Clinical Psychology and Psychotherapy
Clin. Psychol. Psychother. (2016)
Published online in Wiley Online Library (wileyonlinelibrary.com). **DOI:** 10.1002/cpp.2039

Emotional Processing Theory Put to Test: A Meta-Analysis on the Association Between Process and Outcome Measures in Exposure Therapy

Christian Rupp,[1] Philipp Doebler,[1†] Thomas Ehring[2] and Anna N. Vossbeck-Elsebusch[1*]

[1] *Institute of Psychology, Westfälische Wilhelms-University Münster, Münster, Germany*
[2] *Department of Psychology, LMU Munich, Munich, Germany*
[†] *Present address: Institute of Psychology and Education, University of Ulm, Ulm, Germany*

In order to test the predictions derived from emotional processing theory (EPT), this meta-analysis examined correlations between outcome of exposure therapy and three process variables: initial fear activation (IFA), within-session habituation (WSH) and between-session habituation (BSH). Literature search comprised a keyword-based search in databases, a reverse search and the examination of reference lists. Of the 21 studies included in the analyses, 17 provided data concerning IFA (57 endpoints, total $N = 490$), five concerning WSH (7 endpoints, total $N = 116$) and eight concerning BSH (22 endpoints, total $N = 304$). Owing to this data structure, analyses were performed using robust variance estimation with random-effects models being assumed *a priori*.

Results indicated that WSH and BSH are positively related to treatment outcome. By contrast, the statistical association between IFA and outcome of exposure was not confirmed, whereas our moderator analysis suggested that physiological process measures lead to higher correlations than non-physiological ones. The results for IFA and BSH were affected by selective reporting. In sum, our results do not specifically strengthen EPT while matching other theoretical perspectives such as inhibitory learning and reality testing. Further research is needed to provide recommendations concerning the best way of delivering exposure therapy.

Key Practitioner Message:

- This meta-analysis examined three variables of emotional processing theory (EPT).
- Initial fear activation was not linearly related to outcomes of exposure therapy.
- Habituation within and between sessions were shown to correlate with outcome.
- Outcome reporting bias was shown to play a crucial role in this meta-analysis.
- Results do not specifically support EPT.

Keywords: Emotional Processing Theory, Exposure Therapy, Initial Fear Activation, Habituation

The efficacy of exposure therapy for anxiety disorders, posttraumatic stress disorder (PTSD) and obsessive–compulsive disorder (OCD) has been demonstrated in a large number of randomized controlled trials and meta-analyses (e.g., Bisson *et al.*, 2007; Hofmann & Smits, 2008; Mitte, 2005; Powers, Halpern, Ferenschak, Gillihan, & Foa, 2010). It is considered to be the treatment of first choice for pathological fear and avoidance (e.g., Emmelkamp & Ehring, 2014). Yet, the working mechanisms of effective exposure therapy are still a subject of debate, with conditioning, learning and cognitive appraisal theories being the most prominent approaches (Telch, Cobb, & Lancaster, 2013). A seminal and frequently cited model to explain how exposure treatment for anxiety disorders works is emotional processing theory (EPT) put forward by Foa and Kozak (1986). Referring back to Lang's (1977) analysis of cognitive structures underlying emotions, the authors suggest that fear is represented as a memory structure that can be understood as a programme for escape and avoidance. Importantly, EPT specifies how this fear memory can be modified through emotional processing. Fear memory is suggested to consist of three types of information: information about the feared stimulus, about the individual's verbal, behavioural and physiological response to the stimulus, and about the *meaning* ascribed to both the stimulus and the

*Correspondence to: Dr. Anna Vossbeck-Elsebusch, Westfälische Wilhelms-Universität Münster, Institute of Psychology, Fliednerstraße 21, 48149 Münster, Germany.
E-mail: annavoss@uni-muenster.de

Clinical Psychology & Psychotherapy

individual's responses. According to Foa and Kozak, there are two key requirements for emotional processing (and as a consequence therapeutic change) to take place. First, the fear memory needs to be activated so that second, new information, which is incompatible with the original fear memory, can be integrated into this memory structure

A challenge to empirical tests of EPT is the fact that the processes described above are not directly observable. However, Foa and Kozak (1986) have made specific and testable suggestions regarding three key indicators of successful emotional processing in exposure treatment. They regard initial fear activation (IFA) as an indicator of activation of the fear structure, as well as both (2) within-session habituation (WSH) of fear, and (3) between-session habituation (BSH) of fear as indicators of the integration of corrective information. Importantly, Foa and Kozak consider physiological parameters of fear activation and habituation, such as heart rate (HR) and skin conductance (SC), to be the most important indicators. They are assumed to be the basis of all other facets of fear (see also Lang, 1977; 1979). According to EPT, the initial activation of the fear structure is a necessary requirement for change. In turn, WSH is assumed to reflect the incorporation of corrective information about the reaction to the stimulus, or in other words, the learning experience of being confronted with the stimulus without being fearful or anxious, so that the preexisting association between stimulus and response in weakened. Second, WSH is assumed to provide corrective information for the common belief that fear will never decline unless one escapes from the frightening situation. Both aspects provide information that is incompatible with the fear structure and therefore facilitates its modification (Foa & Kozak, 1986).

