opposed to separate models. These findings further suggested that the simultaneous impact of correlated outcomes can be adequately addressed between different responses, which is overestimated or underestimated when examined separately.

Bangladesh is involved in high fertility, malnutrition, and communicable diseases, particularly environmental sanitation and personal hygiene. Every year, many more are crippled, blinded, or otherwise disabled from six major diseases that are preventable through immunization [1]. Vaccines and their tools are readily available for all six diseases, relatively inexpensive and effective in saving lives. These six vaccine-preventable diseases are measles, pertussis (whooping cough), tetanus, polio, tuberculosis, and diphtheria [2]. The Bangladesh government has different immunization programs like DPT, BCG, and MR. The importance of the immunization program can be understood better if we remember that about 3.7 million babies are born in Bangladesh every year [3]. About 2400 young children die every day, accounting for 100 deaths every hour. Four percent of death among infants under age1 occur in Bangladesh [4]. Of the 107.9 million population of Bangladesh, about 18 million are children below five years. The risk of death is higher in these ages [5]. Here, tetanus alone accounts for 223000 child deaths and measles for 20000 to 40000 deaths every year (Govt. of Bangladesh, 1985s) [6]. Although vaccination is more effective in reducing neonatal mortality, the actual practice of vaccination is still low [7]. Government immunization service statistics for 1984 showed that the national coverage for BCG was 1.5%, followed by D.P.T. (3 doses; 14%), polio (3 doses;1.1% and measles (0.9%), (Government of Bangladesh, 1985 s) [8]. Nowadays acceptability of immunization is higher, but its coverage remains low [9]. According to the report of WHO and UNICEF, the national coverage for BCG was reported to be 98% in 2014 Government immunization service, followed by D.P.T. (3 doses; 94%), polio (3 doses; 94%), and measles (4.35%). It is clear that the percentage of vaccination increased in 2014 over 1984 [10]. But BDHS data in 2004 showed that 58.5% of respondents were completely immunized, and 63.8% of respondents were completely immunized, as reported by BDHS data in 2014 [11].

Mother education plays an important role in determining the health conditions for children. As far as education is concerned, policymakers need to focus on the mother's education because mothers are more concerned with their child's health. [12]. Therefore, educating mothers about immunizing their children against these serious diseases will improve national health issues. [13]. It helps the children to be mentally healthy and active. From BDHS 2014, 85.54% of women have no education, and only 15.42% are educated. It is also observed that only 37.77% of women complete the vaccination, and the remaining percentage of women do not complete the vaccination [14]. The main reason behind this is the lack of public awareness about the importance of immunization and the lack of mother education [15]. Since most of the women in Bangladesh have no education, thus, most of the women complete all immunization. So, the main goal of mother education is for better nutrition and health of children [16]. Therefore, mother education is correlated with immunization. Mother education is associated with reduced risk of incomplete immunization and reduced risk of other child health outcomes such as malaria, malnutrition, and mortality [17]. Childhood immunization is a preventive measure that is administered after birth. It was capable of producing antibodies against specific diseases [18].

Most studies have ignored the interdependence between these two related outcomes and considered them separate events, leading to misleading results due to the obvious relationship between them. Therefore, both mother education and immunization are accounted for the overlapping effects from one event to the other. Our study fits two sets of models, including a separate model and a joint model. Each of these hierarchical logistic regressions addresses intra-class correlation. The motivation for a hierarchical logistic regression model is to address the random effect of the 600 regions in Bangladesh. The joint modelling is capable of addressing the correlation of the two responses with the same set of covariates. We first provided a review of joint modelling and focused on two binary outcomes, mother education, and immunization. We shared 2014 Bangladesh Demographic and Health Information as a tool to address mother education and immunization. Finally, we provided a clear evaluation of the advantage of joint modelling over separate modelling and compared the coefficients in both model estimates.

