

Suggestibility in functional neurological disorder: A meta-analysis

Lillian Wieder, BSc¹, Richard J. Brown, PhD, ClinPsyD^{2,3}, Trevor Thompson, PhD⁴, & Devin B. Terhune, PhD¹

¹ *Department of Psychology, Goldsmiths, University of London, London, UK*

² *School of Health Sciences, University of Manchester, Manchester, UK*

³ *Complex Trauma and Resilience Research Unit, Greater Manchester Mental Health NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, UK*

⁴ *School of Human Sciences, University of Greenwich, London, UK*

Corresponding author:

Devin B. Terhune, PhD
Department of Psychology
Goldsmiths, University of London
8 Lewisham Way
New Cross, London, UK SE14 6NW
d.terhune@gold.ac.uk

Word count: 2,965 words

Abstract

Objective Responsiveness to direct verbal suggestions (suggestibility) has long been hypothesized to represent a predisposing factor for functional neurological disorder (FND) but previous research has yielded conflicting results. The aim of this study was to quantitatively evaluate whether FND patients display elevated suggestibility relative to controls via meta-analysis.

Methods Four electronic databases were searched in November 2019, with the search updated in April 2020, for original studies assessing suggestibility using standardized behavioural scales or suggestive symptom induction protocols in FND patients and controls. The meta-analysis followed Cochrane, PRISMA, and MOOSE guidelines. Data extraction and study quality coding were performed by two independent reviewers. Standardized suggestibility scores and responsiveness to symptom induction protocols were used to calculate standardized mean differences (*SMDs*) between groups.

Results Of 26,643 search results, 19 articles presenting 11 standardized suggestibility datasets (FND: $n=316$; control: $n=360$) and 11 symptom suggestibility datasets (FND: $n=1285$; control: $n=1409$) were included in random-effects meta-analyses. Meta-analyses revealed that FND patients displayed greater suggestibility than controls on standardized behavioural scales (*SMD*, 0.48 [95% CI, 0.15, 0.81]) and greater responsiveness to suggestive symptom induction (*SMD*, 1.39 [95% CI, 0.92, 1.86]). Moderation analyses presented mixed evidence regarding the extent to which effect sizes covaried with methodological differences across studies. No evidence of publication bias was found.

Conclusions These results corroborate the hypothesis that FND is characterized by heightened responsiveness to verbal suggestion. Atypical suggestibility may confer risk for FND and be a cognitive marker that can inform diagnosis and treatment of this condition.

Introduction

Functional neurological disorder (FND) is characterized by impaired motor or cognitive functioning that resembles neurological pathology but is not adequately explained by it.¹ FND has a prevalence of 4-12 per 100,000^{2,3} and represents ~16% of clinical referrals and visits to epilepsy clinics.⁴ It is associated with considerable diagnostic delays and frequent misdiagnosis,⁵ which add to the already significant psychological, social, and economic impact of the condition.⁶

FND has long been hypothesized to be characterized by elevated responsiveness to direct verbal suggestions (suggestibility).⁷ Suggestibility is theorized to confer vulnerability for FND⁸ through aberrant meta-awareness of intentions,⁹⁻¹² the capacity for suggestions to trigger automatized behavioural routines or mental representations,⁸ and/or a tendency to form precise (symptom) priors that override motor and perceptual systems.^{13,14} The use of suggestion to provoke FND symptoms is widely used to aid diagnosis¹⁵ and functional symptoms are responsive to suggestion-based treatments, such as hypnosis and placebo.¹⁶ In addition, suggestibility has been shown to predict prognosis¹⁷ and response to treatment¹⁸ in FND patients. Conditions with germane symptom profiles, such as dissociative disorders, are also characterized by elevated hypnotic suggestibility.¹⁹

Despite these various strands of evidence, the empirical association between suggestibility and FND seems to be highly variable.^{20,21} In order to quantify the evidence for elevated suggestibility in FND patients, we conducted a random-effects meta-analysis of controlled studies of suggestibility on standardized behavioural scales and in response to suggestive symptom induction protocols. Secondary analyses investigated moderating influences on patient-control differences, such as experimenter blindness,²² methodological quality, the inclusion of a hypnotic induction,²³ and symptom provocation method.¹⁵

Method

Eligibility criteria

Inclusion criteria included (1) English language; (2) full paper in a peer-reviewed journal; (3) patient sample characterized by FND/symptoms, encompassing conversion disorder (DSM), dissociative neurological disorder (ICD), specific functional neurological syndromes (e.g., non-epileptic seizures), and conditions where functional neurological symptoms are a diagnostic feature (i.e. DSM-IV somatization disorder;

Briquet's syndrome); (4) inclusion of a control group; and either (5) use of a standardized behavioural measure of direct verbal suggestibility²³, or (6) assessment of symptom induction through suggestion (e.g., suggestive seizure induction¹⁵).

