Open Access Maced J Med Sci **electronic publication ahead of print,**
published on July 17, 2018 as https://doi.org/10.3889/oamjms.2018.296

ID Design Press, Skopje, Republic of Macedonia
Open Access Macedonian Journal of Medical Sciences.
https://doi.org/10.3889/oamjms.2018.296
eISSN: 1857-9655
Basic Science

A Bayesian Analysis With Informative Prior on Disease Prevalence for Predicting Missing Values Due To Verification Bias

Abdollah Hajivandi[1], Hamid Reza Ghafarian Shirazi[2*], Seyed Hassan Saadat[3], Mohammad Chehrazi[4]

[1]Bushehr University of Medical Sciences, Bushehr, Iran; [2]Social Determinant of Health Research Center, Yasuj University of Medical Sciences, Yasuj, Iran; [3]Behavioral Sciences Research Center, Lifestyle Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran; [4]Department of Epidemiology and Reproductive Health, Reproductive Epidemiology Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran

Citation: Hajivandi A, Ghafarian Shirazi HR, Saadat SH, Chehrazi M. A Bayesian Analysis with Informative Prior on Disease Prevalence for Predicting Missing Values Due To Verification Bias. Open Access Maced J Med Sci. https://doi.org/10.3889/oamjms.2018.296

Keywords: Bayesian inference; Verification Bias; Informative Prior; MCMC Simulation; missing data

***Correspondence:** Hamid Reza Ghafarian Shirazi. Social Determinant of Health Research Center, Yasuj University of Medical Sciences, Yasuj, Iran. E-mail: gshr3@yahoo.com

Received: 01-May-2018; **Revised:** 21-Jun-2018; **Accepted:** 03-Jul-2018; **Online first:** 17-Jul-2018

Funding: This research was finantialy supported by the Royan Institute

Competing Interests: The authors have declared that no competing interests exist

Abstract

AIM: Verification bias is one of the major problems encountered in diagnostic accuracy studies. It occurs when a standard test performed on a non-representative subsample of subjects which have undergone the diagnostic test. In this study we extend a Bayesian model to correct this bias.

METHODS: The study population is patients that have undergone at least two repeated failed IVF/ICSI (in vitro fertilization/intra cytoplasmic sperm injection) cycles. Patients were screened using ultrasonography and those with polyps were recommended for hysteroscopy. A Bayesian modeling was applied on mechanism of missing data using an informative prior on disease prevalence. The parameters of the model were estimated through Markov Chain Monte Carlo methods.

RESULTS: A total of 238 patients were screened, 47 of which had polyps. Those with polyps were strongly recommended to undergo hysteroscopy, 47/47 decide to have a hysteroscopy and in 37/47 polyps confirmed. None of the 191 patients with no polyps detected in ultrasonography underwent a hysteroscopy. A model using Bayesian approach was applied with informative prior on polyp prevalence. False and true negatives were estimated in the Bayesian framework. The false negative was obtained 14 and 177 true negatives were obtained, so sensitivity and specificity was estimated easily after estimating the missing data. Sensitivity and specificity were equal to 74% and 94% respectively.

CONCLUSION: Bayesian analyses with informative prior seem to be powerful tools in the simulation of experimental space.

Introduction

Advances in medical technology have provided doctors with the various ways of diagnostic methods to identify patients. Diagnostic accuracy studies help physicians in selecting the most appropriate test to evaluate the patient's clinical situation and decision making about treatment by examining the characteristics of these tests. The sensitivity and specificity of diagnostic tests are assessed by comparing the results of evaluated tests with result of the standard test that is performed on the same patients. Generally the most accurate test for diagnosis is considered as the gold standard. For evaluation of the diagnostic test in best situation both tests are performed for all study subjects. One of the problems often encountered in these studies is that due to it being costly or invasive. The standard test is not performed on all the study subjects. This seems reasonable in the clinic setting but in a diagnostic test study it creates verification bias [1]. Verification bias is a selection bias and occurs when the standard test performed on a non-representative subsample of study subjects which have already undergone the diagnostic test. This occurs, for example, when inclusion probabilities for the subsample are

