



COMMENTARY

# The Decline of Effect Size in Psi Research: An Evidence-Based Commentary on Tressoldi and Storm (2024)

**Dick J. Bierman**

Universities of Amsterdam, Utrecht  
& Groningen, The Netherlands  
d.j.bierman@mac.com

**James Spottiswoode**

Beverly Hills, USA  
james.spottiswoode@gmail.com

## ABSTRACT

We analyzed all available meta-analytic databases covering 10 protocols that are being used in psi research for a decline in effect size over years. From the total of 10 protocols seven showed a significant decline. An analysis using the first 20 publications (or the maximum number of publications) for each protocol gave a clear overall decline in effect size with publication sequence number with a specific protocol ( $\rho = -0.176$   $N = 185$   $P = 0.008$ ). The apparent contradiction of strong declines of Effect Size vs. Year when analyzed per protocol and no decline over-all is explained by the repeating pattern of rise and fall of results obtained by the different protocols with the more sensitive protocols with a larger mean effect size appearing later in years.

We also independently found a very robust decline in effect size with increasing number of trials,  $N$ , or power of the psi studies ( $\rho = -0.397$ ,  $N = 873$ ,  $P < 10^{-30}$ ). Since the power of studies also turned out to increase over the years we analyzed the temporal decline of ES over years controlling for  $N$  and found that the temporal declines were not a consequent of the independent decline effect of ES with increasing  $N$ .

<https://doi.org/10.31275/20264003>

GOLD OPEN ACCESS



Creative Commons License 4.0.  
CC-BY-NC. Attribution required.  
No commercial use.

## KEYWORDS

Decline Effect, meta-analysis, psi research.

## INTRODUCTION

For a background of research efforts into declines we refer to the introduction in the Tressoldi & Storm (hereafter TS) article (Tressoldi & Storm, 2024). The decline effect has been discussed in “psi” research since the early times of experimental investigations (Bierman, 2003; Colborn, 2007). The availability of meta-analytic databases allowed TS to do a robust analysis over a part of published peer reviewed psi studies ( $N = 406$ ), concluding that apart from the “Remote Viewing” protocol this idea of

such a decline was a myth. Indeed the overall analysis using all available experimental data did not show a statistically significant decline.

The goal of this response is not to give an explanation for these declines but to correct the false conclusions that TS drew in their JSE articles. We also refer to the seminal article by Protzko and Schooler (Protzko & Schooler, 2017) where they give several potential explanations for decline effects found in mainstream psychology research. One of the most prominent among those is the “regression to the mean (of 0)” with a

good second place for “publication bias”. No wonder that declines have been a hot topic in psi research circles.

## METHOD

All meta-analytic databases that the authors were aware of, including the databases originally used by TS, were integrated in a single database of all known psi studies published in peer reviewed journals. The resulting database is twice as large as the total of databases used in the original TS analysis. (N-studies=874)

## Included Meta-Analyses

### Addition Criteria

We added any meta-analytic database of published psi experiments that we could locate, resulting in augmenting the 5 databases that were used by TS with five new databases that used the following protocols:

1. deviceAC: Studies like Telephone Telepathy, SMS telepathy & email telepathy. (Sheldrake et al, 2025)
2. DMILS: Direct Mental influence on Living systems (Schmidt et al, 2004)
3. Remote Helping studies (Schmidt, 2012)
4. Remote Staring studies (Schmidt et al, 2004)
5. rngPK: Psychokinesis on RNG’s studies (Bösch et al, 2006)

While the originally used protocols used by TS all were meant to investigate anomalous cognitive phenomena like telepathy, clairvoyance or precognition, the added protocols for the updated analysis also cover psychokinesis on material or biological systems, also known as Mind over Matter phenomena.