Exclusion criteria included (1) studies in which suggestion was used to aid diagnosis; (2) case studies/series or non-empirical papers; (3) overlapping/insufficient data; (4) use of interrogative suggestibility scales, which capture a different form of suggestibility characterized by high compliance²⁴; and (5) studies of patients with functional somatic syndromes not specifically characterized by functional neurological symptoms (e.g., fibromyalgia).

Search strategy

PubMed, PsycINFO, Web of Science and Academic Search Complete databases were searched in November 2019 for eligible studies using terms relating to suggestibility and FND (see **online supplemental content 1**) and then integrated into a single database. The search was repeated in April 2020 but yielded no new studies. The reference lists of all eligible studies (and relevant review papers) were manually searched to identify additional studies. Authors were contacted when data were unavailable or to clarify ambiguities in methodology.

Study selection

Two raters (LW and a second rater) independently screened and assessed all studies for their eligibility using a two-stage procedure. First, all titles and abstracts were screened and articles not meeting eligibility criteria were rejected. Second, all remaining papers were reviewed to establish a final list of articles. Discrepancies at either stage were resolved in consultation with a third reviewer (DBT) and sometimes a fourth reviewer (RJB). Authors of eligible studies were contacted to address any questions regarding insufficient data. Of 5 author groups contacted, 4 (80%) provided data sufficient to permit study inclusion or answered queries that justified exclusion.

Data extraction

The two outcome types (standardized suggestibility and symptom suggestibility) were measured using continuous and categorical measures, respectively. After exclusion of two studies with overlapping data,

data from eligible studies were extracted and coded independently by LW and the second rater using a data extraction form including: (i) study details (title, year, geographical location), (ii) diagnosis, (iii), diagnosis method, (iv) demographics (sample size and gender and age distributions), (v) study design details (suggestibility type [standardized or symptom], inclusion of a hypnotic induction, scale, administration method [live or recorded presentation], and scoring method [self or experimenter]), provocation method [see **Table 2**], and experimenter blindness [blind, unblind, or unreported]), and (vi) descriptive statistics (*Ms* and *SDs* [standardized suggestibility] or response counts [symptom suggestibility]). Symptom suggestibility response counts include only responses identified as typical for the respective patient (typicality reported: $k=8$; unreported: $k=3$). When studies reported results for more than one provocation method ($k=2$), the rounded mean was used. Two studies included data for both standardized and symptom suggestibility. There was 91% agreement between the two raters and discrepancies were resolved with a third reviewer (DBT).

Study quality

A 13-item scale was developed to assess study quality (see **online supplemental content 1**). Items were adapted from an earlier meta-analysis that included items based on Cochrane criteria and PRISMA recommendations²⁵ and a range of other methodological criteria such as experimenter blindness. LW and the second rater independently rated each item categorically (0=criterion not met, 1=criterion met) and a summed total was computed for each study. Agreement between raters was 90% ($\kappa=.73$) and discrepancies were resolved with DBT.

Meta-analysis and meta-regression

Individual study effect sizes included between-group differences in suggestibility that were computed with standardised mean differences (*SMDs*; Hedges' *gs*) using Review Manager (RevMan v. 5.3, 2014; The Nordic Cochrane Centre, The Cochrane Collaboration, Copenhagen). Symptom suggestibility data consisted of binary outcomes that were used to compute odds ratios, which were subsequently transformed to *SMDs*²⁶ in MATLAB (v. 2017b, MathWorks, Natick, MA). *SMDs* were coded such that positive values reflected greater suggestibility in FND patients than controls.

Publication bias was assessed by examining funnel plots of effect sizes against standard errors for asymmetry, as might occur due to a small number of studies with small or non-significant effect sizes. We

also tested for asymmetry using Egger's bias test²⁷ where $p < .05$ is indicative of asymmetry. We report revised effect sizes correcting for asymmetry using the trim-and-fill method²⁸ and funnel plots in

Figure 3.

Random-effects meta-analyses were performed in JASP (v. 0.8.6, 2019; JASP Team, Netherlands). Outlier detection was made on the basis of studentized residuals ($> |3.3|$).²⁹ There were no outliers in the standardized suggestibility data (range: -1.63 to 1.33) or the symptom suggestibility data (range: -2.01 to 1.26). Moderating effects were assessed using meta-regression analyses whenever data were available for at least 2 studies at each level of a categorical moderator and at least 10 studies for continuous moderators.³⁰ Moderators included five categorical measures (experimenter blindness, hypnotic induction, symptom provocation method [suggestion vs. placebo; symptom studies], symptom typicality [not reported vs. report; symptom studies]; and control type [non-clinical vs. clinical; standardized studies]) and one continuous measure (methodological quality).

