Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2016; 2(1): 20-24.

Review Article

Statistical methods in the meta-analysis of prevalence of human diseases

Ke-Sheng Wang[1*], Xuefeng Liu[2]

[1] Department of Biostatistics and Epidemiology, College of Public Health, East Tennessee State University, Johnson City, TN, USA
[2] Department of Systems Leadership and Effectiveness Science, School of Nursing, University of Michigan, Ann Arbor, MI, USA

ARTICLE INFO

Received 12.04.2015
Revised 22.09.2015
Accepted 15.01.2016
Published 15.03.2016

Available online at:
http://jbe.tums.ac.ir

Key words:
human diseases;
prevalence;
meta-analysis;
software;
heterogeneity;
publication bias

ABSTRACT

Background & Aim: A meta-analysis refers to the statistical synthesis of results from a series of studies and has been used to estimate pooled prevalence of human diseases. In this study, we review some statistical issues regarding the meta-analysis of the prevalence of human diseases such as statistical software and programs, transformations of prevalence rate, assessment of heterogeneity, and publication bias.

Introduction

A meta-analysis refers to the statistical synthesis of results from a series of studies. If an effect size is consistent across the series of studies, the meta-analysis procedures can be used to report that the effect is robust across the samples and to estimate the magnitude of the effect more precisely than we could obtain in any of the individual studies (1). The prevalence is defined as the number of persons with a health event of interest (e.g., disease) present in the population at a specific time divided by the total number of persons in the population at that time (2). Prevalence is a proportion, which can be categorized as point prevalence (total number of existing health events in a population at a given time point) and period prevalence (total number of existing health events in a population during a specified time period). Several software and programs have been used in the estimation of the pooled prevalence of human diseases. For example, the comprehensive meta-analysis software (Biostat, Englewood, NJ, USA, www.Meta-analysis.com) has been used to estimate the pooled prevalences of depressive symptoms in patients with chronic obstructive pulmonary disease (3), and HIV among high school and college student in China (4). Furthermore, Chen et al. (5) determined the prevalence of coinfection with either hepatitis C virus or hepatitis B virus in patients infected with HIV and Jayawardena et al. (6) discussed the prevalence of pre-diabetes and diabetes in South Asia using the Review Manager (RevMan

* Corresponding Author: Ke-Sheng Wang, Postal Address: Department of Biostatistics and Epidemiology, College of Public Health, East Tennessee State University, PO Box 70259, Lamb Hall, Johnson City, TN 37614-1700, USA.
Email: wangk@etsu.edu

computer program, Cochrane Collaboration, London, UK, www.cc-ims.net/RevMan). Mitchell et al. (7) summarized the pooled prevalence of depression, anxiety, and adjustments disorders and Luppa et al. (8) analyzed the age- and gender-specific prevalence of depression in latest-life using the statistical package StatsDirect (StatsDirect Ltd., England, www.statsdirect.com). The MetaXL is a new, freely available, software program for the meta-analysis in Microsoft Excel (Australia, www.epigear.com) and has been used to estimate the multiple category prevalence of multiple sclerosis (9). General purpose statistical package, such as SAS, SPSS, and STATA, has no inherent support for meta-analysis. However, some macros have been developed to perform meta-analysis of prevalence. For example, several studies have used the STATA statistical software package (StataCorp LP, College Station, Texas, USA) (5, 10-12), a general mixed effects linear models approach using SAS (PROC GLIMMIX macro for SAS, Cary NC, USA) (13, 14), and macros using SPSS software (SPSS, Inc., Chicago, IL, USA) (15, 16). Furthermore, some studies conducted Bayesian logistic meta-regression with the help of WINBUGS software (WinBUGS software, MRC Biostatistics Unit, Cambridge, UK) (17, 18). The basic features and similarities and differences among these software and programs can be found in the book "Introduction to Meta-Analysis" by Borenstein and Hedges (1) and the website: www.Meta-analysis.com.

