Türkiye Klinikleri Biyoistatistik Dergisi
Turkiye Klinikleri Journal of Biostatistics
Turkiye Klinikleri J Biostat. 2022;14(1):14-21

ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH
DOI: 10.5336/biostatic.2022-88046

Evaluating the Power of Meta-Analyses for Mean Differences: Methodological Research

Ortalama Farklar İçin Meta-Analizlerin Gücünün Değerlendirilmesi: Metodolojik Araştırma

Nazlı TOTİK DOĞAN[a], Zeliha Nazan ALPARSLAN[a]

[a]Department of Biostatistics, Çukurova University Faculty of Medicine, Adana, Türkiye

ABSTRACT Objective: Meta-analysis is a method that summarizes the results of various independent studies on the same subject and is used to obtain more reliable and accurate results. An important issue that is a sign of quality in meta-analyses, which is at the top of the evidence pyramid, is power, and the power of a meta-analysis can be evaluated under the two model approaches: fixed effect or random effect, with the existence of heterogeneity. Despite this importance, power details are not included in many meta-analysis studies. Whatever the model, power is influenced by effect size, the smallest sample size among studies, level of heterogeneity and the number of articles included in the meta-analysis. This study aims to determine the minimum sample size for a given number of articles in order to reach an acceptable power for a meta-analysis. **Material and Methods:** The power values for different numbers of articles (2, 5, 7, 10, 15, 18) and minimum of sample sizes in articles were determined, taking into consideration different effect sizes (0.1-0.2, 0.3-0.4, 0.5-0.7) under fixed and random effect models of low, medium and high heterogeneity. **Results:** In order for a meta-analysis study with a higher heterogeneity and/or a lower effect size to reach an acceptable power, the minimum sample size in the included articles should be high. **Conclusion:** This study, which considers the factors affecting power, can provide guidance to clinicians on how to calculate power according to heterogeneity and what is necessary or sufficient to achieve an acceptable quality in their studies.

Keywords: Meta-analysis; power analysis; effect size; fixed effect model; random effect model

ÖZET Amaç: Meta-analiz, birbirinden bağımsız benzer konulardaki çalışmaları bir araya getirerek çalışmaların sonuçlarındaki çeşitliliği açıklamak, daha güvenilir ve doğru sonuçlara ulaşılmasını sağlamak amacıyla kullanılan bir yöntemdir. Kanıt değeri yüksek olan meta-analiz çalışmalarında, nitelik göstergesi olan en önemli husus güçtür ve bu güç, heterojenliğin varlığına göre sabit etki ya da rastgele etki modeli olmak üzere kullanılan iki yaklaşım altında değerlendirilir. Bu öneme rağmen birçok meta-analiz çalışmasında güç detaylarına yer verilmemektedir. Yapılan güç değerlendirmeleri etki büyüklüğü, makalelerdeki minimum örnek büyüklüğü ve çalışma sayıları faktörlerinden etkilenir. Bu çalışmada, belirli etki büyüklüğünde kabul edilebilir bir güce ulaşmak adına makalelerdeki minimum örnek büyüklüğünün ve makale sayısının belirlenmesi amaçlanmıştır. **Gereç ve Yöntemler:** Düşük, orta ve yüksek heterojenliğe sahip rastgele etki ve sabit etki modelleri altında farklı etki büyüklükleri (0,1-0,2, 0,3-0,4, 0,5-0,7) dikkate alınarak farklı makale sayıları (2, 5, 7, 10, 15, 18) ve makalelerdeki örnek büyüklüklerinin minimumu değerlendirilerek güç değerleri belirlenmiştir. **Bulgular:** Heterojenlik düzeyinin daha fazla olduğu ve/veya düşük etki büyüklüğüne sahip bir meta-analiz çalışmasına dâhil edilen makalelerdeki örnek büyüklüklerinin minimumunun yüksek olması ile ancak kabul edilebilir bir güce ulaşılabilmektedir. **Sonuç:** Gücü etkileyen faktörlerin göz önünde bulundurulduğu bu çalışma, heterojenlik durumuna göre nasıl güç hesabı yapılabileceği ve kabul edilebilir bir kaliteye ulaşmak adına nelerin gerekli veya yeterli olduğu konusunda klinisyenlere rehberlik sağlayabilir.