Results

Study inclusion

A PRISMA diagram showing study selection is presented in **Figure 1**. Principal features of these studies, including diagnosis criteria and procedures, are presented in **Tables 1** and **2**.

**** Figure 1 about here ****

**** Table 1 about here ****

**** Table 2 about here ****

Study and participant characteristics

The 19 papers included 11 standardized suggestibility studies (FND: $n=316$, control: $n=360$) and 11 symptom suggestibility studies (FND: $n=1285$, control: $n=1409$). The standardized studies were published between 1984 and 2009 and were conducted in the US ($k=5$), the Netherlands ($k=4$), and the UK ($k=2$). The symptom studies were published between 1994 and 2018, and were conducted in the US ($k=7$), Puerto Rico ($k=1$), France ($k=1$), India ($k=1$) and the UK ($k=1$). The mean percentage of females for standardized studies was 86.13 for FND ($k=8$) and 45.71 for controls ($k=7$), whereas for symptom studies ($k=6$) it was 77.17 for FND and 61.33 for controls. In the standardized studies, the mean age for FND was 31.22 ($SD=8.01$) ($k=6$),

and for controls it was 36.96 ($SD=5.09$) ($k=7$), whereas for the symptom studies ($k=6$), mean age for FND was 33.41 ($SD=8.79$), and 37.20 ($SD=14.38$) for controls.

Details of the types of standardized scales and provocation methods as well as the use of a hypnotic induction are provided in **Table 1** and **Table 2**, respectively. A hypnotic induction was used in 9 standardized studies and 3 symptom studies. All symptom studies included verbal suggestion but varied in their use of various nocebo procedures.

Methodological quality criteria

Ratings for each study on each of the 13 methodological quality criteria items are shown in **online supplemental content 1, Tables 1 and 2**. Although some of the criteria were met by the majority of studies, multiple criteria were not reliably met. Only 5 of 22 studies (23%) reported that the experimenter was blind to group, 18 (82%) described the inclusion/exclusion criteria, 16 (73%) described the diagnosis procedure and criteria in adequate detail, 10 (45%) described participant characteristics and 3 (14%) exhibited demographic comparability between patients and controls. In the standardized studies, 7 of 11 (64%) described the scale and procedure in adequate detail, and only 1 (9%) included a measure to correct for compliance. In the symptom studies, 8 of 11 (73%) clearly described the provocation method.

Meta-analysis of standardized suggestibility

Meta-analysis of 11 standardized behavioural suggestibility studies found that patients with an FND exhibited greater suggestibility than controls, $SMD=0.48$ [0.15, 0.81], $Z=2.84$, $p=.004$ (see **Figure 2**). Positive results were observed in 8 studies with a high inconsistency of effect sizes across studies, $I^2=73\%$. Jackknife analysis in which each study effect was sequentially omitted and the analysis re-performed indicated that the group difference was reliably significant (SMD range: 0.40-0.56).

**** Figure 2 about here ****

Meta-analysis of symptom suggestibility

Meta-analysis of 11 symptom suggestibility studies found that patients with FND displayed greater responsiveness than controls, $SMD=1.39$ [0.92, 1.86], $z=5.77$, $p<.001$ (see **Figure 2**). Overall positive responses were observed in 49% of patients with FND and 6% of controls (sensitivity=49%, specificity=94%). Ten of the 11 studies exhibited positive $SMDs$ although there was substantial

inconsistency in effect sizes across studies, $I^2=99\%$. Jackknife analysis revealed that the group difference remained significant after omitting each study in a sequential manner (*SMD* range: 1.25 to 1.54).

Meta-analysis of standardized vs. symptom suggestibility

The weighted effect size for symptom studies was significantly greater than that for standardized studies, $z=2.60$, $p=.009$. This difference remained stable, $z=2.29$, $p=.022$, after removing the two studies included in both data sets. When the two data sets were aggregated, the cumulative standardised effect size was slightly less than 1, *SMD*=0.96 [0.62, 1.29].

Publication bias

Egger's test did not suggest asymmetry in the distribution of effect sizes in the standardized studies, $z=0.69$, $p=.49$, or in the symptom studies, $z=1.50$, $p=.13$. A trim and fill estimate produced only a slight reduction in effect sizes for standardized studies, $\Delta SMD=-.05$, and symptom studies, $\Delta SMD=-.10$ (see **Figure 3**).

