A Meta-Reanalysis of Dream-ESP Studies: Comment on Storm et al. (2017)

Matthew C. Howard

Mitchell College of Business, The University of South Alabama, USA

Summary. Storm, Sherwood, Roe, Tressoldi, Rock and Risio (2017) performed a meta-analysis of laboratory studies of ESP in dreams, finding a moderate overall effect (r = .20). Certain statistical decisions, such as primarily reporting unweighted effects, may have inflated these observations. In the current article, I perform an inverse-variance weighted meta-analysis using the database provided by Storm et al., finding a much smaller overall effect (r = .07). I also find evidence that a significant relationship exists between effect size and sample size, suggesting that the prior results may have been primarily driven by large effects found in small-n studies. I suggest that future researchers of ESP in dreams should produce more large-n studies, which could alleviate many notable criticisms of the current psi literature.

Keywords: Dreaming, extrasensory perception (ESP), meta-analysis

1. Introduction

Storm et al. (2017) supported that laboratory studies of extrasensory perception (ESP) in dreams produce a significant meta-analytic effect, seemingly no matter the studied ESP mode (telepathy, clairvoyance, precognition), research design (Maimonides Dream Laboratory [MDL] studies, Rapid Eye Movement [REM] sleep studies), or other characteristics (author, year). The authors should be commended for their continued investigations into such a sensitive topic, ESP, as well as their adherence to transparent research practices. While both the Meta-Analytic Reporting Standards (MARS) as well as Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommend researchers provide their created database, all too few actually follow these guidelines. Storm et al.'s meta-analysis has notable strengths.

When initially reading Storm et al., I was struck by two unusual features of the meta-analysis. First, the reported meta-analytic effect sizes are very large for psi research. Prior meta-analyses, even when supporting psi, have reported very small effects for retroactive influences (Hedges' g = .09; Bem et al., 2015), precognition (r = .02; Honorton & Ferrari, 1989), mental intention on dice rolls (r = .01; Radin & Ferrari, 1991), and conscious ESP (r = .01; Storm et al., 2012). Storm et al. reported much larger effects for telepathy (r = .22), clairvoyance (r = .18), and precognition (r = .17) in laboratory dream ESP studies. Effects of this size are considered moderate by recent guidelines (Bosco et al., 2015; Gignac & Szodorai, 2016; Paterson et al., 2016) and should be "noticeable to the naked eye of a careful observer" (Gignac & Szodorai, 2016, p. 74). Given that psi research is still heavily doubted in most academic outlets, it is safe to say that psi effects are not typically noticeable to the naked eye of the careful observer.

Second, Storm et al. primarily report unweighted meta-analytic effects, which is generally atypical for meta-analyses. Most meta-analyses weight effect sizes by their associated sample size or inverse variance (Cheung, 2015; Lipsey & Wilson, 2001; Schmidt & Hunter, 2014), which allows a larger study to have a stronger influence on the meta-analytic results than a smaller study. In the majority of Storm et al.'s results, a study with a sample size of two has an equal weight as a study with a sample size of two-hundred. Likewise, other modern meta-analytic approaches could have been applied, such as various outlier identification methods, inverse-variance weighted meta-regressions, and calculations of within-group Qs.

This latter observation may be the source of the former observation, and Storm et al.'s results may be notably different when weighting effect sizes and applying other modern meta-analytic approaches. For this reason, the current article recalculates Storm et al.'s effects using their Appendix A to test the robustness of their results.

2. Methods

Storm et al. provided sample sizes, effect sizes, and partial coding decisions for each of their included studies in Appendix A. Their efforts allow other researchers to reanalyze their results, and thereby no search strategies or coding decisions are discussed in the current article. Readers should refer to Storm et al. for information regarding these aspects of the meta-analysis. I did not contact Storm et al. about any aspect of their meta-analysis, because I wanted to remain as impartial as possible when reanalyzing their meta-analytic effects.

