Statistical power and effect sizes in psychology are decreasing over time

Felix Singleton Thorn, Paul Dudgeon and Fiona Fidler
The University of Melbourne

Background: Statistical power describes the probability of a statistical test returning statistically significant results given a specific alternative hypothesis. In order to know whether the average statistical power of psychology research is changing over time, we need to establish how the average statistical power of psychology is changing for a consistent effect size (study 1) and whether effect sizes are changing over time (study 2).

Study 1: Estimating the statistical power of psychology research

Using a systematic review protocol we identified 46 studies published from 1962 to 2017 that calculated the statistical power of statistical tests in over 8000 articles from various areas of psychology research at Cohen's (1988) effect size benchmarks (e.g., correlations of $r = .1$, $r = .3$, and $r = .5$). **Analysis:** Using mixed effects meta-regression (including random effects for area of research, study, and individual estimates) we estimated the average statistical power of psychology research and the change in statistical power over time.

Results: The average statistical power of psychology research at Cohen's (1988) benchmarks is extremely low for 'small' effects, .23 (95% CIs [.17, .29]), somewhat low for 'medium' effects, .62 (95% CI [.54, .70]), and only acceptably high for 'large' effects, .80 (95% CI [.68, .92]). **Change over time:** Average statistical power at these benchmarks has remained relatively stable over time, with an estimated change per year of -0.000 (95% CI [-0.003, 0.003]), 0.001 (95% [-0.002, 0.004]), and -0.001 (95% [-.002, 0.001]) at the small, medium, and large benchmarks respectively.
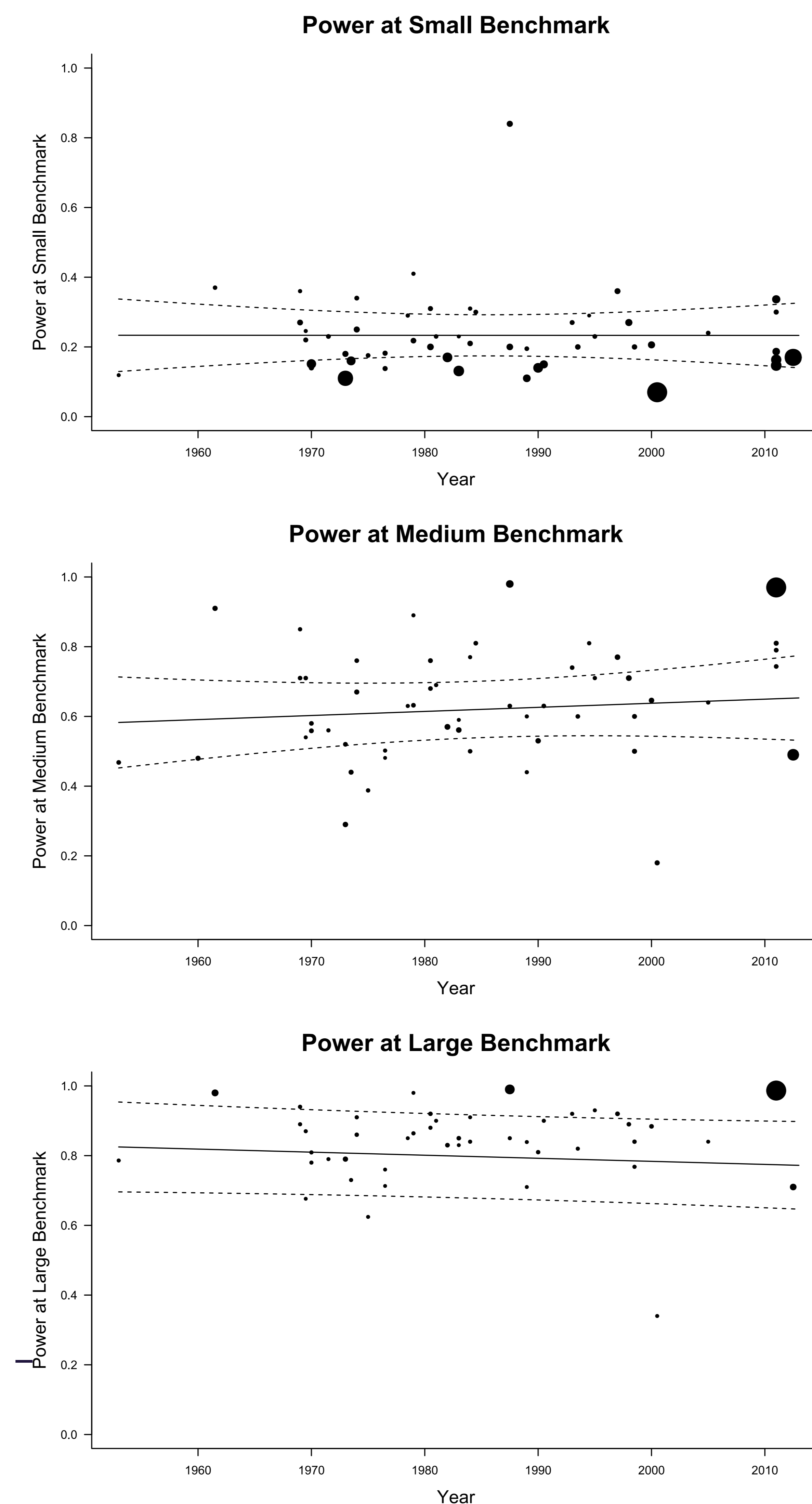

Figure 1. Statistical power over time

Study 2: Estimating the change in effect sizes over time

Using a database of over 130,000 APA formatted statistical tests from over 9,000 articles published in 5 APA journals from 1985 to 2013 (Nuijten, Hartgerink, van Assen, Epskamp, & Wicherts, 2015), we converted all statistical tests results to a common effect size and estimated the average effect change over time using mixed effects meta-regression (with random effects for effect, article, and journal).