**** Figure 3 about here ****

Meta-regression of standardized and symptom suggestibility

Given the observed inconsistency in the magnitude of effects, a set of meta-regression analyses considered whether group differences were moderated by binary and continuous predictors pertaining to study methodologies (see **Table 3**). Effect sizes were significantly larger when a hypnotic induction was included in standardized studies, but not in symptom studies. Effect sizes were also larger in studies that reported whether suggested symptoms were typical for the patient relative to those that did not report this information; this implies that effect sizes are not inflated by the inclusion of atypical symptoms in response rates for suggestive symptom induction. By contrast, effect sizes were not significantly related to the type of control (clinical vs. non-clinical; standardized studies) or suggestive induction protocol (suggestion vs. nocebo [suggestion and sham]; symptom studies). Experimenter blindness did not significantly moderate group differences with numerical differences in opposing directions for standardized and symptom studies. Similarly, methodological quality related to effect sizes in opposing directions: greater quality was significantly associated with lower effect sizes in standardized studies, but with larger (albeit non-significantly) effect sizes in symptom studies. These effects were primarily driven by procedure description. Standardized effect sizes were smaller in studies that met criteria for clarity of inclusion/exclusion criteria:

$z=-2.54, p=.011$; diagnostic procedure: $z=-2.82, p=.005$; and scale administration procedure: $z=-2.28, p=.022$ (all other $ps>.09$). In contrast, symptom effect sizes were larger in studies that met criteria for clarity of diagnostic procedure, $z=3.23, p=.001$ (all other $ps>.09$).

Discussion

The results of this meta-analysis suggest that FND patients display elevated suggestibility on standardized behavioural scales and in response to suggestive symptom induction protocols, consistent with theoretical predictions to this effect.^{7, 20, 21} There was no evidence for publication bias although there was considerable heterogeneity in both data sets, which was partly explained by methodological variability.

These findings are consistent with models proposing responsiveness to suggestion as a vulnerability factor for FND and those attributing functional symptoms to precise symptom priors.^{8, 13} Recent theoretical work conceptualizes functional symptoms as arising from an interaction between multiple factors including an autonomic arousal response, inhibitory deficits that increase the likelihood of body misrepresentation, and symptom priors that collectively trigger the automatic activation of rogue symptom representations or conditioned actions.⁸ Suggestibility may confer heightened sensitivity to symptom-specific cues or dissociative responses to stressors. Moreover, suggestibility has been proposed to reflect a generalized tendency to form precise priors that override various motor and perceptual systems,³¹ which is thought to be a key process in FND¹³ and symptom reporting more generally.¹⁴ This aligns with research showing that hypnotic suggestibility predicts symptom severity in FND patients.³² Observed links between hypnotic suggestibility and emotional responsiveness to social cues,²³ highlight the potential role of suggestibility in symptom modeling or triggering through social observation.⁸ The perception that functional symptoms are extra-volitional may be further augmented by aberrant meta-awareness of intentions, as observed in both FND^{11, 12} and high hypnotic suggestibility.^{9, 10}

Although the two forms of suggestibility moderately covary,³³ FND patients were more responsive to symptom-specific ($SMD=1.39$ [0.92, 1.86]) than to standardized suggestions ($SMD=0.48$ [0.15, 0.81]), implying selectively greater suggestibility for functional symptoms. Indeed, the sole non-significant provocation study³⁴ administered suggestions for generic symptoms that did not necessarily mirror patients' symptom profiles. Previous research similarly found that highly dissociative individuals and dissociative and acute stress disorder patients are more responsive to suggestions for dissociative experiences (e.g.,

amnesia).¹⁹ Symptom suggestibility in FND may thus partly reflect a kindling process whereby symptoms become more responsive to verbal suggestion over time.⁸

Atypical suggestibility in FND patients complements research showing greater suggestibility in germane conditions, such as dissociative and stress disorders^{19,20} and somatoform disorders.³⁵ Many of these conditions are characterized by detachment and/or compartmentalization symptoms.³⁶ Variability in such symptoms may account for heterogeneity in the suggestibility profiles of FND patients, with heightened responsiveness to suggestions being specific to, or more pronounced among, patients with marked dissociative symptomatology.^{19,37} Moderation analyses indicated that the administration of a hypnotic induction was associated with greater standardized suggestibility among FND patients. This is consistent with the proposal that individuals with compartmentalization symptoms benefit more from a hypnotic induction¹⁹ although the mechanistic basis of this effect remains unclear.²³

These effects attest to the efficacy of suggestion in the diagnosis of FND.¹⁵ Suggestive symptom induction displayed high specificity (94%) although sensitivity is poor (49%), indicating that this technique is insufficient as a standalone diagnostic procedure. The inclusion of sham methods or a hypnotic induction were not associated with greater discrimination of FND patients and controls relative to verbal suggestion alone but warrant further attention. Suggestive symptom induction is likely to be especially valuable in suggestible populations such as adolescent and elderly samples²³ or patients with comorbid dissociative or stress disorder diagnoses.^{19,20} It may also inform diagnosis of comorbid non-epileptic seizures in epilepsy patients³⁸ and medication prescription in FND patients.³⁹ Insofar as suggestibility is a positive predictor of outcome with suggestion-based treatments,¹⁸ these results also support greater incorporation of suggestion techniques in treatment protocols, which show promising results in randomized-controlled trials.¹⁶ However, they also highlight the need to control for suggestion, or consider its role, in diagnostic and treatment procedures, particularly those that evoke strong response expectancies.