One change was made to the dataset of Storm et al., however. Storm et al. included Watt (2014) in their analyses. Due to a letter to the editor (Mörck, 2015), Watt and Valášek (2015) provided additional observations from Watt (2014) to "bring them out of the file drawer" (p. 106). I want to clarify that Watt (2014) appeared to not engage in unethical research practices, as Watt (2014) provided an initial justification for not including these participants that was justifiably reconsidered. The provision of these additional

Corresponding address:
Matt C. Howard, Assistant Professor, Marketing and Quantitative Methods, The University of South Alabama, 337 Mitchell College of Business, Mobile, Alabama, 36695, USA.
Email: Mhoward@SouthAlabama.edu

Submitted for publication: September 2018
Accepted for publication: October 2018

observations speaks well towards Watt's (2014) willingness to be transparent in the research process. I included these additional observations and recalculated the effect size of Watt (2014).

2.1. Analyses

Analyses were calculated using Comprehensive Meta-Analysis V3 and R 3.4.4. Eight methods were applied to find outliers and influential cases, but I primarily considered three in determining these studies: studentized deleted residuals, Cook's distance, and covariance rations (Viechtbauer & Cheung, 2010). Additional information and R syntax regarding these outlier analyses are provided in Supplemental Material A. After deleting excessive outliers and/or influential cases, I calculated estimates of publication bias. The associated analyses were fail-safe k, Egger's test, and random-effects trim-and-fill method. I also attempted to perform a weight-function model analyses to detect publication bias, but too few studies were included in the current dataset to obtain reliable results, especially for subgroup analyses.

Meta-analytic effects were calculated using a random-effects model with inverse-variance weights, and effects sizes are reported as correlation coefficients (r). A concern with sample-size or inverse-variance weighting is the possibility of one or two large-n studies entirely overshadowing the effects of all other studies. For this reason, I performed sensitivity analyses. I rescaled all studies with a sample size of two standard deviations above the mean (147) to this upper limit, and I recalculated all analyses. No effect size differed by more than .01, and all inferenced remained the same. Thus, it is believed that no excessively large-n studies overpowered the current results, and I therefore report the meta-analytic results without rescaling these studies.

An even more sophisticated approach to meta-analyses is the three-level meta-analysis, which applies a multilevel modeling approach to account for multiple effect sizes from the same sample. Because this was not an issue in Storm et al.'s dataset, I chose not to apply three-level meta-analytic methods. Inverse-variance weighted random-effects meta-regressions were conducted to probe the effect of dichotomous study characteristics (e.g. MDL, REM). Within-group Qs were calculated ($Q_{total} - Q_{between} = Q_{within}$) to compare multiple-category study characteristics (author, publication outlet). This process can indicate whether a statically significant difference exists among many specified groups, and it is analogous to ANOVA.

3. Results

3.1. Outlier Identification

Storm et al. identified two outliers, Studies #2 and #47 (Storm et al., 2017, Appendix A). The initial outlier analyses identified (at least) five possible outliers. In order of Cook's distance values, these were Studies #25, #43, #19, #2, and #47. I chose to remove all five outliers from further analyses. After their removal, the outlier analyses indicated that one large outlier still remained (Van de Castle, 1971). I removed this study, and I provide further justifications for my decision in Supplemental Material B. All results of the current meta-analysis are reanalyzed with Van de Castle (1971) in Supplemental Material C. Also, all analyses did not include Studies #11 and #48 because random-effects meta-analyses with inverse-variance weights cannot include studies with sample sizes of two. This resulted in a dataset of 44 studies.

3.2. Publication Bias

Storm et al. reported a fail-safe N of 110, which suggests that 110 unpublished null results must be discovered for the overall observed effect to no longer be statistically significant. The current analyses found a similar result: the fail-safe N was 93 for the analysis of all studies. However, the fail-safe N was noticeably smaller when the studies were separated by ESP mode: telepathy (0), clairvoyance (7), and precognition (0). This suggests that the current interpretations could be noticeably swayed by unpublished studies.

The publication bias analyses indicated that biases may be present in the dataset. Egger's test was statistically significant for the overall analyses and marginally significant for telepathy studies. The trim-and-fill method suggested that studies may be missing for the overall analyses as well as the study of each ESP mode. Figure 1 provides a visual representation of the trim-and-fill results for the overall analysis, which indicates that the overall effect (provided below) is no longer statistically significant when including implied missing studies.