## Decline Effect Estimates

We use the non-parametric Spearman’s rho correlation between Effect Size (ES) and Year of Publication throughout to quantify the decline. Parametric analyses require assumptions about the underlying distributions of the variables used. In the 5 protocols analyzed by TS both the distributions of Effect Size as well as the distribution of Publication Years are not normal (ES: Shapiro-Wilk  $W = 0.859$ ,  $P < 0.001$ ; YEAR: Shapiro-Wilk  $W = 0.950$ ,  $P < 0.001$ ). So it is questionable to use parametric measures for decline like the slope of linear regression measure that TS used.

## RESULTS

### Temporal Decline of Effect Size Over the Years

In Table 1 we present a survey of the correlations per protocol.

As can be seen 7 out of 10 protocols show declines. Interestingly when all data are pooled there is a weak incline! In the discussion we will show how both these apparently conflicting results can be true.

### Decline of Effect Size with Increasing Sample Size (or Power)

In Table 2 we present the results of another hitherto not often discussed decline, the decline of Effect Size with increasing statistical power of a study (represented by N. the sample size of the study.)

The mean Sample Size (or power) of studies within a protocol does increase over the years which may give a

**Table 1.** Correlation Coefficients for Effect Size Versus ‘Year of Publication’.

Protocol	Psi type	Years	N Studies	Mean ES	Rho	P (one-tailed)
device-AC	AC	[2003-2023]	26	.162	-0.62	< 0.001
DMILS	PK	[1976-2000]	40	.177	-0.36	0.01
Dreams	AC	[1966-2014]	50	.196	-0.28	0.02
Forced Choice	AC	[1987-2022]	137	.015	0.10	0.88
Ganzfeld	AC	[1974-2020]	113	.137	-0.05	0.30
Presentiment	AC	[1997-2018]	62	.262	0.03	0.59
Remote Helping	PK	[1995-2006]	11	.127	-0.64	0.02
Remote Staring	PK	[1989-1998]	15	.317	-0.7	0.002
Psychokinesis on RNG	PK	[1969-2004]	380	.020	-0.26	< 0.001
Remote Viewing	AC	[1974-2021]	40	.425	-0.45	< 0.002
ALL PROTOCOLS	mixed	[1966-2023]	874	.098	+0.07	0.98

AC = Anomalous Cognition studies PK = Psychokinesis studies



**Table 2.** Correlation Coefficients for Effect Size Versus Total Number of Trials in the Study.

Protocol	No. of Studies	Rho	P one-tailed*
device-AC	26	-0.44	0.01
DMILS	40	-0.24	0.07
Dreams	50	-0.24	0.05
Forced Choice	137	-0.06	0.24
Ganzfeld	113	-0.02	0.42
Presentiment	62	0.1	0.78
Remote Helping	11	-0.54	0.04
Remote Staring	15	-0.5	0.03
Psychokinesis on RNG	380	-0.33	< 0.001
Remote Viewing	40	-0.45	< 0.002
ALL PROTOCOLS**	874	-0.40	< 10 <sup>8</sup>

\* one-tailed testing is allowed because the directional decline hypothesis was pre-registered. (<https://osf.io/29fvy>; see H1 predicting a decline).

hint as to how to explain the decline over year of the Effect Size as indirectly caused by the temporal incline of mean sample size and the decline of ES with larger sample size. The correlations showing the increase in power over years are given in Table 3.

## DISCUSSION

### Why the Difference Between TS Analysis and the Extended Current Analysis?

What is the explanation for the different outcomes of TS's analysis and ours? Firstly TS used a parametric method to quantify the decline. This is dubious as both underlying distributions are far from normal with the distribution of Effect Sizes being strongly skewed and having an extreme outlier (the ES value of 1.6 in one of the RV studies is, more than 5 standard deviations from the mean). Secondly, in their analysis they used only meta-analytic databases with protocols intended to measure different aspects of Anomalous Cognition (AC). We added five protocols from which only one (deviceAC, like phone telepathy) measured AC, the others measured different aspects of Psychokinesis on biological and material systems. We compared the correlation coefficients for both types of psi to see if possible the decline was mostly or only present in studies that measured aspects of PK. We compared the correlation coefficients obtained for both types of psi. The mean correlation coefficients of Effect Size with Year are -0.21 for Anomalous Cognition and -0.49 for PK studies. The correlations or AC protocols do not differ from the correlations

