

RESEARCH ARTICLE

Misclassification Adjustment of Family History of Breast Cancer in a Case-Control Study: a Bayesian Approach

Rahmatollah Moradzadeh¹, Mohammad Ali Mansournia^{1*}, Taban Baghfalaki², Reza Ghasvand³, Mohammad Reza Noori-Daloui⁴, Kourosh Holakouie-Naieni^{1*}

Abstract

Background: Misreporting self-reported family history may lead to biased estimations. We used Bayesian methods to adjust for exposure misclassification. **Materials and Methods:** A hospital-based case-control study was used to identify breast cancer risk factors among Iranian women. Three models were jointly considered; an outcome, an exposure and a measurement model. All models were fitted using Bayesian methods, run to achieve convergence. **Results:** Bayesian analysis in the model without misclassification showed that the odds ratios for the relationship between breast cancer and a family history in different prior distributions were 2.98 (95% CRI: 2.41, 3.71), 2.57 (95% CRI: 1.95, 3.41) and 2.53 (95% CRI: 1.93, 3.31). In the misclassified model, adjusted odds ratios for misclassification in the different situations were 2.64 (95% CRI: 2.02, 3.47), 2.64 (95% CRI: 2.02, 3.46), 1.60 (95% CRI: 1.07, 2.38), 1.61 (95% CRI: 1.07, 2.40), 1.57 (95% CRI: 1.05, 2.35), 1.58 (95% CRI: 1.06, 2.34) and 1.57 (95% CRI: 1.06, 2.33). **Conclusions:** It was concluded that self-reported family history may be misclassified in different scenarios. Due to the lack of validation studies in Iran, more attention to this matter in future research is suggested, especially while obtaining results in accordance with sensitivity and specificity values.

Keywords: Misclassification - bias - Bayesian assessment - self-report

Asian Pac J Cancer Prev, 16 (18), 8221-8226

Introduction

Data collection from family history of diseases is very important in both clinical and research studies. Family history data are essential to estimate a family aggregation, obtain its effects on disease, discover heterogeneity, observe existence of multiple phenotypes in a family, and estimate the number of genes involved in a disorder (Szatmari and Jones, 1999).

Having a family history of cancer is an important risk factor for many cancers, including breast cancer (Ebrahimi et al., 2002; Holakouie-Naieni et al., 2007; Mahouri et al., 2007; Hassanzadeh et al., 2012; Zare et al., 2013; Hosseinzadeh et al., 2014; Tehranifar et al., 2015), which its family history is often evaluated via self-report (Veisy et al., 2015). However, validity of self-reported data is often problematic, as it is believed that the participants in the study tend to be underestimated, or may deny the existence of family history of the disease. So, people with a Family History of Breast Cancer (FHBC) in first-degree relatives are less likely to report it. Reliability of self-report questionnaires may be influenced by the characteristics of the participants. Consistency of self-reported data on

family history may be affected by some features such as healthy adults, gender, ethnicity, and education level. Also, validity and/or reliability of self-reported health data may be affected by depression, disability and cognitive disorders (Marrie et al., 2008). Given these considerations and keeping in mind the increase in the number of studies determining the family history status of breast cancer, it is essential to identify valid and reliable measures.

According to previous studies, although specificity of self-reported FHBC in first-degree relatives are high, but its sensitivity may be very low (Szatmari and Jones, 1999). Misclassification of exposure variables related to under-reporting is a major problem in many studies (Prescott and Garthwaite, 2005) and the definition of exposure may alter the effects of exposure (Rothman et al., 2008; Keil et al., 2014). Ignoring misclassification may lead to bias estimations, therefore, it should be considered while both designing and analyzing the study (Prescott and Garthwaite, 2005).

Recently, various methods were published to correct potential biases in observational studies. But for several reasons, such as complexity of the procedures and lack of need for them, these methods are rarely used or reported.

¹Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, ⁴Department of Medical Genetics, School of Medicine, Tehran University of Medical Sciences, ²Department of Statistics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran, ³Oslo Center for Biostatistics and Epidemiology, Institute of Basic Medical Sciences, University of Oslo, Norway *For correspondence: mansournia_ma@yahoo.com, holakoik@hotmail.com

However, these reasons may not justify disregarding the observed potential bias between exposure-outcome relationships (Gelder et al., 2014b).

