Journal of Scientific Exploration, Vol. 30, No. 1, pp. 10–15, 2016 0892-3310/16

RESEARCH ARTICLE

Prospective Statistical Power: Sample Size Recommendations for the Investigation of the Main Parapsychological Phenomena

Patrizio Tressoldi

Dipartimento di Psicologia Generale, Università di Padova, Italy
patrizio.tressoldi@unipd.it

Submitted February 25, 2015; Accepted December 19, 2015; Published March 15, 2016

Abstract—The aim of this paper is to offer a practical guideline for researchers investigating parapsychological phenomena to choose appropriate sample sizes to achieve a statistical power equal to or above 0.80. The availability of different meta-analyses related to different parapsychological phenomena allow a sufficient estimation of the expected effect sizes, which usually range from small to very small. With these measures, it is possible to estimate the numerosity of sample sizes necessary to achieve a level of statistical power that can facilitate the replication of different parapsychological phenomena. I discuss ways to deal with the investigation of phenomena with very small effect sizes requiring very large sample sizes.

Keywords: statistical power—sample size—replication—effect size—extrasensory perception—interaction at distance

Introduction

In science, replication of experimental findings is a *sine qua non* for supporting the evidence of all phenomena. At present, the problem of replication is a hot topic in psychology, neuroscience, and medical fields (see Ioannidis 2005, Yong 2012, Pashler & Wagenmakers 2012, Begley & Ioannidis 2015), and there are multiple initiatives and proposals on how to overcome this increasing distrust toward the scientific methodology used in these disciplines (for example, Simons, Holcombe, & Spellman 2014; http://validation.scienceexchange.com).

When studying anomalous phenomena that do not seem to fit with the dominant scientific paradigms, multiple independent replications is an ever-greater requirement. All evidence currently available supporting the reality of so-called parapsychological phenomena, both related to the acquisition of information beyond the range of sensory organs and to mind interaction at a distance to physical and biological targets, is still deemed to be "exceptional claims" that must be supported by "exceptional evidence."

The aim of this methodological paper is to help researchers interested in the investigation of parapsychological phenomena in planning their experiments by suggesting an estimate for the minimum number of participants necessary to achieve a statistical power of at least .80 to detect the phenomena they are examining.

From a statistical point of view, a problem remains of how to devise a sufficient statistical power for "identifying" or "capturing" a given phenomenon. Statistical power (for a more complete description, see Faul et al. 2007 and Tressoldi & Utts 2015) is the power to detect a given phenomenon after defining the probability of risk of accepting the hypothesis that it is true with a given effect size (*ES*) when it is not. Statistical power depends on three numerical values: (1) the probability of type I error, that is the probability of accepting the existence of a phenomenon when it is not, typically set to .05; (2) the size(s) of the sample(s) used for the test; and (3) an *ES* parameter indexing the actual degree of deviation from the probability of "non-existence" in the underlying population.

In most experimental designs, the accepted probability of making a Type I error is $\alpha = .05$ and the desired power is not less than .80. However, in order to define how to obtain such a level of power, it is necessary to specify the *ES* of the phenomena being identified. It is intuitive that the smaller the phenomenon, the greater the sensitivity needed to detect it. This analogy is similar to the signal/noise relationship. The smaller the signal, the stronger the means must be to detect it in the noise.

Power analysis should be used prospectively, that is before starting an experiment to calculate the minimum sample size required so that one can increase the probability to detect an effect of a given size. The definition of a satisfactory level of statistical power is not restricted to the so-called frequentist statistical approach which derives from the pioneering theories of Fisher and Neyman-Pearson (see Neyman 1937), but is also relevant for the Bayesian statistical approach. With this approach, the prior beliefs, expressed in probabilistic terms about the existence of a given phenomenon, are updated using the data obtained in a given study. In principle, a study that was underpowered could still be used in such an updating process, but it is unlikely to lead to a major change in beliefs (Kennedy 2015, Kruschke 2014).

The recent updates of the accumulated evidence related to many parapsychological phenomena obtained by meta-analyses clearly show that all these phenomena have small *ESs* (Storm, Tressoldi, & DiRisio, 2010, 2012, Mossbridge, Tressoldi, & Utts 2012, Schmidt 2012, Bem et al.

2015, Roe, Sonnex, & Roxburgh 2014, Baptista, Derakhshani, & Tressoldi 2015). The consequence of these results are straightforward: To achieve a statistical power of at least .80, it is necessary to plan the recruitment of a high number of participants.