Limitations

The principal limitations of these data concern methodological variability across studies. Methodological quality was significantly or descriptively related to effect sizes in both data sets albeit in opposite directions. Among standardized studies, older studies that did not precisely specify inclusion/exclusion criteria and/or diagnostic and scale administration procedures tended to exhibit larger effect sizes; in contrast, precise

specification of diagnostic procedures was associated with larger effect sizes in symptom studies. In most studies, the operator administering the assessment was not blind to patient group, which may inflate effect sizes²², although there was no evidence for an experimenter effect in the standardized studies. The sole blind symptom study had a lower effect size than the remainder of the studies but was still large in magnitude ($SMD=0.93$ [0.61, 1.26]). Most studies included clinical controls (e.g., epilepsy patients), which raises issues regarding generalizability although effect sizes did not significantly relate to control type. Symptom suggestibility estimates are also confounded by baseline symptom frequency, which is not incorporated into these assessments.¹⁵ This potentially renders patients with high symptom frequency at an increased risk of false positive responses, although there is evidence that this is not the case.⁴⁰ The studies also varied in whether successful responses to symptom induction protocols were contingent upon the typicality of the response, which accounted for variability in effect sizes. Standardized studies were limited insofar as only one controlled for compliance. In addition, standardized suggestibility scales include a disproportionate number of suggestions for dissociative and functional symptoms (e.g., paralysis),^{20, 23, 32} raising the question of whether elevated suggestibility in FND generalizes beyond these symptoms. Collectively, these findings underscore the need for optimization and standardization of suggestive symptom induction protocols,¹⁵ compliance-correction,²³ and more diverse suggestion batteries.

Conclusions

This meta-analysis corroborates the long-held view that FND is characterized by elevated suggestibility. Increased suggestibility has direct implications for the risk factors underlying this condition, the use of suggestion to aid diagnosis, the utility of suggestion-based treatments for functional symptoms, and heterogeneity within this population.

Acknowledgements

The authors thank Che Ofusasia, Goldsmiths, University of London, for contributing to the data coding and evaluation.

Contributors

All authors conceived the project. LW carried out the database searches and data coding with assistance from RJB, TT, and DBT. LW and DBT performed the meta-analysis with assistance from RJB and TT. LW and DBT drafted the initial manuscript. All authors reviewed and approved the final version of the manuscript.

Funding

This study was supported by Bial Foundation bursary 70/16 (DBT) and a Gyllenbergs Foundation fellowship (DBT).

Competing interests

No, there are no competing interests for any author

Patient consent for publication

Not required.

Provenance and peer review

Not commissioned; externally peer reviewed.

References

1. Espay AJ, Aybek S, Carson A, et al. Current concepts in diagnosis and treatment of functional neurological disorders. *JAMA Neurol* 2018;75:1132-1141.
2. Carson AJ, Brown R, David AS, et al. Functional (conversion) neurological symptoms: Research since the millennium. *J Neurol Neurosurg Psychiatry* 2012;83:842-850.
3. Benbadis SR, Allen Hauser W. An estimate of the prevalence of psychogenic non-epileptic seizures. *Seizure* 2000;9:280-281.
4. Stone J, Carson A, Duncan R, et al. Who is referred to neurology clinics? The diagnoses made in 3781 new patients. *Clin Neurol Neurosurg* 2010;112:747-751.
5. Reuber M. The etiology of psychogenic non-epileptic seizures: Toward a biopsychosocial model. *Neurol Clin* 2009;27:909-924.
6. Gelauff J, Stone J, Edwards M, Carson A. The prognosis of functional (psychogenic) motor symptoms: a systematic review. *J Neurol Neurosurg Psychiatry* 2014;85:220-226.
7. Janet P. The major symptoms of hysteria. London, UK: Macmillan & Co., 1907.
8. Brown RJ, Reuber M. Towards an integrative theory of psychogenic non-epileptic seizures (PNES). *Clin Psychol Rev* 2016;47:55-70.
9. Lush P, Naish P, Dienes Z. Metacognition of intentions in mindfulness and hypnosis. *Neuroscience of Consciousness* 2016:1-10.
10. Terhune DB, Hedman LRA. Metacognition of agency is reduced in high hypnotic suggestibility. *Cognition* 2017;168:176-181.
11. Baek K, Donamayor N, Morris LS, et al. Impaired awareness of motor intention in functional neurological disorder: Implications for voluntary and functional movement. *Psychol Med* 2017;47:1624-1636.
12. Jungilligens J, Wellmer J, Schlegel U, Kessler H, Axmacher N, Popkirov S. Impaired emotional and behavioural awareness and control in patients with dissociative seizures. *Psychol Med* 2019:1-9.
13. Edwards MJ, Adams RA, Brown H, Parees I, Friston KJ. A Bayesian account of 'hysteria'. *Brain* 2012;135:3495-3512.
14. Van den Bergh O, Witthoft M, Petersen S, Brown RJ. Symptoms and the body: Taking the inferential leap. *Neurosci Biobehav Rev* 2017;74:185-203.