Although, exposure misclassification correction using Frequentist methods are also possible, but Bayesian methods are preferable (Greenland, 2008).

In this study, we used Bayesian methods to adjust the exposure misclassification. Bayesian methods under consideration was introduced by MacLehose et al (MacLehose et al., 2009). We attempted to explore the relationship between breast cancer and its family history amongst first-degree relatives after adjusting for confounders, regardless the exposure misclassification and also, to compare the results with posterior odds ratios after adjusting for confounder variables and a range of exposure misclassification scenarios.

Materials and Methods

Study population

We have used a hospital-based case-control study to identify breast cancer risk factors amongst Iranian women. The findings of this study have been published in detail, previously (Ghiasvand et al., 2011; Ghiasvand et al., 2012). Cases were recruited from Mottahari Breast Clinic of Shiraz University of Medical Sciences. This center collects data from about 80% of all incident breast cancer patients treated in main hospitals of Shiraz city. Eligible cases were women with an incident histopathologically confirmed breast cancer, diagnosed at 50 years of age or older. Most cases (93%) were interviewed within six months after diagnosis. We consider all the individuals with complete data in this study. Also eleven case patients from other provinces have been deleted from the study because we could not find any matched controls for them. Controls were frequency-matched with cases on five-year age groups and province of residence. Controls were primarily selected from healthy female visitors accompanying patients referred to the Faghihi hospital for general surgery (60%), urology (24%) and cardiovascular (16%) diseases. A total of 1090 controls were selected, but 92 women (8%) refused to participate.

For both cases and controls, face-to-face interviews were performed. Controls were interviewed from May through August 2009. Interviews were conducted by two trained female nurses (one for cases and one for controls), and the time of interviews was similar for cases and controls. None of the interviewers were aware of the study hypotheses.

Statistical analysis

Many researchers have proposed probabilistic bias analysis methods to consider bias (Lash and Fink, 2003; Fox et al., 2005; Greenland, 2005; Greenland and Lash, 2008; Lash et al., 2009; MacLehose and Gustafson, 2012 January). These methods serve to adjust the main effect estimate and propagate uncertainty surrounding the bias parameter, incorporating this uncertainty into the variance estimate of the adjusted main effect. The approach taken in probabilistic bias analysis is to repeatedly draw a random sample from the bias parameter distribution(s) and use

those sampled parameters to adjust the effect estimate. The resulting distribution will be “bias-adjusted” main effect (Fox et al., 2005; MacLehose and Gustafson, 2012 January).

Bayesian analysis is based on the posterior distribution, although samples from it can be generated using Markov Chain Monte Carlo (MCMC) algorithm (MacLehose and Gustafson, 2012 January). The Monte-Carlo sensitivity analysis procedure (Fox et al., 2005) is computationally intensive, while Bayesian analysis with MCMC algorithms can be implemented quickly (Chu et al., 2006).

Unlike deterministic maximum-likelihood algorithms, the MCMC methods are some stochastic procedures that repeatedly generate random samples. The process of generating the random samples in MCMC return to the role of the Markov chain, and the process of generating summary statistics from the generated random samples return to Monte Carlo integration (Hamra et al., 2013a).

Details of Bayesian method have been described in the following.

The exposure variable in the study is FHBC in the first degree relatives (fh^{true}), where true family histories ($fh^{reported}$) are unobserved values, which we’ve estimated in this study.

Breast Cancer is the outcome variable.

Z is a vector of possible confounders: age at menarche (Less than 12 years, 12-15 and more than 15 years), menopause status (before, after), parity (yes, no), past use of oral contraceptives (yes, no), age at first pregnancy (Less than 25 years, equal and more than 25 years and nulliparous) and history of breastfeeding (yes, no) and body mass index (BMI) (as a continuous variable).

We treated fh^{true} as measured with error by $fh^{reported}$ and used information from previous research on the sensitivity and specificity of self-reported family history data to produce corrected estimates.