3.3. Primary Analyses - Replications

The analysis of overall effects included 1,708 trials, which produced an inverse-variance weighted correlation of .07 (z-value = 2.857; p = .004; 95% CI [.02, .12]). Although this result is almost one-third of the previously reported overall effect size (.20), it is nevertheless statistically significant. This effect is no longer statistically significant when including the implied missing studies from the trim-and-fill analysis (point estimate r = .049, 95% CI [-.01, .11]). The analysis of telepathy studies produced an inverse-variance weighted correlation of .07 (z-value = 1.463; p = .144; 95% CI [-.03, .17]). This effect is one-third of the originally reported effect size (.22), and it is not statistically significant. The analysis of clairvoyance studies produced an inverse-

Table 1. Results of Publication Bias Analyses

	k	I^2	Fail Safe k	Egger's Test β_0	Egger's Test t	Implied Missing Left of Mean	Implied Missing Right of Mean
Overall	44	0	93	.466	2.137*	9	0
Telepathy	21	0	0	.398	1.914†	6	0
Clairvoyance	12	26.713	7	-.535	.532	0	3
Precognition	9	27.487	0	.573	.834	1	0

Note: † p < .10, * p < .05, ** p < .01, *** p < .001.

Table 2. Primary Meta-Analytic Findings

	Number of studies	*Sample size*	Inverse-variance weighted effect size	95% Confidence interval	z-value	p-value	Cochran's Q (df)
Overall	44	1,708	.07	.02, .12	2.857	.004**	39.184 (43)
Small n	24	202	.20	.03, .36	2.299	.022*	13.393 (23)
Medium n	13	345	.13	-.01, .27	1.820	.069†	18.683 (12)
Large n	7	1,161	.04	-.02, .10	1.305	.192	2.199 (6)
Telepathy	21	449	.07	-.03, .17	1.463	.144	8.505 (20)
Small n	17	140	.14	-.07, .33	1.309	.191	6.921 (16)
Medium n	3	106	.11	-.09, .30	1.075	.282	.686 (2)
Large n	1	203	.03	-.11, .17	.410	.682	0
Clairvoyance	12	236	.19	.03, .35	2.252	.024*	15.009 (11)
Small n	4	34	.31	-.10, .63	1.495	.135	2.000 (3)
Medium n	8	202	.17	-.03, .36	1.664	.096†	12.673 (7)
Large n	0	0	-	-	-	-	-
Precognition	9	823	.01	-.06, .13	.796	.426	11.032 (8)
Small n	3	28	.39	-.18, .76	1.365	.172	3.337 (2)
Medium n	2	37	-.05	-.64, .57	-.150	.881	3.954 (1)
Large n	4	758	.03	-.04, .10	.863	.388	1.600 (3)

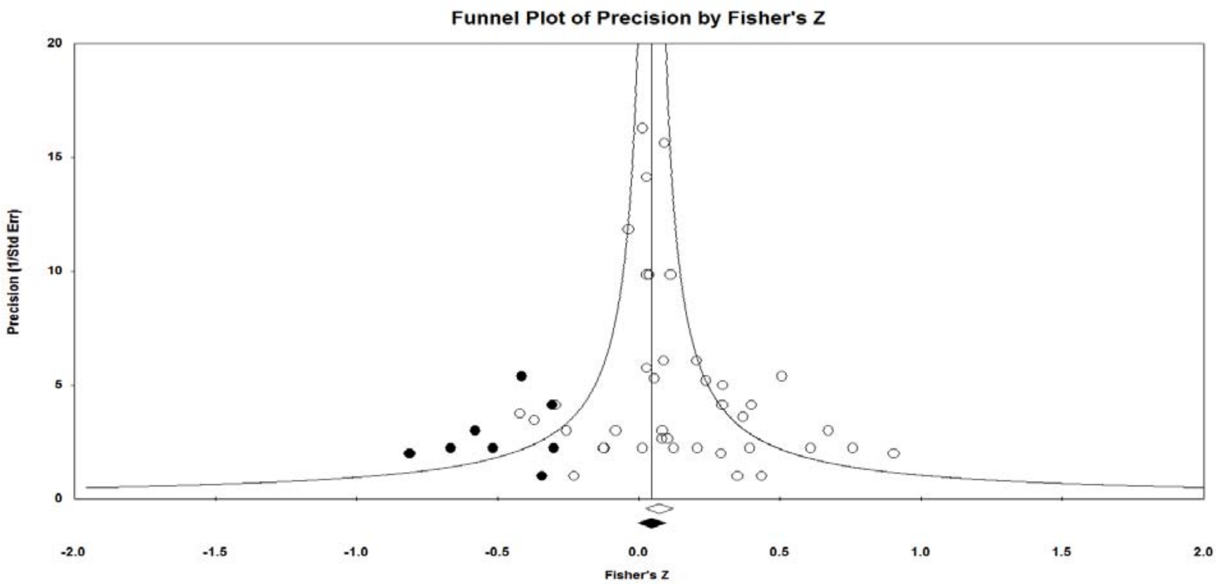
Note: Small n = studies with 15 or fewer participants; Medium n = studies with 15 to 99 participants; Large n = studies with 100 or more participants.
† p < .10, * p < .05, ** p < .01, *** p < .001.