dependent on the initial stage results and/or on a covariate related to the disease status. This bias is usually divided into two types: differential bias and partial bias. Partial verification bias occurs when standard test is only performed on patients who have a positive diagnostic test result. Differential verification bias occurs when standard test which perform on patients with positive test result, is different to the standard test that is performed on patients who have a negative result [1] [2]. In most studies, there is no result for gold standard test for some people and any kind of verification bias are considered from the viewpoint of missing data. So there is no result for gold standard test for some people. Models that are built based on this view are dependent on the process of missing data. Begg and Greenes were the first researchers who tried to correct the partial verification bias [3]. They used the diseased proportion of the sample which had been used to perform the gold standard test on. Their models were based on the conditional independency assumption; this assumption means that missing process or sample selection to done gold standard only dependent on the test result and not on the true disease status. On the other hand, missing event of a sample given the test result is independent of true disease status. Some researchers think this assumption is unrealistic and in some cases this can be misleading [4] [5]. But Kosinski [6] tried to overcome this problem by considering that in addition to the test result, the missing process was dependent on the true disease status and unobserved information about the disease [5] [7]. Obviously, having the full data we could compare the different models in predicting the results of the gold standard test. Estimation of the models was based on Frequentist and maximum likelihood approach. Problems in these methods are described in Martinez and Buzoianu [5] [7]. It seems that problems of these methods are due to a new variable called v which is defined as the missing process. In this article we have covered the missing process and problems associated with it which have not been studied before. We eliminated the variable V using a new definition and fitted a new Bayesian model that is not dependent on V. In section 2, we proposed the detailed model in the Bayesian framework to estimate sensitivity and specificity. In section 3, a real data example is described and the Bayesian model is applied and the results are presented. Finally, In Section 4 we discuss the model and its results.

Let that missing process occurs simultaneously with the result of the diagnostic test [7]. On the other hand, if the test result was positive, the gold standard test should be performed but if the result was negative the gold standard test is not performed. The test process is depicted below:

$$Test\ Result \begin{cases} positive \{ Gold\ standard \\ negative\ \{ Miss \end{cases}$$

Therefore, the probability of missing is zero for a positive test result and 1 for a person with a negative result. If T represents the diagnostic test result and V represents the missing process, the following probability exists.

$$P(V|T) = \begin{cases} 1\ V \neq T \\ 0\ V = T \end{cases}$$

$$P(V|T) = \frac{P(V,T)}{P(T)} = 1 \xrightarrow{yields} P(V,T) = P(T) \quad (1)$$

If G represents the gold standard result then the aim is to find a function to predict the gold standard status for individuals with a negative diagnostic test result. Thus, sensitivity and specificity could be estimated. The model is started with following probability:

$$P(G|T,V) = \frac{P(G,T,V)}{P(T,V)} = \frac{P(T)\,P(G|T)P(V|G,T)}{P(T,V)} \quad (2)$$

Where according to the definition 1, probability [2] exists in the following form:

$$P(G|T,V) = \frac{P(T)P(G|T)P(V|G,T)}{P(T)} = P(G|T)P(V|G,T) \quad (3)$$

As we considered (1) T and V to occur simultaneously, information about the events T and G gives full information about V. therefore the relationship [3] can be rewritten as bellow:

$$P(G|T,V) = P(G|T)$$

So, it can be concluded that the missing process is independent of the disease status given the diagnostic test status (it can be assumed that the disease status will be clarified by the gold standard test), but the disease status has direct impact on the missing process. For example, in someone who is suffering from a disease, the probability of a positive diagnostic and the missing probability are low. Table 1 is a cross table which illustrates the diagnostic test and gold standard frequencies.