We conducted Bayesian uncertainty analyses conditional on prior hypotheses generated from published studies. Briefly, three models are jointly considered: an outcome, an exposure, and a measurement model, which allowed simultaneous imputation of the true family history exposure status and estimation of its effect on the risk of the breast cancer.

For our analysis, we used the directed acyclic graph (DAG) shown in Figure 1.

Figure 1. Directed Acyclic Graph (DAG) for the relationship between self-reported family history status and breast cancer.

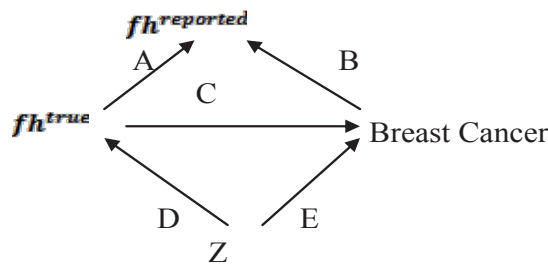


Figure 1. Directed Acyclic Graph (DAG) for the Relationship between Self-reported Family History Status and Breast Cancer

The outcome model specified the probability of breast cancer as a function of fh^{true} (arrow C) and the other covariates Z (arrow E), a measurement model that specified the probability of $fh^{reported}$ as a function of fh^{true} (arrow A) and breast cancer (arrow B), and an exposure model that specified the probability of fh^{true} as a function of the covariates Z (arrow D) (MacLehose et al., 2009).

Outcome Model

The outcome model which is used in the study is given by

$$\text{logit}(\Pr(BC_i=1)) = \beta_0 + \beta_1 fh_i^{true} + Z_i' \theta$$

Where β_0 , β_1 and θ are the unknown parameters.

In the Bayesian approach, the prior distributions are defined for these unknown parameters.

We used a non-informative $N(0,)$ for the prior distribution of the intercept term (β_0), and informative priors for other coefficients in the model (Table 1). Also, we have used non-informative priors $N(0,1000)$ (according to Gelman et al. (Gelman et al., 2003) and prior distributions $N(0,1.38)$ (according to Hamra et al. (Hamra et al., 2013b; Greenland and Mansournia, 2015)) for other regression coefficients in the model.

Exposure model

For the true exposure variable, we consider the following exposure model:

$$\text{logit}(\Pr(fh_i^{true}=1)) = \omega_0 + Z_i' \omega$$

Where the prior distribution $N(0,10\ 000)$ is used for intercept coefficient, and prior distributions $N(0,1)$, $N(0,10)$, and $N(0,1.38)$ are used for other regression coefficients in this model (MacLehose et al., 2009).

Measurement Model

For the reported exposure, we consider the following measurement model:

$$\begin{aligned} \text{PR}(fh_i^{reported}=1) \\ = \alpha_0 fh_i^{true} (1-BC_i) + \alpha_1 (1-fh_i^{true}) (1-BC_i) + \alpha_2 fh_i^{true} BC_i + \alpha_3 (1-fh_i^{true}) BC_i \end{aligned}$$

In this model α_0 is the sensitivity of reported family history (fh) among controls, α_1 is the false positive rate (FPR) of reported family history among controls, α_2 is the sensitivity of reported family history (fh) among cases and α_3 is the false positive rate (FPR) of reported family history among cases.

The exposure and measurement models were used to impute values of true family history variable in a way similar to that used with missing data techniques. These imputed values were then used to estimate the associations between FHBC and breast cancer.

In order to estimate the effects of under-reporting (misclassification) in the exposure of FHBC and breast cancer, we implemented two models that specified α_0 , α_1 , α_2 and α_3 in the measurement model.

Similar to MacLehose et al. (2009), in model 1, sensitivity and specificity are equal to 1.00 (that is no misclassification), which lead to a standard Bayesian logistic regression model.

As mentioned in MacLehose et al (2009), model

2 is based on the assumption that the sensitivities and false positive rates used in the measurement model are not exactly known. For the prior distribution of the sensitivity and false positive rate, a beta distribution was chosen with prior parameters selected to reflect a priori beliefs concerning reported family history. The values for sensitivity and false positive rate among cases and controls were assumed not to be correlated.

