Epidemiology & Health Data Insights
(eISSN 3080-8111)

Methodological Paper

Reporting Practices in Epilepsy Research: An Overview and Tutorial

Ruslan Akhmedullin[1], Gaziz Kyrgyzbay[2], Darkhan Kimadiev[2], Zhasulan Utebekov[2], Ayana Ablayeva[1]

[1]*Department of Medicine, School of Medicine, Nazarbayev University, Astana, Kazakhstan*
[2]*Epileptology Centre, RSE Medical Centre Hospital of the President's Affairs Administration of the Republic of Kazakhstan, Astana, Kazakhstan*

Received: Oct 09, 2025
Accepted: Oct 29, 2025

Corresponding author's email:
ruslan.akhmedullin@nu.edu.kz

Abstract:

Background. This study examined the reporting practice of subgroup effects of meta-analytic research published in the leading journal of the ILAE.

Methods. We selected studies that used ratio measures and employed subgroup analyses. Subgroup differences were calculated as the difference between the log-transformed estimates over the square root of the sum of the squared standard errors. For the calculated test scores, a corresponding two-tailed p-value was calculated using the standard normal cumulative distribution function. The authors also conducted additional analyses accounting for multiple comparisons.

Results. The literature search identified 55 publications, of which 14 (25 %) were included. Neither study used a formal test to compare the subgroups. The number of reported subgroup estimates ranged from 2 to 20, and the number of pairwise comparisons ranged from 1 to 53. Overall, there were 187 comparisons, resulting in a median log difference of 0 (IQR 0.12 – 0.17) and a range from -2.92 to 2.32. The median p-value was 0.54 (IQR 0.21-0.85) with 18 (9%) comparisons showing p-values lower than the conventional significance level, whereas 6 and 21 contrasts were $0.05 < p \leq 0.10$ and $0.10 < p \leq 0.20$, respectively. Seven (4%) comparisons resulted in a p-value lower than the corrected significance level when adjusted for multiple comparisons.

Conclusion. There was a lack of compliance with the reporting guidelines. The findings from the subgroup analyses were commonly interpreted without employing a formal test. There is need to emphasize the importance of adherence to established reporting standards when presenting the subgroup effects.

Keywords: Epilepsy; Meta-Analysis; Subgroup Analysis; Reporting Bias

Introduction

A meta-analysis provides improved precision and the opportunity to resolve disputes arising from contradicting claims (1, 2). The effect size in a meta-analysis is computed given the variability in a primary study. This inevitably brings a degree of uncertainty, affecting the implications and generalization of the findings. Researchers often seek to conduct a post-hoc analysis to determine the subgroup effect related to the clinical characteristics of a study population.

Subgroup claims can be misleading, especially when certain assumptions are overlooked (1, 3). Existing evidence suggests that the interpretation of findings from subgroup analyses requires particular caution and should not be mentioned in the conclusions (1, 3, 4). The identification of the varying intervention effects requires a comparison of the subgroup estimates. Frequently, inferences are drawn from the absolute differences between subgroups (1, 5), contrasting their significance test results (1, 3-6), visual inspection of confidence intervals overlap, or other contrasts that lack control over multiple comparisons (1, 6, 7). These approaches, however, barely reveal group-wise differences in effects or lack thereof.

There are general principles that should be considered when analyzing and interpreting subgroup effects (3, 8). However, the quality of reporting of statistical findings remains poor (4). Previous works provided detailed discussions of issues arising from subgroup analyses referring to clinical trials (3-6). However, it has rarely been the subject of interest in the context of meta-analysis. We conducted this study, to explore the current reporting practice in meta analytic research and discuss potential issues when interpreting the subgroup effects.

Materials and Methods

Search strategy and study selection.