Table 1: Diagnostic test and gold standard result

	Positive Gold standard	Negative Gold standard
Positive test	y_{11}	y_{12}
Negative test	y_{21}	y_{22}

Probability $P(G|T)$could be obtained by the following:

$$P(G^+|T^+) = \frac{P(T^+|G^+)P(G^+)}{P(T^+|G^+)P(G^+) + (1 - P(T^-|G^-))P(G^-)} = P_{11}$$

$$P(G^+|T^-) = \frac{(1 - P(T^+|G^+))P(G^+)}{P(T^-|G^-)P(G^-) + (1 - P(T^+|G^+))P(G^+)} = P_{21}$$

$$P(G^-|T^+) = \frac{(1-P(T^-|G^-))P(G^-)}{P(T^+|G^+)P(G^+) + (1-P(T^-|G^-))P(G^-)} = P_{12}$$

$$P(G^-|T^-) = \frac{P(T^-|G^-)P(G^-)}{(1-P(T^+|G^+))P(G^+) + (P(T^-|G^-))P(G^-)} = P_{22}$$

Where, $P(T^+|G^+)$ and $P(T^-|G^-)$ are the sensitivity (SN) and specificity (SP) of the diagnostic test respectively. A reasonable initial model is to assume that the number of each cell $\{y_{ij}\}_{i,j=1,2}$ has the following distribution

$$y_{ij} \sim Binomial(P_{ij}, n_i), \sum_i n_i = N \;\; i,j = 1,2$$

$$P_{11} + P_{12} = 1$$

$$P_{21} + P_{22} = 1$$

$$SN = \frac{y_{11}}{P(G^+)N} \quad (4)$$

$$SP = \frac{[P(G^-)N] - y_{12}}{P(G^-)N} \quad (5)$$

$$P(G = 0) = 1 - P(G = 1)$$

$$P(G = 1) \sim beta(a, b)$$

Therefore, the posterior probability function for disease prevalence is:

$$P(P(G^+)|y_{ij}) \propto \binom{n_i}{y_{ij}} P_{ij}^{\,y_{ij}} (1-P_{ij})^{(n_i - y_{ij})} \frac{1}{B(a,b)} P(G^+)^{a-1}(1-P(G^+))^{b-1} \quad (6)$$

Proof of the formulas is given in Appendix A.

Repeated IVF failures are one of the problems of infertility centers. Structural abnormalities in uterine cavity such as fibroids and mullerian anomalies can play an important role in the failure of embryo implantation during IVF cycles [8]. Repair of uterine cavity pathologies has been suggested as a therapeutic action to improve the results of ART cycles in these individuals. Hysteroscopy is considered as a gold standard test for evaluating the uterine cavity in infertile patients. This method allows direct observation of the uterus and cervix, Thus Increasing the accuracy of diagnosis. This can also highlight uncertain results of other diagnostic methods [9]. Uterine anomalies are usually well diagnosed with hysteroscopy [10]. A number of studies reported that sensitivity, specificity, positive predictive value and negative predictive value of transvaginal sonography is similar to the results of hysteroscopy [8]. Some studies also found no strong correlation between the result of transvaginal sonography and hysteroscopy [11]. It is recommended that all women undergoing IVF candidates, before doing IVF placed under hysteroscopy [12]. But given that hysteroscopy is an invasive procedure, maybe it is not performed on all patients who assessed using transvaginal sonography, and done only for patients who have had positive results of transvaginal sonography. Consequently verification bias occurs.

At First we have analyzed the data obtained from 140 patients admitted to the Royan Institute, and transvaginal sonography and hysteroscopy was performed for every one of them, these patients have at least two Repeated IVF/ICSI cycles which have failed. With regard to Hysteroscopy as the gold standard, we calculated the sensitivity and specificity of vaginal sonography in detecting polyps with the frequentist method, then with assuming that Hysteroscopy has been performed only on patients who have positive results of vaginal sonography. We estimated the sensitivity and specificity of vaginal sonoghraphy with Bayesian approach and compared these estimations with sensitivity and spasticity of the actual data.