variance weighted correlation of .19 (z-value = 2.252; p = .024; 95% CI [.03, .35]). This effect is similar to the originally reported effect (.18), and it is statistically significant. Lastly, the analysis of precognition studies produced an inverse-variance weighted correlation of .04 (z-value = .796; p = .426; 95% CI [-.06, .13]). This effect is one-fourth of the originally reported effect (.17), and it is not statistically significant.

The significant effect for clairvoyance studies was further investigated for robustness. The effect would no longer be statistically significant (r = .15 – 19, p > .05) when removing any of five studies (Dalton et al., 1999; Dalton et al., 2000; Kanthamani & Khilji, 1990; Roe et al., 2007; Sherwood et al., 2000). This suggests that this significant effect is somewhat precarious.

I also tested whether the meta-analytic effects differed by the ESP mode via calculating a between-group Q. The ESP mode did not have a significant effect (Q_B = 4.358, df = 2, p = .113), suggesting that the meta-analytic effect sizes do not differ by the ESP mode.

Figure 1. Funnel Plot of Observed and Imputed Studies



Note: White circles represent observed studies. Black circles represent imputed studies. The point estimate of the observed values was .072 (95% CI [.02, .12]), whereas the point estimate of the adjusted values is .049 (95% CI[-.01, .11]).

Table 3. Meta-Analytic Findings Separated by Author and Publication Outlet

Analyses Separated by Author	Number of studies	*Sample size*	Inverse-variance weighted effect size	95% Confidence interval	z-value	p-value	Cochran's Q (df)
Braud	2	66	.12	-.13, 36	.950	.342	.657 (1)
Child	2	15	.63	.09, .88	2.219	.027*	.194 (1)
Dalton	3	76	.38	.16, .57	3.292	.001**	.610 (2)
Foulkes	2	16	.00	-.55, .55	.003	.997	.153 (1)
Harley	2	40	.00	-.52, .52	.000	1.000	2.987 (1)
Hearne	6	42	.08	-.31, .45	.406	.685	.510 (5)
Kanthamani	3	34	.28	-.11, .59	1.429	.153	.777 (2)
Krippner	11	294	.07	-.05, .19	1.095	.273	7.484 (10)
Luke	2	411	-.00	-.10, .09	-.079	.937	.220 (1)
Markwick	2	200	.08	-.07, .21	1.055	.291	.276 (1)
Roe	5	138	.06	-.12, .23	.650	.516	3.202 (4)
Other	4	376	.06	-.15, .26	.527	.598	6.901 (3)
Analyses Separated by Publication Outlet	Number of studies	*Sample size*	Inverse-variance weighted effect size	95% Confidence interval	z-value	p-value	Cochran's Q (df)
Annual Convention of the Parapsychological Association	4	50	.30	-.00, .56	1.937	.053†	.836 (3)
Dream Telepathy: Experiments in nocturnal ESP	2	16	.00	-.55, .55	.000	1.000	.155 (1)
European Journal of Parapsychology	2	120	.07	-.12, .26	.742	.458	1.026 (1)
Experimental Medicine & Surgery	2	20	.00	-.55, .55	.002	.999	1.361 (1)
International Journal of Parapsychology	2	19	.15	-.38, .60	.533	.594	.118 (1)
Journal of the American Society for Psychical Research	3	48	.52	.25, .71	3.571	<.001***	.467 (2)
Journal of Parapsychology	2	259	.28	-.25, .68	1.051	.293	2.923 (1)
Journal of the Society for Psychical Research	12	200	.03	-.13, .18	.325	.746	5.687 (11)
Perceptual and Motor Skills	2	16	.00	-.55, .55	.003	.997	.153 (1)
Research in Parapsychology	7	298	.09	-.07, .25	1.079	.281	8.671 (6)
Publication outlet with only one represented study	6	662	.02	-.06, .10	.566	.571	3.127 (5)

