Biostatistics and Biometrics
Open Access Journal
ISSN: 2573-2633

Research Article
Volume 1 Issue 1 - March 2017
DOI: 10.19080/BBOAJ.2017.01.555555

Biostat Biometrics Open Acc J

Which is the Preferred Measure of Heterogeneity in Meta-Analysis and Why? A Revisit

Mona Pathak[1], Sada Nand Dwivedi[1]*, SVS Deo[2], V Sreenivas[1], Bhaskar Thakur[1]

[1]*Department of Biostatistics, All India Institute of Medical Sciences, India*

[2]*Department of Surgical Oncology, IRCH, All India Institute of Medical Sciences, India*

Submission: December 20, 2016; **Published:** March 27, 2017

***Corresponding author:** Sada Nand Dwivedi, Department of Biostatistics, All India Institute of Medical Sciences, India, Tel: 919810571956; FAX: 91-11-26588663/641; Email: dwivedi7@hotmail.com

Abstract

Heterogeneity assessment is unavoidable under meta-analysis as it guides the path to choose appropriate synthesizing method. Its consideration further strengthens the drawn inferences. In the literature, there are various measures of heterogeneity, such as, Q statistic, H^2 statistic, HM_2 statistic, R^2 statistic and I^2 statistic. The present work mainly aimed to appraise all existing methods and find the preferred measure of heterogeneity in meta-analysis. Various heterogeneity measures were compared on the basis of theoretical criteria as well as their analytical results on the observed data from a systematic review of randomized controlled trials comparing neoadjuvant chemotherapy with adjuvant chemotherapy in the treatment of breast cancer patients. Theoretical criteria involve dependence on the extent of heterogeneity, invariance from scale and sample size and its interpretation. I^2 statistic is a sample size and scale invariant measure, which increases with increased heterogeneity. Unlike other methods of measuring extent of heterogeneity, I^2 has finite upper bound as well. Further, I^2 has precise confidence interval among all measures of heterogeneity assessment. In addition, it has appealing interpretation in terms of the proportion of the total variability that is because of heterogeneity alone. With all these virtues, I^2 is a preferred measure of heterogeneity in meta-analysis.

Keywords: Evidence based medicine; Randomized Controlled Trials; Meta-analysis; Heterogeneity Measures

Abbreviations: NACT: Neoadjuvant Chemotherapy; ACT: Adjuvant Chemotherapy; FEM: Fixed effect method; REM: Random effect method; RCT: Randomized controlled trial

Introduction

Meta-analyses are getting popular day by day in the present era of evidence based medicine. It involves statistical techniques to synthesize the effect size of the considered outcome from various comparative studies [1]. It synthesizes effect size of an intervention from the studies addressing the same problem under consideration. Even if only randomized controlled trials are considered, they may differ in their study designs, interventions, population characteristics and the measured outcomes. In other words, heterogeneous nature of the included studies is often possible. This may increase the between study variability in the effect size. This extent of variability which excess the extent because of sampling variation alone is known as heterogeneity [1,2]. The statistical heterogeneity can also be interpreted as the excess of between study variations over within study variance. It means that there may be genuine difference between the study effect sizes, it may not be because of chance alone [3]. Heterogeneity assessment in meta-analysis is very important as it guides the path to choose one of the synthesizing methods as well as interpretation of results. To be more specific, used method of synthesis i.e., fixed effect method, random effect method and recently proposed weighted least square method [4] is guided by observed extent of heterogeneity in a particular meta-analysis. So measurement of heterogeneity is an integral part of a meta-analysis. The non-incorporation of heterogeneity in a meta-analysis may reduce the confidence of related recommendation [3].