Since there was expert belief about disease prevalence but not about sensitivity and specificity, we therefore formulated those in terms of disease prevalence. Furthermore, to account for the uncertainty of prevalence, we used the Beta distributions as informative. Hyper-parameters $(a \; and \; b)$ were determined by subjective percentiles technique [13]. The prior information was provided by a gynecologists and a midwife independently. Each expert provided the best guess and a 90% prior credible interval for the prevalence. In particular we wanted upper boundaries and lower bound, without consulting the Literature, of the true value of the polyp prevalence in the population. The across-expert average of these quantities is listed below:

$$P(p < 0.33) = 0.95$$

$$P(p < 0.001) = 0.05$$

The averaged credible interval for prevalence is used to determine the hyperparameters a and b.

$$p \sim beta(0.5981227, 5.610896)$$

In the present setting, for instance, a Monte Carlo sample of 100000 draws from the prior which gives $P(0.001 < p < 0.33) \approx 0.9$, so the prior information is being represented as desired.

Prior distribution of prevalence must be truncated from 0.15 because generated numbers below this cut point cause sensitivity and specificity to become negative. Therefore, the truncated beta distribution has been truncated between 0.15 and 1. There is no truncated distribution in OpenBugs; the I (.,.) operator is used only to denote censored observation [14]. However, when all parameters in a prior distribution are observed, the I (.,.) operator can be used for modeling truncated prior distributions. If there are unknown parameters the inferences will be wrong [15]. OpenBugs remove this ambiguity between truncation and censoring by introducing the truncation operator T (.,.) [16].

Therefore, the prior density for $\{p\}_{T(0.15,1)}$ is

$$P(p) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}$$

We apply the model to real data that was introduced previously. In total 238 patients were undergoing ultrasound tests for the diagnosis of polyps. 191 people had negative test results and 47 had positive results. The gold standard test was then done for two groups. The results are showed in Table 2.

Table 2: Results of diagnostic and gold standard tests

	G^+	G^-
T^+	**37**	**10**
T^-	**13**	**178**

Thus, the sensitivity, specificity, positive and negative predictive value is easily calculated as follows:

$$SN = \frac{37}{50} = 0.74$$

$$SP = \frac{178}{188} = 0.946$$

$$PPV = \frac{37}{47} = 0.787$$

$$NPV = \frac{178}{191} = 0.932$$

But the problem will be started when the gold standard test is not performed on people with negative diagnostic test results. So Table 2 comes in the form below:

Table 2: A contingency table that depict missing mechanism

	G^+	G^-
T^+	**37**	**10**
T^-	***NA***	**NA**

NA: Not Available.

In Bayesian approach, model [6] is posterior density function of disease prevalence. The parameter of the model [1] was estimated using simulation from the posterior distribution. The Markov Chain Monte Carlo (MCMC) method using Adaptive Metropolis Block was applied [13]. Algorithm was run in OpenBugs 3.1.2 environment [14] [15]. After 10000 times iteration the chain was converged to posterior distribution and 5000 initial samples were discarded as Burn-in period. Graphical methods such as autocorrelation function, posterior density and trace plot were used for checking convergence of chains [16] (not shown). The program was run in OpenBugs given in Appendix 2. Bayesian Results after sampling (simulation) were shown in Table 3.