Pooled Prevalence of Human Diseases

The prevalence has two features: (1) it is always between 0 and 1 (inclusive) and (2) the sum over categories always equals 1. Therefore, it is assumed that prevalence follows a binomial distribution (9). The pooled prevalence is an average of the individual study results weighted by the inverse of their variances using a fixed/random effects model (19).

According to Barendregt et al. (9), the variance of prevalence can be expressed as $Var(p) = p(1-p)/N$, where p is the prevalence proportion, and N is the population size. Then, the pooled prevalence can be expressed as follows:

$$p = \frac{\sum_i \frac{p_i}{Var(p_i)}}{\sum_i \frac{1}{Var(p_i)}}$$

With standard error (SE),

$$SE(p) = \sqrt{\sum_i \frac{1}{Var(p_i)}}$$

The confident interval of the pooled prevalence can be expressed as follows:

$$CI_Y(p) = P \pm Z\alpha_{/2} SE(p)$$

Where, $Z_{\alpha/2}$ is the appropriate factor from the standard normal distribution for the desired confidence percentage (e.g., $Z_{0.025} = 1.96$).

The inverse variance method works fine for the prevalence proportions around 0.5. However, it arise problems when the proportions get closer to 0 or 1. For example, when the proportion becomes small or big, the variance of the study is squeezed toward 0. To normalize the distribution of the effect size, the natural logarithm (log) of prevalence has been used to estimate the pooled prevalence (3, 20) with the variance-based method (19). After the meta-analysis, the effect sizes should be converted back into prevalence (3). However, the commonly used logit transformation cannot succeed in stabilizing the variance. Barendregt et al. (9) proposed that the double arcsine transformation is preferred over the logit transformation, and the method is also suitable to meta-analysis of multiple category prevalence.

Assessment of heterogeneity

The pooled prevalence estimates and 95% confidence intervals, stratified by study setting and other covariates, can be determined by fixed/random effects meta-analysis using the inverse variance method (19) based on the heterogeneity test result that can be assessed using Cochrane's Q statistics, I^2 (21), and τ^2 statistics (22). The Q statistic is reported with χ^2 and P values. For Q statistics, due to the low power of this test, a minimum cut-off P value of 0.10 has been established as a threshold of heterogeneity (4, 10, 21, 23). I^2 lies between 0% and 100% and τ^2 show the variance between studies. The I^2 statistic is reported as a

percentage with increasing values indicating greater heterogeneity between estimates of individual studies. For example, the general interpretation of I^2 values is: 0-40% might indicate low heterogeneity, 30-60% may represent moderate heterogeneity, 50-90% may represent substantial heterogeneity, and 75-100% may indicate considerable heterogeneity (Cochrane Handbook for Systematic Reviews of Interventions. http://www.cochrane-handbook.org) (24). Several studies have used I^2 to assess heterogeneity with thresholds of $\geq$ 25%, $\geq$ 50% and $\geq$ 75% indicating low, moderate and high heterogeneity, respectively (3, 4, 12, 23, 25).

In the fixed effects model, it is assumed that all of the observed differences between studies are due to chance, where the inverse variance is used for the weighted method. A random effects model is recommended for the meta-analysis of prevalence when heterogeneity is observed in prevalence estimates across studies (19) as a fixed effects model is likely to produce misleading results in the presence of significant heterogeneity (26). For the random effects model, it is assumed that each study may have a different underlying effect. This model leads to relatively more weight being given to smaller studies and to wider confidence intervals than the fixed effects models.

We can also test the heterogeneity by conducting a meta-regression analysis. Meta-regression can be used to estimate the extent to which measured covariates could explain the observed heterogeneity in prevalence estimates across studies. The regression coefficients (β) indicate the average difference in prevalence proportion for one category compared to the other. Effects of individual covariates can be examined first in univariate models and then in a multivariate model constructed in a step-wise fashion. When evidence is found of heterogeneity in the prevalence estimates between studies, meta-regression with Z-test can be used to identify moderators which might contribute to the heterogeneity of prevalence (27-30).

