ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH

DOI: 10.5336/biostatic.2018-62212

Bayesian Meta-Analysis of Prevalence: Alzheimer's Disease in Europe

Prevalansın Bayesci Meta-Analizi: Avrupa'da Alzheimer Hastalığı

Esin AVCI[a]

[a]Department of Statistics,
Giresun University
Faculty of Art and Science,
Giresun, TURKEY

Received: 29.08.2018
Received in revised form: 18.10.2018
Accepted: 18.10.2018
Available Online: 07.12.2018

Correspondence:
Esin AVCI
Giresun University
Faculty of Art and Science,
Department of Statistics, Giresun,
TURKEY/TÜRKİYE
esinavci@hotmail.com

This study was presented orally at 4th International Researchers, Statisticians and Young Statisticians Congress, 28-30 April 2018, İzmir, Turkey.

ABSTRACT Objective: Statistically, prevalence is defined as the frequency of disease at a given time in a particular population. Meta-analysis can be a useful in estimating prevalence more precisely. Meta-analysis is a statistical method that combines the results of studies on a determined topic in order to derive an average estimate. In medical and statistical literatüre, the meta-analysis of prevalence is often considered to be a frequentist rather than an Bayesian approach. Frequentist meta-analysis uses two main statistical models: fixed-effect and random-effect models. Under the fixed-effect model, the effect size is assumed to be the same for all studies, whereby all differences in the observed effects are due to sampling errors. By contrast, under the random-effects model the true effect may change from study to study. In Bayesian meta-analysis, one has to gather the data from the selected studies, choose informative or non-informative prior, model the posterior, and run simulations in order to assess parameters of the posterior distribution. In addition to direct probability statements on different scales and predictions, the conflict between fixed- and random-effects meta-analysis are handled through the Bayesian approach. **Material and Methods:** The frequentist and Bayesian meta-analysis applied to data obtained from Niu et al. (2017) in order determine the prevalence of Alzheimer 's disease in Europe. **Results:** The Bayesian approach gave a narrower confidence interval and smaller relative error than the frequentist approach. **Conclusion:** More accurate prevalence estimates are derived from the Bayesian approach.

Keywords: Meta-analysis; Bayesian approach; prevalence; Alzheimer's Disease

ÖZET Amaç: İstatistiksel olarak prevalans, belirli bir popülasyonda belirli bir zamanda hastalık sıklığı olarak tanımlanır. Prevalansı daha kesin tahmin etmek için meta-analiz yararlı bir yöntem olabilir. Meta-analiz, ortalama bir tahminde bulunmak için belirli bir konu üzerine yapılan çalışmaların sonuçlarını birleştiren istatistiksel bir yöntemdir. Medikal ve istatistiksel literatürde prevelansın meta-analizi genellikle Bayes yaklaşımdan çok frekanscı yaklaşımda ele alınmaktadır. Frekanscı meta-analizi, sabit ve rastgele etki modeli olmak üzere iki ana istatistik modeli kullanmaktadır. Sabit etki modeli altında, etki büyüklüğünün tüm çalışmalar için aynı olduğu ve gözlemlenen etkilerdeki tüm değişimlerin örnekleme hatasından kaynaklandığı varsayılmaktadır. Rastgele etkiler modeli altında ise gerçek etki, çalışmadan çalışmaya değiştiği varsayılmaktadır. Bayesci meta analizinde, seçilen çalışmalardan elde edilen sonuçların toplanması, bilgilendirici veya bilgilendirici olmayan önsellerin seçilmesi ve sonsal dağılımın modellenmesi ile parametrelerin değerlendirilmesi gerekmektedir. Bayesci yaklaşım, farklı etkiler ve tahminler için doğrudan olasılık ifadelerinin yanı sıra, sabit ve rastgele etki model arasında seçim yapmayı ortadan kaldırmaktadır. **Gereç ve Yöntemler:** Avrupa'da Alzheimer Hastalığı prevalansını belirlemek için, Niu ve ark. (2017) çalışmasından elde ettiği verilere frekansçı ve Bayesci yaklaşım ile meta-analizi uygulanmıştır. **Bulgular:** Bayesci yaklaşım frekanscı yaklaşımdan daha dar bir güven aralığı ve küçük göreceli bir hata vermiştir. **Sonuç:** Daha kesin prevalans tahminleri Bayes yaklaşımından türetilmiştir.