To identify meta-analyses, we manually searched the database of Epilepsia, the leading journal of the International League Against Epilepsy (ILAE) (9). The literature suggests that the number of published meta-analyses has notably increased over the last decade (10), with ratio measures being the most employed measures of association (11). Therefore, the authors screened only publications from January 2015 to August 2025 and selected only studies that employed ratio measures such as Odds Ratio (OR), Risk Ratio (RR), or Incidence Rate Ratio (IRR). Studies that did not report subgroup analyses or ratio measurements were excluded.

Data extraction.

A standardized data extraction form was developed using Microsoft Office. For the publications included, appropriate research details, such as the year of publication, measure of effect, effect size, and confidence intervals, were extracted. This was a brief literature review and was not preregistered. The extracted data and analytic code are available within the publication.

Statistical analysis.

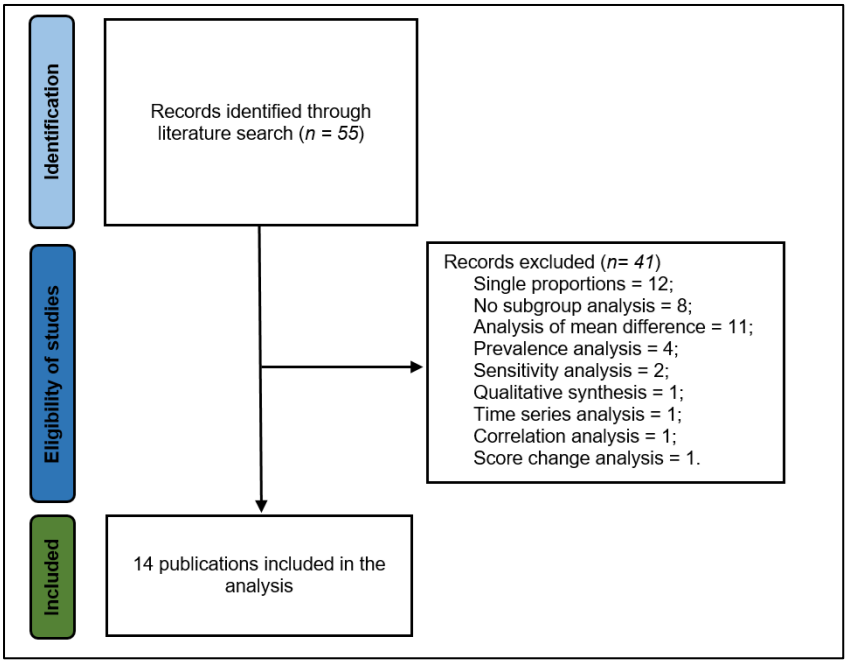
Unlike previous PRISMA statements (12), the latest PRISMA guidelines (13) provide an extended checklist with a clear requirement to report on how subgroup effects were analyzed, as well as the corresponding p-values and 95% CI for an estimate. We used log-transformed estimates of effect sizes. When confidence intervals were not reported directly, they were calculated using the conventional approach (14). The Standard Errors (SE) were approximated by dividing the width of the confidence interval by 3.92, and the difference between the subgroups was calculated as follows: (β1 - β2) / [sqrt(SE1)2 + (SE2)2] (15, 16). We then calculated a corresponding two-tailed *p*-value as $p = 2\Phi(-|z|)$, where Φ is the standard normal cumulative distribution function. We also performed additional analysis to account for the familywise error rate by employing a Bonferroni correction as α/n, where α and n are significance level and the number of comparisons conducted within a specific meta-analysis, respectively. Data were reported as percentages for categorical variables and as median and interquartile range (IQR) for continuous values. The corresponding code used for data analysis is available in the appendix (Supporting Information, Analysis R code).

Results

Overall, the literature search identified 55 publications, 41 of which did not meet the inclusion criteria. The excluded articles represented an analysis of single proportions (n=12, 22%), did not report subgroup analysis (n=8, 15%), and calculated standardized mean difference (n=6, 11%), mean difference (n=5, 9%), prevalence analysis (n=4, 7%), sensitivity analysis (n=2, 4%), qualitative synthesis (n=1, 2%), time series analysis (n=1, 2%), correlation analysis (n=1, 2%), and analysis of change in a score (n=1, 2%). Consequently, our final

analysis included 14 (25%) studies (17-30) (Figure 1, Supplemental Table 1).