Methods

ES *Estimates*

ES estimates of different parapsychological phenomena were chosen from the more recently available meta-analyses which summarized the updated accumulated evidence related to both extra-sensory perception (ESP) and mental interaction at a distance to different targets, both human and biological. These meta-analyses are presented in Table 1.

Sample Size Estimation

Sample size estimates for power = 0.80 and 0.90 were computed using the freeware G*Power software v.3.1.9.2 (Faul et al. 2007). In the "Test family" menu, we chose the more common statistics used with the different research protocols which are presented in Table 1. In the "Statistical test" menu, we chose the more common comparisons, e.g., paired or unpaired difference; in the "Type of power analysis" menu, we chose "A priori: Compute required sample size, were given α, power, and effect size"; in the "Input Parameters" windows, we always inputted "Tail(s)" = One; α = 0.05 (for the application of focused or confirmatory hypotheses) *ES* point estimate, and their lower and upper confidence intervals drawn from the different meta-analyses.

Discussion

As shown in Table 1, when investigating ESP with different procedures based on free response protocols, it seems not difficult to achieve a satisfactory statistical power with samples ranging from 20 to approximately 100 participants. On the contrary, for the investigation of ESP using classical forced-choice protocols, it seems quite impossible to achieve a satisfactory statistical power given the necessity to recruit approximately more than 10,000 participants.

For the rest of the phenomena, from the implicit psychophysiological and behavioral anticipations to the remote mental interaction with biological and human targets, the number of participants necessary to recruit to achieve a satisfactory statistical power is quite high, but not impossible to achieve.

If we examine all studies included in the meta-analyses, many of them have a statistical power (well) below the level of .80, with all the

TABLE 1

Sample Size Estimates to Achieve a Statistical Power of 0.80 and 0.90 for the Investigation of the Main Parapsychological Phenomena Given the Expected Effect Sizes and Using the More Common Statistical Tests

Phenomena	Statistical test	*ES* [95%CI]	Power = 0.80 sample size	Power = 0.90 sample size
ESP with Ganzfeld [1]	Exact. Proportion: difference from constant°	0.14 [0.09,0.18]	72 [45-161]	97 [59-219]
ESP with Remote Vision [2]	Exact. Proportion: difference from constant°	0.24 [0.20,0.28]	26 [20-36]	36 [26-49]
ESP with Dream [2]	Exact. Proportion: difference from constant°	0.14 [0.06,0.22]	72 [32-348]	97 [42-479]
Forced-Choice ESP [3]	Exact. Proportion: difference from constant°	0.01 [0.006,0.011]	11756 [9728-32448]	16266 [32448-44969]
Psychophysiologial Anticipation [4]	Paired sample t-test —Mann Whitney	0.21 [0.15,0.27]	142 [87-277]	196 [120-382]
Behavioral Anticipation [5]	Paired sample t-test —Mann Whitney	0.11* [0.08,0.14]	513 [317-968]	710 [439-1340]
Distant Mental Interaction —Human Targets [6]	Independent samples t-test	0.19 [0.15,0.24]	688# [432-1102]	952# [598-1524]
Distant Mental Interaction —Biological Targets [6]	Independent samples t-test	0.20 [0.17,0.23]	620# [470-858]	858# [650-1188]
Distant Intention Effects [7]	Paired sample t-test —Mann Whitney	0.11 [0.01,0.22]	513 [130-61827]	710 [179-85640]
Remote Staring [7]	Paired sample t-test —Mann Whitney	0.12 [0.02,0.22]	431 [130-15458]	597 [179-21411]

° 0.25, * only fast-thinking protocols, # group1 + group2

[1] Storm et al. 2010, Tressoldi 2011; [2] Baptista et al. 2015, Storm et al. (submitted); [3] Storm et al. 2012, Tressoldi 2011; [4] Mossbridge et al. 2012; [5] Bem et al. 2015; [6] Roe et al. 2015; [7] Schmidt 2012

consequences outlined in the Introduction. How does one proceed when it is difficult or quite impossible to achieve a satisfactory statistical power?

Among the possible solutions to this problem, one is to try to increase the expected *ES* by recruiting selected participants. As demonstrated by Baptista et al. (2015) in all free-response protocols, e.g., ESP in a ganzfeld environment or ESP using RV techniques, selected participants, that is participants who have experience with these kind of tasks and are very committed to succeed, obtain almost a double *ES* with respect to non-expert participants.

