Dealing With Sparse Data Bias in Medical Sciences: Comprehensive Review of Methods and Applications

Mohammad Hossein Panahi[1], Kazem Mohammad[2], Razieh Bidhendi Yarandi[3], Fahimeh Ramezani Tehrani[4]

[1] Department of Epidemiology, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[2] Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

[3] Department of Biostatistics, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran

[4] Reproductive Endocrinology Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Received: 10 May 2020; Accepted: 18 Oct. 2020

Abstract- This study aims to illustrate the problem of (Quasi) Complete Separation in the sparse data pattern occurring medical data. We presented the failure of traditional methods and then provided an overview of popular remedial approaches to reduce bias through vivid examples. Penalized maximum likelihood estimation and Bayesian methods are some remedial tools introduced to reduce bias. Data from the Tehran Thyroid and Pregnancy Study, a two-phase cohort study conducted from September 2013 through February 2016, was applied for illustration. The bias reduction of the estimate showed how sufficient these methods are compared to the traditional method. Extremely large measures of association such as the Risk ratios along with an extraordinarily wide range of confidence interval proved the traditional estimation methods futile in case of sparse data while it is still widely applying and reporting. In this review paper, we introduce some advanced methods such as data augmentation to provide unbiased estimations.

Acta Med Iran 2020;58(11):591-598.

Keywords: Bayesian method; Complete/Quasi-complete separation; Data augmentation; Penalization methods; Sparse data bias

Introduction

Sparse data bias happens in rare-events in medical data, causing distorting the results. Occasionally, we face an error as "the maximum likelihood estimates do not exist" when running the binary regression models. In this case, an extremely large value and an unusually wide range of confidence intervals were obtained for the estimated parameter. It occurs in a situation called "Separation". This phenomenon is defined as a state when some values of the predictor are associated with only one outcome value. For example, if all cases of preeclampsia occurred in women with gestational diabetes mellitus (1,2). Despite the existence of different statistical techniques to avoid this problem (3-8), many recently-published medical research works are subjected to sparse data bias, the very big value of OR as well as extraordinary wide 95%CI because of large SD proved the existence of small samples and sparse data (9-19).

In this paper, we aim at illustrating this phenomenon in practice and providing a comprehensive overview of remedial methods and a catalog of efficient references to make medical researchers familiar with these techniques through inspiring examples. We pointed out these methodological approaches with lucid language and presented them through hypothetical and real data. As an applied example, data from the Tehran Thyroid and Pregnancy Cohort Study, conducted from September 2013 through February 2016, was applied for illustration.

Illustration of the problem: when the model fails and how to identify the complete or Quasi-Complete Separation

Consider the data pattern illustrated in Figure1-Right. A binary logistic regression through SPSS provided a warning: "The log-likelihood value is approaching zero, there may be a complete separation in the data, the maximum likelihood (MLE) estimates do not exist." Results showed a big Odds Ratio (OR) and also the inestimable upper limit of 95% CI for OR due to the large

Corresponding Author: R. Bidhendi Yarandi
Department of Biostatistics, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran
Tel: +98 2122180146, Fax: +98 2122180146 E-mail address: razi_bidhendi@yahoo.com

estimated standard error (Table 1A). Moreover, in a weaker condition called quasi-complete separation, there is not perfect discrimination for all observations (Figure1-Left). After fitting the logistic regression model, a warning signal regarding quasi-complete separation was reported "A quasi-complete separation may exist in the data, the maximum likelihood estimates do not exist" (Table 1B). In practice, Quasi-complete separation is more common than complete separation. To check the existence of an empty cell, which is a sufficient condition for the separation drawing a 2 by 2 table of each predictor with Y is highly recommended.