The beta distribution has two parameters; b_1 and b_2 . For sensitivities, b_3 is the number of women reported family history truly and b_2 is the number of women reported untruly family history. For FPRs, b_1 is the number of women reported family history but in truth not having family history, and b_2 is the number of women reported no family history and in truth not having family history (MacLehose et al., 2009). Also, we used uniform priors (beta(1,1)) for the coefficients in the measurement model.

All models were fitted using the Bayesian methods, which were run to achieve convergence. The convergence was checked with the Gelman-Rubin diagnostic test (Gelman and Rubin, 1992).

After the burn-in period, the iterations of the MCMC algorithm are random draws from the posterior distributions of interest; the mean was exponentiated to obtain the odds ratio of interest. We exponentiated the 2.5th and 97.5th percentiles of the random draws to obtain 95% Posterior Credible Intervals (CRIs)(van Gelder et al., 2014a).

We used R, STATA 12.0 and OpenBUGS softwares for the data analysis.

Results

In the present study, a total of 880 cases with breast cancer and 998 controls were included in the study.

In logistic regression analysis, odds ratio for the relationship between FHBC in the first degree relatives and breast cancer adjusted for covariates is 2.55 (95% CRI: 2.55 , 3.35).

In Bayesian analysis, first, we stated no misclassification model (model 1). Then model 1 for different values of the outcome model priors was run. Similar to MacLehose et al. (MacLehose et al., 2009), priors for outcome model parameters from previous studies were selected (Table 1). In status A, the results of Bayesian analysis (Table 2) showed that odds ratio of the relationship between FHBC in the first degree relatives and breast cancer was 2.98 (95% CRI: 2.41, 3.71). In B status, priors for outcome model parameters were changed and non-informative priors were used. After that, in order to consider Gelman et al. (Gelman et al., 2003) recommendation, $N(0, 1000)$ prior was used. OR of the aforementioned relationship was 2.57 (95% CRI: 1.95, 3.41). In C status $N(0, 1.38)$ as priors for outcome model parameters were used. OR of the relationship between FHBC in first degree relatives and breast cancer was 2.53 (95% CRI: 1.93, 3.31).

In misclassification model (model 2), it is essential to define the prior distributions for the unknown parameters of the outcome, exposure and measurement models.

The priors of the parameters of the outcome model were selected from different perspectives, including

previous literature and non-informative priors (N(0,1000) and N(0,1.38)). The priors of unknown parameters of the exposure model (g) defined in three scenarios, N(0,1), N(0,10) and N(0,1.38). Also, in the similar manner, in all

Table 1. The Outcome Model Priors Based on Previous Literatures

Variable name	Prior	
	OR*	95% CI
Family history		
No	1	
Yes	1.94	1.35, 2.78
Age at menarche1		
more than 15 years	1	
Less than 12 years	1.2	.74, 1.95
Age at menarche2		
more than 15 years	1	
12-15 years	1.3	.82, 1.99
Menopause status		
after	1	
before	3.15	2.32, 4.25
Parity		
yes	1	
no	1.85	1.25, 2.74
Past use of oral contraceptives		
no	1	
yes	1.1	.79, 1.53
Age at first pregnancy		
Less than 25 years	1	
Equal and more than 25 years and nulliparous	0.64	.45, .92
History of breastfeeding		
no	1	
yes	0.39	.27, .56
Body Mass Index (BMI) (continuous)	1.01	.99, 1.02

* OR: Odds Ratio; ** CI: Confidence Interval

Table 2. Priors for the Outcome and Exposure Models and Related OR (CRI 95%) for Models

status	Outcome model priors	Exposure model priors	OR*	95%CRI**
A	According to Table 1	-	2.98	2.41, 3.71
B	N(0,1000)	-	2.57	1.95, 3.41
C	N(0,1.38)	-	2.53	1.93, 3.31
D	According to Table 1	N (0,1)	2.64	2.02, 3.47
E	According to Table 1	N (0,10)	2.64	2.02, 3.46
F	N (0,1000)	N (0,1)	1.6	1.07, 2.38
G	N (0,1000)	N (0,10)	1.61	1.07, 2.40
H	N (0,1.38)	N (0,1)	1.57	1.05, 2.35
I	N (0,1.38)	N (0,1.38)	1.58	1.06, 2.34
J	N (0,1.38)	N (0,10)	1.57	1.06, 2.33
K	According to Table 1	N(0,1)	3.74	2.66-5.27
L	N(0,1.38)	N(0,10)	1.02	0.10-9.87
M	N(0,1.38)	N(0,1)	0.99	0.10-9.60