According to EPT, the incorporation of new information in the fear structure due to WSH leads to a decline in the anticipation of fear and distorted threat appraisals such as losing control or suffering a heart-attack, which is then reflected by BSH, i.e., the reduction of initial fear across various exposure sessions. Additionally, BSH is suggested to strengthen the belief that one is indeed able to cope with the frightening situation (Foa & Kozak, 1986). A practical implication of EPT is that exposure therapy should be delivered in a way that IFA, WSH and BSH are facilitated.

The predictions of the original EPT regarding an association between the three process variables (IFA, WSH and BSH) and treatment outcome have been tested empirically in a number of studies. Existing narrative reviews of these studies arrive at slightly different conclusions. For example, Craske *et al.* (2008) conclude that, taking all evidence together, 'the premises of EPT are only weakly supported' (p. 11). In contrast, Crits-Christoph, Gibbons, and Mukherjeed (2013) postulate in their review of the empirical literature on EPT that relatively high correlations between IFA, WSH and BSH and exposure treatment outcome are generally supportive of EPT. Also, it is worth mentioning that in a reformulation of the original theory, Foa, Huppert, and Cahill (2006) removed WSH as an indicator of emotional processing as some studies had failed to show an association between WSH and treatment outcome (e.g., Jaycox, Foa, & Morral, 1998; Kamphuis & Telch, 2000; Telch, Valentiner, Ilai, Petruzzi, & Hehmsoth, 2000). However, their conclusion to omit WSH as an indicator of emotional processing was not based on a systematic quantitative analysis. In an effort to provide an examination of the theory's original formulation including all three process variables, we explicitly decided to include WSH in our meta-analysis.

In sum, the existing reviews concerning the role of IFA, WSH and BSH are limited by their narrative nature. The aim of the current meta-analysis therefore was to provide a systematic quantitative analysis of the association between IFA, WSH and BSH on the one hand and therapeutic outcome on the other hand. In order to explain possible heterogeneity in the results, we examined a number of potential moderator variables, which are (in descending order of importance according to the predictions of EPT, Foa & Kozak, 1986) the type of process measure, the type of outcome measure, outcome reporting bias (ORB), diagnosis, the type of exposure, the purity of exposure treatment and the duration of treatment. On a descriptive level, we also coded the mean age of study participants, the percentage of females and the drop-out rate.

METHOD

Inclusion Criteria

Suitable studies for inclusion in this meta-analysis were selected using the following criteria.

(1) The study population was required to have a primary diagnosis of an anxiety disorder, either full-blown or at a subclinical level. Studies investigating the extinction of fear conditioned in the laboratory were excluded.
(2) Exposure had to be the core element of treatment. Exposure *in vivo*, exposure *in sensu*, exposure in virtual reality and written emotional disclosure (WED) were accepted. Interventions combining an element of exposure with an instructed use of coping skills (e.g., systematic desensitization) or secondary tasks (e.g., EMDR) were excluded.
(3) Inclusion further required that a correlation between at least one of the three process variables and at least one measure assessing outcome of exposure treatment (see below) had to be reported and that

sufficient statistical information (correlation coefficient and corresponding sample size) could be retrieved from the text or from a table. A broad range of operationalizations for each of the three process variables was accepted. In the case of IFA, this comprised physiological measures such as HR (peak, mean, reactivity and variability), SC (peak, mean, reactivity, spontaneous fluctuations), self-report ratings (such as the Subjective Units of Distress Scale [SUDS] originally proposed by Wolpe, 1969), external ratings of fear (e.g., on the basis of facial expressions) and several parameters of behavioural approach tests (BAT). Another important requirement was that IFA had been measured within a situation involving a confrontation with the feared stimulus or situation that was intense enough to activate the fear structure (as proposed by Foa and Kozak, 1986). Corresponding operationalizations of process measures were accepted in terms of WSH and BSH, whereas there were no *a priori* criteria formulated concerning the exact way of calculating the particular difference scores (i.e., concerning the particular sessions used for computing these two process measures).