Materials and Methods

Joint modelling was used for analyzing correlated data. When the outcome processes are correlated, joint modelling has been empirically demonstrated to lead to improvement of efficiency and prediction, as well as reduced bias. A less biased estimate leads to a more accurate estimate. In this case, we say that the estimate based on the joint model is less biased than the separate model. Previous works on the joint modelling of correlated outcomes have largely focused on continuous outcomes or a combination of continuous and discrete outcomes. Dunson proposed a Bayesian approach for joint modelling binary and continuous subunit-specific outcomes and illustrated this approach with a developmental toxicity data example. Liang et al. proposed the joint modelling and analysis of longitudinal data with informative observation times via latent variables. Song et al. [15] investigated joint models for a time-to-event and a longitudinal response through a likelihood-based approach that only assumes that the random effects have a smooth density. In light of other studies, joint modelling is not a novel approach in statistics; however, reporting joint modelling of two correlated binary outcomes is somewhat rare. A study by Ghebremichael [6] identified risk factors for two binary outcomes of HSV-2 and HIV-1 infections using a joint response model, where accommodated the interdependence between the two infections. Guo et al. proposed separate and joint modelling of longitudinal and event time data for correlated variables. Additionally, Fang et al. recommended the joint and separate modelling for a correlated binary outcome, i.e., HIV knowledge and contraceptive use. We followed Fang in applying a similar method. We accounted for the possible intra-class correlation commonly seen in hierarchical data. In particular, we modelled two correlated responses including education and immunization. Each response was taken from the same subject, thus causing correlation. We accounted for the hierarchical structure because subjects are nested within enumeration areas (EAs) and correlated.

Data

We used data obtained from the 2014 Bangladesh Demographic and Health Survey (BDHS). It is the sixth national demographic and health survey on maternal and child health in Bangladesh. Bangladesh has seven administrative divisions, including Barisal, Chittagong, Dhaka, Khulna, Rajshahi, Rangpur, and Sylhet. The survey data delineated a national representation of the entire population living in non-institutional dwelling units in Bangladesh. The primary sampling unit consisted of enumeration areas (EAs). Each EA had an average of about 120 households, and there were about 600 EAs. These EAs were selected with the probability proportional to size. There were 207 clusters in urban areas and 393 in rural areas. The survey data consisted of 7,530 child information on a hierarchical structure. The two binary responses included mother education or immunization. The covariates related to demographic status included current mother age, mother education, religion, region, mother mobility, mother affiliation, economic status. The covariates also included the media exposure, such as having a radio, TV, and reading magazine. Table 1 provides the descriptive statistics of these covariates. Notably, some covariates such as education and economic status were measured on an ordinal scale. Furthermore, there were two educational categories, including lack of education and education. There was an index number to measure family income for each household, ranging from poor, middle, and rich. Notably, religion played an important role in the respondents' beliefs, where 90% of the sample were being found to be Muslim. Geographically, approximately 69.31% of respondents lived in rural areas compared to 30.69% who lived in urban areas. Specifically, 84.31% of women are educated compared to 15.69% of women who had never received any education. Regarding living conditions, approximately 59.53% of households had a television at home, but only 7.39% and 33.08 of them had a radio and reading magazines, respectively. Based on the economic status, 40.14% of the population fell into the poor category, followed by the middle class (21.05%) and rich (38.81%). Mother mobility means going to a health center alone or with younger children. Herein approximately 71.47% of women go to health centers with them. Mother affiliation is one of the most important covariates, where 28.73% of women belonged to different development organizations.

Table 1

Summary of Statistics

Variable	Code	Level	Frequency	Percent
Education	0	No education	1161	15.69
	1	Education	6238	84.31
Age	0	15-19	1071	14.47
	1	20-29	4573	61.80
	2	30-49	1755	23.72
Immunization	0	No	4191	58.75
	1	Yes	3208	41.25
Economic status	0	Poor	3041	40.14
	1	Middle	1435	21.05
	2	Rich	2923	38.81
Mother mobility	0	No	2070	28.53
	1	Yes	5329	71.47
Religion	0	Non-Islam	598	07.58
	1	Islam	6801	92.42
Region	0	Urban	2345	30.69
	1	Rural	5054	69.31
Media exposure	0	No	4519	62.11
	1	Yes	2880	37.89
Mother affiliation	0	No	5119	71.27
	1	Yes	2280	28.73
Division	1	Barisal	863	12.33
	2	Chittagong	1421	18.87
	3	Dhaka	1301	18.16
	4	Khulna	810	10.76
	5	Rajshahi	901	11.96
	6	Rangpur	910	12.08
	7	Sylhet	1193	15.84