* OR: Odds Ratio; **CRI: Credible Interval

Table 3. The Measurement Model Priors for the Case and Control Groups

	Sensitivities priors		FPR* priors	
	b1	b2	b3	b4
case	695	45	76	1302
control	62	25	14	1115

*FPR False Positive Rate

situations, according to MacLehose et al. (MacLehose et al., 2009), priors for intercept of the outcome and exposure models were N(0, 10⁶) and N (0, 10⁴), respectively.

Priors of the measurement model, sensitivities and FPR in the case and control groups, were selected from Tehranifar et al and Jurek et al. (Jurek et al., 2009; Tehranifar et al., 2015) studies (Table 3) and beta(1,1). Therefore, we had 10 statuses including D, E, F, G, H, I and J (Table 2) and K, L and M (Table 2), that is, we could run 10 analyses based on Bayesian method.

In simulation with priors specification of D status, adjusted OR for misclassification and confounder variables between FHBC and breast cancer was 2.64 (95% CRI: 2.02, 3.47). In simulation with priors specification of E status, adjusted OR between FHBC and breast cancer was 2.64 (95% CRI: 2.02, 3.46).

In other simulations, in F and G statuses, in simulation with non-informative priors for outcome model, adjusted OR for misclassification and confounder variables were 1.60 (95% CRI: 1.07, 2.38) and 1.61 (95% CRI: 1.07, 2.40), respectively.

Similarly, in H, I and J situations, according to the priors N (0,1.38) for outcome model, the results were 1.57(95% CRI: 1.05, 2.35), 1.58(95% CRI: 1.06, 2.34) and 1.57 (95% CRI: 1.06, 2.33), respectively.

In accordance with the priors in K, L and M situations and the uniform distributions for sensitivities and specificities (beta(1,1)), the results were 3.74(95% CRI: 2.66, 5.27), 1.02(95% CRI: 0.10, 9.87) and 0.99(95% CRI: 0.10, 9.60), respectively.

All results in all scenarios showed that all ORs were statistically significant. All the results, taking into account the assumptions were acceptable, which will be explained in the discussion.

Discussion

All analyses of epidemiological studies should be seen as part of a sensitivity analysis, because we can't claim that assumptions are absolutely correct. Statisticians only consider inferential possibilities and do not pay attention to the inferences. Bias analysis focuses on the inferences which are not pointless and unreasonable (Greenland, 2009). On the other hand, any article should not be considered as a candidate for bias analysis. When no useful conclusion can't be running by conventional analysis, bias analysis is not necessary (Greenland, 2009).

Quantitative bias analysis prepare an estimate of the uncertainty caused by systematic errors and can provide a useful guide for further investigation (Lash et al., 2014).

In this study, we assumed that a family history of the disease in the first degree relatives is a candidate for bias due to misclassification. It is possible for a person to be asked whether she has a family history of breast cancer or not, and she does not assume the need for giving correct answer to this question. So they underestimate the question and may give inaccurate answers.

We consider the following limitations for our study. We assumed differential misclassification for the exposure variable. That is, sensitivity and specificity in the case and control groups are different from each other. As

another limitation for the study, the potential errors in other variables were ignored. In addition to the variables studied, other confounding variables may not be included in this study, which may affect the results. As another assumption, the amounts of previous literature priors were established for the present study. Moreover, it was assumed that the probability density of the external validation data is similar to those exists in this study. Because of the lack of validation data on FHBC in the first degree relatives in Shiraz (or in Iran as a whole), we used priors outside of Iran, therefore we assumed them as similar. A major challenge in data analysis applying Bayesian methods is appropriate priors determination. In this study, previous literature and non-informative and weakly informative priors were used for the coefficients of outcome and exposure models. Non-informative and weakly informative priors were selected as recommended by Gelman et al. (Gelman et al., 2003) and Hamra et al. (Hamra et al., 2013b; Greenland and Mansournia, 2015), respectively. Given the abovementioned assumptions and the accuracy of the bias model, we corrected odds ratios for misclassification of exposure variable on breast cancer data using Bayesian methods described by MacLehose et al (MacLehose et al., 2009).