Table 3: Estimation of parameters simulated by MCMC algorithm in Bayesian framework

Parameters	Posterior mean	Standard deviation	2.5 credible interval	97.5 credible interval	Start	Sample
Sensitivity	0.7417	**0.1092**	0.5139	0.925	**5001**	**10000**
Specificity	0.9272	0.0411	0.8289	0.9842	**5001**	**10000**
PPV	**0.7858**	0.0586	0.6578	0.8868	**5001**	**10000**
NPV	0.9269	0.03828	0.83	0.9783	**5001**	**10000**

Discussion

Verification bias is a type of selection bias that occurs when the standard test performed on a non-representative subsample of study subjects which have already undergone the diagnostic test. In most studies there is no result for gold standard test for some people. There are several ways to solve this problem. For example we assume that results of the gold standard test are negative for all these people. This approach is clinically inappropriate. Since there is a possibility of error in diagnostic test and consequently negative predictive value may not be 1. This method greatly increased the sensitivity and specificity values and therefore overestimation can occur. Another way is results of previous tests for people who have had the gold standard test to be replaced. The estimates from this strategy could be subject to bias, because the conditions governing the present situation may be different from any of the previous experiments. In this paper a new statistical model was presented based on conditional probabilities where the model parameters, sensitivity and specificity, was estimated from the Bayesian approach [17]. The conditional probabilities are the same positive and negative predictive values which provide information about the model. The data missing process is different from other corresponding studies, meaning the gold standard results do not exist for individuals with a negative test. The assumption of conditional independence that was used by Begg and Greenes is quite applicable in case of missing process of current data. When a patient's diagnostic test is specified then the missing state is determined automatically. Thus the disease status is independent of the missing process given knowing the diagnostic test result. On the other hand, the actual status of the disease increases or decreases the chances of missing. So the model can be defined based on conditional probabilities $P(G \mid T)$. To construct better models and predict the missing values more accurately, information must enter the study according to an expert's opinion because the probability distribution does not give enough information about the missing values. This could be done with considering the disease prevalence in the population as uncertain and attributing the probability distribution to it. This can give useful information to

the model about the overall distribution of the gold standard. The proposed model is for diagnostic test ultrasound against the gold standard hysteroscopy in detection of polyps. But whether or not it can be applied to other tests, should be studied to allow the estimation of the parameters, after the probability of disease combined with the distribution of diagnostic tests. Finally, we hope stronger and more powerful Models in the development of diagnostic tests can be developed in the future.

Appendix A:

Proofs of formula (4) and (5):

$$\frac{y_{11}+y_{21}}{N} = P(G=1)$$

$$SN = \frac{y_{11}}{y_{11}+y_{21}} = \frac{y_{11}}{N.P(G=1)}$$

$$\frac{y_{12}+y_{22}}{N} = P(G=0)$$

$$SP = \frac{y_{22}}{y_{12}+y_{22}} = \frac{(N.P(G=0)-y_{12})}{N.P(G=0)}$$

Appendix B: Openbugs codes

In order to run the model with informative beta distribution for prevalence of disease the following code was used:

```
model{
PG1~dbeta(2,12) T(0.1541400874,1)
SN<-(0.7858*10)/(0.2142*238*PG1)
SP<-((238*(1-PG1))-10)/(238*(1-PG1))
for(i in 1:a){
for(j in 1:b){
y[i,j]~dbin(p[i,j],n[i])
}}
p[1,1]<-(SN*PG1)/((SN*PG1)+((1-SP)*(1-PG1)))
p[2,2]<-(SP*(1-PG1))/((SP*(1-PG1))+((1-SN)*PG1))
p[2,1]<-((1-SN)*PG1)/(((1-SN)*PG1)+(SP*(1-PG1)))
p[1,2]<-((1-SP)*(1-PG1))/(((1-SP)*(1-PG1))+(SN*PG1))
sensitivity<-y[1,1]/(y[1,1]+y[2,1])
specificity<-y[2,2]/(y[2,1]+y[2,2])
}
```

Acknowledgements

We thank Royan Institute for financial support and Dr. Kazem Mohammad for their helpful comments which helped improve this paper.