15. Popkirov S, Gronheit W, Wellmer J. A systematic review of suggestive seizure induction for the diagnosis of psychogenic nonepileptic seizures. *Seizure* 2015;31:124-132.
16. Deeley Q. Hypnosis as therapy for functional neurologic disorders. *Handb Clin Neurol* 2016;139:585-595.
17. Gambini O, Demartini B, Chiesa V, Turner K, Barbieri V, Canevini MP. Long-term outcome of psychogenic nonepileptic seizures: The role of induction by suggestion. *Epilepsy Behav* 2014;41:140-143.
18. Moene FC, Spinhoven P, Hoogduin KA, van Dyck R. A randomized controlled clinical trial of a hypnosis-based treatment for patients with conversion disorder, motor type. *The International journal of clinical and experimental hypnosis* 2003;51:29-50.
19. Terhune DB, Cardeña E. Dissociative subtypes in posttraumatic stress disorders and hypnosis: Neurocognitive parallels and clinical implications. *Current Directions in Psychological Science* 2015;24:452-457.
20. Bell V, Oakley DA, Halligan PW, Deeley Q. Dissociation in hysteria and hypnosis: Evidence from cognitive neuroscience. *Journal of Neurology, Neurosurgery, and Psychiatry* 2011;82:332-339.
21. Brown RJ, Reuber M. Psychological and psychiatric aspects of psychogenic non-epileptic seizures (PNES): A systematic review. *Clin Psychol Rev* 2016;45:157-182.
22. Holman L, Head ML, Lanfear R, Jennions MD. Evidence of experimental bias in the life sciences: Why we need blind data recording. *PLoS Biol* 2015;13:e1002190.
23. Terhune DB, Cleeremans A, Raz A, Lynn SJ. Hypnosis and top-down regulation of consciousness. *Neurosci Biobehav Rev* 2017;81:59-74.
24. Gudjonsson GH. Interrogative suggestibility and compliance. In: Ridley AM, Gabbert F, La Rooy DJ, eds. *Suggestibility in legal contexts: Psychological research and forensic implications*. Chichester, UK: Wiley-Blackwell, 2013: 45-61.
25. Thompson T, Terhune DB, Oram C, et al. The effectiveness of hypnosis for pain relief: A systematic review and meta-analysis of 85 controlled experimental trials. *Neurosci Biobehav Rev* 2019;99:298-310.
26. Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. *Psychol Bull* 1995;117:167-178.

27. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-634.
28. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;56:455-463.
29. Viechtbauer W, Cheung MW. Outlier and influence diagnostics for meta-analysis. *Res Synth Methods* 2010;1:112-125.
30. Higgins JPT, Green S. *Cochrane handbook for systematic reviews of interventions*. West Sussex: Wiley, 2008.
31. Martin JR, Pacherie E. Alterations of agency in hypnosis: A new predictive coding model. *Psychol Rev* 2019;126:133-152.
32. Roelofs K, Hoogduin KA, Keijsers GP, Naring GW, Moene FC, Sandijck P. Hypnotic susceptibility in patients with conversion disorder. *J Abnorm Psychol* 2002;111:390-395.
33. Khan AY, Baade L, Ablah E, McNerney V, Golewale MH, Liow K. Can hypnosis differentiate epileptic from nonepileptic events in the video/EEG monitoring unit? Data from a pilot study. *Epilepsy Behav* 2009;15:314-317.
34. Chabrol H, Peresson G, Clanet M. Lack of specificity of the traditional criteria for conversion disorders. *Eur Psychiatry* 1995;10:317-319.
35. Wickramasekera IE. Somatization: Concepts, data, and predictions from the high risk model of threat perception. *J Nerv Ment Dis* 1995;183:15-23.
36. Holmes EA, Brown RJ, Mansell W, et al. Are there two qualitatively distinct forms of dissociation? A review and some clinical implications. *Clin Psychol Rev* 2005;25:1-23.
37. Roelofs K, Keijsers GP, Hoogduin KA, Naring GW, Moene FC. Childhood abuse in patients with conversion disorder. *Am J Psychiatry* 2002;159:1908-1913.
38. Kutlubaev MA, Xu Y, Hackett ML, Stone J. Dual diagnosis of epilepsy and psychogenic nonepileptic seizures: Systematic review and meta-analysis of frequency, correlates, and outcomes. *Epilepsy Behav* 2018;89:70-78.
39. Gogia B, Rai PK, Matthys SA, et al. Use of noninvasive induction techniques in the diagnosis of PNES. *Epilepsy Behav* 2019;99:106491.