Among the other solutions, we suggest disclosing the problem of statistical power and ignoring *p* values altogether, focusing more on *ESs* and their estimation by using confidence intervals in line with the so-called "statistical reform" movement endorsed recently by the editor of *Psychological Science* (Eich 2014), underlying their importance for meta-analyses as suggested by Braver, Thoemmes, and Rosenthal (2014). For further suggestions, see Tressoldi and Utts (2015) and Tressoldi and Giofré (2015).

To summarize, the take-home message of this methodological paper is: Before speculating about the theoretical reasons underlying the unreliability of evidence of most if not all parapsychological phenomena, we must exclude the possibility that it may be due to the neglect of statistical power.

References Cited

Baptista, J., Derakhshani, M., & Tressoldi, P. E. (2015). Explicit Anomalous Cognition: A Review of the Best Evidence in Ganzfeld, Forced-choice, Remote Viewing, and Dream Studies. In *Parapsychology: A Handbook for the 21st Century* edited by E. Cardeña, J. Palmer, and D. Marcusson-Clavertz, Jefferson, NC: McFarland.

Begley, C. G., & Ioannidis, J. P. (2015). Reproducibility in science improving the standard for basic and preclinical research. *Circulation Research*, *116*(1):116–126.

Bem, D., Tressoldi, P., Rabeyron, T., & Duggan, M. (2015). A Meta-Analysis of 90 Experiments on the Anomalous Anticipation of Random Future Events. *F1000Research*, *4*:1188. (doi: 10.12688/f1000research.7177.1) Available at the Social Science Research Network: http://ssrn.com/abstract=2423692

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science*, *9*(3):333–342.

Eich, E. (2014). Business not as usual. *Psychological Science*, *25*(1):3–6.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2):175–191.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8):e124.

Kennedy, J. E.(2015). Beware of inferential errors and low power with Bayesian analyses: Power analysis is needed for confirmatory research. *Journal of Parapsychology, 79*(1):53–64. https://jeksite.org/psi/confirm_bayes.pdf

Kruschke, J. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press.

Mossbridge, J., Tressoldi, P. E., & Utts, J. (2012). Predictive physiological anticipation preceding seemingly unpredictable stimuli: A meta-analysis. *Frontiers in Psychology, 3*:390. doi:10.3389/fpsyg.2012.00390

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability, *Philosophical Transactions of the Royal Society of London A*, *236*:333–380.

Pashler, H., & Wagenmakers, E. J. (2012). Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? *Perspectives on Psychological Science*, *7*(6):528–530.

Roe, C. A., Sonnex, C., & Roxburgh, E. C. (2015). Two meta-analyses of noncontact healing studies. *EXPLORE: The Journal of Science and Healing*, *11:*11–23.

Schmidt, S. (2012). Can we help just by good intentions? A meta-analysis of experiments on distant intention effects. *The Journal of Alternative and Complementary Medicine*, *18*(6): 529–533. doi:10.1089/acm.2011.0321

Simons, D. J., Holcombe, A. O., & Spellman, B. A. (2014). An introduction to registered replication reports at *Perspectives on Psychological Science*. *Perspectives on Psychological Science*, *9*(5):552–555.

Storm, L., Tressoldi, P. E., & Di Risio, L. (2010). Meta-analyses of free-response studies 1992–2008, assessing the noise reduction model in parapsychology. *Psychological Bulletin*, *136:*491–494. doi:10.1037/a0019840

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):243–274.

Storm, L., Sherwood, S. J., Roe, C. A., Tressoldi, P. E., Rock, A. J., & Di Risio, L. (in press). On the Correspondence between Dream Content and Target Material under Laboratory Conditions: A Meta-Analysis of Dream-ESP Studies, 1966–2014.

Tressoldi, P. E. (2011). Extraordinary claims require extraordinary evidence: The case of non-local perception, a classical and Bayesian review of evidences. *Frontiers in Psychology*, *2*:117. doi:10.3389/fpsyg.2011.00117

Tressoldi, P. E., & Giofré, D. (2015). The pervasive avoidance of prospective statistical power: Major consequences and practical solutions. *Frontiers in Psychology*, *6:*726. Available at Social Science Research Network:
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2579268

Tressoldi, P., & Utts, J. (2015). Statistical Guidelines for Empirical Studies. In *Parapsychology: A Handbook for the 21st Century* edited by E. Cardeña, J. Palmer, & D. Marcusson-Clavertz, Jefferson, NC: McFarland.

Yong, E. (2012). Nobel laureate challenges psychologists to clean up their act. *Nature,* October 3. doi:10.1038/nature.2012.11535