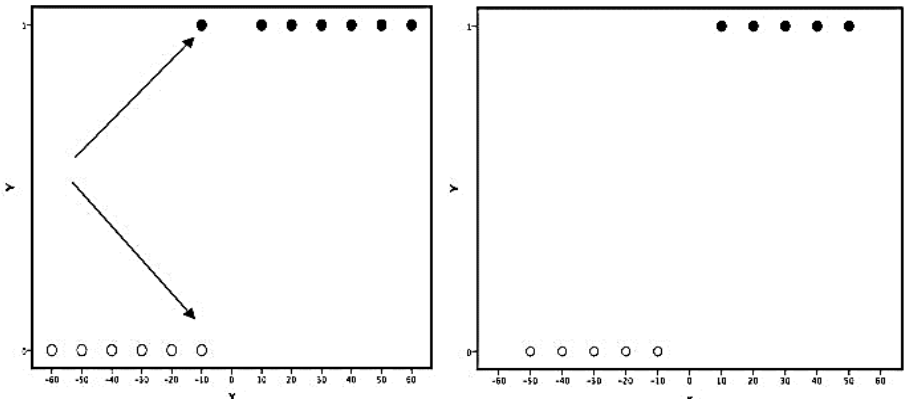

Figure 1. An illustration of Complete (Right) and quasi-Complete (Left) Separation

Table 1 A. SPSS output obtained from fitting a logistic regression to estimate OR (95% CI) in the presence of complete separation

Parameter	B	SE	Hypothesis Test			OR= Exp (B)	95% Wald CI for OR	
			Wald Chi^2	Df	Sig.		Lower	Upper
(Intercept)	-1.469E-9	10072.2351	.000	1	1.000	1.000	.000	.[a]
X	1.913	1007.2235	.000	1	.998	6.772	.000	.[a]

Table1 B. SPSS output obtained from fitting a logistic regression to estimate OR (95% CI) in the presence of quasi-complete separation

Parameter	B	SE	Hypothesis Test			OR= Exp (B)	95% Wald CI for OR	
			Wald Chi^2	Df	Sig.		Lower	Upper
(Intercept)	18.341	9607.2831	.000	1	.998	92299888.36	.000	.[a]
X	1.834	960.7283	.000	1	.998	6.259	.000	.[a]

Dependent Variable: Y, Model: (Intercept), x, [a]Set to system missing due to overflow

Materials and Methods

Remedies for the non-existence of MLEs due to separation

Penalization approaches

Firth penalization

Firth 1993 firstly presents a bias reduction of the MLE method (20). It is like adding 0.5 to each cell of a 2x2 table (21). Technically, it is proportional to the square root of the determinant of the Fisher information matrix for the parameters of interest. Firth Penalization is not always a problem-solving technique due to the largely biased estimate, especially in the case of ratio estimates such as odds ratio or relative risk (7,22). Table 2 illustrates (Quasi) complete separation. Remedy through Firth's penalization was presented there. Firth penalization tried to shrink OR toward zero while its confidence interval is still extraordinarily wide (23-25).

Data augmentation

Data Augmentation (DA) is a method that adds any prior information as a penalty to the actual data to obtain the unbiased MLEs in the form of so-call pseudo information augmentation. It applies inverse-variance weighted averaging for estimation (26). This technique provides an effective remedy to treat the bias caused by data sparseness (27-30). Compatibility of the prior and data is of great importance; DA fails in case of incompatibility causes misleading results (5,31). It means that they should be homogenous enough to average their estimates by information weighting. P-value from the statistic 'standardized' difference between frequentist

estimate and prior estimate tests Compatibility of Prior and data. We provided an illustrating example in the Appendix.

Table 2. Firth penalization approach illustration

A: Quasi-complete separation		**B: Firth penalization for Quasi-complete separation**	
10	9	10.5	9.5
1	12	1.5	12.5
(OR=13.3, 95%CI: 1.4 to 123.9)		(OR=9.2, 95%CI: 1.29 to 66.7)	
C: Complete separation		**D: Firth penalization for Complete separation**	
10	9	10.5	9.5
0	12	0.5	12.5
(OR=∞, 95%CI: Not available)		(OR=27.6, 95%CI: 1.4 to 533.2)	