To begin with, Assessment of heterogeneity is basically done using visual interpretation through graphs and/or using the mathematical measures. For exploring heterogeneity, a number of graphical measures are available such as Forest plot, Galbraith radial plot, Z-score plot, funnel plot and method described by Baujat et al. [5]. Although they are easily understandable even by non-statistical researches, they indicate only presence or absence of heterogeneity. However, as obvious, testing absence of heterogeneity and its quantification is beyond the scope of graphical methods [2]. So to overcome this problem, mathematical methods to assess and measure heterogeneity were developed. To tests absence of heterogeneity, Q statistic, Z^2_{WLS}, Z^2_{WLS} R, Z^2_X, and Likelihood ratio test discussed Takkouche,

et al. [6]. They have reported that Q statistic is best among them. Hence, only Q statistic was included in this study for comparison. But these measures do not quantify the extent of heterogeneity. To cope up this problem, Higgins and Thompson reported three measures, H^2 statistic, R^2 statistic and I^2 statistic, which quantify the magnitude of heterogeneity [7]. With a different argument, Mittlböck & Heinzl suggested to use Hm_2, which also quantifies the magnitude of heterogeneity [8]. In spite of the above mentioned facts, in the current literature, Q statistic and I^2 statistic prevail to be most widely used measures of heterogeneity. Computation modules regarding these two measures are available in the standard statistical software for meta-analysis like Stata, R and Review Manager [9].

Keeping in view of the availability of various methods to assess absence of heterogeneity and/or quantify extent of heterogeneity, as obvious, need of comparing these methods arises [7,8,10] . As a first attempt in this regard Higgins & Thompson [7] compared H^2, R^2 and I^2 statistics in case of various measures of effect sizes (odds ratio, risk ratio, hazards ratio and standardized mean difference) using observed data. Another study [10] compared Q and I^2 on simulated data in case of only quantitative outcome, i.e., mean difference effect sizes in form of hedges 'd' and glass's 'g'. Further, one more study [7] compared H^2_M with I^2 on simulated data again only in case of quantitative outcomes. It may be worthwhile to mention that two of the above mentioned three studies used simulated data that may often provide non-convincing and distracted results. To summarise, there is no study comparing all the methods quantifying heterogeneity on observed measures of effect sizes. Also, earlier studies did not focus on finding out a preferred heterogeneity measure. Therefore, to find out preferred heterogeneity measure, the present study was aimed to revisit again comparing all the measures of quantifying heterogeneity along with Q statistic, using observed data on clinically more relevant measures of effect sizes like hazard ratio and risk ratio.

Materials and Methods

The dataset and methods of heterogeneity assessment are described in successive paragraphs:

Data

The observed dataset used in application of heterogeneity measures was derived from a systematic review performed to assess the effectiveness of neoadjuvant chemotherapy [11], duly approved by institutional ethics committee. Neoadjuvant chemotherapy (NACT) is the type of chemotherapy administered before the loco-regional treatment; however chemotherapy administered after the loco-regional treatment is known as adjuvant chemotherapy (ACT). This systematic review involved all randomized controlled trials comparing NACT with ACT in histologically proven breast cancer patients and measuring at least one of the considered outcomes, amely overall survival (time from randomization to death), disease free survival (time from randomization to recurrence or death), recurrence free survival (time from randomization to recurrence only), Loco-regional recurrence (time from randomization to Local and/or regional recurrence), Local recurrence (time from randomization to recurrence of ipsilateral breast, chest wall and local area), regional recurrence (time from randomization to recurrence to axilla and regional lymph nodes), Distal recurrence (time from randomization to metastases to other part of the body), and breast conserving surgery. To begin with, PubMed and Cochrane databases identified a total of 1239 records. Out of them, a total of 17 randomized controlled trials were found to be eligible comparing at least one of the considered outcomes. Since each considered outcome is in form of time to event, except breast conserving surgery, considered effect size in the analysis is hazards ratio and risk ratio respectively. Further, multiple outcomes were considered for sake of assessing the heterogeneity measures under varying set of condition under meta-analysis.