(4) The particular process variable was required to be correlated with a measure of treatment outcome. Either, this was a pre-post difference score concerning standardized clinical symptom scales (both external and self-rating) or concerning any other physiological or subjective measure of fear taken in the fear-provoking situation. Alternatively, post-scores were also accepted as outcome measures provided that the pre-scores had been partialled out.
(5) In order to find as many studies as possible, the minimum sample size per study was required to be $N = 5$.
(6) There was no restriction concerning the type of publication. As to language, our search was restricted to English, German, French and Spanish publications.

Literature Search

The systematic literature search comprised four successive steps. (1) We first performed a keyword-based search in established databases (PsycINFO, PSYNDEX and ERIC). The search applied the search term presented in Appendix A and included all publications up to February 2014. (2) The keyword-based search was followed by a reverse search on the basis of the review by Craske *et al.* (2008) and the chapter by Crits-Christoph *et al.* (2013). Thus, we checked all studies that subsequently cited these two documents using *Google Scholar*. (3) The next step involved searching the reference lists of the 18 articles included until then plus the two documents on which the reverse search had been performed previously. (4) Finally, the search was supplemented by written correspondence with the authors of those studies meeting most inclusion criteria, but either lacked crucial statistical information or were not accessible via the common pathways. For economic reasons, this step only concerned documents published in the year 2003 or later, based on the consideration that research findings should be filed for ten years. This final step of search did not lead to the inclusion of any further document. Figure 1 displays the search process and the results.

Moderator Variables

A set of 10 potential moderator variables were defined *a priori*, and all selected studies were coded using these variables. All studies were coded by the first author using an *a priori* designed coding sheet and coding guide. We planned to analyse the effect of all moderator variables via meta-regression, but in meta-regression, the Satterhwaite approximation of the degrees of freedom of the regression coefficients for the random-effects model should be approximately at least four (Tipton, 2014), as otherwise inference can be unreliable. Since this criterion was fulfilled only in case of the first three moderators described below, we refrained from analyzing the remaining seven moderators that were not reliable enough from a methodological point of view.

The first moderator variable was the type of process measure with two factor levels being distinguished: *physiological* (i.e., any HR or SC parameters) and *non-physiological*, meaning self-ratings of fear or anxiety and external ratings of fear (made by the therapist or independent assessors). The second moderator was the type of outcome measure. Here, we distinguished between measures based on standardized clinical symptom scales, reflecting the gold standard of measuring treatment outcome in treatment research, and any other physiological or non-physiological measures gathered in the fear-provoking situation itself. Accordingly, the two factor levels were labelled *symptoms* and *situations*. The two moderators were chosen with reference to the finding of an asynchrony between various indicators of fear measured during exposure therapy, with very low correlations emerging between different types of measures (Gerew, Romney, & Leboeuf, 1989; Grey, Rachman, & Sartory, 1981; Leitenberg, Agras, Butz, & Wincze, 1971; Nesse *et al.*, 1985). Also, *type of process measure* was chosen with reference to Foa and Kozak's (1986) specific assumption that physiological activation is a more valid indicator of fear than self-report measures.

The third moderator was *ORB*. It was designed to account for the varying extent to which this sort of bias (i.e., selective reporting of certain effect sizes at the expense of others, Sterne, Egger, & Moher, 2008) was