**Table 3.** Correlation Between Mean Number of Trials Per Study, Thus Power, and the Year of Publication.

Protocol	No. of Studies	Rho	P two-tailed
device-AC	26	0.26	0.20
DMILS	40	0.2	0.22
Dreams	50	0.54	5.2x10 <sup>-5</sup>
Forced Choice	137	-0.03	0.73
Ganzfeld	113	0.25	0.008
Presentiment	62	0.29	0.022
Remote Helping	11	0.23	0.50
Remote Staring	15	0.48	0.070
Psychokinesis on RNG	380	0.39	3.11x10 <sup>-15</sup>
Remote Viewing	40	0.54	0.0003

for PK protocols according to the independent samples Mann-Whitney U test (N=10, W=15, z= -1.492, P=0.136). Interestingly when we compare the mean declines obtained by the analyzers Tressoldi and Storm with the declines obtained in the new protocols analyzed by Bierman & Spottiswoode (which is only slightly different from the comparison AC vs. PK-declines) then the declines show larger differences (N = 10, W = 37, Z = 1.984, P = 0.047 two tailed).

### Can the Decline Over Years be Explained by the Increase in Power Over the Years or Vice Versa?

The strong and surprising findings relating to the Effect Size as function of power of the experiment and to the increase of power within each used protocol over the years suggest that possibly the increase in power is the source of the decline effect over years.

To test the idea that decline over years was indirectly caused by incline in power, we calculated non parametric partial correlations coefficients for all the protocols by controlling for the variable N (Table 4).

The absolute values of most of the partial correlations (partial rho) are smaller than the uncorrected correlations (rho) but the number of negative correlations as well as the number of significant ones remains the same. We conclude that the decline of ES over Years cannot be attributed to the incline in power over Years

### Why is There Not an Overall Protocols Decline?

An apparent inconsistency and possible source of incorrect conclusions is the fact that each of the protocols (except Presentiment) does decline and still, over-all there is even a weak incline. How is that possible? This apparent

inconsistency has been known in the statistical literature as Simpson's Paradox. In Figure 1 we have given an idealized example how this paradox can occur.

Basically, after publication of the first successful results of a new protocol, a number of replications follow with the effect size declining over the subsequent years within that studies using that protocol. In the meantime, another new protocol arises and declines also, but later in time. In this case the declines obscure each other.

To illustrate that this is not just a speculative model we used the first 20 publications within each protocol, or as many as there were available, and calculated the correlation between effect size and the rank number in the sequence of publications. This can be visualized as a shift of the results for different protocols so that the first publication in each protocol is now on the same x-value of the plot. A significant

negative correlation ( $\rho = -0.176$   $N = 185$   $P = 0.008$ ) was found (see Figure 2). This time the decline was found pooling all protocols together. The outcome of this exercise was not sensitive for the number of studies we selected. Twenty is a reasonable choice because then all protocols contribute about the same amount to the analysis.