Methods

Before understanding individual measures of heterogeneity being compared in present study, it will be helpful to get familiarized with required terminology and notation:

Let-

k: Number of the studies included in meta-analysis

n_i : Sample size of i^{th} study

y_i : Estimate of the parameter ϑ_i

σ_{w_i} : Within study variance of ith study

τ^2 : Between study variance

w_i: Weight associated with i^{th} study under fixed effect method (FEM) while computing pooled effect size

w_i^*: Weight associated with i^{th} study under random effect method (REM) while computing pooled effect size.

In spite of various forms of weights, in the present article, it was being considered as usually precision (i.e., reciprocal of the variance). In FEM weights are just based on the within study variance, however reciprocal of addition of both i.e., within as well as between study variance is used as weights in REM. The pooled summary estimate is reported as the weighted average:

$$\mu_F = \frac{\sum w_i y_i}{\sum w_i} \quad \text{Where} \quad w_i = \sum 1/\sigma_i^2 \qquad (1)$$

The variance of this pooled effect estimate under FEM

How to cite this article: Mona P, Sada Nand D, SVS D, V Sreenivas, Bhaskar T. Which is the Preferred Measure of Heterogeneity in Meta-Analysis and Why? A Revisit. Biostat Biometrics Open Acc J. 2017;1(1): 555555. DOI: 10.19080/BBOAJ.2017.01.555555.

Where $v(f)=\frac{1}{\sum w_i}$ (2)

$\mu_R=\frac{\sum w_i^* y_i}{\sum w_i^*}$ Where $w_i^*=\sum\frac{1}{\sigma_i^2+\tau^2}$

The variance of this pooled effect estimate under REM

$$v(r)=1/(\sum w_i^*)$$

Criteria of preferred measure of heterogeneity

The Higgins & Thompson [7,8,10] have suggested that a good measure of heterogeneity ought to satisfy the following characteristics:

a. Dependence on the extent of heterogeneity: it should be a monotonically increasing function of between study heterogeneity.

$$f(\mu,\tau'^2,\sigma^2,k)>f(\mu,\tau^2,\sigma^2,k) \text{ whenever } \tau'^2>\tau^2$$

b. Scale invariance: It should not change with the change in scale of measurement.

$f(a+b\mu,b^2\tau'^2,b^2\sigma^2,k)>f(\mu,\tau^2,\sigma^2,k)$ for any a, b

c. Sample size invariance: It should be independent of number of studies included in the meta-analysis.

$$f(\mu,\tau^2,\sigma^2,k')>f(\mu,\tau^2,\sigma^2,k)$$

d. Interpretation: It should have easier interpretation.

The performance of all the measures was compared on the basis of above mentioned theoretical criteria and their performance on observed data discussed later.

Q statistic

Q statistic to test the heterogeneity in the meta-analysis was devised by Cochrane long back for combining the results of various experiments [12]. The Q statistic is defined as the weighted sum of square of deviation of individual effect size from pooled effect size computed by fixed effect method. Further, weights used in calculating pooled effect size are retained. This measure follows Chi-square distribution with k-1 degrees of freedom testing the null hypothesis of non-heterogeneity among studies.

$$Q=\sum w_i(y_i-\hat{\mu}_F)^2 \quad (3)$$

Being a sum of squares, the theoretically Q ranges from zero to infinite. Q statistic suffers with low power. In other words, this statistic performs better if large numbers of studies are included in the meta-analysis. Practically, most of the meta-analysis deals with small number of studies hence it was suggested to use significance level of 10% instead of 5% [2,13]. Further, Q statistic is sometimes over powered in case of large sample size and reveals a small between study variability as significant. To specify, the magnitude of Q statistic does not give any idea about the extent of heterogeneity [7].