Other penalization approaches

Smoothing methods using penalized likelihood as the examples of regularization methods are the alternative approaches to overcome the sparse-data issue. Although we do not go through the daunting discussion of these methods, introducing them can be useful. Imagine our object is to estimate the coefficient of a logistic regression model. The penalized likelihood function defined as, $L^*(\beta) = L(\beta) + \lambda(\beta)$ where $\lambda(.)$ is a penalty function. The different forms of λ provide a variety of penalization methods (33). Consider a penalty function as $\lambda(\beta) = \lambda \sum|\beta_i|^q$; For example, a quadratic form of the penalty function (q=2) was applied for models with the standardized explanatory variables. This is also referred to as the L2-norm method, which is analogs of ridge regression for linear normal-response models (33,34). Using the Absolute penalty function (q=1) instead of the quadratic form introduces the L1-norm regularization. This penalty method is also referred to as the LASSO in ordinary least squares regression. The L0-norm takes $\lambda(\beta)$ to be proportional to the number of nonzero β_i, (q=0). Each form has the own pros and cons; for example, in LASSO, compare to the quadratic form, the more unstable coefficients may be shrunk toward 0 and thus eliminated from the model, so LASSO is useful when the goal is to reduce the dimension of the model. Its disadvantage is when shrinking the truly-large estimates toward zero, which can be highly biased (35). In these approaches, the degree of smoothing depends on the penalty parameter and the choice of which is the matter of bias/variance trade-off.

Bayesian approach modeling

Bayesian inference provides a powerful methodology to deal with the separation problem (36). Contrary to panelized maximum likelihood techniques, which apply the traditional estimating approach by adjusting the result through a penalty, the Bayesian method simulates a distribution function named posterior, which is proportional to the likelihood of the data and prior distribution. In this case, the penalized likelihood estimates are the posterior modes for the Bayesian approach using prior distributions. In other words, a prior distribution on the parameters could be considered as a penalty. For example, posterior modes for the Bayesian approach using Jeffrey's prior distribution (beta with parameters α=0.5, β=0.5) provide similar results as Firth's penalized estimate (38). With prior distribution Beta (α=0.5, β=0.5) and likelihood binomial (y, n), posterior distribution would be Beta (α=y + 1, β=n-y + 2). Posterior simulation through sampling was firstly initiated by the Monte Carlo integration (MCI) approach; later, its concept was developed in the Importance of Sampling. Contrary to MCI on which samples are treated evenly, a "weight," which shows the importance of a sample, is allocated to each generated sample through an important function. As Bayes rule, Posterior distribution $f(\theta|y) \propto f(y|\theta)f(\theta)$ in which $f(y|\theta)$ is the likelihood of observed data and $f(\theta)$ prior distribution of θ. Weighted priors can simulate the posterior distribution through this approach. It should be considered that the contribution of generated samples to the estimation of the posterior depends on how much it is supported by data. It means that the prior distribution should not be far from the likelihood. With the emergence of MCMC, posterior simulation methods entered a new stage and tackled most flaws of the traditional methods. Different classes of prior distribution could be applied in this case. Although the prior selection is a matter of conflict since the results of Bayesian analyses may be sensitive to its choice. Sensitivity analysis is introduced to overcome this issue (38). It is worth mentioning that DA and MCMC with the same prior information yield similar results. Contrary to MCMC, which has a more complex algorithm of analysis, DA introduces a simpler, more fast-running, and understandable approach (4). The

normal and log-F distributions are the most common prior families for logistic coefficients (39). Normal priors for βj are symmetric, while Log-F (m, m) prior provides a more flexible tool, which is the natural conjugate-prior family for the logistic regression. It is like adding data with m/2 successes on m trials (4). The variance of the prior also has a profound influence on the background information it carries; prior distributions with large variance provide weak background information.

Applied example: effect of urine iodine status during pregnancy on stillbirth

In this study, data were extracted from the Tehran Thyroid and Pregnancy Study, a two-phase cohort study conducted from September 2013 through February 2016. Details of the study protocol have previously been published (40,41). The healthy pregnant women were divided into two groups according to the urine iodine concentration (UIC) status during pregnancy: UIC <150 μg/L and UIC ≥150 μg/L (reference group), and pregnancy outcomes were evaluated for them. We analyzed stillbirth, which was subjected to sparsity in the levels of UIC. STATA codes are available in the Appendix.