Anahtar kelimeler: Metaanaliz; güç analizi; etki büyüklüğü; sabit etki modeli; rastgele etki modeli

Power calculations have always been an important part of statistical planning and have become a necessary component for most research. Just as it is in other types of studies, power is important in planning and interpreting meta-analyses. Meta-analysis studies are used to combine different findings from independent studies, a standardized effect size can be obtained by combining results from different studies.

Correspondence: Nazlı TOTİK DOĞAN
Department of Biostatistics, Çukurova University Faculty of Medicine, Adana, Türkiye
E-mail: nazlitotik.biostat@gmail.com

Peer review under responsibility of Turkiye Klinikleri Journal of Biostatistics.

Received: 13 Jan 2022 ***Received in revised form:*** 17 Feb 2022 ***Accepted:*** 10 Mar 2022 ***Available online:*** 20 Apr 2022

Although the results from the meta-analyses have been widely used in recent years, power studies are scarce. Meta-analysis studies, which are famous for being at the top of the evidence pyramid, are frequently used, especially by clinicians. However, power calculations are often overlooked. Taking into account the expected size of the overall effect, the number of included studies, and their sample sizes, a power analysis for a meta-analysis will provide the researcher with a statistically meaningful outcome through the assessment of power.[1-4] The aim of this study is to analyze the power of meta-analyses under varying conditions [effect size, minimum sample sizes in articles (MSSA), and number of studies] related to determining the minimum sample size for a given number of articles. It is also within the scope of the goal to reach a study as a guide for clinicians on how to obtain a power calculation according to the heterogeneous situation and what the essential or sufficient circumstances are to reach an acceptable power by addressing these conditions affecting power.

META-ANALYSIS AND POWER CALCULATION

Meta-analysis is a quantitative method that combines the effect size estimates of a series of studies to obtain a common effect size estimate. Meta-analysis, by estimating the effect size in each study and combining these estimates, produces estimates of effects synthesized with greater statistical power than the individual studies. Therefore, it is more likely that a meta-analysis will obtain different impacts of meaningful effects, relationships, and study differences in research. A meta-analysis also provides estimates of the best effect size for power analysis, so, even if the true effect is not large, it will be less likely that future studies will be designed with insufficient power. Two meta-analysis models have been developed for the purpose of making conclusions about the influence parameters of observed studies: Fixed effect model and random effect model. Power calculation details are summarized below for these 2 models.[5-7]

Power Calculation for the Fixed Effect Model:

Under the fixed effect model, it is assumed that all studies have a common (true) effect size, and the differences in observed effects are due to sampling error. The term "common effect model" can be used as a more descriptive term when referring to the fixed effect model in the meta-analysis. That is, since there is only one true effect, the fixed effect model uses a single effect size. Since the sample size of the studies cannot be infinite, a sample error occurs, and in this case, the true effect is not the same as the observed effect.[8] The observed effect T_i for any study is expressed as

$$T_i = \theta + \varepsilon_i$$

with the true effect size θ and the sample error ε_i.

Each study in the meta-analysis is often based on a different sample size, and since the estimates of studies with larger sample sizes are better than those with smaller sample sizes, a weight is calculated for each effect size to take into account the sample size when the mean effect size is used.[9-11] The weight assigned to each study in a fixed-effect meta-analysis is

$$w_i = \frac{1}{v_i},$$

where

$$v_i = \frac{n_{1i} + n_{2i}}{n_{1i} n_{2i}} + \frac{d_i^2}{2(n_{1i} + n_{2i})}$$

is the within-study variance for study i. n_{1i} and n_{2i} are the sample size of case and control group in the ith study, respectively. For studies that use two groups (case and control) standardized mean difference is estimated as