Note: † $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Next, I reanalyzed the effect of MDL. The 11 MDL studies produced an inverse-variance weighted effect size of .07 (z-value = 1.095; p = .273; 95% CI [-.05, .19]), which is less than one-fourth of the originally reported effect (.33). The 33 non-MDL studies produced an inverse-variance weighted effect size of .07 (z-value = 2.640; p = .008; 95% CI [.02, .13]), which is half of the originally reported effect (.14). The confidence interval of the non-MDL studies is entirely contained in the MDL studies, indicating that there is not a statistically significant difference in the two groups. This is further supported by the meta-regression that failed to reach statistical significance (Int = .073, B = -.005, SE = .068, p = .941; 95% CI [-.14, .13]).

I also reanalyzed the effect of REM. The 18 REM studies produced an inverse-variance weighted effect size of .06 (z-value = 1.107; p = .268; 95% CI [-.05, .17]), which is one-fourth of the originally reported effect (.24). The 26 non-REM studies produced an inverse-variance weighted effect size of .09 (z-value = 2.481, p = .013, 95% CI [.02, .16]), which is smaller than the originally reported effect (.16). Because the confidence intervals greatly overlapped, there is not a statistically significant difference between the two groups. This is further supported by the meta-regression (Int = .074, B = -.011, SE = .063, p = .856; 95% CI [-.14, .11]).

I performed a meta-regression to identify the impact of publication year, which did not produce a statistically significant result (Int = 2.853; B = -.001, SE = .001, p = .351; 95% CI [-.004, .002]). An additional analyses determined whether effect sizes differed by the author (Table 3). Studies were grouped in the 12 author categories as defined by Storm et al., and the results showed that this grouping approach is not significant (QB = 15.213, df = 11, p = .173).

3.4. Additional Analyses

Analyses were conducted to probe the effect of sample size on study results. Three groups were created that logically appeared in the database. The first included studies with a sample size over 99 (k = 7), the second included studies with a sample size of 15 to 99 (k = 13), and the third included studies with a sample size below 15 (k = 24). The large-n studies produced a very small effect that did not reach statistical significance (r = .039; z-value = 1.305; p = .192; 95% CI [-.04, .09]); the moderate-n studies produced a small effect that was marginally significant ($r = .131$; z-value = 1.820; $p = .069$; 95% CI [-.01, .27]); and the small-n studies produced a moderate effect that was statistically significant ($r = .199$; z-value = 2.299; $p = .022$; 95% CI [.03, .36]). The Begg and Mazumdar rank correlation test, which analyzes the relationship between standard errors and effect sizes, was marginally significant (Kendall's $\tau = .165$, $z = 1.568$, $p = .058$). Because this test is often underpowered (Borenstein, 2005), this result suggests that a notable relationship exists between sample size and effect size in the dataset.

Analyses were conducted to determine whether effect sizes differed by publication outlet. Eleven categories were identified, and the results showed that grouping the studies by outlet did not produce a significant effect ($Q_B = 14.660$, $df = 10$, $p = .145$). However, the 34 studies published in parapsychology journals produced an inverse-variance weighted effect size of .11 (z-value = 3.299; $p = .001$; 95% CI [.04, .17]), whereas the 10 studies published in non-parapsychology journals produced an inverse-variance weighted effect size of .02 (z-value = .543; $p = .587$; 95% CI [-.06, .10]). The moderate overlap in confidence intervals suggests that this difference may be significant, and a meta-regression produced a marginally significance result for this effect (Int = .021, B = .088, SE = .051, $p = .083$; 95% CI [-.01, .19]).