40. Popkirov S, Jungilligens J, Gronheit W, Wellmer J. Diagnosing psychogenic nonepileptic seizures: Video-EEG monitoring, suggestive seizure induction and diagnostic certainty. *Epilepsy Behav* 2017;73:54-58.

Figure 1. PRISMA flowchart of study selection process

Figure 2. Forest plots of Standardised Mean Differences (*SMDs*) (with 95% confidence intervals) from (left) 11 standardized suggestibility studies and (right) 11 symptom suggestibility studies. Marker sizes reflect study weights with smaller markers denoting smaller weights.

Figure 3. Funnel plots of standardized mean differences (*SMDs*) from (left) 11 standardised suggestibility studies and (right) 11 symptom suggestibility studies. Filled circles denote individual study effect sizes and empty circles denote estimated missing individual effect sizes due to publication bias imputed using the trim and fill method. Summary *SMDs* [95% CI] using the trim-and-fill method were *SMD*=0.43 [0.10, 0.76] (standardized suggestibility studies) and *SMD*=1.29 [0.84, 1.75] (symptom suggestibility studies).

Table 1. Characteristics of included studies measuring standardized suggestibility

Source	Diagnostic criteria		FND			Controls			Scale	Hypnotic induction
	Formal	Procedure	Diagnosis	n (% female)	Age [M (SD)]	Diagnosis	n (% female)	Age [M (SD)]		
Khan et al., 2009	-	MI	NES	24 (-)	- (-)	ES	16 (-)	- (-)	HIP	Y
Brown et al., 2008	DSM-IV	CR	SD	19 (95)	40.9 (14.1)	DYS	17 (59)	45.4 (11.1)	BSS	N
Roelofs et al., 2002	DSM-IV	MI	CD	50 (84)	37.2 (11.9)	mAD	50 (82)	36.4 (11.1)	SHSS:C	Y
Litwin and Cardena, 2001	-	MI	NES	10 (100)	30.5 (9.9)	ES	31 (45)	35.2 (8.9)	SHCS	Y
Moene et al., 2001 / Spinhoven et al. 1991*	DSM-III-R	CR	CD	96 (-)	- (-)	NCC	82 (57)	40.1 (14.3)	SHCS	Y
Barry et al., 2000	-	MI, CR	NES/ES	36 (-)	- (-)	ES	22 (-)	- (-)	HIP	Y
Goldstein et al., 2000	-	MI, CR	NES	20 (80)	34.4 (12.4)	NCC	20 (45)	36 (8.5)	CIS	N
Kuyk et al., 1999	-	MI	NES	20 (80)	25 (-)	ES	17 (18)	37 (-)	SHCS	Y
Kuyk et al., 1995	-	MI, CR	NES	6 (50)	19.3	ES	7 (14)	28.6 (-)	SHCS	Y
Bliss, 1984a	-	CR	CD	18 (100)	- (-)	NCC	49 (-)	- (-)	SHSS:C	Y
Bliss, 1984b	-	CR	BS	17 (100)	- (-)	NCC	49 (-)	- (-)	SHSS:C	Y

Notes. DSM=Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association); MI=Medical investigations; CR=case note review for ruling out organic causes; PDF=positive diagnostic features; NES=non-epileptic seizures; SD=somatization disorder; CD=conversion disorder; ES=epileptic seizures; BS=Briquet's syndrome; DYS=dystonia; mAD=mixed affective disorders; NCC=non-clinical controls; HIP = *Hypnotic Induction Profile*; BSS = *Barber Suggestibility Scale*; SHSS:C = *Stanford Hypnotic Susceptibility Scale: Form C*; SHCS = *Stanford Hypnotic Clinical Scale*; CIS = *Creative Imagination Scale*; - = Not reported; * = Control data were drawn from Spinhoven et al. (1991).