H^2 statistic

H^2 statistic is defined as the relative excess in Q over its degree of freedom:

$$H^2=\begin{cases}\frac{Q}{k-1} if Q\geq(k-1)\\ 1 if\ Q<(k-1)\end{cases} \quad (4)$$

But, Higgnis & Thompson (2002) suggested to use H statistic, square root of H^2, because it is easy for clinicians to understand standard deviation and confidence interval than variance. They also defined H as the estimated residual standard deviation from the slope of the un-weighted least squares regression line through origin on Galbrith plot [7,14] i.e., plotting yi √wi against √wi:

$$H^2=Q/(k-1)=\sum(y_i\sqrt{w_i}-\hat{\mu}_F\sqrt{w_i})^2/(k-1) \quad (5)$$

And the confidence interval is:

$$\exp(\ln H\pm z_\alpha se[\ln H]) \quad (6)$$

Where $se[\ln\ln H]=\begin{cases}\frac{1}{2}\cdot\frac{\ln\ln(Q)-\ln\ln(k-1)}{\sqrt{2Q}-\sqrt{2k-3}} if Q>k\\ \sqrt{\frac{1}{2(k-2)}\left(1-\frac{1}{3(k-2)}\right)} if Q\leq k\end{cases}$ (7)

H statistic also increases with increasing τ^2 and does not depend on sample size. The value of H=1 represents complete homogeneity. So theoretically it ranges from one to infinite. Further, for assessing significance of heterogeneity, if the 95% CI of H statistic does not involve the null value (i.e., one), heterogeneity becomes significant at 5% level of significance. There is no universal rule to grade H^2 in mild, moderate or severe [7].

H_M^2 statistic

Usually the between study variance is provided by equating Q statistic to its expected value [15,16]:

$$Q=\tau^2\left(\sum_{i=1}^k w_i-\frac{\sum_{i=1}^k w_i^2}{\sum_{i=1}^k w_i}\right)+(k-1) \quad (8)$$

$$\Rightarrow \tau^2=\frac{Q-(k-1)}{\left(\sum_{i=1}^k w_i-\frac{\sum_{i=1}^k w_i^2}{\sum_{i=1}^k w_i}\right)} \quad (9)$$

Further, within study variance estimate [6] is:

$$\hat{\sigma}_W^2=\frac{(k-1)\sum_{i=1}^k w_i}{\left(\left(\sum_{i=1}^k w_i\right)^2-\sum_{i=1}^k w_i^2\right)} \quad (10)$$

How to cite this article: Mona P, Sada Nand D, SVS D, V Sreenivas, Bhaskar T. Which is the Preferred Measure of Heterogeneity in Meta-Analysis and Why? A Revisit. Biostat Biometrics Open Acc J. 2017;1(1): 555555. DOI: 10.19080/BBOAJ.2017.01.555555.

From equation 4, 8, 9 and 10:

$$H^2 = \frac{Q}{k-1} = \frac{\tau^2}{\hat{\sigma}^2} + 1 = \frac{\tau^2 + \hat{\sigma}^2}{\hat{\sigma}^2} \quad (11)$$

So Mittlböck & Heinzl [7] suggested using as access of between study variance over within study variance as:

$$H_M^2 = \frac{\tau^2}{\hat{\sigma}^2} = H^2 - 1 = (Q - df)/df \quad (12)$$

Obviously H^2 and H_m^2 share analogous properties. More specifically unlike (1, max), H_M^2 varies from (0, max), so it ranges from zero to infinite.

R^2 statistic

Since heterogeneity occurs because of relative access of between study variation over within study variation, the ratio of between study variance to within study variance may be a measure of heterogeneity [7], represented as R^2. Being ratio of two variances, it describes the relative inflation in the confidence interval for a summary estimate pooled by random effect model compared with a fixed effect model. As obvious, increase in between study variance will further raise the R^2 value. As variance of pooled effect estimate under REM will always be greater than that of FEM, R^2 ranges from 1 to infinite. In the case of homogeneity of effect sizes, the same pooled effect estimate and their confidence interval under both the methods i.e. FEM and REM will provide R^2=1. Being a ratio of two variances, it is a unit free measure. Under any of the method, the estimated variance of pooled effect size is sum of the weights, so computationally R^2 can also be represented as:

$$H_M^2 = \frac{\tau^2}{\hat{\sigma}^2} = H^2 - 1 = (Q - df)/df \quad (13)$$

I^2 statistic

The I^2 statistic measures the proportion of total variability that is due to heterogeneity rather than chance.