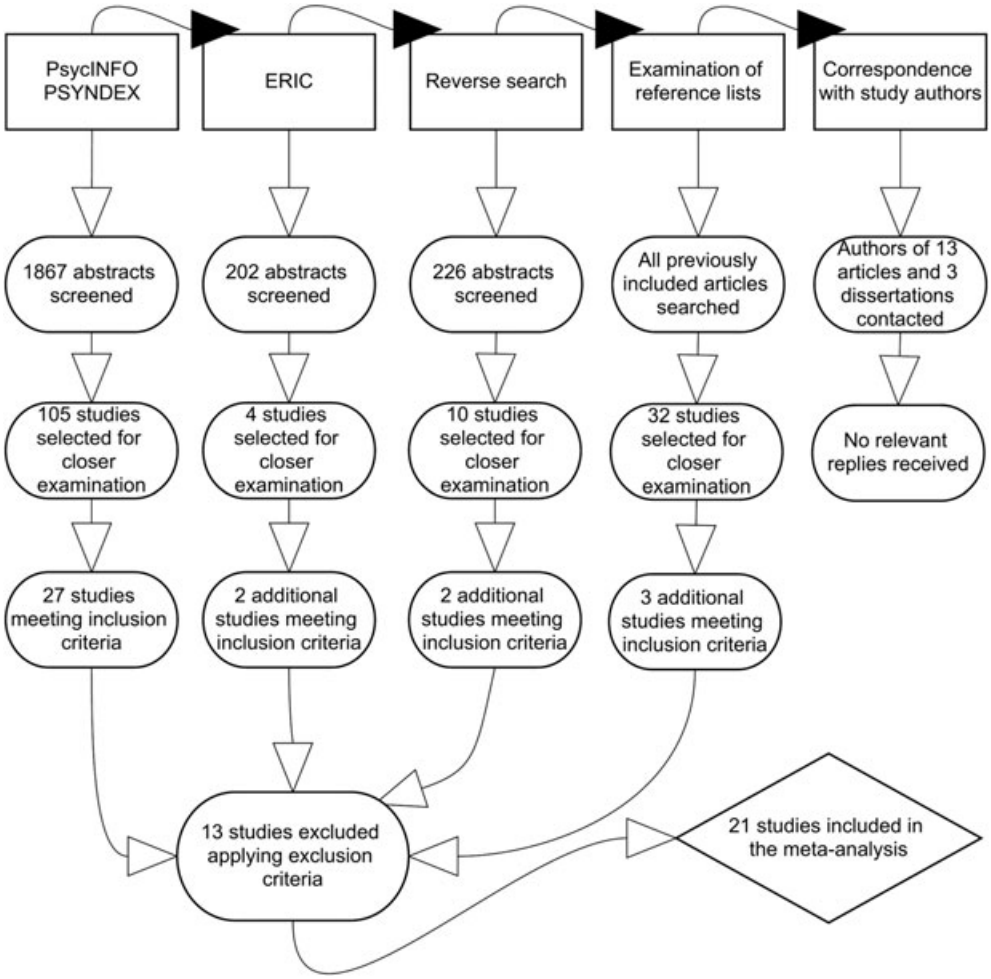

Figure 1. Flowchart displaying the successive literature search process. While the solid arrows indicate the sequence of search steps, the outlined arrows show how search results from each step were processed

evident in the included studies. Accordingly, the three factor levels were labelled *clearly evident*, *possible* and *none*, reflecting the amount of evidence for the omission of results, e.g., when a correlation coefficient had obviously been computed but was not reported due to non-significance (a case that was labelled 'clearly evident'). The factor level *none* was used for all cases in which no such suspicion was justified. In order to maximize statistical power through dichotomization, the factor levels *none* and *possible* were summarized for the analyses.

Although not being subject to meta-regression, the moderators *diagnosis* and *type of exposure* were nevertheless included in the forest plots to provide descriptive information. With regard to *diagnosis*, we applied the levels *phobic* (including specific phobias, agoraphobia and social phobia) and *other*, referring to all other anxiety disorders following DSM-IV-TR (APA, 2000), including subclinical forms. Thus, we also included PTSD and OCD although they are no longer categorized as anxiety disorders in DSM-5 (APA, 2013). In terms of *type of exposure*, we differentiated between exposure *in sensu* (including WED) and exposure *in vivo* (including exposure in virtual reality). Apart from these five moderators, we also considered the *purity of exposure* as reflected by the degree to which treatment was, for example, amended by elements of cognitive therapy, and the *duration of treatment*, reflecting the total time participants were exposed to fear-provoking stimuli or situations. The other three moderators were the *percentage of females* among the study participants, the participants' *mean age* and the *dropout rate*.

Statistical Analysis

Main Analyses

All subsequent analyses were performed with the help of the statistical software R, using the packages *robumeta*

(Fisher & Tipton, 2014), *metafor* (Viechtbauer, 2010) and *metasens* (Schwarzer, Carpenter, & Rücker, 2014). All correlation coefficients were transformed to the Fisher's *z* scale prior to data analysis. All analyses employed anα-value of 5%. Based on the preceding literature, a large amount of heterogeneity was expected. We therefore computed separate random-effects models for each of the three process variables (IFA, WSH, BSH) and calculated established statistics (Q, I^2, τ^2) for all three meta-analyses.

Due to a varying number of effect sizes per study (referred to as *endpoints* from now on), the analysis had to deal with statistical dependencies within the data. In consideration of these dependencies, all analyses except for those dealing with ORB or publication bias (PB) were performed by means of *robust variance estimation* (RVE; Tipton, 2014; Hedges, Tipton & Johnson, 2010). Additionally, all analyses applied the *small-sample adjustments* Tipton (2014) suggests for meta-analyses integrating less than 40 studies. Importantly, RVE does require the imputation of the correlation ρ (*rho*) of effect sizes within a study. We used the default value of 0.8 for all analyses subsequently reported, and the impact of the value set for this parameter was examined by sensitivity analyses as suggested by Fisher and Tipton (n.d.). As in regular random effect modelling, in RVE, the statistics Q, I^2 and τ^2 refer to the amount of variation or variance between *studies*, not between individual *endpoints*.