$$d_i = \frac{\bar{X}_{1i} - \bar{X}_{2i}}{s_i}$$

where $\bar{X}_{1i}$ and $\bar{X}_{2i}$ are the sample means in the case and control groups and s is the within-groups standard deviation. For the calculation of v_i, the ratio of the total case and total control sample sizes in the meta-analysis study is applied to the sample size of the article with the smallest sample size. In the present study, where the case and control ratio is taken to be 1:1, $n_{1i} = n_{2i} = n_i$ and

$$v_i = \frac{1}{n_i}\left(2 + \frac{d_i^2}{4}\right)$$

The weighted mean effect size and variance are expressed as

$$\bar{T}_{.} = \frac{\sum_{i=1}^{k} w_i T_i}{\sum_{i=1}^{k} w_i} \quad \text{and} \quad v_{.} = \frac{1}{\sum_{i=1}^{k} w_i}$$

If the v_i values are thought to be nearly equal, then $v_{.} = v/k$ where v is the common value of the v_i and k is the number of articles. Nevertheless, it is important to know that if the v_i values are not the same, and $\bar{v}$ is the average of the $\bar{v}_i$, $\bar{v}/k$ will be larger than $v_{.}$, and using $\bar{v}/k$ in place of $v_{.}$ in power calculations will cause an underestimate of the statistical power.[1]

Under hypothesis $H_0: \theta_1 = \theta_2 = \cdots = \theta_0$ with $SE(\bar{T}_{.}) = \sqrt{v_{.}}$

$$Z = \frac{\bar{T}_{.} - \theta_0}{\sqrt{v_{.}}}$$

has the standard normal distribution when $\theta = \theta_0$. If $\neq \theta_0$, Z has a normal distribution with $\theta \sim N(\lambda, 1)$ where $\lambda = (\theta - \theta_0)/\sqrt{v_{.}}$ Thus the power of the one-tailed test is given by

$$p = 1 - \Phi(c_\alpha - \lambda) \tag{1}$$

and the power of the two-tailed test is given by

$$p = 1 - \Phi(c_{\alpha/2} - \lambda) + \Phi(-c_{\alpha/2} - \lambda) \tag{2}$$

where $\Phi(x)$ is the standard normal cumulative distribution function and c_α corresponds to the critical Z value.[1,12]

Power Calculation for the Random Effect Model:

The random effect model holds the assumption that the true effect sizes are different, and these differences in participation or in the implementation of interventions cause different effect sizes. With the observed effect for any study T_i^* , the deviation of the true effect size from the general mean $(\mu - \theta), \zeta_i$, and the sample error $\varepsilon_i = T_i^* - \theta$ representing the variation between the true effect size of the study, the observed effect size is:

$$T_i^* = \mu + \zeta_i + \varepsilon_i$$

With the variance within the study (v_i) and the variance between the studies (τ^2), the variance of a study under the random effects model can be calculated. τ^2 is the estimated as ϑ^2

$$\vartheta^2 = \frac{\sum_{i=1}^{k} w_i^* T_i^{*2} - \frac{\left(\sum_{i=1}^{k} w_i^* T_i^*\right)^2}{\sum_{i=1}^{k} w_i^*} - (k-1)}{\sum w_i^* - \frac{\sum w_i^{*2}}{\sum w_i^*}}$$

Under the random effect model, the weight of a study (w_i^*) in the meta-analysis is $\frac{1}{v_i^*}$, where v_i^* is the sum of the variances between the studies for the study ($v_i^* = v_i + \vartheta^2$). The weighted mean effect size is calculated as:

$$\bar{T}_{.}^{*} = \frac{\sum_{i=1}^{k} w_i^* T_i^*}{\sum_{i=1}^{k} w_i^*}$$

and the weighted effect size variance as:

$$v_{.}^{*} = \frac{1}{\sum_{i=1}^{k} w_i^*}$$

Similar to the fixed effect model $v_{.}^{*} = v^*/k$ and $\lambda^* = (\theta - \theta_0)/\sqrt{v_{.}^{*}}$ Thus the power of the one-tailed test is given by

$$p = 1 - \Phi(c_\alpha - \lambda^*) \quad (3)$$

and the power of the two-tailed test is given by

$$p = 1 - \Phi\left(c_{\alpha/2} - \lambda^*\right) + \Phi\left(-c_{\alpha/2} - \lambda^*\right) \quad (4)$$