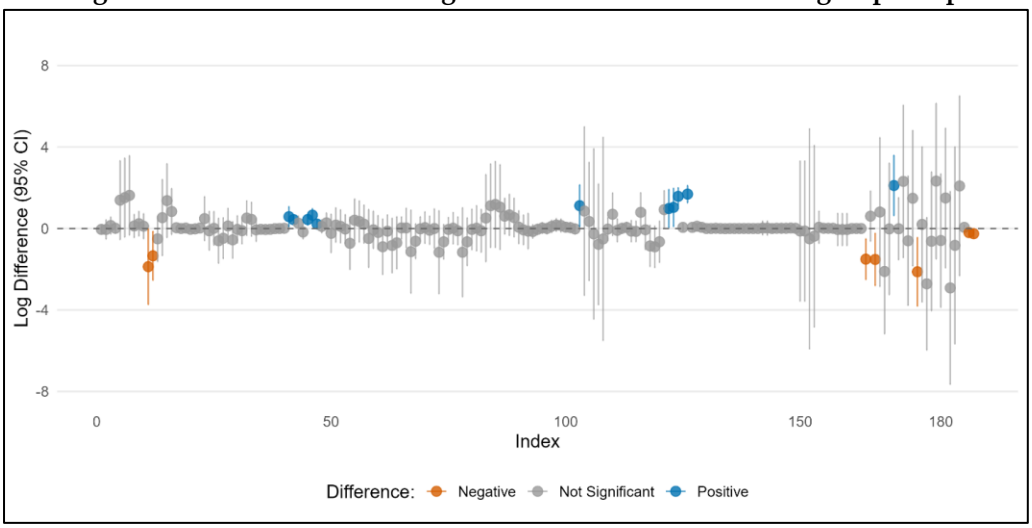
Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram.



The included studies mostly used RR (n=7, 50%), followed by OR (n=6, 43%), and IRR (n=1, 7%) as measures of effect. Ten studies (71%) (17, 19-21, 23, 25-27, 29, 30) mentioned a planned subgroup analysis in the Methods section, and two (14%) (17, 29) incorporated individual participant data (IPD). However, neither study used a formal test to compare the subgroup effects. Seven (50%) studies (19, 21, 22, 24, 27, 29, 30) used subgroup analysis to explore heterogeneity, and eight (57%) studies (17, 20-22, 25, 26, 28, 30) mentioned subgroup findings in the conclusions.

The number of reported subgroup estimates ranged from 2 to 20, whereas the number of corresponding pairwise comparisons ranged from 1 to 53 (Supplemental Table 1). There were 187 unique subgroup comparisons that resulted in a median log-difference of 0 (IQR 0.12 – 0.17), ranging between -2.92 and 2.32 (Figure 2).