Anahtar Kelimeler: Meta-analizi; Bayes yaklaşımı; prevelans; Alzheimer hastalığı

Meta-analysis aimes to combine and summarize various, mutually exclusive studies. One of the advantages of meta-analysis is to enlarge the sample size.[1] Meta-analysis was first applied in 1904 by Karl Pearson to synthesize several independent typhoid vaccine studies.[2] Scientists had developed meta-analysis after the 1980s, whereby meta-analysis became a statistical technique.[3,4,5]

Recently, meta-analysis has been used to estimate the pooled prevalence of depressive symptoms in patients with chronic obstructive pulmonary disease,[6] and with HIV among high school and college students in China.[7] Furthermore, Anchala et al.[8] estimated the prevalence of hypertension in India, Chen et al.[9] determined the prevalence of coinfection with either hepatitis C or B in patients infected with HIV, and Jayawardena et al.[10] discussed the prevalence of pre-diabetes and diabetes in South Asia. Studies on Bayesian meta-analysis of prevalence have increased over the past five years. Hussain et al.[11] estimated the local prevalence of hepatitis B in Pakistan, Liu et al.[12] determined the prevalence of hepatitis B infection rates among Chinese volunteer blood donors, Song et al. estimated the prevalence of vitamin A deficiency in Chinese children, Maheu-Giroux et al. estimated the lifetime and point prevalence of vaginal fistula symptoms in 19 sub-Saharan Africa countries and Xin et al. determined the prevalence of schistosomiasis japonica in the lake regions of China's Hubei province.[13-15]

Frequentist and Bayesian meta-analysis have been applied to data obtained from Niu et al. (2017) in order to determine the prevalence of Alzheimer's disease in Europe. The analysis was carried out using R software. "meta" and "MCMCpack" packages used in order to perform both frequentist and Bayesian approach of meta-analysis, respectively.

MATERIAL AND METHODS

Statistically, prevalence is defined as the frequency of disease at a given time in a particular population. Meta-analysis can be a useful in estimating prevalence more precisely.[16] In medical and statistical literature meta-analysis of prevalence is often considered in frequentist approach rather than Bayesian approach.

Frequentist meta-analysis uses two main statistical models: fixed-effect and random-effect model. The fixed-effect model assumes that the effect size is the same for all studies, and that the observed difference between effects are due to sampling error. The random-effects model assumes the effect size may vary from study to study.[1]

One of the measures of heterogeneity that is not sensitive to either the scale of the effect size and the number of studies is, I^2 wherby, I^2 is the proportion of true heterogeneity to total variation in observed effects.[1]

A forest plot displays the results of the individual studies as well as the overall estimate from the meta-analysis. Each horizontal line on a forest plot represents an individual study with the result plotted as a box, and with the 95% confidence interval of the result displayed as the line.[1]

The Bayesian approach is based on Bayes' theorem on conditional probabilities. After developing high quality statistical software packages, the mathematical and computational difficulty was simplified using Markov Chain Monte Carlo (MCMC) methods. The main difference between the Bayesian and frequentist approach is that the Bayesian does not solely depend on likelihood. It also combines initial beliefs or knowledge (prior) with information from the data (likelihood) and produces updated beliefs (posterior).[17] Compared to frequentist methods, Bayesian methods offers a number of practical advantages.[18,19,20] In Bayesian meta-analysis, one has to gather the data from selected studies, choose informative or non-informative prior, model the posterior, and run simulations in order to assess the parameters of the posterior distribution. In addition to direct probability expression on various scales and predictions, the conflict between fixed- and random-effects meta-analysis is handled using the Bayesian approach.[21]

Based on various statistical points of view, there are two main methodological options to estimate that can true prevalence: the Bayesian as well as the frequentist approach.