References

1. Whiting P, Rutjes AWS, Reitsma JB, Glas AS, Bossuyt PMM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy. Annals of Internal Medicine. 2004; 140(3):189. https://doi.org/10.7326/0003-4819-140-3-200402030-00010 PMid:14757617

2. Lu Y, Dendukuri N, Schiller I, Joseph L. A Bayesian approach to simultaneously adjusting for verification and reference standard bias in diagnostic test studies. Stat Med. 2010; 29(24):2532-43. https://doi.org/10.1002/sim.4018 PMid:20799249

3. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics. 1983; 207-15. https://doi.org/10.2307/2530820 PMid:6871349

4. de Groot JA, Janssen KJ, Zwinderman AH, Bossuyt PM, Reitsma JB, Moons KG. Correcting for partial verification bias: a comparison of methods. Annals of epidemiology. 2011; 21(2):139-48. https://doi.org/10.1016/j.annepidem.2010.10.004 PMid:21109454

5. Martinez EZ, Alberto Achcar J, Louzada-Neto F. Estimators of sensitivity and specificity in the presence of verification bias: A Bayesian approach. Computational statistics & data analysis. 2006; 51(2):601-11. https://doi.org/10.1016/j.csda.2005.12.021

6. Kosinski AS, Barnhart HX. Accounting for nonignorable verification bias in assessment of diagnostic tests. Biometrics. 2003; 59(1):163-71. https://doi.org/10.1111/1541-0420.00019

7. Buzoianu M, Kadane JB. Adjusting for verification bias in diagnostic test evaluation: a Bayesian approach. Statistics in medicine. 2008; 27(13):2453-73. https://doi.org/10.1002/sim.3099 PMid:17979150

8. Ayida G, Chamberlain P, Barlow D, Kennedy S. Uterine cavity assessment prior to in vitro fertilization: comparison of transvaginal scanning, saline contrast hysterosonography and hysteroscopy. Ultrasound Obstet Gynecol. 1997; 10(1):59-62. https://doi.org/10.1046/j.1469-0705.1997.10010059.x PMid:9263425

9. Crosignani PG, Rubin BL. Optimal use of infertility diagnostic tests and treatments. The ESHRE Capri Workshop Group. Hum Reprod. 2000; 15(3):723-32. https://doi.org/10.1093/humrep/15.3.723 PMid:10686227

10. Seinera P, Maccario S, Visentin L, DiGregorio A. Hysteroscopy in an Ivf-Er Program: Clinical experience with 360 infertile patients. Acta obstetricia et gynecologica Scandinavica. 1988; 67(2):135-7. https://doi.org/10.3109/00016348809004185 PMid:3176927

11. Golan A, Eilat E, Ron el R, Herman A, Soffer Y, Bukovsky I. Hysteroscopy is superior to hysterosalpingography in infertility investigation. Acta obstetricia et gynecologica Scandinavica. 1996; 75(7):654-6. https://doi.org/10.3109/00016349609054692 PMid:8822660

12. Golan A, Ron-El R, Herman A, Soffer Y, Bukovsky I, Caspi E. Diagnostic hysteroscopy: its value in an in-vitro fertilization/embryo transfer unit. Human Reproduction. 1992; 7(10):1433. https://doi.org/10.1093/oxfordjournals.humrep.a137589 PMid:1291572

13. Liu J, Gustafson P, Cherry N, Burstyn I. Bayesian analysis of a matched case–control study with expert prior information on both

the misclassification of exposure and the exposure–disease association. Statistics in medicine. 2009; 28(27):3411-23. https://doi.org/10.1002/sim.3694 PMid:19691019

14. Spiegelhalter D, Thomas A, Best N, Lunn D. WinBUGS Version 1.4 User Manual, 2005.

15. Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: Evolution, critique and future directions. Statistics in medicine. 2009; 28(25):3049-67. https://doi.org/10.1002/sim.3680 PMid:19630097

16. Curtis SM. BUGS code for item response theory. Journal of Statistical Software. 2010; 36(1):1-34.

17. Gelman A, Carlin J, Stern H, Rubin DB. Bayesian data analysis. New York: Chapman and Hall, 1995.