$$I^2 = \tau^2 / (\sigma^2 + \tau^2) \quad (14)$$

$$I^2 = \begin{cases} \frac{Q-(k-1)}{Q} x100\% \; for Q > (k-1) \\ 0 \; for Q \le (k-1) \end{cases} \quad (15)$$

As evident from equation 15, I^2 is usually measured as the percentage so ranges from zero to 100. I^2 also increases with increasing value of between study variance. Being the ratio of variances, this is a unit free measure. As the measure involves τ^2 in both, numerator as well as denominator, the effect of number of studies on this measure is neutralized. The 95% confidence interval for the I^2 can be obtained from the 95% limits of H^2 as there is relation between I^2 and H^2:

$$I^2 = (H^2 - 1)/H^2 \quad (16)$$

For grading the heterogeneity based on I^2 value, it is categorised at 25%, 50% and 75% as low, moderate and high heterogeneity respectively [3]. Most widely used Software packages for meta-analysis such Stata and Revman have incorporated Q and I^2 statistics only. To facilitate the calculation of H^2, H_m^2 and R^2, a Stata program was developed (appendix).

Comparison of the Measures of Heterogeneity

In the present article, the above discussed mathematical measures of heterogeneity were compared on the basis of their theoretical properties discussed earlier as well as their obtained performance on the observed data.

Application of heterogeneity measures on observed data

The performance of heterogeneity measures was compared using their precision analysis on an observed dataset. For this, the 95% confidence intervals were obtained through bootstrap using 10,000 replications regarding each outcome. For H^2, 95% confidence intervals were also calculated on the basis of standard error of log (H) discussed earlier. The 95% Confidence interval of HM^2 and I^2, were also calculated on the basis relationship with H2. Further, the significance of heterogeneity was observed if the 95% confidence interval of these heterogeneity measures did not include the null value of the respective measure. To mention here, intuitively the null value for each of I^2 and Hm^2 is zero, whereas that for H^2 and R^2 is 1.

Results

A systematic review of randomized controlled trials, comparing neoadjuvant chemotherapy to adjuvant chemotherapy in the breast cancer patients, found 17 eligible studies measuring at least one of the considered outcomes (i.e., overall survival, disease free survival, relapse free survival, loco-regional recurrence, local recurrence, regional recurrence, distal recurrence and breast conserving surgery). These outcomes were reported by 14, 6, 12, 10, 9, 4, 12 and 9 studies respectively. Since, these outcomes have varying sample size as well as between study variance (τ^2); it may make the comparison more generalizable. As evident from Table 1, overall survival, local recurrence and regional recurrence were extracted from homogeneous group of studies (i.e., τ^2=0). Q statistics revealed that disease free survival was from homogeneous group of studies (p=0.237). In addition, as per I^2 statistic, 26% of the total variation was due to heterogeneity alone, but it was not significant (95% CI: 0.00, 58.19). Further, total variance was 1.36 times of the within study variance (H^2=1.36). In other words, between study variance

How to cite this article: Mona P, Sada Nand D, SVS D, V Sreenivas, Bhaskar T. Which is the Preferred Measure of Heterogeneity in Meta-Analysis and Why? A Revisit. Biostat Biometrics Open Acc J. 2017;1(1): 555555. DOI: 10.19080/BBOAJ.2017.01.555555.