In the frequentist approach the meta-analysis of prevalence is based on the inverse variance method. The pooled proportion is an average that is calculated by weighting each study of their variance via a fixed/ random effect model.[16] The individual study weights, in terms of an equation, is as follows;

$$Var(p) = \frac{p(1-p)}{N} \quad (1)$$

where N is the size of population and p is the proportion of each study.

Based on the inverse variance method the estimates of the pooled proportion is given by,

$$P = \frac{\sum_i \frac{p_i}{Var(p_i)}}{\sum_i \frac{1}{Var(p_i)}} \quad (2)$$

with standard error,

$$SE(P) = \sqrt{\sum_i \frac{1}{Var(p_i)}} \quad (3)$$

In the Bayesian meta-analysis of prevalence, prevalence is handled as a random variable. Binomial distribution is convenient, given that the prevalence ranges from 0 to 1. The likelihood function is given by:

$$p(y|p) = \binom{n}{y} p^y (1-p)^{n-y} \quad (4)$$

For the non-informative prior distribution, Beta(α,β) distribution is the best candidate. The posterior distribution of prevalence is proportional to that of the beta distribution.[22] The posterior distribution is given by

$$p(p|y) \sim Beta(y + \alpha, n - y + \beta) \quad (5)$$

In the Bayesian meta-analysis, heterogeneity can be handled more efficiently by simultaneously incorporating variations at all levels.[23]

DATA ANALYSIS AND RESULTS

Alzheimer's disease is widespread form of dementia defined by developing memory and cognitive deterioration.[24]

The meta-analysis containes eight studies including several number of patients ranging in number from 365 to 7528. Table 1. summarizes the total number of Alzheimer's disease and of patients over eight (8) studies. Both the frequentist and Bayesian approaches were applied in order to estimate the prevalence of Alzheimer's disease in Europe (Table 2).

In frequentist meta-analysis, the DerSimonian-Laird method was used in order to calculate the inverse-variance weighting method as well as estimate the between-study variance (τ^2). The result for the random effect model was summarized using Forest plot as shown in Figure 1.

TABLE 1: Alzheimer's disease and total numbers of patients in 8 studies.

No	Study	Alzheimer's disease (event)	Total number of Alzheimer's disease (n)
1	Manubens et al. (1995)	119	1019
2	Ott et al. (1995)	339	7528
3	Prencipe et al. (1996)	50	968
4	Ferini-Strambi et al. (1997)	27	673
5	Obadia et al.(1997)	82	1068
6	Salamon et al.(1999)	128	4123
7	Tsolaki et al.(1999)	20	365
8	Tola-Arribas et al.(2013)	143	2170
	Total	**908**	**17914**

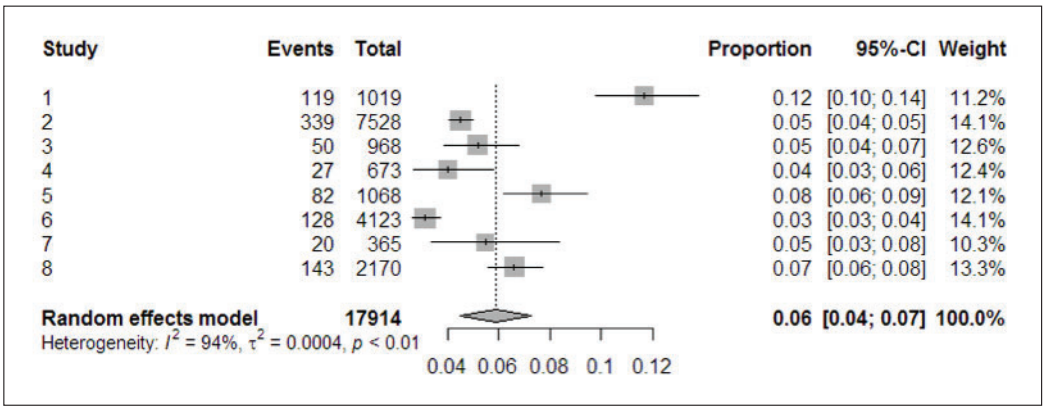

FIGURE 1: Forest plot of Prevalence of Alzheimer disease.