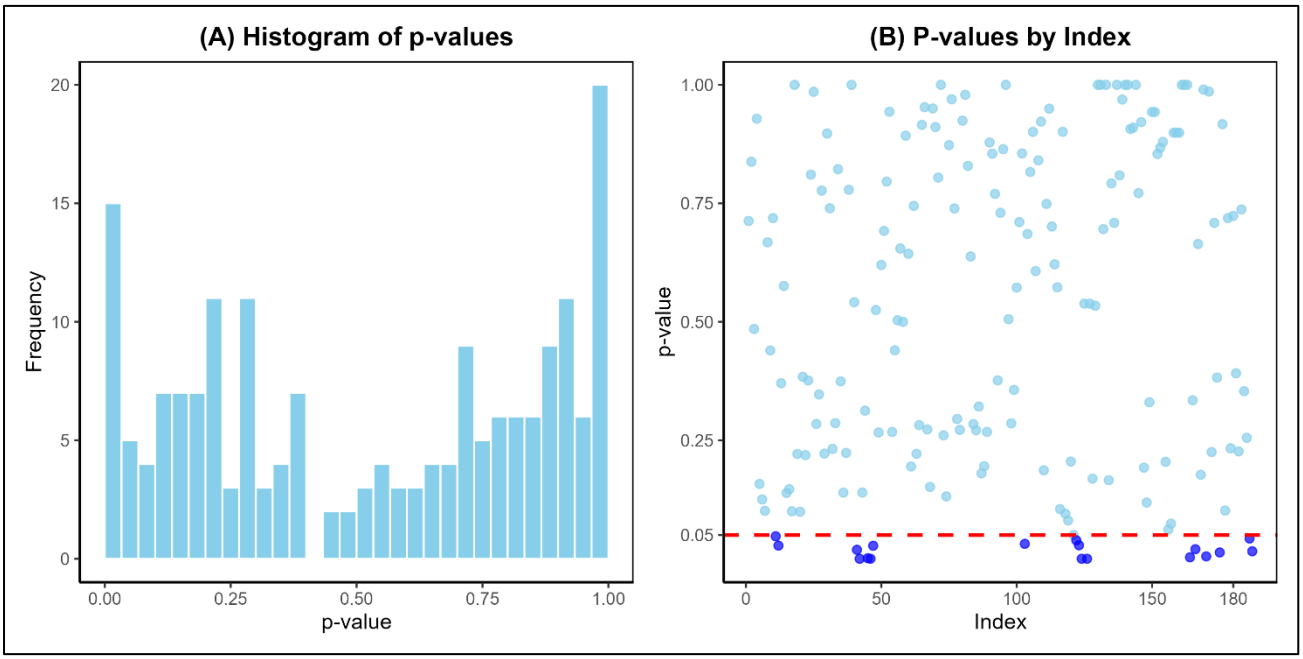
Figure 2. Distribution of the log-scale difference between subgroup comparisons.



CAPTION: Index stands for the unique comparison; log-difference was calculated as (β1 - β2); 95% CI was calculated as log-difference ± 1.96 SE; SE was the square root of [(SE1)2 + (SE2)2].

The corresponding median p-value was 0.54 (IQR 0.21-0.85) (Figure 3A). Of the 187 comparisons, 18 (9%) revealed a p-value lower than the conventional level of significance, whereas 6 (3%) and 21 (11%) contrasts were $0.05 < p \leq 0.10$ and $0.10 < p \leq 0.20$, respectively (Figure 3B). Furthermore, when adjusting for multiple comparisons within a study, only seven (4%) comparisons resulted in a p-value lower than the corrected significance level.

Figure 3. Distribution of p-values comparing log-scale subgroup differences.



CAPTION: Index stands for the unique comparison; (A) Histogram of p-values for log-transformed differences between subgroups; (B) Scatter plot of individual p-values by index, with a horizontal line at the conventional significance threshold (0.05); The p-values for each index were calculated using the test score and standard normal distribution.

Discussion

This study identified several issues in the current reporting practices in meta-analytic research. Our main findings are following:

- A quarter of all meta-analyses did not use interaction tests when comparing subgroups;
- A post-hoc comparisons showed statistical evidence in 9% of cases;
- Tests corrected for multiple comparisons showed that 4% of the contrasts had statistical evidence of differences;

The interpretation of subgroup effects was mostly ignorant of the appropriate statistical tests. Most subgroup differences did not reveal statistical evidence. There is also a lack of corresponding statements in the Methods section and inappropriate mentions in the Conclusions section regarding subgroup effects.