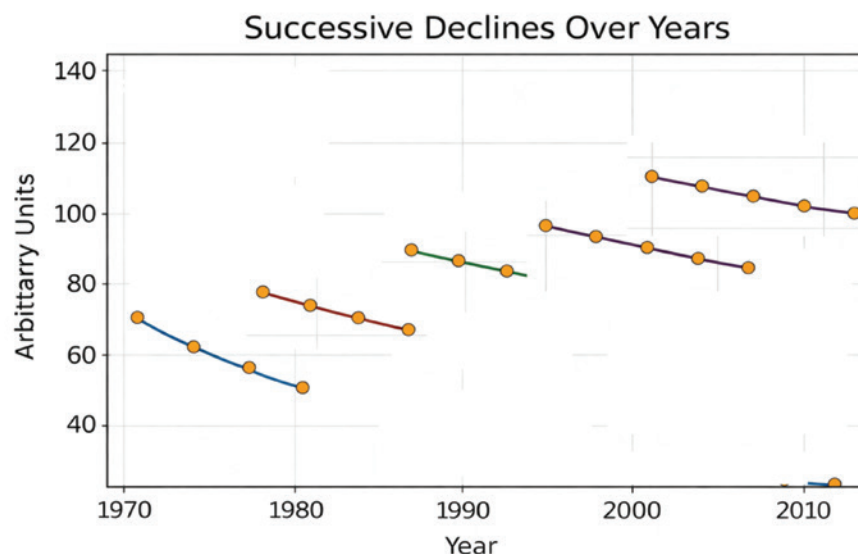
When increasing that number the  $\rho$  becomes just larger. The resulting plot was fitted to a negative potential with an offset. We found a good fit with the offset around  $ES=0.059$ . Remarkably this value is quite close to the Effect Size value that we found for the Ganzfeld database after correction for all the possible Questionable Research Practices, including publication bias. That value was 0.07! (Bierman, Spottiswoode & Bijl, 2016)

## CONCLUSION

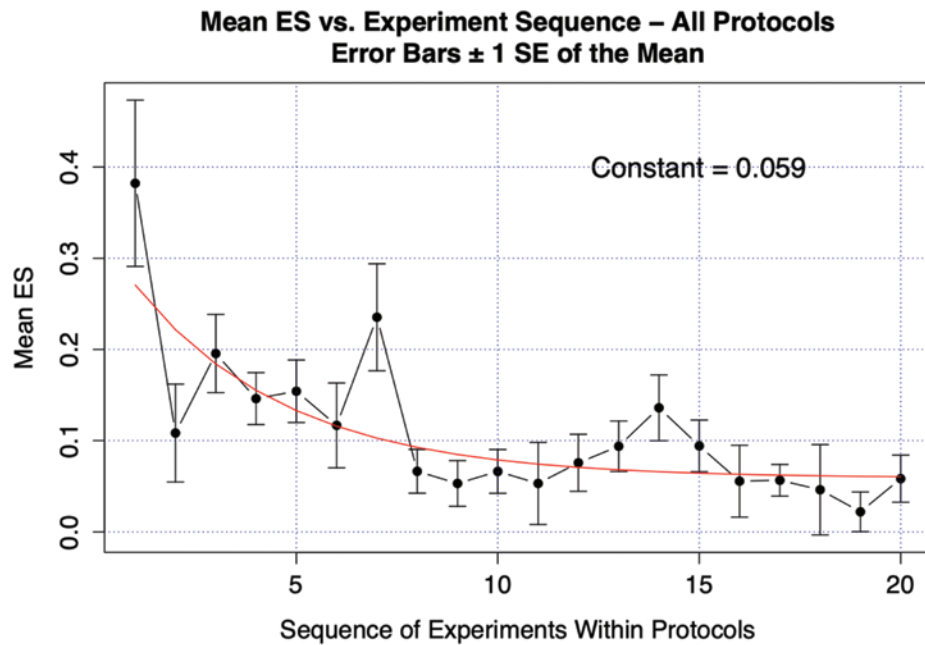
We come to a different conclusion from TS in spite of about half the used data being identical. One of the reasons that was probably instrumental in the final conclusion by TS that the decline is a Myth is that, when all experiments are taken together, there is no decline over years. That may have set the stage for the declaration of decline effects to the status of 'Myth' but using ALL available meta-analytic, ES decline effects cannot be labeled as Myth. Moreover, what TS failed to see is that even their own within protocol analyses gave at least one protocol with considerable decline (RV) and another with a marginal decline (Dream studies, TS  $P = 0.054$ ). Interestingly when using a non parametric method to quantify the