was 0.36 times of the within study variance (H_M^2=0.36). R^2 statistic revealed that heterogeneity has inflated the confidence interval under random effect method, 1.62 times in comparison to that under fixed effect method. Likewise similar result was observed for loco-regional recurrence, relapse free survival and distal recurrence. Whereas, in the case of breast conserving surgery with highest heterogeneity among all the considered outcomes, all the measures including Q statistic reported that they were extracted from heterogeneous group of studies (p<0.001). I^2 revealed that 90% of the total variance, inflating the confidence interval under REM three times that under FEM (R=3), is because of heterogeneity alone. All the heterogeneity measures were on same line and consistently increase with increasing between study variance. As described in Table 1, I^2 has narrowest confidence interval. The width of respective 95% bootstrap confidence interval for I^2, H^2, R^2 and H_M^2 were 0.32, 0.36, 0.95 and 0.36 under overall survival, 0.58, 0.99, 6.66 and 0.99 under disease free survival, 0.47, 0.76, 2.56 and 0.76 for loco-regional recurrence, 0.60, 1.29, 2.80 and 1.29 under relapse free survival; 0.62, 1.43, 3.4 and 1.43 under Distal recurrence; and, the highest width for breast conserving surgery as 0.95, 18.56, 20.16 and 18.56. The similar pattern has been observed under test based confidence interval. To among all the measures of heterogeneity, I^2 statistic has consistently precise confidence interval regardless of the considered outcomes, between study variance and sample size.

Table 1: Summary statistics of heterogeneity measures with respective 95% bootstrap confidence interval and *test based Confidence Interval.

Outcome	No. of Studies	τ^2	Q (p-value)	I^2	H^2	R^2	H^2M
Overall Survival	14	0	11.52	0	1	1	0
			-0.567	(0.00, 0.32)	(1.00, 1.36)	(1.00, 1.95)	(0.00, 0.36)
				(0.00, 0.55)*	(1.00, 2.22)*		(0.00, 1.22)*
Disease Free Survival	6	0.012	6.79	0.26	1.36	2.63	0.36
			-0.237	(0.00, 0.58)	(1.00, 1.99)	(1.00, 7.66)	(0.00, 0.99)
				(0.00, 0.69)*	(1.00, 3.26)*		(0.00, 2.26)*
Local Recurrence	9	0	8.01	0	1	1	0
			-0.433	(0.00, 0.51)	(1.00, 1.82)	(0.20, 2.31)	(0.00, 0.82)
				(0.00, 0.65)*	(1.00, 2.84)*		(0.00, 1.84)*
Regional Recurrence	4	0	0.31	0	1	1	0
			-0.959	(0.00, 0.00)	(1.00, 1.00)	(1.00, 1.00)	(0.00, 0.00)
				(0.00, 0.84)*	(1.00, 6.53)*		(0.00, 5.53)*
Loco-regional Recurrence	10	0.017	10.56	0.15	1.17	1.07	0.17
			-0.307	(0.00, 0.47)	(1.00, 1.76)	(1.00, 3.56)	(0.00, 0.76)
				(0.00, 0.56)*	(1.00, 2.28)*		(0.00, 1.28)*
Relapse free Survival	12	0.027	18.12	0.39	1.65	2.45	0.65
			-0.079	(0.00, 0.60)	(1.00, 2.29)	(1.00, 3.80)	(0.00, 1.29)
				(0.00, 0.69)*	(1.00, 3.25)*		(0.00, 2.25)*
Distal Recurrence	12	0.037	20.63	0.47	1.88	2.85	0.88
			-0.037	(0.00, 0.62)	(1.00, 2.43)	(1.00, 4.40)	(0.00, 1.43)
				(0.00, 0.73)*	(1.00, 3.66)*		(0.00, 2.66)*
Breast Conserving Surgery	9	0.036	80.27	0.9	10.03	12.04	9.03
			0	(0.00, 0.95)	(1.00, 19.56)	(1.84, 22.00)	(0.00, 18.56)
				(0.83, 0.94)*	(5.98, 16.71)*		(4.98, 15.71)*