In Figure 1, the prevalence was %6 (95% CI, 0.04 to 0.07). Significant heterogeneity was observed between-study (Q=110.13, p<0.0001) with $I^2 = 94\%$, which means that 94% of observed variance was obtained from the real difference between studies and moreover can be explained by study-level covariates.

In the Bayesian meta-analysis, each study prevalence (p_i) is handled as a random variable and the number of Alzheimer's disease (y_i) is assumed to follow binomial distribution with n_i and p_i.

$$y_i \sim Binomial(n_i, p_i) \tag{6}$$

where n_i was the total number of participants investigated and p_i was the corresponding prevalence. In this study, 908 out of 17914 patients had Alzheimer 's disease.

The non-informative prior distributions were used in the Bayesian analysis on the assumption that there is no prior information on prevalence. Hence, the prior distribution of prevalence followed the beta(1,1) and beta(0.5,0.5) distribution (Figure 2). Beta(1,1) distribution gives equal weight to all possible values of prevalence. Beta(0.5,0.5) or Jeffrey's prior is defined in terms of Fisher information, and expresses a belief

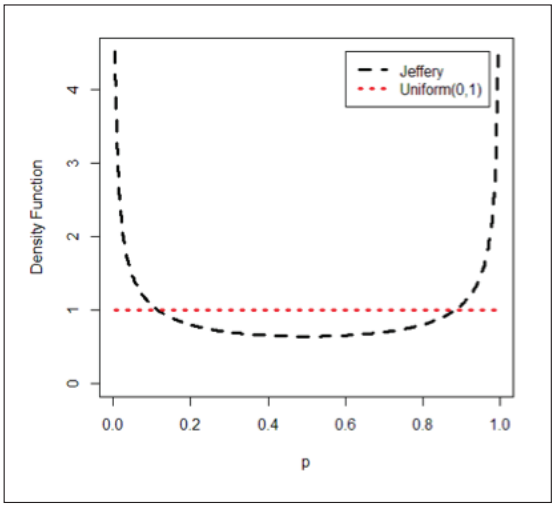

FIGURE 2: Prior distribution plot.

that true prevalence is more likely to be at either extreme end of the distribution rather than anywhere in the center.[17] For all of the parameters, non-informative prior was specified.

The Markov Chain Monte Carlo simulation was used in order to estimate the posterior distribution of prevalence. The R statistical software "MCMCpack" package was used to perform the simulation. An iterative process was carried beyond 25.000, whereby the estimate proved to be very stable. In order to reduce any potential bias, the first 2.500 samples were discarded as burn-in. Figure 3 displayed posterior densities and convergence history for prevalence. Figure 3 a and b displayed the posterior densities of prevalence obtained from beta(1,1) and beta(0.5,0.5) prior, respectively. Each prevalence was symmetric in shape in the posterior densities. The trace plots displayed nice oscillograms around a horizontal line without any trend. The Markov chain mixed well and was presumably sampling from the stationary distribution.

The median and its 95% credible intervals of the posterior distribution of prevalence were reported as summary estimates.

The relative error of each method was computed in order to assess the accuracy of the three estimates, and given in Table 2. In the frequentist approach, the pooled prevalence was estimated as %6 (95% CI, 0.04 to