Statistical Analysis

Let Y_{ij} be a binary random variable and take the values of one or zero that denotes the i^{th} outcome (i = 1, 2) of the i^{th} (j = 1, 2, . . ., n = 7,530) subject, with i = 1 for educated women and i = 2 for knowledge of immunization. The two binary outcomes give rise to a bivariate binary response vector $Y = (Y_{1j}\ Y_{2j})$ with a corresponding data matrix of covariates $(X_{1j},\ X_{2j})$ where $X_{1j} = (1,\ X_{ij1}, \dots\dots,\ X_{ijp})$ and corresponding vectors of regression coefficients were shown (β_1, β_2). Z_1 is the matrix of random effects with corresponding coefficients γ. Covariates included mother age (age), economic status, religion, region, mother mobility, media exposure. Each woman has two outcomes (education or immunization), and education and immunization knowledge outcomes are correlated. We fitted joint hierarchical logistic regression models with shared random effects, accounting for the correlation of clusters and divisions. The shared random intercept accounted for the correlation between the two responses of the same individual. This random intercept captures the unobserved factors specific to each individual that may influence education and immunization knowledge. Regarding the additional random intercept, we presented the joint model as follow:

$$logit[P_i | X_{ij}, \beta_i, u_j] = I_{1j}(\beta_0 + \beta_1 X_{1jk} + \cdots \beta_p X_{pjk} + \gamma_{0_{clu}} + \gamma_{1_{clu}} Z_{jk} + \gamma_{0_{div}} + \gamma_{1_{div}} Z_{j'k'})$$
$$+ I_{2j}(\beta_0 + \beta_1 X_{1jk} + \cdots \beta_p X_{pjk} + \gamma_{0_{clu}} + \gamma_{1_{clu}} Z_{jk} + \gamma_{0_{div}} + \gamma_{1_{div}} Z_{j'k'})$$

where I_{ij} is the indicator function for the outcome, I, $U_j$are the random intercepts assumed to be normally distributed with mean zero and variance of σ_U^2 P_i represents P_{Edu}, (when i = 1) and P_{Immu} (when i = 2). These estimators are consistent and asymptotically normally distributed. We assumed that there were unobserved latent factors involving in the provision of random effects between immunization and education by the same subject. Thus, the two outcomes are correlated within each subject. We achieved this through a shared random intercept; this random intercept is intended to capture the unobserved factor within each subject and may influence the responses. Conditioned on the shared random intercept, the joint responses of a woman are assumed to be independent. By assuming conditional independence, the probability function of shared responses can be obtained. Following Ghebremichael, we used a maximum likelihood approach to estimate the regression and covariance parameters. As the integrals do not have closed-form, we applied an approximation based on Gaussian quadrature, as found to be consistent and asymptotically normally distributed.

Results

In this study, we fitted two sets of models, including separate and joint model. In the separate models, we used hierarchical logistic regression to validate the presence of indeed variation within an individual (cluster [EA], and division). We fitted a three-level hierarchical logistic regression with random intercepts and random slopes to address these issues of intraclass correlation. If we performed a Hosmer–Lemeshow test, within each level and our test results suggested the need for a multilevel logistic model. We found that the division and the cluster variance are significantly different from zero because of a p-value less than .05. We presented results in Table 2 with the odds ratio of the parameter estimates. We found that age, media exposure, and economic status are statistically significant in influencing immunization. More specifically, we observed that the women in the age group 2 (30-49) are more conscious about the immunization of their baby than the age group 0 (15-19) and age group 1 (20-29). Furthermore, media exposure and the region have a great impact on the completeness of immunization.

Table 2

Comparison of Joint Model with Separate Models

Joint model			Separate model			
			Immunization		Education	
variable	Odds ratio	p-value	Odds ratio	p-value	Odds ratio	p-value
Intercept	0.55	< .001	0.44	< .001	3.17	< .001
Age1	0.90	.034	1.22	< .001	0.52	< .001
Age2	0.71	< .001	1.15	.030	0.14	< .001
Religion1	1.01	.749	1.01	.646	1.16	.955
Region1	1.05	.163	1.00	.934	1.59	< .001
Economic Status1	1.11	.025	1.08	.323	2.05	< .001
Economic Status2	1.11	< .001	1.33	< .001	3.19	< .001
Mother.mobility1	0.96	.327	1.25	< .001	1.03	.792
Mother.affliation1	1.04	.290	0.86	< .001	1.13	.771
Media.exposure1	1.14	< .001	1.19	.044	2.83	< .001