Results: This analysis suggests that the average effect size reported in psychological research is decreasing by -0.004 (95% CI [-0.005, -0.004]) Fisher Z units per year, representing an estimated correlation coefficient decrease of -0.1 from 1985 to 2013.

However, it is unclear whether these results apply to the focal analyses or not. Reanalyzing the data examining just the first reported effect shows a smaller yearly decrease of -0.003 (95% CI [-0.004, -0.002]), and reanalyzing the largest reported effect in each paper shows a slight increase over time of 0.0015 (95% CI [0.0002, 0.0028]).
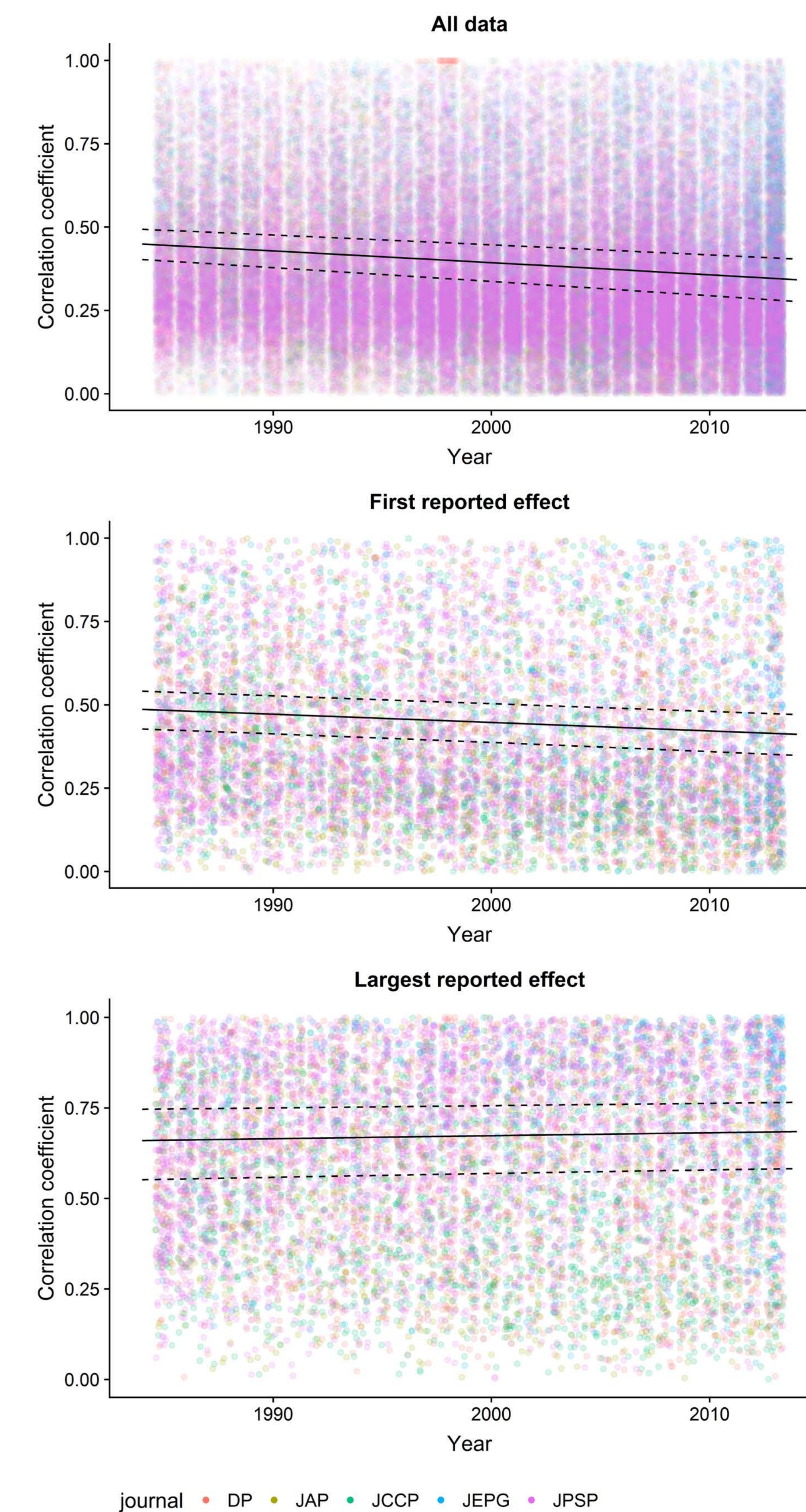

Figure 2. Effect sizes over time

Conclusion

The average statistical power of psychology research at Cohen's effect size benchmarks is lower than desirable and has been relatively stable over time. However, effect sizes reported in published psychology research appear to be becoming smaller over time. This suggests that the average statistical power of statistical tests reported in psychology journals is in fact decreasing. Future research is required to establish whether this result applies to tests of articles' focal research questions, or just at the level of all reported statistical tests.

To read the full articles go to osf.io/g7chn or follow the QR code

References:
Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, New Jersey: Erlbaum.

Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2015). The prevalence of statistical reporting errors in psychology (1985–2013). Behavior Research Methods, 1-22. doi:10.3758/s13428-015-0664-2