Where $\Phi(x)$ is the standard normal cumulative distribution function.[1,2,13] The possible values of the variance within the studies can be obtained in the same way as in the use of the fixed effect model. This variance has been suggested to be used as a mark for heterogeneity by Hedges and Pigott with $\left(\frac{V_i^*}{3}, \frac{2V_i^*}{3}, V_i^*\right)$ corresponding to low-level heterogeneity (LLH), medium-level heterogeneity (MLH) and high-level heterogeneity (HLH) level respectively.[1,14]

MATERIAL AND METHODS

For both fixed and random effect models, the mean and standard deviations of power values for different numbers of articles and MSSA were generated randomly 1,000 values considering the defined intervals for effect size. Fixing the number of articles as 2, 5, 7, 10, 15, and 18, MSSA was calculated for acceptable power with these numbers of articles. In the scenario reported in this study, the case/control ratio was taken as 1:1. The same assessment can be made for different case/control ratios. R 4.1.1 (R programming languages/Project packages "Rmisc", "ggplot2") was used to reach the analysis results in this study.

ASSESSMENT FOR THE FIXED EFFECT MODEL

Fixed effect model power evaluations were made with different effect sizes and different numbers of articles. The effect size classification was adapted from Cohen's (1992) effect size levels (small, medium, and large coinciding with 0.20, 0.50 and 0.80 respectively).[15] Equations (1) and (2) were used for the power calculation under the fixed effect model. Effect sizes and the number of articles were taken as shown in Table 1 and power evaluations were made.

ASSESSMENT FOR THE RANDOM EFFECT MODEL

Using τ^2 as a heterogeneity measure, a heterogeneity was added to the scenario at three levels (low, medium, and high corresponding to 0.33, 0.67, and 1 respectively).

Thus, power evaluations were made in different heterogeneity situations with different effect sizes and different numbers of articles. Equations (3) and (4) provide power calculations for different effect sizes and the number of articles.

TABLE 1: MSSA at effect sizes of 0.1-0.2, 0.3-0.4 and 0.5-0.7 for the fixed effect model, LLH, MLH and HLH random effect models for 80% power.

Scenarios			Effect size	Number of article					
				2	5	7	10	15	18
Type of effect model	Fixed effect model		0.1-0.2	590	250	180	130	80	70
			0.3-0.4	110	46	32	22	14	12
			0.5-0.7	38	16	12	8	6	4
	Random effect model	LLH	0.1-0.2	800	330	250	170	120	90
			0.3-0.4	142	60	42	28	20	16
			0.5-0.7	48	21	14	11	7	6
		MLH	0.1-0.2	1000	400	300	200	150	110
			0.3-0.4	175	68	50	35	26	20
			0.5-0.7	60	24	19	13	9	7
		HLH	0.1-0.2	1200	500	350	240	170	140
			0.3-0.4	215	85	58	42	30	24
			0.5-0.7	73	32	22	16	10	8