Table 3. Baseline characteristics in categories of the endpoint

Outcome/Factors		Still Birth	
		No.	Yes
UIC Level	<150 mg/dl	532	0
	=>150 mg/dl (Ref.)	494	2
Hypertension	Yes	9	0
	No.	1010	2
Age, mean (SD)		27.2(5.3)	20.2(6.9)
BMI, mean (SD)		24.8(4.4)	21.4(1.7)

Table 4. Different classes of prior information median(95%limit)

Priors	Exact prior median OR	95% prior limit OR
1: Normal (ln (1), 0.5)	1	(0.25, 4) *
2: Normal (ln (1), 1.38)	1	(0.10, 10) **
3: Normal (ln (2), 0.5)	2	(0.50, 8) *
4: Normal (ln (2), 1.38)	2	(0.20, 20) **
5:log-F (ln (1), df1=2000, df2=2, scale parameter=1)	1.44	(0.27, 39.50) **
6:log-F (ln (1), df1=2000, df2=2, scale parameter=.5)	1.20	(0.52, 6.28) *
7:log-F (ln (2), df1=2000, df2=2, scale parameter=1)	2.88	(0.54, 78.99) **
8:log-F (ln (2), df1=2000, df2=2, scale parameter=.5)	2.40	(1.04, 12.57) *

*Informative priors with lower limits, **non-informative priors with wider limits

Table 5. Results of multiple logistic regression through different approaches

Factors/ Analysis	UIC	Hypertension	Age	BMI
Ordinary Logistic Regression	NA	NA	NA	.68 (.36, 2.1)
¥ DA with normal prior1	.65 (.19, 2.3),	1.0 (.26, 3.8)	.74(.50, 1.1)	.87 (.57, 1.3)
¥ DA with normal prior2	.41 (.07, 2.5)	1.0 (.11, 8.6)	.72 (.48, 1.1)	.87 (.57, 1.3)
¥ DA with normal prior3	1.2 (.33, 4.0)	1.3 (.80, 2.0)	.76 (.52, 1.1)	.86 (.57, 1.3)
¥ DA with normal prior4	.65 (.11, 3.8)	1.2 (.60, 2.5)	.74 (.49, 1.1)	.87 (.57, 1.3)
¥ DA with log-F prior5	.43 (.09, 2.0)	1.0 (.08, 12.7)	.75 (.48, 1.1)	.87 (.58, 1.3)
¥ DA with log-F prior6	1.4 (.55, 3.3)	1.3 (.89, 1.8)	.83 (.62, 1.1)	.84 (.54, 1.3)
¥ DA with log-F prior7	.75 (.17, 3.0)	1.2 (.59, 2.5)	.74 (.49, 1.1)	.87 (.57, 1.3)
¥ DA with log-F prior8	1.4 (.55, 3.3)	1.3 (.89, 1.8)	.85 (.63, 1.1)	.84 (.55, 1.3)
¥¥ Bayesian with normal prior1	.65(.19, 2.3)	.99 (.19, 3.9)	.74(.50, 1.1)	.86 (.56, 1.3)
¥¥Bayesian with normal prior2	.42 (.09, 2.6)	1.0 (.09, 8.8)	.72 (.48, 1.1)	.87 (.57, 1.3)
¥¥Bayesian with normal prior3	1.2 (.33, 4.0)	1.0 (.80, 2.0)	.76 (.50, 1.0)	.86 (.57, 1.3)
¥¥Bayesian with normal prior4	.61 (.11, 3.2)	1.0 (.80, 2.0)	.70 (.49, 1.1)	.87 (.57, 1.3)

¥ Odds Ratio (95% Profile Likelihood Confidence Interval)
¥¥Odds Ratio (95% credible Interval)

Results

Table 3 provides a data illustration. We applied multiple

logistic regression through DA and Bayesian approaches to estimate the effect of UIC levels adjusted by hypertension, age, and BMI on stillbirth. Various prior information was selected to reflect their effects. We suppose that a positive but not strong association was expected for all covariates (42). We translated this background information into normal and Log-F priors for the log odds-ratio (Table 4). Regarding the positive association between exposures and stillbirth, Log-F prior as an asymmetric prior could better reflect the available background information (7,31). Results showed a reasonable shrink of ORs and 95%CI plausible ranges. Although UIC was not found significant, the risk of stillbirth was higher in mothers with lower UIC than 150 mg/dl when we applied informative priors Normal (ln (2), 0.5), log-F (ln (1), df1=2000, df2=2, scale parameter=.5) and log-F (ln (2), df1=2000, df2=2, scale parameter=.5), (Table 5). In addition, both methods provided almost the same result in normal priors. UIC variable was sensitive to the selection of prior distribution, which ranges from negative to positive effect while other variables were almost stable. Log-F prior was not supplied by the "bayesmh" STATA command.