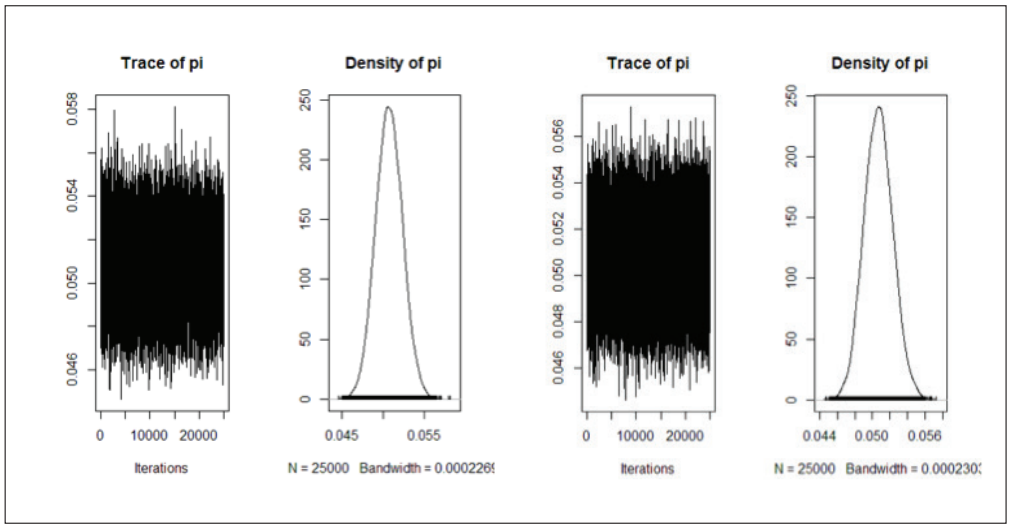

FIGURE 3: Trace and density distribution plots of prevalence: (a) Beta(1,1) prior and (b) Beta(0.5,0.5).

TABLE 2: Results of frequentist and Bayesian meta-analysis.

Method	Pooled Prevalence	LB	UB	Range	Relative Error
Frequentist	0.0589	0.0450	0.0729	0.0279*	0.16196
Bayes (Beta(1,1))	0.0507	0.04757	0.05398	0.00641	0.00059
Bayes (Beta(0.5,0.5))	0.0507	0.04755	0.05395	0.00640	0.00059

0.07). For both priors (beta(1,1) and beta(0.5,0.5)) the estimated pooled prevalence lead to similar numerical values in the Bayesian approach, %5 (95% CI, 0.047 to 0.054). The range of pooled estimates of each prevalence were obtained from both the frequentist and Bayesian approachs as 0.02, 0.00641 and 0.00064, respectively. The narrowest relative error was obtained from the Bayesian approach.

CONCLUSION

The meta-analysis is a beneficial method when it comes to estimating the pooled prevalence of human diseases. While the frequentist approach provides simplicity and ease of use, the Bayesian approach provides complete information about all parameters because it takes into account all of the sources of variations.

In this study, the frequentist and Bayesian meta-analysis methods were applied to data obtained from Niu et al. (2017) in order to determine the prevalence of Alzheimer's disease in Europe. According to the narrow confidence interval as well as to a smaller relative error, the Bayesian approach was better than the frequentist approach, given that it estimated prevalence considerably more accurately.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides and/or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and/or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

This study is entirely author's own work and no other author contribution.

REFERENCES

1. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Overview. Introduction to Meta-Analysis. 1st ed. Chichester, UK: Wiley; 2009. p.17.
2. Pearson K. Report on certain enteric fever inoculation statistics. Br Med J 1904;2(2288):1243-6.
3. Gully SM, Incalcaterra K, Joshi A, Beaubien J. A meta-analysis of team-efficiency, potency, and performance: interpendence and level of analysis as moderators of obserbed relationships. J Appl Psychol 2002;87(5):819-32.
4. Smith ML, Glass GV. Meta-analysis of psychotherapy outcome studies. Am Psychol 1977;32(9):752-60.
5. Cohen LM, Manion L, Morrison K. Critical theory and curriculum research. Research Methods in Education. 6th ed. London: Routledge; 2007. p.30.
6. Zhang MW, Ho RC, Cheung MW, Fu E, Mak A. Prevalence of depressive symptoms in patients with chronic obstructive pulmonary disease: a systematic review, meta-analysis and meta-regression. Gen Hosp Psychiatry 2011;33(3):217-23.
7. Li Y, Xu J, Reilly KH, Zhang J, Wei H, Jiang Y, et al. Prevalence of HIV and syphilis infection among high school and college student MSM in China: a systematic review and meta-analysis. PLoS One 2013;8(7):e69137.