Women use different media such as radio, television, and magazine are more conscious about immunization. Here, we observed that the richest women are more inclined to complete all immunization. Regarding mother affiliation, we found that belonging to different organizations increased the likelihood of awareness about immunization. We did not find religion a significant factor for immunization, even though media exposure, economic status, region, and age can be related to immunization. In the joint model, the women's living status did not affect them, and it did not play an important role in the separate model. This may result from the fact that if we fit the joint model, education has an impact that applied to the wealthy and the poor family together; therefore, the economic status did not play an important role in immunization. But fitting the separate model showed the important role of economics. Most people in Bangladesh believe in Islam. There was no effect for immunization in both joint and separate models. Based on the data presented herein, media exposure has played an important role in joint and separate models (Table 2). We found similar driving factors for immunization as those obtained for education. More exposure to the media is likely to lead to greater immunization.

Education played a paramount role in the likelihood of immunization. Urban women were more likely to know about immunization than rural women. In addition, women who were belonging to different organizations were more likely to be conscious about immunization. Unlike education, immunization was positively driven by economic status; wealthy families were more likely to know about immunization than poor families. These findings focus on the crucial importance of economic status and media exposure on immunization. Furthermore, religion did not play an important part in immunization knowledge. The results regarding education from a separate hierarchical logistic regression model revealed that estimates obtained from age, media exposure, economics status, and region were significant with the same incremental impact. Estimates obtained for religion, mother mobility, and mother affiliation were found to be insignificant, as revealed by the separate model. In the joint model, age, economic status, and media exposure are significant. Additionally, we found that education had a strong divisional random effect. As for immunization knowledge, we found the same incremental significance for all covariates, except mother affiliation and religion, as revealed by separate models. In the separate model, women who were less educated were likely to have less knowledge about immunization. However, the estimate from the joint model showed the opposite story; educated women likely to have more knowledge about immunization. Estimates from the joint model made intuitive sense because the majority of women population in Bangladesh live in rural areas and have not been highly educated, resulting in a lack of knowledge about immunization. In rural areas, most mothers do not complete all immunization. In urban areas, mothers were more conscious of complete immunization. Moreover, women from less wealthy families were less likely to have TV, radio, and magazines, limiting their exposure to immunization media.

Ignoring the random effects of division and EA would underestimate the effect of education. For example, if we had modelled education without any adjustment for divisional and EA random effects, it can be concluded that education would likely increase education by 0.078. However, using a model that considers changes in different geographical divisions and sampling units (EA),, the magnitude of education on immunization increased to 0.146, which means an increase in probability.. More importantly, after considering the correlation between immunization and education, the importance of education decreased to 0.027. This means that the immunization prevention program has a spillover effect on education. Thus, what seems to be an impact from education can be partially contributed to one's exposure to immunization. As for modelling of the immunization, joint models revealed smaller values in magnitude than separate models. These findings substantiated the need for jointly modelling two correlated outcomes.

Discussion

Bangladesh is a developing country. In Bangladesh, 60% of people live in rural areas and 40% of people in urban areas. Since the majority of women live in rural areas, the percentage of educated women is low. For this reason, the women who live in rural areas are not conscious about immunization and do not complete all immunization. However, women who live in urban areas are more educated than women in rural areas. They were more conscious about immunization and completed all immunization. Therefore, more educated women were more inclined to complete all immunizations. For our purpose, we fitted a series of statistical models using the BDHS2014 data.

We found that media exposure was positively associated with education but negatively associated with immunization. In addition, economic status was positively associated with education and immunization. Women with media exposure had more knowledge about immunization. Education was found to increase the likelihood of the knowledge of immunization. We found that exposure was significantly involved in influencing immunization knowledge, but not education. Moreover, economic status had a significant impact on both education and immunization. These two findings showed that education and immunization were relatively correlated with each other. On the other hand, mother affiliation had a significant influence on immunization, but not education. Therefore, media exposure, mother affiliation, and economic status are often correlated with education and knowledge about immunization. The prevalence rates of education and knowledge about immunization in Bangladesh were 93%

and 62%, respectively, indicating a likelihood of overlap between these two events. Thus, we used a joint response model that accommodated the interdependence between education and knowledge about immunization.