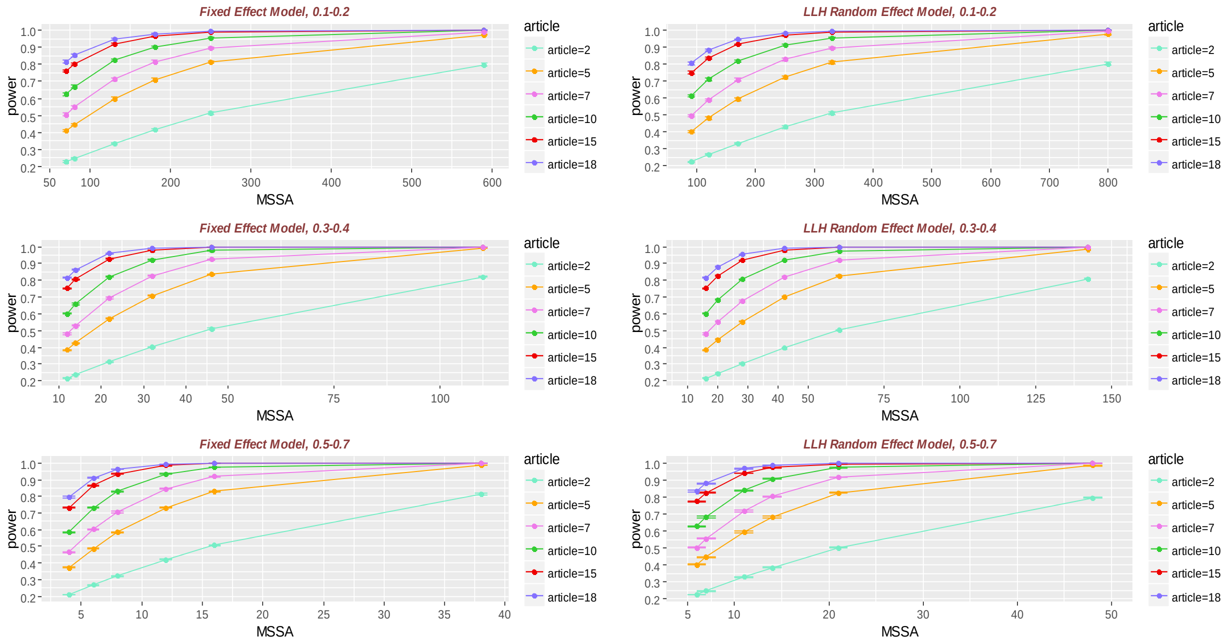
MSSA: Minimum of sample sizes in articles; LLH: Low-level heterogeneity; MLH: Medium-level heterogeneity; HLH: High-level heterogeneity.

RESULTS

The MSSA values required to achieve 80% power for different effect sizes and different numbers of articles under the possible effect models are provided in Table 1. Figure 1 and Figure 2 give the power values attained for different effect sizes and different numbers of articles pertaining to different MSSA for fixed and varying heterogeneity random effect models. As an example for Table 1, it is seen that if the number of articles included is 5, one needs MSSA to be 46 for an acceptable power for the fixed effect case with an effect size of 0.3-0.4. For LLH, MLH, and HLH random effect models with the same effect size, MSSA's of 60, 68, and 85 are needed. Under the fixed effect model, when the effect size is 0.1-0.2 and given 18 studies, an acceptable power was obtained when MSSA was 70. It can be seen from the table that for this effect size, when the number of articles is 2, the MSSA is 590 for 80% power. Similar interpretations can be made from Table 1 for different effect sizes and number of articles. Thus, if the effect size is very low (for example, in the 0.1-0.2 range), it is expected that, for an acceptable power, either the MSSA or the number of articles should be high. It is seen that the value of MSSA decreases with an increasing number of articles and/or increasing effect size. As can be seen from the continuous graphs with 3 different effect size groups, either an increase in the number of articles or a high effect size is required to achieve higher power in a meta-analysis (Figure 1 and Figure 2).