8. Anchala R, Kannuri NK, Pant H, Khan H, Franco OH, Di Angelantonio E, et al. Hypertension in India: a systematic review and meta-analysis of prevalence, awareness, and control of hypertension. J Hypertens 2014;32(6):1170-7.
9. Chen JJ, Yu CB, Du WB, Li LJ. Prevalence of hepatitis B and C in HIV-infected patients: a meta-analysis. Hepatobiliary Pancreat Dis Int 2011;10(2):122-7.
10. Jayawardena R, Ranasinghe P, Byrne NM, Soares MJ, Katulanda P, Hills AP. Prevalence and trends of the diabetes epidemic in south Asia: a systematic review and meta-analysis. BMC Public Health 2012;12:380.
11. Hussain S, Bakhtiar FK, Khan MH. A Bayesian meta-analysis and systematic review on global and local prevalence of hepatitis B. IJSRM 2014;2(12):1846-55.
12. Liu GC, Sui GY, Liu GY, Zheng Y, Deng Y, Gao YY, et al. A Bayesian meta-analysis on prevalence of hepatitis B virus ınfection among Chinese volunteer blood donors. PLoS One 2013;8(11):e79203.
13. Song P, Wang J, Wei W, Chang X, Wang M, An L. The prevalence of vitamin A deficiency in Chinese children: a systematic review and Bayesian meta-analysis. Nutrients 2017;9(12):1285.
14. Maheu-Giroux M, Filippi V, Samadoulougou S, Castro MC, Maulet N, Meda N, et al. Prevalence of symptoms of vaginal fistula in 19 sub-Saharan Africa countries: a meta-analysis of national household survey data. Lancet Glob Health 2015;3(5):e271-8.
15. Xin X, Zhu HP, Yu CH, Xu XJ, Li RD, Qiu J. A Bayesian approach to estimate the prevalence of schistosomiasis Japonica infection in the hubei province lake regions, China. Int J Environ Res Public Health 2013;10(7):2799-812.
16. Barendregt JJ, Doi SA, Lee YY, Norman RE, Vos T. Meta-analysis of prevalence. J Epidemiol Community Health 2013;67(11):974-8.
17. Bolstad WM. Bayesian Inference for Binomial Proportion. Introduction to Bayesian Statistics. 2nd ed. Hoboken, N.J: Wiley-Interscience; 2007. p.141-77.
18. Gardner IA. The utility of Bayes' theorem and Bayesian inference in veterinary clinical practice and research. Aust Vet J 2002;80(12):758-61.
19. Greiner M, Gardner IA. Application of diagnostic tests in veterinary epidemiologic 353 studies. Prev Vet Med 2000;45(1):43-59.
20. O'Hagan A. Bayesian statistics: principles and benefits. Wageningen UR Frontis Series 2004;3:31-45. http://library.wur.nl/frontis/bayes/03_o_hagan.pdf.
21. Zečic D. Frequentist and Bayesian meta-analysis: an example of a combined model. Hammamet, Tunisia: 13th International Research/Expert Conference" Trends in the Development of Machinery and Associated Technology" TMT; 2009. p.745-8.
22. Spiegelhater DJ, Abrams KR, Myles JP. An Overview of the Bayesian Approach. Bayesian Approaches to Clinical Trails and Health-Care Evaluation. 1st ed. Chichester; Hoboken, NJ: John Wiley & Sons; 2002. p.57-64.
23. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the cochrane database of systematic reviews. Int J Epidemiol 2012;41(3):818-27.
24. Niu H, Álvarez-Álvarez I, Guillén-Grima F, Aguinaga-Ontoso I. Prevalence and incidence of Alzheimer's disease in Europe: a meta-analysis. Neurologia 2017;32(8):523-32.