In the model without misclassification (model 1), after applying the priors selected from literature as well as non-informative and weakly informative priors (means = 0 and variances = 1000 and 1.38), FHBC in the first degree relatives of the cases with breast cancer were, respectively, 2.98, 2.57 and 2.53 times more than those of the control group, after adjustment for confounders. On the other hand, in logistic regression analysis, OR of exposure-outcome association adjusted for confounders was 2.55 (95% CRI: 2.55, 3.35). This finding is similar to the results of Bayesian analysis using non-informative and weakly informative priors. So, the exposure-outcome relationship was relatively strong. Similar results were found in the studies conducted by Holakouee et al (Holakouie-Naieni et al., 2007) (OR= 2.96, 95% Confidence Interval(CI): 1.46, 5.99) and Sepandi et al (Sepandi et al., 2014) (OR= 1.94, 95% CI: 1.35, 2.78). But the exposure-outcome odds ratio in Mahouri et al, study (Mahouri et al., 2007 Nov-Dec) was 9.07 (95% CI: 4.06, 12.26). This difference may be due to the different settings of the studies, or it may be as a result of systematic errors occurred in the exposure measurement. However, it should be noted that such odds ratio (9.07, 95% CI: 4.06-12.26) is unusual for the association between FHBC and breast cancer.

The results of the misclassification model of exposure (model 2) showed some variations. If the priors of the outcome model were chosen from previous studies and the prior mean of the exposure model to be considered as 0 and its variances as 1 and 10, respectively, then the odds ratios will be 2.64(95% CRI: 2.02, 3.47) and 2.64(95% CRI: 2.02, 3.46). Therefore, the odds ratios obtained are similar and lower than those found in the model 1 (2.98). This difference is due to the misclassification existed in self-reported FHBC.

In another situation, the value of 1000 is the proper option for the prior variance of the outcome model. If the prior variance of outcome model is 1000 and the prior

variances of exposure model are 1 and 10, then, the odds ratios will be 1.60 (95% CRI: 1.07, 2.38) and 1.61 (95% CRI: 1.07, 2.40), respectively. Considering the sensitivity and specificity, it is remarkably lower than the odds ratio of model 1(2.57, 95% CRI: 1.95, 3.41) (prior variance of outcome model = 1000). Also, its effect remains significant and credible interval is narrower.

If the prior variance of outcome model is 1.38 and the priors variances of exposure model are equal to 1, 1.38 and 10, then the odds ratios will be 1.57(95% CRI:1.05, 2.35), 1.58 (95% CRI: 1.06, 2.34) and 1.57(95% CRI: 1.06, 2.33), respectively. These ORs are almost similar. But comparing with the odds ratio obtained from this prior in model 1(OR= 2.53, 95% CRI: 1.93, 3.31), less effect was indicated, again. This is due to the misclassification existed in the exposure variable.

When uniform distribution used for sensitivity and specificity values, ORs in different values of weakly informative priors for outcome and exposure models were insignificant and their CRIs were too wide. This is due to the use of uniform prior distribution for coefficients of measurement model.

In order to compare the calculated odds ratios applying Bayesian methods found in this study to those found in the other studies on adjusted OR using probabilistic bias analysis, the study conducted by Jurek et al. (Jurek et al., 2009) may be adopted. In no misclassification model, the relationship between FHBC and the outcome in their study was 1.63 (95% uncertainty limits: 1.63, 1.63), which is much less than those found in this study. While, in misclassification model, in different scenarios for sensitivity and specificity values with uniform and triangular distributions, adjusted OR were 2.27 (95% CI: 1.33, 6.01), 1.77(95% CI: 1.25, 3.93), and 1.99 (95% CI: 1.37, 4.21) (Jurek et al., 2009).