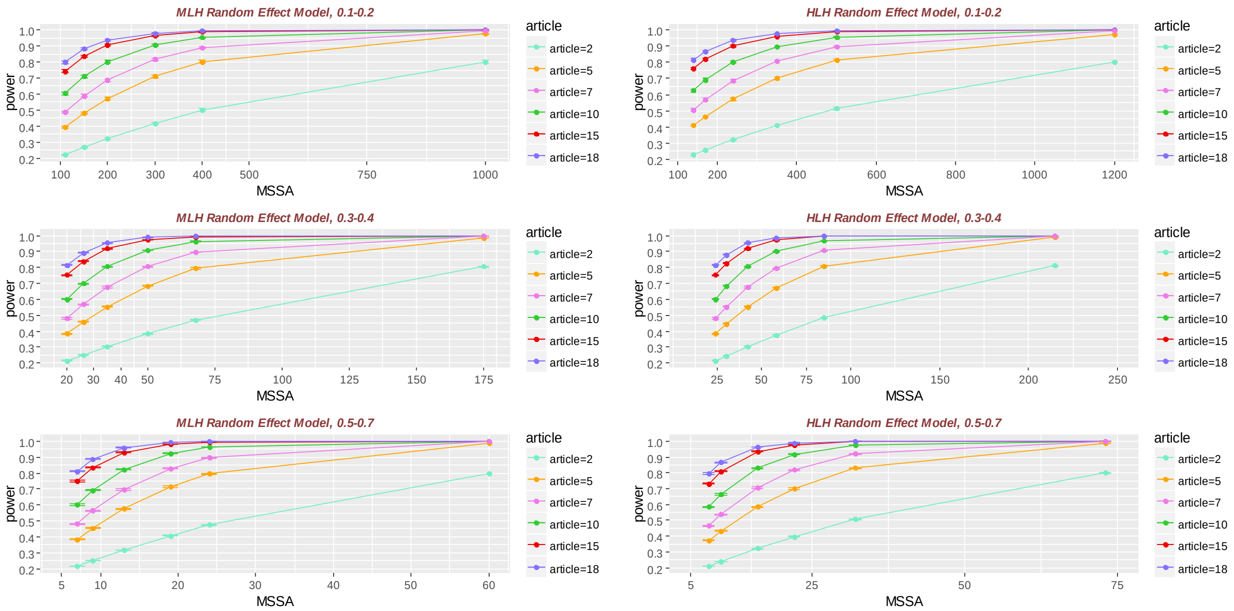
DISCUSSION

Meta-analysis is a statistical method applied to integrate the results of different studies in order to be able to calculate effect sizes with statistical techniques and to provide stronger estimates of the true effect size.[16] With the advancement of evidence-based disciplines, there is an increasing interest in the use of meta-analysis, which allows researchers to combine findings from independent research on the same topic and to find a final answer to the research question concerned.[17] Power calculations of meta-analysis can be made for fixed and random effect models. Although there are not many studies on power evaluations, the importance of this issue has been emphasized by many researchers.[2] In this study, given a certain number of



LLH: Low-level heterogeneity; MSSA: Minimum of sample sizes in articles.

FIGURE 1: Power values for different effect sizes under the fixed and LLH random effects model.



MLH: Medium-level heterogeneity; HLH: High-level heterogeneity; MSSA: Minimum of sample sizes in articles.

FIGURE 2: Power values for different effect sizes under the MLH and HLH random effects model.

articles, the required MSSA values to obtain an acceptable power of the meta-analysis study were obtained with different effect sizes in the fixed effect and random effect models. It was seen that an increase in the effect size or a reduction in the level of heterogeneity provided a reduction in the MSSA values required to achieve acceptable power. For power calculations based on the value range of effect sizes, under the same scenario (i.e., fixing the model and number of articles), an increase in the effect size provided the reduction of the MSSA required for acceptable power.[16] It has also been found that the number of articles has an effect on the power values. The increase in the number of articles has improved the power values, and an additive increase in MSSA may provide higher power. Consideration of the fixed effect, LLH, MLH, and HLH random effect models leads to the conclusion that as the heterogeneity increases, an acceptable power can be attained by a higher number of articles, larger effect size, or higher MSSA. As for heterogeneity, the MSSA required for acceptable power under the random effect model is higher than that obtained under the fixed effect model at the same conditions. This is because both the between-studies and within-study variances are considered in the random effect model. Similar findings were obtained for the MLH and HLH random effect models. Therefore, any degree of heterogeneity makes a significant difference in power calculations, a conclusion parallel with the literature.[18]

LIMITATIONS OF THE STUDY AND FUTURE DIRECTION

There are some limitations to this study. The first of these is the number of articles to be included in the study, with certain values as an example. For more than 18 articles, the acceptable power has already been reached. If it is desired to make a full calculation for the number of articles not included in this study, it can also be calculated by using the formula. Other limitations are that only the mean difference was evaluated as the effect size and the case/control ratios other than 1:1 case/control ratio were not included in this study. Ratios different from the 1:1 case/control ratio and effect sizes other than the mean difference (odds, correlation, etc.) may be evaluated in future studies.