Table 2. Characteristics of included studies measuring symptom suggestibility

Source	Diagnostic criteria		FND			Controls			Method		Responsiveness: <i>n</i> (%)	
	Formal	Procedure	<i>Diagnosis</i>	<i>n</i> (%) female	Age [<i>M</i> (SD)]	<i>Diagnosis</i>	<i>n</i> (%) female	Age [<i>M</i> (SD)]	Suggestion/noceb o	Hypnotic induction	FND	Control
Kandler et al., 2018	-	MI, CR	NES	416 (-)	- (-)	ES	568 (-)	- (-)	VS, HV, PS	N	120 (29%)	40 (7%)
Chen-Block et al., 2016	-	MI, CR	NES/ES	466 (70)	- (-)	ES	581 (54)	- (-)	VS, HV, PS, AP, ST	N	240 (52%)	15 (3%)
Goyal et al., 2014	-	MI	NES	140 (71)	21.38 (10.83)	ES	50 (34)	21.18 (9.98)	VS, PS, VTF, TC, MS, SI	N	84 (60%)	0 (0%)
Arain et al., 2009	-	MI	NES	24 (67)	- (-)	ES	54 (65)	- (-)	VS, HV	N	8 (33%)	6 (11%)
Khan et al., 2009	-	MI	NES	24 (-)	- (-)	ES	16 (-)	- (-)	VS	Y	11 (46%)	2 (13%)
Kenney et al., 2007	FW	PDF	PT	12 (-)	42.5 (11)	ET	33 (-)	56.8 (17)	VS, HV, VTF	N	6 (50%)	5 (15%)
Martinez-Taboas, 2002	-	CR, PDF	NES	8 (88)	32 (9.2)	ES	5 (100)	30 (6.2)	VS, ARS	Y	8 (100%)	0 (0%)
Barry et al., 2000	-	MI, CR	NES/ES	47 (-)	- (-)	ES	22 (-)	- (-)	VS	Y	36 (77%)	1 (5%)
Chabrol et al, 1995	DSM- IV	CR, PDF	CD	15 (-)	44 (14)	mND	40 (-)	53 (17)	VS	N	6 (40%)	17 (43%)
Lancman et al., 1994	-	MI	NES	93 (77)	26.7 (-)	ES	20 (70)	28.2 (-)	VS, AP	N	72 (77%)	0 (0%)
Walczak et al., 1994	-	MI	NES	40 (90)	33.9 (11.6)	ES	20 (45)	34 (11.7)	VS, SI	N	33 (83%)	5 (25%)

Notes. DSM=Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association); MI=Medical investigations; CR=case note review for ruling out organic causes; PDF=positive diagnostic features; Fahn and Williams diagnostic criteria; NES=non-epileptic seizures; ES=epileptic seizures; PT=psychogenic tremor; CD=conversion disorder; ES=essential tremor; mND=mixed neurological disorders; VS = verbal suggestion; HV=hyperventilation; PS=photic stimulation; AP=alcohol patch; ST=subjective trigger; VTF=vibrating tuning fork; TC=temple compression; MS=moist swab; SI=saline injection; ARS=age regression suggestion; - = Not reported.

Table 3. Meta-regression analyses for standardized and symptom suggestibility studies (*SMD* [95% CIs] (*k*))

Moderator		Analysis groups		<i>Z</i>	<i>p</i>	<i>I</i> ²
Suggestibility type						
Hypnotic induction	No	Yes				
	Standardized	-0.37 [-0.82, 0.09] (2)	0.66 [0.34, 0.97] (9)	2.74	.006	61%
	Symptom	1.30 [0.76, 1.85] (8)	1.66 [0.67, 2.64] (3)	0.62	.54	100%
Experimenter blindness	Unblind	Blind				
	Standardized	0.46 [0.03, 0.89] (7)	0.54 [-0.09, 1.16] (4)	0.20	.84	76%
	Symptom	1.44 [0.94, 1.93] (10)	0.93 [0.61, 1.26] (1)			
Suggestive symptom induction method		Suggestion	Nocebo			
	Symptom	1.20 [0.11, 2.28] (4)	1.50 [0.93, 2.08] (7)	0.69	.49	100%
Suggestive symptom typicality		Not reported	Typical			
	Symptom	0.60 [-0.14, 1.34] (3)	1.70 [1.14, 2.26] (8)	2.07	.038	100%
Control sample type		Non-clinical	Clinical			
	Standardized	0.58 [-0.08, 1.24] (4)	0.42 [0.01, 0.82] (7)	-0.40	.69	76%
Methodological quality						
	Standardized	12-item scale		-2.21	.027	66%
	Symptom	10-item scale		1.76	.079	99%

Records identified through
database searching
($n = 26,642$)

Additional records identified
through other sources
($n = 1$)

Records after duplicate removal
($n = 20,194$)

Records screened
($n = 20,194$)

Records excluded ($n = 19,575$)

Abstracts and/or full-text
articles assessed for
eligibility ($n = 619$)

Full-text articles excluded ($n = 600$)
Case studies: 212
Non-clinical sample: 171
Non-eligible clinical sample: 74
No control group: 126
Insufficient data: 5
Duplicate data: 2
Suggestion used for diagnosis: 10

Studies included in
quantitative synthesis
(meta-analysis) ($n = 22$,
from 19 papers)