Based on the results of the present study, exposure-outcome effects, applying the misclassification model, were toward the null, and the adjusted ORs for misclassification in different priors were 2.64, 1.60, 1.61, 1.57, and 1.58. Also, in all other situations, except for a situation, the corrected odds ratios for misclassification were less than the ORs derived from the model without misclassification. In the exception situation, OR was more than that derived from the model without misclassification within which the priors of the measurement model had uniform distributions and the priors of the outcome model were selected from previous literature and the prior variances of the exposure model were considered to be 1.

In the present study, in no misclassification models, credible intervals were wider than the uncertainty limits found in Jurek et al (Jurek et al., 2009) study. In their study, uncertainty limits width was 0.65 but in our study, CRI width for previous literature priors, 1000 and 1.38 prior variances were 1.30, 1.46 and 1.41, respectively. This may be due to the more sample size in Jurek et al. (Jurek et al., 2009) study. However, in the misclassification model of the present study, all of CRIs were narrower than uncertainty limits found in Jurek et al. (Jurek et al., 2009) study.

Finally, due to the lack of validation studies in Iran, more attention should be paid to this matter in future

research, especially to sensitivity and specificity. As different diagnostic tests for breast cancer do not have the same accuracy at different ages, it is suggested to use bias analysis for correcting outcome misclassification. Also, it is recommended to consider the other sources of bias in future studies, and then it could be claimed that the measure of association obtained are closer to the causal measure.