Discussion

In this study, we reviewed sufficient statistical approaches to deal with the sparse data problem. Bayesian and DA methods as the key approaches were conducted in the applied example. Although the results of both methods were almost the same, DA was considered as a more tangible and easy-to-applied method, especially for researchers less familiar with the statistical method. It has the advantages of being simple and fast-running (4). In these methods, we seek a balance between prior information in the form of expert knowledge or belief and evidence from data. Achieving the right balance is a challenging issue. Actually, we need a strong enough prior to support weak evidence that usually comes from insufficient data, while non-informative prior does not effectively tackle the problem of separation. In this view, it can be valuable to have some flexibility in the location, scale, and shape of the prior. We thus considered two basic families of priors for logistic coefficients: the normal and the generalized Log-F distributions. We suppose that a positive but not strong association was expected for all covariates. We translated this background information into normal and Log-F priors for the log odds-ratio. Regarding the positive association between exposures and endpoint information, Log-F prior as an asymmetric prior could better reflect the available background information (7,39,43).

Extremely large measures of association such as the odds ratio along with an extraordinarily wide range of confidence interval proved the traditional estimation methods futile in case of sparse data while it is still widely applying and reporting. The biased estimation provides misleading information inappropriate for clinical decision making. Bayesian and data augmentation as the advanced methods provide an unbiased estimation.

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors declared no Conflict of interest.

Appendix

Data augmentation illustrating example

Consider a hypothetical 2 by 2 contingency table (Appendix-Table 1), provided Ln (OR)=Ln (13) =2.6, Var (Ln (OR)) = 1.29, 95% CI OR= (1.4 to 123.9).

Appendix-Table 1. A Hypothetical 2 by 2 Contingency Table

	Treated	Non-Treated
Diseased	a=10	b=9
Healthy	c=1	d=12
Total	N_1=11	N_2=21

Ln (OR)=Ln (13) =2.6, Var (Ln (OR)) = 1.29, 95% CI OR= (1.4 to 123.9)

Prior information for OR with 95% limits between $\frac{1}{2}$ and 2 was considered. Mean and variance of prior for Ln (OR) are estimated as followings;

$$\text{prior mean Ln(OR)} = \text{average of 95\% limits} = \frac{(\text{Ln}\left(\frac{1}{2}\right) + \text{Ln}(2))}{2} = 0$$

$$\text{prior variance Ln(OR)} = \left(\frac{\text{Width of interval in Ln(RR)units}}{\text{Width of interval in standard devision units}}\right)^2 = \left(\frac{\left|\text{Ln}\left(\frac{1}{2}\right) - \text{Ln}(2)\right|}{2 * 1.96}\right)^2 = 0.13$$

Therefore, a normal prior with mean 0 and variance 0.13 was defined. The contribution of prior and data information to estimate posterior mean and variance could be assessed through their inverse variances equaling $\frac{1}{\text{Variance of prior}} = \frac{1}{0.13} = 8.3$ and $\frac{1}{\text{variance of LnOR}} = \frac{1}{1.29} = 0.78$ which showed prior information dominated data information by nearly 11 times. Posterior mean and variance for Ln (OR) could be estimated as the following weighted averaging rule of thumb; Posterior mean for $\ln(\text{OR}) = \frac{\frac{\text{Mean of Prior}}{\text{Variance of prior}} + \frac{\text{lnOR}}{\text{Variance of LnOR}}}{\frac{1}{\text{Variance of prior}} + \frac{1}{\text{Variance of LnOR}}} = \frac{\frac{0}{0.13} + \frac{2.6}{1.29}}{\frac{1}{0.13} + \frac{1}{1.29}} = 0.24$, Posterior variance for $\ln(\text{OR}) \approx \frac{1}{\frac{1}{\text{Variance of prior}} + \frac{1}{\text{Variance of LnOR}}} = \frac{1}{\frac{1}{0.13} + \frac{1}{1.29}} = 0.12$ and 95% posterior CI for OR $\approx \exp\left(\text{posterior mean} \pm 1.96 * (\text{Posterior variance})^{\frac{1}{2}}\right) = \exp\left(0.24 \pm 1.96 * (0.12)^{\frac{1}{2}}\right) = \exp(-0.44, 0.91) = (0.64, 2.48)$.