4. Discussion

Several implications of these results should be highlighted. Most effects of Storm et al. become notably smaller in the current reanalysis, some becoming one-third and one-fourth of the originally reported result. By modern standards (Gignac & Szodorai, 2016; Paterson et al., 2016), the overall effect of laboratory dream ESP studies was small (.07); the effect of telepathy studies was small (.07); the effect of clairvoyance studies was moderate (.19); and the effect of precognition studies was very small (.04). The overall effect as well as the effect of clairvoyance studies were statistically significant, whereas the effect of telepathy and precognition studies were not statistically significant. Thus, these results do not definitively support or not support the existence of psi in laboratory studies of ESP in dreams.

These results are more precarious than previously believed, however. Perhaps the most obvious influence on these results is the sample size of the original study, as large-n studies produced a very small effect (.04), medium-n studies produced a small effect (.13), and small-n studies produced a moderate effect (.20). The ESP mode with the smallest effect, precognition studies, was the ESP mode with the most large-n studies, likely due to the relative ease of performing these studies compared to telepathy and clairvoyance studies (as noted by Storm et al.). Likewise, the only ESP mode with a significant effect, clairvoyance studies, was also the only ESP mode without any large-n studies, suggesting that its notably larger observed result may have been caused by study characteristics rather than a substantive clairvoyance effect.

Moving forward, authors should strive to conduct large-n studies of ESP in dreams. Doing so could reduce the likelihood of spurious results, but also defend the study of ESP from arguments against its validity. As of now, authors could argue that most dream ESP researchers "cherry-pick" their results by collecting many small samples and publishing the significant findings. This argument is further supported by the trim-and-fill analysis; the overall effect was no longer significant when imputing studies that were implied to be missing. By performing more large-n studies, psi researchers could make a better case defending this argument; although file drawers could contain many studies with sample sizes in the 10s, it is less likely for these draws to contain many studies with sample sizes in the 100s.

Also, the significant effect of clairvoyance would become non-significant by removing any of five studies. While grouping studies by author did not produce a significant result, it should be highlighted that three of these five studies were the three studies performed by Dalton (all medium-n). Also, Child and Krippner produced multiple outstanding effects in small-n studies, including effect sizes of .68, .72, and two of .94. Proponents of psi may argue that these authors unlocked the key to producing large effects in dream ESP studies. Critics of psi may argue that these authors cherry-picked their samples to report. Regardless, additional large-n studies by these authors (and others) could clarify whether these effects are substantive.

Additionally, the small observed effect of dream ESP studies in the current reanalysis has two primary implications. First, it places the effect size of laboratory dream ESP studies in the range of typical psi studies rather than Ganzfield experiments. Storm et al. logically associated dream ESP studies with Ganzfield experiments, which have produced meta-analytic effect sizes typically ranging from .13 to .24. The current observed effects, especially when limited to large-n studies, were closer to the aforementioned meta-analyses with effects of .01 to .05. Future research should consider the theoretical relation of dream ESP studies to general psi dynamics.

Second, it causes dream ESP studies to fall into the age-old debate of psi. Proponents of psi often argue that small effects (e.g. $r = .01$-$.05$) are meaningful reflections of reality, and psi should influence the natural environment without being noticeable to the naked eye; otherwise, it would be a known phenomenon. Critics of psi argue that these effects are purely publication and experimenter biases; if these biases are indeed present, then they could produce effects sizes of .01 to .05 when studying complete randomness (e.g. making predictions in the absence of psi). Regardless of the cause, it should also be recognized that proponents and critics of psi may have differing opinions because they read different studies, as the average reported effect size varied between parapsychology and non-parapsychology outlets. This difference may partially explain the seemingly never-ending, fervent arguments expressed by those on both sides of the psi debate, and the current reanalysis seems to be (unfortunately) unable to resolve this debate.