Running numerous subgroup analyses to explore heterogeneity remains questionable (1, 31). Seven (50%) studies used subgroup analysis to explore the causes of heterogeneity (19, 21, 22, 24, 27, 29, 30). On the one hand, the existing PRISMA statement (13) needs the specification of methods to analyze heterogeneity, implying a subgroup analysis or meta-regression. However, the literature suggests that reliable inferences can only be drawn from pre-specified analyses (1). Furthermore, explorations of heterogeneity upon the heterogeneity is observed (which was mostly the case) can only enable the generation of hypotheses rather than making any conclusive claims (1). This also calls into question the evaluation of heterogeneity when the number of primary studies is limited (1, 32). These controversies might affect reporting practices and highlight the importance of clarifying when stratification analyses are appropriate and when they are not.

There is a suspicion that researchers may selectively report subgroup analyses, making it difficult to understand how many "less interesting" findings were left out (4, 6, 31). Hence, when conducting a test with a 95% confidence level, the one remains with a 5% chance of a Type I error. For instance, one study (17) had 30 comparisons. Conducting such a number of tests at the

conventional significance level suggests a 79% probability of at least one contrast falsely rejecting the null. Thus, under the assumption of a 5% Type I error rate, we would anticipate 1.5 ≈ 2 falsely observed "statistically significant" differences. Indeed, the crude comparison revealed two indices with statistical evidence of difference, which is exactly the same as we would expect by random chance. However, when applying the Bonferroni correction for this study (0.05/30), no contrasts were lower than the adjusted significance level. This underscores the importance of transparent reporting of the number of comparisons made, as well as correcting for multiple comparisons to avoid inflation of Type I error.

Most subgroup analyses from systematic reviews are limited by between-study comparisons (3, 8). However, this is an exclusion for the meta-analysis of IPD. Two studies incorporated IPD (17, 29). However, the authors did not formally test the differences between the meta-analysis of aggregated data and the meta-analysis of IPD, which may limit the implications of the findings.

This study represents a brief overview of a current reporting practice, providing a tutorial for comparing subgroup effects. However, our literature search was restricted to a single journal. Therefore, our findings may only be applicable to publications in the field of epilepsy research. Our results are solely based on available reports; some studies may not have fully disclosed data on the outcomes of subgroup analyses. We also focused only on studies that used ratio measures, which represented 25% of all meta-analyses. It would be interesting to obtain data on other measures of association and determine how they would affect our findings and suggestions. Finally, we observed a handful of comparisons that provided statistical evidence of differences. However, a large effect with a large p-value may have more practical importance than a small effect with a small p-value (33). For instance, our (exponentiated) differences ranged from 0 to 10, indicating the possibility of meaningful effects, albeit not being "statistically significant." Furthermore, included studies differed in research areas. Therefore, our findings are by no means to be interpreted as lacking in practical relevance; rather, we aimed to draw attention to the importance of formal statistical tests when inferring from subgroups.

Conclusion

We searched the content of leading ILAE journal over the past decade. We found that all the findings from the subgroup analyses were interpreted separately without employing appropriate statistical tests. The vast majority of our subgroup comparisons resulted in weak evidence against the null hypothesis of no difference. Medical journals may require both researchers and reviewers to be aware of the importance of adhering to reporting principles to improve the implications of their findings.

Supplementary Materials

Supplementary Materials include:

- R code used for data analysis
- Supplemental Table 1. Characteristics of included studies.

Supplementary file available via: https://www.journalehdi.com/suppfile/730/Supplemental-materials.docx

Acknowledgments

Author Contributions: Conceptualization: R.A.; methodology and data curation: G.K., D.K.; formal analysis: R.A. and Z.U.; writing – original draft preparation: R.A., G.K.; writing – review and editing; Z.U.

Data availability statement: Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Statement of Ethics: The study approval was not required, since this was an analysis of published studies.

Conflict of Interest Statement: The authors declare that they have no competing interests.