The width of CI obtained from this method was 66 times narrower than the ordinary approach. Also, the value of posterior mean which is closer to prior mean than to data showed the influence of prior as well. DA adds data generated from prior to the actual dataset automatically and then uses ordinary likelihood method to estimate parameters.

The number of pseudo-data generated by Prior estimate, which shows the contribution of prior in estimating posterior, calculated as the followings; for simplicity we consider OR$\cong RR$ for small sample size (44). Estimated $RR = \frac{\frac{a}{N_1}}{\frac{b}{N_2}} \overset{N_1=N_2}{\Longleftrightarrow} \frac{a}{b} = 1 \Rightarrow a = b$, and estimated variance for $Ln(RR) \approx \frac{1}{a} + \frac{1}{b} \overset{a=b}{\Longleftrightarrow} \frac{2}{a} = 0.12 \Rightarrow a = b \approx 17$, means that 17 observations were added to the exposed and non-exposed cases to estimate the posterior OR.

Compatibility of the prior and data is of great importance, DA fails in case of incompatibility causes misleading results (5,31). It means that they should be homogenous enough to average their estimates by information weighting. P-value from the statistic 'standardized' difference between frequentist estimate and prior estimate tests Compatibility of Prior and data. For our example, $\frac{(MeanLnOR - Mean\ Prior)}{(Variance\ LnOR - variance\ Prior)^{\frac{1}{2}}} = \frac{(Ln(13)-0)}{(1.29+0.12)^{\frac{1}{2}}} = 2.16 \overset{Normal\ table}{\Longrightarrow} P_{value} = 0.031$ that compatibility hypothesis is rejected.

Software

Firth penalization

STATA command "firthlogit", SAS software "FIRTH option in PROC LOGISTIC" and R-Package "logistf" could be applied to run the Firth penalization analysis

Data augmentation and bayesian approaches

The "Penlogit" is a STATA command for the approximate Bayesian logistic regression using penalized likelihood estimation via data augmentation

STATA software packages installation code

ssc install penlogit
ssc install bayesmh

STATA command

penlogit outcome_stillbirth UI_150 hypertension Age BMI, lfprior(UI_150 ln(2) 2000 1 1 hypertension ln(2) 2000 1 1 Age ln(2) 2000 1 1) ppl(UI_150 hypertension Age BMI) or

penlogit outcome_stillbirth UI_150 hypertension Age BMI, lfprior(UI_150 ln(2) 2000 1 .5 hypertension ln(2) 2000 1 .5 Age ln(2) 2000 1 .5) ppl(UI_150 hypertension Age BMI) or

penlogit outcome_stillbirth UI_150 hypertension Age BMI, lfprior(UI_150 ln(1) 2000 1 1 hypertension ln(1) 2000 1 1 Age ln(1) 2000 1 1) ppl(UI_150 hypertension Age BMI) or

penlogit outcome_stillbirth UI_150 hypertension Age BMI, lfprior(UI_150 ln(1) 2000 1 .5 hypertension ln(1) 2000 1 .5 Age ln(1) 2000 1 .5) ppl(UI_150 hypertension Age BMI) or

penlogit outcome_stillbirth UI_150 hypertension Age BMI, nprior(UI_150 ln(1) .5 hypertension ln(1) .5 Age ln(1) .5) ppl(UI_150 hypertension Age BMI) or

penlogit outcome_stillbirth UI_150 hypertension Age BMI, nprior(UI_150 ln(1) 1.38 hypertension ln(1) 1.38 Age ln(1) 1.38) ppl(UI_150 hypertension Age BMI) or

penlogit outcome_stillbirth UI_150 hypertension Age BMI, nprior(UI_150 ln(2) .5 hypertension ln(2) .5 Age ln(2) .5) ppl(UI_150 hypertension Age BMI) or