The source of heterogeneity can be the differences due to inadequate sample size, different study design, different populations, different treatment, different adjustments, different statistical analyses, different reporting, etc. To deal with the statistical heterogeneity, following analyses can be conducted such as do subgroup analysis (such as sex, age, and geographical design), exclude the outlying studies, choose another scale or change the effect measure, perform random effects meta-analysis, and meta-regression (4, 5, 11, 24, 30).

Assessment of publication bias

In meta-analysis, the publication bias can be checked using the Begg's funnel plot (5, 10, 12, 31). If publication bias is not present, the funnel plot is expected to be roughly symmetrical. The publication bias can be further checked using three statistical tests. First, the Egger test is a test for asymmetry of the funnel plot (32). This test is based on a linear regression of normalized effect estimate (estimate divided by its SE) against precision (reciprocal of the SE of the estimate). The intercept provides a measure of asymmetry – the larger its deviation from zero, the more pronounced the asymmetry (5, 7, 10, 12). Second, the Harbord's test (33) is similar to Egger's test but uses a modified linear regression method to reduce the false positive rate, which is a problem with the Egger test when there are a large treatment effects, few events per trial or all trials are of similar sizes (8, 10). Third, the Begg and Mazumdar's test (34) tests the interdependence of variance and effect size using rank correlation method (4, 7, 10, 12). In general, the significance is set at a $P \leq 0.05$ (4, 11). However, some authors have suggested that the recommended level is a $P \leq 0.10$ for the test of publication bias (5, 20, 35).

Conclusion

The meta-analysis is a useful tool to estimate the pooled prevalence of human diseases. The double arcsine transformation of single and multi-category prevalence is preferred over the commonly used logit transformation when the prevalence is small or large, with the consequence that such studies get a large weight in the meta-analysis. The Bayesian approach may have some advantages because it can take

into account all sources of variations and reflect these variations in the pooled result. When heterogeneity is observed, several methods can be used to do further analyses such as performing subgroup analysis, excluding the outlying studies, and conducting random effects analysis and meta-regression. Studies with small sample size may cause bias. It is suggested that some studies with small sample size may be excluded from the meta-analysis when bias is observed by graphic or statistic test method.

Acknowledgments

None.

References

1. Borenstein M, Hedges LV. Introduction to meta-analysis. Hoboken, NJ: John Wiley & Sons; 2009.
2. Gordis L. Epidemiology. Philadelphia, PA: Elsevier/Saunders; 2009.
3. 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.
4. 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.
5. 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.
6. 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.
7. Mitchell AJ, Chan M, Bhatti H, Halton M, Grassi L, Johansen C, et al. Prevalence of depression, anxiety, and adjustment disorder in oncological, haematological, and palliative-care settings: a meta-analysis of 94 interview-based studies. Lancet Oncol 2011; 12(2): 160-74.
8. Luppa M, Sikorski C, Luck T, Ehreke L, Konnopka A, Wiese B, et al. Age- and gender-specific prevalence of depression in latest-life--systematic review and meta-analysis. J Affect Disord 2012; 136(3): 212-21.
9. Barendregt JJ, Doi SA, Lee YY, Norman RE, Vos T. Meta-analysis of prevalence. J Epidemiol Community Health 2013; 67(11): 974-8.
10. Nath A, Agarwal R, Malhotra P, Varma S. Prevalence of hepatitis B virus infection in non-Hodgkin lymphoma: a systematic review and meta-analysis. Intern Med J 2010; 40(9): 633-41.
11. Ma YQ, Mei WH, Yin P, Yang XH, Rastegar SK, Yan JD. Prevalence of hypertension in Chinese cities: a meta-analysis of published studies. PLoS One 2013; 8(3): e58302.
12. Matcham F, Rayner L, Steer S, Hotopf M. The prevalence of depression in rheumatoid arthritis: a systematic review and meta-analysis. Rheumatology (Oxford) 2013; 52(12): 2136-48.
13. Krob HA, Fleischer AB, D'Agostino R, Haverstock CL, Feldman S. Prevalence and relevance of contact dermatitis allergens: a meta-analysis of 15 years of published T.R.U.E. test data. J Am Acad Dermatol 2004; 51(3): 349-53.
14. Paul M, Riebler A, Bachmann LM, Rue H, Held L. Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations. Stat Med 2010; 29(12): 1325-39.
15. Lipsey M, Wilson DB. Practical meta-analysis. New York, NY: Sage Publications; 2001.
16. van Emmerik-van OK, van de Glind G, van den Brink W, Smit F, Crunelle CL, Swets M, et al. Prevalence of attention-deficit hyperactivity disorder in substance use disorder patients: a meta-analysis and meta-regression analysis. Drug Alcohol Depend 2012; 122(1-2): 11-9.
17. Rudnicka AR, Jarrar Z, Wormald R, Cook DG, Fletcher A, Owen CG. Age and gender variations in age-related macular