**Table 4.** Non-Parametric Correlations and Partial Correlation Between ES and Year.

Protocol	Rho	Partial Rho
device-AC	-0.617*	-0.581*
DMILS	-0.364*	-0.331*
Dreams	-0.282*	-0.190
Forced Choice	0.105	0.103
Ganzfeld	-0.045	-0.043
Presentiment	0.024	-0.005
Remote Helping	-0.635*	-0.622*
Remote Staring	-0.698*	-0.605*
Psychokinesis on RNG	-0.261*	-0.155*
Remote Viewing	-0.449*	-0.267*



**Figure 1.** Idealized Example of a History of Experimental Publications Effect Sizes Over the Years.



**Figure 2.** Mean ES Over All 10 Protocols as a Function of the Publication Sequence Number within the Respective Protocol.

decline like we did, the dream studies also decline strongly. And when using nonparametric methods, in 4 out of the 5 protocols that were analyzed by TS there is a negative non-parametric correlation between ES and Year. Using the word ‘Myth’ even seems dubious with only the original 5 protocols when using non-parametric methods. However, we conclude after using all available meta-analytic data and non-parametric methods to quantify the decline, that generally there is a definite decline over years within protocols, with the remarkable exception of the presentiment protocol.

Another reason for the conclusion by TS that the decline effect is a Myth may be that TS restricted themselves to AC experiments. We showed here that AC experiments do have a weaker decline than PK experiments.

We do not go into speculations about how to explain the decline of psi research effect sizes over the years (within the protocols) but we recommend as a guideline to study the paper by Protzko and Schooler (2014). Apart from the potential “explanations” given in that paper we think it may be fruitful to consider fundamental physical postulates that prevent the anomalies to be used to create a paradox.

## REFERENCES

Bierman, D. J. (2003). On the nature of anomalous phenomena: Another reality between the world of subjective consciousness and the objective world of physics?

In *The physical nature of consciousness* (pp. 269-292). John Benjamins Publishing Company. <https://doi.org/10.1075/aicr.29.12bie>

Bierman, D. J., Spottiswoode, J. P., & Bijl, A. (2016). Testing for questionable research practices in a meta-analysis: An example from experimental parapsychology. *PloS one*, *11*(5), e0153049. <https://doi.org/10.1371/journal.pone.0153049>

Bösch, H., Steinkamp, F., & Boller, E. (2006). Examining psychokinesis: The interaction of human intention with random number generators--a meta-analysis. *Psychological bulletin*, *132*(4), 497. <https://doi.org/10.1037/0033-2909.132.4.497>

Colborn, M. (2007). The decline effect in spontaneous and experimental psychical research. *Journal of the Society for Psychical Research*, *71*(886), 1-22.

Kievit, R. A., Frankenhuys, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson’s paradox in psychological science: a practical guide. *Frontiers in psychology*, *4*, 513. <https://doi.org/10.3389/fpsyg.2013.00513>

Protzko, J., & Schooler, J. W. (2017). Decline effects: Types, mechanisms, and personal reflections. *Psychological science under scrutiny: Recent challenges and proposed solutions*, 85-107. <https://doi.org/10.1002/9781119095910.ch6>

Schmidt, S., Schneider, R., Utts, J. M., & Walach, H. (2004). Distant intentionality and the feeling of being stared at—Two meta-analyses. *British Journal of Psychology*, *95*, 235–247. <https://doi.org/10.1348/000712604773952449>

- Schmidt, S. (2012). Can we help just by good intentions? A meta-analysis of experiments on distant intention effects. *Journal of Alternative and Complementary Medicine*, 18(6), 529–533. <https://doi.org/10.1089/acm.2011.0321>
- Sheldrake, R., Stedall, T., & Tressoldi, P. (2025). Telecommunication telepathy: A meta-analysis. *Journal of Anomalous Experience and Cognition*, 5(1), 47-69. <https://doi.org/10.31156/jaex.25934>
- Tressoldi, P., & Storm, L. (2024). The myth of the decline effect in psi research: The empirical evidence. *Journal of Scientific Exploration*, 38(3).