penlogit outcome_stillbirth UI_150 hypertension Age BMI, nprior(UI_150 ln(2) 1.38 hypertension ln(2) 1.38 Age ln(2) 1.38) ppl(UI_150 hypertension Age BMI) or

```
##########################
bayesmh outcome_stillbirth UI_150 hypertension Age BMI, likelihood(logit) prior({UI_150}, normal(0,.5))prior({Age}, normal(0,.5))prior({hypertension}, normal(0,.5))prior({BMI}, normal(0,.5)) prior({_cons}, flat)
bayesmh outcome_stillbirth UI_150 hypertension Age BMI, likelihood(logit) prior({UI_150}, normal(0,1.38))prior({Age}, normal(0,1.38))prior({hypertension}, normal(0,1.38))prior({BMI}, normal(0,1.38)) prior({_cons}, flat)
bayesmh outcome_stillbirth UI_150 hypertension Age BMI, likelihood(logit) prior({UI_150}, normal(ln(2),.5))prior({Age}, normal(ln(2),.5))prior({hypertension}, normal(ln(2),.5))prior({BMI}, normal(ln(2),.5)) prior({_cons}, flat)
bayesmh outcome_stillbirth UI_150 hypertension Age BMI, likelihood(logit) prior({UI_150}, normal(ln(2),1.38))prior({Age}, normal(ln(2),1.38))prior({hypertension}, normal(ln(2),1.38))prior({BMI}, normal(ln(2),1.38)) prior({_cons}, flat)
##########################
```

References

1. Albert A, Anderson JA. On the Existence of Maximum Likelihood Estimates in Logistic Regression Models. Biometrika 1984;71:1-10.
2. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med 2002;21:2409-19.
3. Greenland S, Schwartzbaum JA, Finkle WD. Problems due to Small Samples and Sparse Data in Conditional Logistic Regression Analysis. Am J Epidemiol 2000;151:531-9.
4. Sullivan SG, Greenland S. Bayesian regression in SAS software. Int J Epidemiol 2012;42:308-17.
5. Discacciati A, Orsini N, Greenland S. Approximate Bayesian logistic regression via penalized likelihood by data augmentation. Stata J 2015;15:712-36.
6. Lyles RH, Guo Y, Greenland S. Reducing Bias and Mean Squared Error Associated With Regression-Based Odds Ratio Estimators. J Stat Plan Inference 2012;142:3235-41.
7. Greenland S, Mansournia MA. Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. Stat Med 2015;34:3133-43.
8. Heinze G. A comparative investigation of methods for logistic regression with separated or nearly separated data. Stat Med 2006;25:4216-26.
9. Fushihara G, Kamide T, Kimura T, Takeda R, Ikeda T, Kikkawa Y, et al. Factors associated with early seizures after surgery of unruptured intracranial aneurysms. Clin Neurol Neurosurg 2019;178:93-6.
10. Gambhir S, Grigorian A, Ashbaugh A, Spencer D, Ramakrishnan D, Schubl SD, et al. Early Versus Late Pulmonary Embolism in Trauma Patients: Not All Pulmonary Embolisms are Created Similarly. J Surg Res 2019;239:174-9.
11. Kim WH, Kim HJ, Park HY, Park JY, Chae YS, Lee SM, et al. Axillary Pathologic Complete Response to Neoadjuvant Chemotherapy in Clinically Node-Positive Breast Cancer Patients: A Predictive Model Integrating the Imaging Characteristics of Ultrasound Restaging with Known Clinicopathologic Characteristics. Ultrasound Med Biol 2019;45:702-9.
12. Panahi MH, Bidhendi RY. Bias in determining factors associated with early seizures after surgery of unruptured intracranial aneurysms. Clin Neurol Neurosurg 2019;179:66.
13. Yarandi RB, Panahi MH. Is Granulocyte colony-stimulating factor associated with development of aortitis? Cytokine 2019;120:191.
14. Oshima Y, Takahashi S, Tani K, Tojo A. Granulocyte colony-stimulating factor-associated aortitis in the Japanese Adverse Drug Event Report database. Cytokine 2019;119:47-51.
15. Bidhendi Yarandi R, Panahi MH. Bias estimation of predictors and internal validity of the study "Admission characteristics predictive of in-hospital death from hospital-acquired sepsis: A comparison to community-acquired sepsis". J Crit Care 2019;56:321.
16. Bidhendi Yarandi R, Panahi MH. Methodological issues regarding "Decline in ankle-brachial index is stronger in poorly than in well controlled diabetes: Results from the Heinz Nixdorf Recall cohort study". Atherosclerosis 2019;286:179.
17. Bidhendi Yarandi R, Panahi MH. Postnatal nutritional deficit is an independent predictor of bronchopulmonary dysplasia among extremely premature infants born at or <28 weeks gestation: Some methodological issues. Early Human Development. 2019;134:47.
18. Panahi MH, Bidhendi Yarandi R. Is irradiation significantly associated with a higher risk for CVD? Eur Arch Otorhinolaryngol 2019;277:2651.
19. Alipour A, Hoseinabadi TS, Esmaeilzadeh A, Shadkam O. The Effect of Epinephrine Sprayed on the Papilla on Prevention of Post–Endoscopic Retrograde Cholangiopancreatography Pancreatitis (PEP); a Double-