degeneration prevalence in populations of European ancestry: a meta-analysis. Ophthalmology 2012; 119(3): 571-80.
18. Liu GC, Sui GY, Liu GY, Zheng Y, Deng Y, Gao YY, et al. A Bayesian meta-analysis on prevalence of hepatitis B virus infection among Chinese volunteer blood donors. PLoS One 2013; 8(11): e79203.
19. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986; 7(3): 177-88.
20. Horikawa C, Kodama S, Yachi Y, Heianza Y, Hirasawa R, Ibe Y, et al. Skipping breakfast and prevalence of overweight and obesity in Asian and Pacific regions: a meta-analysis. Prev Med 2011; 53(4-5): 260-7.
21. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002; 21(11): 1539-58.
22. Amini N, Alavian SM, Kabir A, Aalaei-Andabili SH, Saiedi Hosseini SY, Rizzetto M. Prevalence of hepatitis d in the eastern mediterranean region: systematic review and meta-analysis. Hepat Mon 2013; 13(1): e8210.
23. Gao X, Cui Q, Shi X, Su J, Peng Z, Chen X, et al. Prevalence and trend of hepatitis C virus infection among blood donors in Chinese mainland: a systematic review and meta-analysis. BMC Infect Dis 2011; 11: 88.
24. Cortese S, Moreira Maia CR, Rohde L, Morcillo-Peñalver C, Faraone SV. Prevalence of obesity in attention-deficit/hyperactivity disorder: study protocol for a systematic review and meta-analysis. BMJ Open 2014; 4(3): e004541.
25. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327(7414): 557-60.
26. Higgins JP, Thompson SG. Controlling the risk of spurious findings from meta-regression. Stat Med 2004; 23(11): 1663-82.
27. Sterne JAC, Bradburn MJ, Egger M. Meta-analysis in Stata. In: Egger M, Davey-Smith G, Altman D, Editors. Systematic reviews in health care: meta-analysis in context. Hoboken, NJ: John Wiley & Sons; 2001.
28. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002; 21(11): 1559-73.
29. Raudenbush SW. Analyzing effect sizes: random effects models. In: Cooper H, Hedges LV, Valentine JC, Editors. The handbook of research synthesis and meta-analysis. New York, NY: Russell Sage Foundation; 2009. p. 295-315.
30. Clancy MJ, Clarke MC, Connor DJ, Cannon M, Cotter DR. The prevalence of psychosis in epilepsy; a systematic review and meta-analysis. BMC Psychiatry 2014; 14: 75.
31. Dear KBG, Begg CB. An approach for assessing publication bias prior to performing a meta-analysis. Statist Sci 1992; 7(2): 237-45.
32. Egger M, Davey SG, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997; 315(7109): 629-34.
33. Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. Stat Med 2006; 25(20): 3443-57.
34. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994; 50(4): 1088-101.
35. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol 2000; 53(11): 1119-29.