Lastly, some results of my reanalysis supported Storm et al.'s conclusions. MDL studies did not significantly differ from non-MDL studies; REM studies did not significantly differ from non-REM studies; and effects were not significantly different for different ESP modes. I echo the suggestions of Storm et al. regarding these study features as well as several other recommendations – particularly the seven points provided in Storm et al.'s conclusion. Thus, while the current results shed considerable doubt that dream ESP studies produce a substantive effect, many of Storm et al.'s suggestions should be adopted in future research – future research that can more decisively determine whether ESP in dreams is real or the result of researcher bias.

References

Bem, D., Tressoldi, P., Rabeyron, T., & Duggan, M. (2015). Feeling the future: A meta-analysis of 90 experiments on the anomalous anticipation of random future events. F1000Research.

Borenstein, M. (2005). Software for publication bias. Publication bias in meta-analysis: Prevention, assessment and adjustments, 193-220.

Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. Journal of Applied Psychology, 100(2), 431.

Cheung, M. (2015). Meta-analysis: A structural equation modeling approach. John Wiley & Sons.

Dalton, K., Steinkamp, F., & Sherwood, S. J. (1999). A dream GESP experiment using dynamic targets and consensus vote. Journal of the American Society for Psychical Research.

Dalton, K., Utts, J., Novotny, G., Sickafoose, L., Burrone, J., & Phillips, C. (2000). Dream GESP and consensus vote: A replication. The Journal of Parapsychology, 64(3), 242.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. Personality and individual differences, 102, 74-78.

Honorton, C., & Ferrari, D. C. (1989). “Future telling”: A meta-analysis of forced-choice precognition experiments, 1935–1987. Journal of Parapsychology, 53(28), 1-308.

Kanthamani, H., & Khilji, A. (1990). An experiment in ganzfeld and dreams: A confirmatory study. In Proceedings of Presented Papers: The Parapsychological Association 33rd Annual Convention (pp. 126-137).

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. Sage Publications, Inc.

Mörck, N.C. (2015). Letter to the Editor. Journal of Parapsychology, 79(1), 127.

Paterson, T. A., Harms, P. D., Steel, P., & Credé, M. (2016). An assessment of the magnitude of effect sizes: Evidence from 30 years of meta-analysis in management. Journal of Leadership & Organizational Studies, 23(1), 66-81.

Radin, D. I., & Ferrari, D. C. (1991). Effects of consciousness on the fall of dice: A meta-analysis. Journal of Scientific Exploration, 5(1), 61-83.

Roe, C., Sherwood, S. J., Farrell, L., Savva, L., & Baker, I. (2007). Assessing the roles of the sender and experimenter in dream ESP research. European Journal of Parapsychology.

Schmidt, F. L., & Hunter, J. E. (2014). Methods of meta-analysis: Correcting error and bias in research findings. Sage publications.

Sherwood, S., Roe, C., Simmonds, C., & Biles, C. (2002). An exploratory investigation of dream precognition using consensus judging and static targets. Journal of the Society for Psychical Research.

Storm, L., Sherwood, S., Roe, C., Tressoldi, P., Rock, A., & Di Risio, L. (2017). On the correspondence between dream content and target material under laboratory conditions: A meta-analysis of dream-ESP studies, 1966-2016. International Journal of Dream Research.

Storm, L., Tressoldi, P. E., & Di Risio, L. (2012). Meta-analysis of ESP studies, 1987-2010: Assessing the success of the forced-choice design in parapsychology. Journal of Parapsychology, 76(2), 243-273.

Van de Castle, R. L. (1971). The study of GESP in a group setting by means of dreams. Journal of Parapsychology, 35, 312.

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta analysis. Research Synthesis Methods, 1(2), 112-125.

Watt, C. (2014). Precognitive dreaming: Investigating anomalous cognition and psychological factors. Journal of Parapsychology, 78(1), 115-125.

Watt, C., & Valášek, M. (2015). Postscript to Watt (2014) on precognitive dreaming: Investigating anomalous cognition and psychological factors. Journal of Parapsychology, 79(1), 105-107.