Conclusion

Based on the findings presented herein, joint models deflate parameters that are otherwise overestimated in separate models. We found that the strength of the relationship was higher in absolute values by using separate models, causing false interpretations of the attributes in the magnitude. To correctly translate the importance of certain covariates, the correlated outcomes caused by common covariates should be taken into account. In this case, after accounting for the correlation between education and knowledge about immunization in simultaneous estimation, we found the strengths of the relationship to be smaller than those in separate models.

The implication regarding public health is straightforward when we have two social programs that aim to have a similar impact, a positive spillover effect from one to the other exists. However, by treating them independently, we may put too much emphasis on the covariates. For example, treating the two separate events, we may recommend increasing the budget for girls to enter school.

Alternatively, if we included the interdependency of the programs, we will realize that the importance of media exposure on education is compensated by the knowledge about immunization. Many developing countries have substantial geographic variations in immunization, although the factors driving these variations are not always understood. Many studies showed that variations in immunization and education remained after accounting for individual and household factors. We found that education and immunization knowledge varied within enumeration areas and across divisions, resulting in significant cluster effects. Although we did not analyze the specific driving factors contributing to geographical variations, our joint model showed that they were the same contextual factors driving both education and immunization knowledge.

Acknowledgments

Specially thanks to Md. Razu Amhed for suggestions and concerns in our research.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding Acknowledgements

Not applicable.

References

[1] Azim T, Khan SI, Haseen F, Huq NL, Henning L, Pervez MM, Chowdhury ME, Sarafian I. HIV and AIDS in Bangladesh. J Health Population Nutrition. 2008 Sep;26(3):311.

[2] Niport M. and Macro International. 2009. Bangladesh Demographic and Health Survey 2007.

[3] Dunson DB, Chen Z, Harry J. A Bayesian approach for joint modeling of cluster size and subunit-specific outcomes. Biometrics. 2003 Sep;59(3):521-30.

[4] Fang D, Sun R, Wilson JR. Joint modeling of correlated binary outcomes: The case of contraceptive use and HIV knowledge in Bangladesh. PloS one. 2018 Jan 19;13(1):e0190917.

[5] Guo X, Carlin BP. Separate and joint modeling of longitudinal and event time data using standard computer packages. Am Statistician. 2004 Feb 1;58(1):16-24.

[6] Ghebremichael M. Joint modeling of correlated binary outcomes: HIV-1 and HSV-2 co-infection. J Applied Statistics. 2015;42(10):2180-91.

[7] Gueorguieva RV, Agresti A. A correlated probit model for joint modeling of clustered binary and continuous responses. J Am Statistical Association. 2001;96(455):1102-12.

[8] Gardiner JC. Joint modeling of mixed outcomes in health services research. SAS Global Forum Paper 2013 (pp. 435-2013).

[9] Khan HA. A hierarchical model of contraceptive use in urban and rural Bangladesh. Contraception. 1997;55(2):91-6.

[10] Hubbard RA, Inoue LY, Diehr P. Joint modeling of self-rated health and changes in physical functioning. J Am Statistical Association. 2009;104(487):912-28.

[11] Hu FB, Goldberg J, Hedeker D, Flay BR, Pentz MA. Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes. Am J Epidemiol. 1998;147(7):694-703.

[12] Irimata KM, Wilson JR. Monte-Carlo Simulation in Modeling for Hierarchical Generalized Linear Mixed Models. InMonte-Carlo Simulation-Based Statistical Modeling 2017 (pp. 255-283). Springer, Singapore.

[13] Khan HR, Shaw E. Multilevel logistic regression analysis applied to binary contraceptive prevalence data. J Data Sci. 2011;9:93-110.

[14] Liang Y, Lu W, Ying Z. Joint modeling and analysis of longitudinal data with informative observation times. Biometrics. 2009;65(2):377-84.

[15] Song X, Davidian M, Tsiatis AA. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. Biometrics. 2002;58(4):742-53.

[16] Stephenson R, Baschieri A, Clements S, Hennink M, Madise N. Contextual influences on modern contraceptive use in sub-Saharan Africa. American J Public Health. 2007;97(7):1233-40.

[17] Wong GY, Mason WM. The hierarchical logistic regression model for multilevel analysis. J AmStatistical Association. 1985;80(391):513-24.

[18] Zhu M. Analyzing multilevel models with the GLIMMIX procedure. SAS Institute Inc. SAS026-2014. 2014.