Blind Randomized Control Trial. Acta Med Iran 2020;58:456-60.
20. Firth D. Bias Reduction of Maximum Likelihood Estimates. Biometrika 1993;80:27-38.
21. Greenland S. Simpson's Paradox From Adding Constants in Contingency Tables as an Example of Bayesian Noncollapsibility. Am Stat 2010;64:340-4.
22. Rahman MS, Sultana M. Performance of Firth-and logF-type penalized methods in risk prediction for small or sparse binary data. BMC Med Res Methodol 2017;17:33.
23. Heinze G, Ladner T. logistiX: Exact Logistic Regression Including Firth Correction. R package version, 2013.
24. Kosmidis I. brglm: Bias reduction in binomial-response generalized linear models. R Foundation for Statistical Computing, 2013.
25. Coveney J. FIRTHLOGIT: Stata module to calculate bias reduction in logistic regression. IDEAS/RePEc search, 2015.
26. Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. Int J Epidemiol 2006;35:765-75.
27. Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. BMJ. 2016;352:i1981.
28. Greenland S, Christensen R. Data augmentation priors for Bayesian and semi-Bayes analyses of conditional-logistic and proportional-hazards regression. Stat Med 2001;20:2421-8.
29. Bedrick EJ, Christensen R, Johnson W. A New Perspective on Priors for Generalized Linear Models. J Am Stat Assoc 1996;91:1450-60.
30. Bedrick EJ, Christensen R, Johnson W. Bayesian binomial regression: Predicting survival at a trauma center. Am Stat 1997;51:211-8.
31. George EPB. Sampling and Bayes' Inference in Scientific Modelling and Robustness. J R Stat Soc Ser A Stat Soc 1980;143:383-30.
32. Simonoff JS. A penalty function approach to smoothing large sparse contingency tables. Ann Appl Stat 1983;11:208-18.
33. Lee A, Silvapulle M. Ridge estimation in logistic regression. Commun Stat Simul Comput 1988;17:1231-57.
34. Le Cessie S, Van Houwelingen JC. Ridge estimators in logistic regression. J R Stat Soc Ser A Stat Soc 1992;41:191-201.
35. Trevor H, Robert T, JH F. The elements of statistical learning: data mining, inference, and prediction. New York, NY: Springer, 2009.
36. Bidhendi-Yarandi R, Mohammad K, Zeraati H, Tehrani FR, Mansournia MA. Bayesian Methods for Clinicians Med J Islam Repub Iran 2020;34:78.
37. Agresti A, Kateri M. Categorical data analysis. New York: Springer, 2011.
38. Hamra GB, MacLehose RF, Cole SR. Sensitivity analyses for sparse-data problems-using weakly informative bayesian priors. Epidemiology (Cambridge, Mass) 2013;24:233-9.
39. Greenland S. Prior data for non-normal priors. Stat Med 2007;26:3578-90.
40. Nazarpour S, Tehrani FR, Simbar M, Tohidi M, Azizi F. Thyroid and pregnancy in Tehran, Iran: objectives and study protocol. Int J Endocrinol Metab 2016;14:e33477.
41. Nazarpour S, Ramezani Tehrani F, Simbar M, Tohidi M, Minooee S, Rahmati M, et al. Effects of Levothyroxine on Pregnant Women With Subclinical Hypothyroidism, Negative for Thyroid Peroxidase Antibodies. J Clin Endocrinol Metab 2017;103:926-35.
42. Mills JL, Ali M, Buck Louis GM, Kannan K, Weck J, Wan Y, et al. Pregnancy Loss and Iodine Status: The LIFE Prospective Cohort Study. Nutrients 2019;11:534.
43. Allison PD. Convergence failures in logistic regression. SAS Global Forum, 2008.
44. Jewell NP. Statistics for epidemiology. UK: Chapman and Hall/CRC, 2004.