CONCLUSION

The study was performed to calculate the MSSA in order to achieve an acceptable power with different effect sizes, number of articles, and heterogeneity levels. Suggestions have been made to avoid a low-power meta-analysis.

The MSSA required for acceptable power for the fixed effect model is lower than that for the random effect model with heterogeneous structure. The effect size, the number of articles, and MSSA should all be taken into account since they are all factors effecting power calculations. If the effect size is high, the number of articles will decrease, and high values for MSSA will not be needed. On the contrary, for an acceptable power, a study with a very low effect size should be carried out with the appropriate number of articles and/or MSSA to avoid waste of resources. To conclude, one should keep in mind that the power of a meta-analysis increases with an increase in the number of articles included, an increase in MSSA, and a decrease in heterogeneity.

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 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

***Idea/Concept:** Zeliha Nazan Alparslan, Nazlı Totik Doğan; **Design:** Zeliha Nazan Alparslan, Nazlı Totik Doğan; **Control/Supervision:** Zeliha Nazan Alparslan; **Data Collection and/or Processing:** Nazlı Totik Doğan; **Analysis and/or Interpretation:** Zeliha Nazan Alparslan, Nazlı Totik Doğan; **Literature Review:** Nazlı Totik Doğan; **Writing the Article:** Nazlı Totik Doğan; **Critical Review:** Zeliha Nazan Alparslan, Nazlı Totik Doğan.*

REFERENCES

1. Hedges LV, Pigott TD. The power of statistical tests in meta-analysis. Psychol Methods. 2001;6(3):203-17. [Crossref] [PubMed]
2. Jackson D, Turner R. Power analysis for random-effects meta-analysis. Res Synth Methods. 2017;8(3):290-302. [Crossref] [PubMed] [PMC]
3. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. 1st ed. Chichester, England: Wiley; 2009. [Crossref]
4. Cook DA, Hatala R. Got power? A systematic review of sample size adequacy in health professions education research. Adv Health Sci Educ Theory Pract. 2015;20(1):73-83. [Crossref] [PubMed]
5. Ellis PD. The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results. 1st ed. Cambridge: Cambridge University Press; 2010. [Crossref]
6. Wolf FM. Meta-Analysis: Quantitative Methods for Research Synthesis. Vol.59. 1st ed. USA: Sage; 1986.
7. Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. Stat Med. 1999;18(3):321-59. [Crossref] [PubMed]
8. Hedges LV, Vevea JL. Fixed-and random-effects models in meta-analysis. Psychological Methods. 1998;3(4):486-504. [Crossref]
9. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods. 2010;1(2):97-111. [Crossref] [PubMed]
10. Gliner JA, Morgan GA, Harmon RJ. Meta-analysis: formulation and interpretation. J Am Acad Child Adolesc Psychiatry. 2003;42(11):1376-9. [Crossref] [PubMed]
11. Hedges L, Olkin I. Statistical Methods for Meta-Analysis. 1st ed. Orlando, FL: Academic Press; 1985. [Crossref]
12. Cohn LD, Becker BJ. How meta-analysis increases statistical power. Psychol Methods. 2003;8(3):243-53. [Crossref] [PubMed]
13. Hedges LV, Pigott TD. The power of statistical tests for moderators in meta-analysis. Psychol Methods. 2004;9(4):426-45. [Crossref] [PubMed]
14. Valentine JC, Pigott TD, Rothstein HR. How many studies do you need? A primer on statistical power for meta-analysis. Journal of Educational and Behavioral Statistics. 2010;35(2):215-47. [Crossref]
15. Cohen J. Statistical power analysis. Current Directions in Psychological Science. 1992;1(3):98-101. [Crossref]
16. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013;14(5):365-76. Erratum in: Nat Rev Neurosci. 2013;14(6):451. [Crossref] [PubMed]
17. Liu J. Statistical power in meta-analysis [Doctoral thesis]. Columbia: University of South Carolina; 2015. Erişim tarihi: 17 Aralık 2019. Erişim linki: [Link]
18. Turner RM, Bird SM, Higgins JP. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. PLoS One. 2013;8(3):e59202. [Crossref] [PubMed] [PMC]