Funding Sources: This work did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Deeks JJ. Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.0 [updated July 2019]. Cochrane; 2019. Available from: www.training.cochrane.org/handbook
2. Collaborative overview of randomised trials of antiplatelet therapy--I: Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. Antiplatelet Trialists' Collaboration. BMJ. 1994 Jan 8;308(6921):81-106. https://doi.org/10.1136/bmj.308.6921.81
3. Sun X, Ioannidis JP, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis: users' guide to the medical literature. JAMA. 2014 Jan 22;311(4):405-11. https://doi.org/10.1001/jama.2013.285063
4. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet. 2000 Mar 25;355(9209):1064-9. https://doi.org/10.1016/s0140-6736(00)02039-0
5. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. BMJ. 2010 Mar 29;340:c117. https://doi.org/10.1136/bmj.c117
6. Altman DG. Clinical trials: subgroup analyses in randomized trials--more rigour needed. Nat Rev Clin Oncol. 2015 Sep;12(9):506-7. https://doi.org/10.1038/nrclinonc.2015.113
7. Ioannidis JP. Microarrays and molecular research: noise discovery? Lancet. 2005 Feb 5-11;365(9458):454-5. https://doi.org/10.1016/S0140-6736(05)17856-4
8. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med. 1992 Jan 1;116(1):78-84. https://doi.org/10.7326/0003-4819-116-1-78
9. Shorvon S. The early history (1909-1961) of Epilepsia, the journal of the International League Against Epilepsy, and its echoes today. Epilepsia. 2007 Jan;48(1):1-14. https://doi.org/10.1111/j.1528-1167.2006.00934.x
10. Hoffmann F, Allers K, Rombey T, Helbach J, Hoffmann A, Mathes T, et al. Nearly 80 systematic reviews were published each day: Observational study on trends in epidemiology and reporting over the years 2000-2019. J Clin Epidemiol. 2021 Oct;138:1-11. https://doi.org/10.1016/j.jclinepi.2021.05.015
11. Papageorgiou SN, Tsiranidou E, Antonoglou GN, Deschner J, Jäger A. Choice of effect measure for meta-analyses of dichotomous outcomes influenced the identified heterogeneity and direction of small-study effects. J Clin Epidemiol. 2015 May;68(5):534-41. https://doi.org/10.1016/j.jclinepi.2014.11.013
12. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015 Jan 1;4(1):1. https://doi.org/10.1186/2046-4053-4-1
13. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021 Mar 29;372:n71. https://doi.org/10.1136/bmj.n71
14. Altman DG, Bland JM. How to obtain the P value from a confidence interval. BMJ. 2011 Aug 8;343:d2304. https://doi.org/10.1136/bmj.d2304
15. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. BMJ. 2003 Jan 25;326(7382):219. https://doi.org/10.1136/bmj.326.7382.219
16. Borenstein M, Higgins JP. Meta-analysis and subgroups. Prev Sci. 2013 Apr;14(2):134-43. https://doi.org/10.1007/s11121-013-0377-7
17. Halliday A, Duncan A, Cheung M, Boston RC, Apiwattanakul M, Camacho X, et al. Second-line immunotherapy and functional outcomes in autoimmune encephalitis: A systematic review and individual patient data meta-analysis. Epilepsia. 2022 Sep;63(9):2214-24. https://doi.org/10.1111/epi.17337
18. Ferlazzo E, Gasparini S, Beghi E, Sueri C, Russo E, Leo A, et al. Epilepsy in

cerebrovascular diseases: Review of experimental and clinical data with meta-analysis of risk factors. Epilepsia. 2016 Aug;57(8):1205-14. https://doi.org/10.1111/epi.13448