Although Q statistic is best among the heterogeneity test [6], but it does not quantify the magnitude of heterogeneity. Further, it depends on number of studies included in the sample size as it suffers with low power in case of small number of studies. However, it has excess power in case of large number of studies included in meta-analysis. Unlike Q statistic, H^2, H_M^2, R^2 and I^2 statistics quantifies the magnitude of heterogeneity. Keeping in the view of observed results and earlier described criteria regarding preferred heterogeneity measure, all these four measures are invariant of the sample size. If weights under consideration are inverse of study level variance (i.e,), H^2 and H_M^2 statistics also become scale invariant. However, R^2 and I^2 are size as well as scale invariant measures regardless of the type of weights under consideration. Theoretically the range of R^2 may vary from 1 to infinite. In contrary to this, I^2 is expressed as percentage (0-100%). Because of finite quantity of I^2, it has appealing interpretation as it measures the proportion of the total variability in the effect size that is because of heterogeneity alone.

How to cite this article: Mona P, Sada Nand D, SVS D, V Sreenivas, Bhaskar T. Which is the Preferred Measure of Heterogeneity in Meta-Analysis and Why? A Revisit. Biostat Biometrics Open Acc J. 2017;1(1): 555555. DOI: 10.19080/BBOAJ.2017.01.555555.

Discussion

There are very few studies comparing the heterogeneity measures under meta-analysis. Long back a study [6] compared various measures to test the absence of heterogeneity and found that Q statistic is best among all these measures. But Q statistics does not quantify the heterogeneity. A comparison on simulated data for effect sizes as hedges'd' and glass's 'g' by reported that I^2 is better than Q because of its interpretation only [10]. Another study compared H^2_M with I^2 again on simulated data in case of quantitative outcomes and argued that H_M^2 may be more intuitive because of its interpretation [8]. It may be worthwhile to mention that these two studies using simulated data may often provide non-convincing and distracted results. Higgins and Thompson proposed H^2, R^2 and I^2 statistics and applied in case of various effect sizes (odds ratio, risk ratio, hazards ratio and standardized mean difference) using observed data. They found that for the ten studies included in meta-analysis, significant value of H^2 at p=0.1, p=0.05 and p=0.01 are 1.64, 1.88 and 2.40 respectively. However, in case of 30 studies, these H^2 values are significant at 1.35, 1.46 and 1.72 (Higgins and Thompson 2002). So there is no universal rule to grade heterogeneity as mild, moderate or severe on the basis of H^2. They reported that all these three measures follow the criteria of good measure of heterogeneity but abstained to comment on preferred measure among them [7]. Since H^2 and H_M^2 share analogues properties, it may not be possible to grade heterogeneity on the basis of HM2 as well. Further, on the basis of R^2 also, heterogeneity cannot be graded. But, I^2 grades heterogeneity as mild moderate and sever at its value as 25%, 50% and 75% respectively.

Keeping in view of the criteria of appropriate measure of heterogeneity described earlier, theoretical characteristics under each measure and related analytical results on various outcomes in the present study, it is amply clear that I^2 only has finite quantification of heterogeneity (0-100%) leading to more logical and appealing interpretation. In contrary, other measures (H^2, R^2, H_M^2) involve infinite range of the quantification of heterogeneity, which is likely to provide unrealistic summaries of uncertainty in heterogeneity measures. In other words, in view of infinite range of these measures, any conventionally suggested criteria for interpretation will be full of subjectivity. The analytical results in the present study have also shown beyond doubt that I^2 provides more precise results. Hence, I^2 may be an obvious preference over other measures of heterogeneity. To obtain sufficient evidence regarding conclusion under the present study, keeping in view of importance of several real data analysis, various outcomes involving obvious variation in number of studies and between study variance, related to effectiveness of neoadjuvant chemotherapy was attempted. However, one may further explore similar analysis in other areas involving other effect sizes like mean difference and regression coefficients.

