

Is the effect small or large?

MEDISIN OG TALL

STIAN LYDERSEN

E-mail: stian.lydersen@ntnu.no

Stian Lydersen, dr.ing. and professor of medical statistics at the Regional Centre for Child and Youth Mental Health and Child Welfare (RKBU Central Norway) at the Department of Mental Health, Norwegian University of Science and Technology.

The author has completed the ICMJE form and declares no conflicts of interest.

How do we quantify the results of a study? Is the effect measured on the original scale or a standardised effect size most relevant?

Reindal and colleagues (1) studied age for onset of independent walking. For children diagnosed with autism spectrum disorder, the mean age (standard deviation) was 14.74 (4.28) months, and for children without autism spectrum disorder, it was 13.76 (2.88) months. The difference was therefore $14.74 - 13.76 = 0.98$ months. This is the effect size measured on the original scale, also called unstandardised effect size. In addition, the authors report a standardised effect size as this difference divided by the standard deviation in the comparison group, that is, $0.98 / 2.88 = 0.34$ (see figure 1). Which of these measures is most relevant?

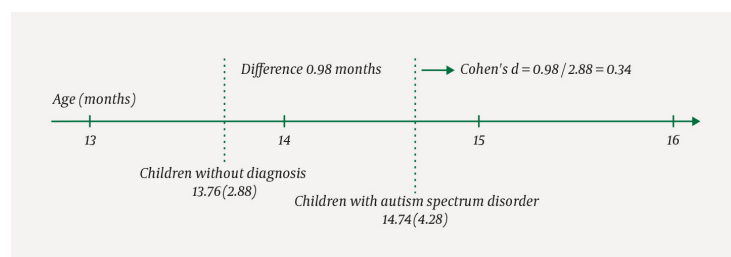


Figure 1 The mean (standard deviation) of age for onset of independent walking among 376 children with autism spectrum disorder and 114 children without this diagnosis (1). The difference was 0.98 months, which corresponds to Cohen's $d = 0.34$.

What is effect size?

The term *effect size* is not precise. Some authors use this term for Cohen's d or a related measure such as Glass's delta or Hedges' g (2). These are the difference between two means, divided by a standard deviation, and are examples of standardised effect sizes. Other examples of standardised effect sizes are the Pearson correlation coefficient, the standardised regression coefficient in linear regression, and partial eta squared in analyses of variance (ANOVA).

In the behavioural sciences, it is not uncommon to report standardised effect sizes. But what role do they actually have? Researchers who report standardised effect sizes usually

refer to the book *Statistical Power Analysis for the Behavioral Sciences* by Jacob Cohen (1923–1998) (3, 4). In this book, Cohen introduces standardised effect sizes as the basis for computing power or sample size in a future study, but he does not discuss other applications of standardised effect sizes.

After a study has been carried out, the choice of a relevant effect size depends on the context. Examples of unstandardised effect sizes are the difference between two means, the unstandardised regression coefficient, the odds ratio, or the risk difference. Several authors recommend in general to report unstandardised effect sizes (5, 6). Further discussions on unstandardised and standardised effect sizes are given in (7) and (8).

Cohen classifies Cohen's d as small, medium, and large if it equals 0.2, 0.5, or 0.8 (4, p. 26). Other authors classify standardised effect sizes in intervals, and partly somewhat differently from Cohen, see for example (4, p. 79–80 and (9, p. 123). Classifying standardised effect sizes can be useful when calculating power or sample size for a future study, but several authors find such classifications to have little relevance for observed effect in a completed study (5, 8).

Unstandardised is easy to understand

A difference in age for onset of independent walking of 0.98 months between two groups is easy to understand. Does the standardised effect size Cohen's $d = 0.34$ provide any additional clinically relevant information? Standardised effect sizes can be useful as a basis for power or sample size calculation for a future study, and they can also be useful input in meta-analyses, but otherwise, standardised effect sizes seem to have little relevance.

REFERENCES:

1. Reindal L, Nærland T, Weidle B et al. Age of first walking and associations with symptom severity in children with suspected or diagnosed autism spectrum disorder. *J Autism Dev Disord* 2019; 49: 1–17. [PubMed][CrossRef]
2. Grissom RJ, Kim JJ. *Effect sizes for research. Univariate and multivariate applications*. 2nd ed. New York, NY: Routledge, 2012.
3. Cohen J. *Statistical power analysis for the behavioral sciences*. 1st ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.
4. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
5. Pek J, Flora DB. Reporting effect sizes in original psychological research: A discussion and tutorial. *Psychol Methods* 2018; 23: 208–25. [PubMed][CrossRef]
6. Baguley T. Standardized or simple effect size: what should be reported? *Br J Psychol* 2009; 100: 603–17. [PubMed][CrossRef]
7. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. *J Exp Psychol Gen* 2012; 141: 2–18. [PubMed][CrossRef]
8. Kelley K, Preacher KJ. On effect size. *Psychol Methods* 2012; 17: 137–52. [PubMed][CrossRef]
9. Campbell MJ, Swinscow TDV. *Statistics at square one*. 11th ed. Wiley-Blackwell, 2009.

Published: 18 February 2020. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.19.0665

© The Journal of the Norwegian Medical Association 2020. Downloaded from tidsskriftet.no