19. Khoudari H, Alabbas M, Tobochnik S, Burneo J, Cox B, Lemus HN. Seizure outcomes after resection of temporal encephalocele in patients with drug-resistant epilepsy: A systematic review and meta-analysis. Epilepsia. 2025 Apr;66(4):945-54. https://doi.org/10.1111/epi.18130
20. Lezaic N, Gore G, Josephson CB, Wiebe S, Jetté N, Keezer MR. The medical treatment of epilepsy in the elderly: A systematic review and meta-analysis. Epilepsia. 2019 Jul;60(7):1325-40. https://doi.org/10.1111/epi.16063
21. Siriratnam P, Foster E, Shakhatreh L, Neal A, Carney PW, Jackson GD, et al. The effect of epilepsy surgery on productivity: A systematic review and meta-analysis. Epilepsia. 2022 Apr;63(4):789-811. https://doi.org/10.1111/epi.17185
22. Guo X, Wang C, Long D, Zhang H, Yin S, Peng X, et al. Clinical phenotypic spectrum of NRXN1 microdeletions and their association with epilepsy: A systematic review and meta-analysis. Epilepsia. 2025 Jun;66(6):2022-35. https://doi.org/10.1111/epi.18291
23. Lattanzi S, Trinka E, Striano P, Zaccara G, Del Giovane C, Nardone R, et al. Cannabidiol efficacy and clobazam status: A systematic review and meta-analysis. Epilepsia. 2020 Jun;61(6):1090-8. https://doi.org/10.1111/epi.16540
24. Rafati A, Jameie M, Amanollahi M, Pasebani Y, Salimi N, Kazemi MHF, et al. Association of new onset seizure and COVID-19 vaccines and long-term follow-up: A systematic review and meta-analysis. Epilepsia. 2024 Oct;65(10):3010-27. https://doi.org/10.1111/epi.18076
25. Mariajoseph FP, Chen Z, Sekhar P, Rewell SS, O'Brien TJ, Antonic-Baker A, et al. Incidence and risk factors of posttraumatic epilepsy following pediatric traumatic brain injury: A systematic review and meta-analysis. Epilepsia. 2022 Nov;63(11):2802-12. https://doi.org/10.1111/epi.17405
26. Kobulashvili T, Kuchukhidze G, Brigo F, Zimmermann G, Höfler J, Leitinger M, et al. Diagnostic and prognostic value of noninvasive long-term video-electroencephalographic monitoring in epilepsy surgery: A systematic review and meta-analysis from the E-PILEPSY consortium. Epilepsia. 2018 Dec;59(12):2272-83. https://doi.org/10.1111/epi.14594
27. Vary O'Neal A, Tamani I, Mendo CW, Josephson CB, Burneo JG, Steven DA, et al. Epilepsy surgery in adults older than 50 years: A systematic review and meta-analysis. Epilepsia. 2024 Jun;65(6):1548-59. https://doi.org/10.1111/epi.17995
28. Johnson BJ, Jung KE, MacKenzie MA, Bah A, Jetté N, Mohamed N, et al. Association of social determinants of health with first antiseizure medication prescription for patients with newly diagnosed epilepsy: A systematic review and meta-analysis. Epilepsia. 2025 May;66(5):1560-72. https://doi.org/10.1111/epi.18251
29. Goel K, Pek V, Shlobin NA, Chen JS, Wang A, Ibrahim GM, et al. Clinical utility of intraoperative electrocorticography for epilepsy surgery: A systematic review and meta-analysis. Epilepsia. 2023 Feb;64(2):253-65. https://doi.org/10.1111/epi.17475
30. Antaya TC, Espino-Alvarado PH, Oiamo T, Wilk P, Speechley KN, Burneo JG. Association of outdoor air and noise pollution with unprovoked seizures and new onset epilepsy: A systematic review and meta-analysis. Epilepsia. 2024 Jul;65(7):1847-67. https://doi.org/10.1111/epi.18023
31. Kaufman JS, MacLehose RF. Which of these things is not like the others? Cancer. 2013 Dec 15;119(24):4216-22. https://doi.org/10.1002/cncr.28359
32. Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. BMJ. 2012 Mar 15;344:e1553. https://doi.org/10.1136/bmj.e1553
33. Lemoine NP, Hoffman A, Felton AJ, Baur L, Chaves F, Gray J, et al. Underappreciated problems of low replication in ecological field studies. Ecology. 2016 Oct;97(10):2554-61. https://doi.org/10.1002/ecy.1508