Acknowledgements

This research was supported by Tehran University of Medical Sciences

References

- Akbari A, Razzaghi Z, Homae F, et al (2011). Parity and breastfeeding are preventive measures against breast cancer in Iranian women. *Breast Cancer*, **18**, 51-5.
- Chu H, Wang Z, Cole S, et al (2006). Sensitivity analysis of misclassification: A graphical and a Bayesian approach. *Ann Epidemiol*, **16**, 834-41.
- Ebrahimi M, Vahdaninia M, Montazeri A (2002). Risk factors for breast cancer in Iran: a case-control study. *Breast Cancer Res*, **4**, R10.
- Fox M, Lash T, Greenland S (2005). A method to automate probabilistic sensitivity analyses of misclassified binary variables. *International J Epidemiol*, **36**, 1370-6.
- Gelman A, Carlin JB, Stern HS, et al 2003. *Bayesian Data Analysis*, London, CRC Press.
- Gelman A, Rubin DB (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457-511.
- Ghiasvand R, Bahmanyar S, Zendehdel K, et al (2012). Postmenopausal breast cancer in Iran; risk factors and their population attributable fractions. *BMC Cancer*, **12**, 414.
- Ghiasvand R, Maram ES, Tahmasebi S, et al (2011). Risk factors for breast cancer among young women in southern Iran. *International J Cancer*, **129**, 1443-9.
- Greenland S (2005). Multiple-bias modelling for analysis of observational data. *J Royal Statistical Society, A*, **168**, 1-25.
- Greenland S (2008). Invited commentary: variable selection versus shrinkage in the control of multiple confounders. *Am J Epidemiology*, **167**, 523-9.
- Greenland S (2009). Bayesian perspectives for epidemiologic research: III. Bias analysis via missing-data methods. *International J Epidemiol*, **38**, 1662-73.
- Greenland S, Lash T (2008). Bias analysis. In 'Modern Epidemiology', Eds Lippincott-Williams-Wilkins, Philadelphia, 345-80
- Greenland S, Mansournia MA (2015). Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Statistics in Medicine*, **34**, 3133-43.
- Hamra G, MacLehose R, Richardson D (2013a). Marcov Chain Monte Carlo: an introduction for epidemiologists. *International J Epidemiol*, **42**, 627-34.
- Hamra GB, MacLehose RF, Cole SR (2013b). Sensitivity analyses for sparse-data problems-using weakly informative bayesian priors. *Epidemiology (Cambridge, Mass.)*, **24**, 233-9.
- Hassanzadeh J, Moradzadeh R, Rajae fard A, et al (2012). A comparison of case-control and case-only designs to investigate gene-environment interactions using breast cancer data. *Iranian Journal of Medical Sciences*, **37**, 112-8.
- Holakouie-Naieni K, Ardalan A, Mahmoodi M, et al (2007). Risk factors of breast cancer in North of Iran: A Case-Control in Mazandaran Province. *Asian Pac J Cancer Prev*, **8**, 395-8.
- Hosseinizadeh M, Eivazi Ziaei J, Mahdavi N, et al (2014). Risk factors for breast cancer in Iranian women: a hospital-based case-control study in tabriz, Iran. *J Breast Cancer*, **17**, 236-43.
- Jurek AM, Lash TL, Maldonado G (2009). Specifying exposure classification parameters for sensitivity analysis: family breast cancer history. *Clinical Epidemiology*, **1**, 109-17.
- Keil AP, Daniels JL, Hertz-Picciotto I (2014). Autism spectrum disorder, flea and tick medication, and adjustments for exposure misclassification: the CHARGE (CHildhood Autism Risks from Genetics and Environment) case-control study. *Environmental Health*, **13**, 3.
- Lash T, Fink A (2003). Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology*, **14**, 451-8.
- Lash TL, Fox MP, Fink AK 2009. Applying quantitative bias analysis to epidemiologic data, berlin, Germany, Springer.
- Lash TL, Fox MP, MacLehose RF, et al (2014). Good practices for quantitative bias analysis. *International Journal of Epidemiology*, **43**, 1969-85.
- MacLehose RF, Gustafson P (2012 January). Is probabilistic bias analysis approximately Bayesian? *Epidemiology*, **23**, 151-8.
- MacLehose RF, Olshan AF, Herring AH, et al (2009). Bayesian Methods for Correcting Misclassification: An Example from Birth Defects Epidemiology. *Epidemiology*, **20**, 27-35.
- Mahouri K, Dehghani-Zahedani M, Zare S (2007 Nov-Dec). Breast cancer risk factors in south of Islamic Republic of Iran: a case-control study. *East Mediterr Health J*, **13**, 1265-73.
- Marrie RA, Cutter G, Tyry T, et al (2008). Smoking status over two years in patients with multiple sclerosis. *Neuroepidemiology*, **32**, 72-9.
- Prescott GJ, Garthwaite PH (2005). Bayesian analysis of misclassified binary data from a matched case-control study with a validation sub-study. *Statistics in Medicine*, **24**, 379-401.
- Rothman KJ, Greenland S, Lash TL (2008). Validity in epidemiologic studies. In 'modern epidemiology', eds lippincott williams and wilkins, Philadelphia, 128-47
- Sepandi M, Akrami M, Tabatabaee H, et al (2014). Breast Cancer Risk Factors in Women Participating in a Breast Screening Program: a Study on 11,850 Iranian Females. *Asian Pac J Cancer Prev*, **15**, 8499-502.
- Szatmari P, Jones MB (1999). Effects of misclassification on estimates of relative risk in family history studies. *Genetic Epidemiology*, **16**, 368-81.
- Tehranifar P, Wu H-C, Shriver T, et al (2015). Validation of family cancer history data in high-risk families: the influence of cancer site, ethnicity, kinship degree, and multiple family reporters. *American Journal of Epidemiology*.
- van Gelder MMHJ, Donders ART, Devine O, et al (2014a). Using bayesian models to assess the effects of under-reporting of cannabis use on the association with birth defects, national birth defects prevention study, 1997-2005. *Paediatric and Perinatal Epidemiology*, **28**, 424-33.
- van Gelder MMHJ, Donders ART, Devine O, et al (2014b). Using bayesian models to assess the effects of under-reporting of cannabis use on the association with birth defects, national birth defects prevention study, 1997-2005. *Paediatric and Perinatal Epidemiology*, n/a-n/a.
- Veisy A, Lotfinejad S, Salehi K, et al (2015). Risk of breast cancer in relation to reproductive factors in north-west of Iran, 2013-2014. *Asian Pac J Cancer Prev*, **16**, 451-5.
- Zare N, Haem E, Lankarani KB, et al (2013). Breast cancer risk factors in a defined population: weighted logistic regression approach for rare events. *J Breast Cancer*, **16**, 214-9.