Conclusion

I^2 statistic is a sample size and scale invariant measure, which increases with increased heterogeneity. Unlike other methods of measuring extent of heterogeneity, I^2 has finite upper bound as well. Further, I^2 has precise confidence interval among all measures of heterogeneity assessment. In addition, it has appealing interpretation in terms of the proportion of the total variability that is because of heterogeneity alone. With all these virtues, I^2 is a preferred measure of heterogeneity in meta-analysis.

Acknowledgement

The authors sincerely acknowledge the All India Institute of Medical Science, New Delhi for providing enquired resources and 'Institute fellowship' to support Ph.D. student, Ms Mona Pathak.

References

1. Michael B, Hedges LV, Higgins PTJ, Rothstein HR (2009) Introduction to Meta-Analysis. Chichester, John Wiley & Sons, UK.
2. Alex JS, Abrams KR, Jones DR, (2000) Methods for Meta-Analysis in Medical Research. Wiley 1: 1-380.
3. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring Inconsistency in Meta-Analyses. BMJ 327(7414): 557-560.
4. Stanley TD, Doucouliagos H (2015) Neither Fixed nor Random: Weighted Least Squares Meta-Analysis. Stat Med 34(13): 2116-2127.
5. Baujat B, Mahé C, Pignon JP, Hill C (2002) A Graphical Method for Exploring Heterogeneity in Meta-Analyses: Application to a Meta-Analysis of 65 Trials. Stat Med 21(18): 2641-2652.
6. Takkouche B, Cadarso Suárez C, Spiegelman D (1999) Evaluation of Old and New Tests of Heterogeneity in Epidemiologic Meta-Analysis. Am J Epidemiol 150(2): 206-215.
7. Higgins JP, Thompson SG (2002) Quantifying Heterogeneity in a Meta-Analysis. Stat Med 21(11): 1539-1558.
8. Mittlböck, M, Heinzl H (2006) A Simulation Study Comparing Properties of Heterogeneity Measures in Meta-Analyses. Stat Med 25(24): 4321-4333.
9. Higgins JPT, Green S (2008) Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. In: Higgins JPT and Green S (Eds.), Chichester England, Hoboken, NJ: Wiley-Blackwell, USA.
10. Huedo Medina TB, Sánchez Meca J, Marín Martínez F, Botella J (2006) Assessing Heterogeneity in Meta-Analysis: Q Statistic or I2 Index? Psychol Methods 11(2): 193-206.
11. Mona Pathak, Sada Nand Dwivedi, Deo SVS, Pramod Kumar Julka, Vishnubhatla Sreenivas (2015) Neoadjuvant Chemotherapy in Treatment of Breast Cancer. PROSPERO.
12. Cochran William G (1954) The Combination of Estimates from Different Experiments. Biometrics 10(1): 101-129.
13. Petitti DB (2001) Approaches to Heterogeneity in Meta-Analysis. Stat Med 20(23): 3625-3633.
14. Galbraith RF (1988) Graphical Display of Estimates Having Differing Standard Errors. Technometrics 30(3): 271-281.
15. Der Simonian R, Kacker R (2007) Random-Effects Model for Meta-Analysis of Clinical Trials: An Update. Contemp Clin Trials 28(2): 105-

How to cite this article: Mona P, Sada Nand D, SVS D, V Sreenivas, Bhaskar T. Which is the Preferred Measure of Heterogeneity in Meta-Analysis and Why? A Revisit. Biostat Biometrics Open Acc J. 2017;1(1): 555555. DOI: 10.19080/BBOAJ.2017.01.555555.

114.

16. DerSimonian R, Laird N (1986) Meta-Analysis in Clinical Trials. Control Clin Trials 7(3): 177-188.

How to cite this article: Mona P, Sada Nand D, SVS D, V Sreenivas, Bhaskar T. Which is the Preferred Measure of Heterogeneity in Meta-Analysis and Why? A Revisit. Biostat Biometrics Open Acc J. 2017;1(1): 555555. DOI: 10.19080/BBOAJ.2017.01.555555.