The three main moderator variables were examined in different ways. Since *type of process measure* and *type of outcome measure* were seen as the most important moderator variables with regard to EPT, these variables were analysed by computing separate meta-analyses for each combination of factor levels. However, our assumption was that endpoints were only comparable by meta-analysis if they corresponded with regard to the type of outcome measure. Therefore, e.g., all IFA endpoints with symptom-based outcome measures were pooled, but symptom-based and situation-based outcome measures were never combined. This procedure led to a total of five subsets labelled IFA-symptoms, IFA-situations, WSH-symptoms, WSH-situations and BSH-symptoms. As noted above, we assumed the different operationalizations of treatment outcome to be non-comparable, which precluded any hypothesis testing concerning possible differences between the two factor levels *symptoms* and *situations*. The impact of *type of process measure* was investigated by RVE meta-regression within each of the subgroups defined by the two factor levels of *type of outcome measure*.

Reporting Bias

As noted above, we also analysed PB and ORB - two different forms of reporting bias (Sterne *et al.*, 2008, p. 298). ORB has long been neglected, although it has been shown to constitute a severe problem concerning the conclusions drawn from meta-analyses and systematic reviews (Kirkham *et al.*, 2010). In contrast to the main analyses, the reporting bias analysis was based on aggregated data (i.e., only one value per study and per process variable), so that studies with multiple endpoints would not dominate. The aggregated values were computed as the arithmetic means of all correlation coefficients extracted from a study. Two further reasons for using aggregated data were the lack of methods for PB or ORB in RVE meta-analysis and the fact that PB methods intend to detect whether *whole studies* (and not single endpoints) are missing.

PB was investigated following the suggestion of Sterne *et al.* (2011) that at least 10 studies should be available. First, funnel plots (using the standard error on the vertical axis) were inspected for asymmetry. In a second step, the random-effects version of the Egger regression test was used (Egger, Smith, Schneider, & Minder, 1997; Harbord, Egger, & Sterne, 2006).

The influence of ORB on the results of the various meta-analyses was first examined by means of RVE meta-regression using *ORB* as a predictor of the effect size. This was not possible for the WSH endpoints, for there were too few of them. The logic behind this was that if selective reporting leads to the omission of insignificant results close to zero, studies *without* clearly evident bias should on average report smaller effect sizes - an effect that has for instance been demonstrated by Furukawa, Watanabe, Omori, Montori, and Guyatt (2007). Due to the aforementioned dichotomization process, *clearly evident* was the factor level used as reference category.

Additionally, the methods of Copas and Jackson (2004) and the Copas (1999) selection model were used to obtain worst case estimates for the amount of downward ORB. For the Copas and Jackson method, we imputed the number of missing studies based on the observed number of studies with clearly evident and/or possible ORB. For the Copas selection model, we followed the suggestions of Schwarzer, Carpenter, and Rücker (2010) and Carpenter, Schwarzer, Rücker and Künstler (2009) with respect to the selection model parameters.

RESULTS

Included Studies

In sum, 21 studies were included in the meta-analysis. Seventeen of these studies reported correlations between IFA and outcome to exposure therapy, comprising 57 individual endpoints. Five studies reported

correlations between WSH and outcome, with a total of seven endpoints, while eight studies included a total of 22 suitable correlations between outcome and BSH. The total sample sizes were $N = 490$ in the case of IFA, $N = 116$ for WSH and $N = 304$ for BSH. Table 1 presents a detailed description of the 21 studies including eight of the 10 moderators. Appendix B displays the references of the studies that were included in the meta-analysis.

Omission of Studies Applying Exclusion Criteria

Thirteen studies meeting the initial inclusion criteria were subsequently excluded following closer examination. The two most prominent reasons for subsequent exclusion were an inappropriate operationalization of the particular process variable and the fact that the process variable was correlated with raw post-scores, from which pre-scores had not been partialled out. The first criterion, for

Table 1. Description of the included studies

Study	Mean age	% female	Dropout rate (%)	Diagnosis	Type of expo.	Other treatment elements	ORB	PV
Alpers and Sell (2008)	48.4	80.0	44.2	Claustrophobia	IV (TA)	No	None	IFA (P, SR)
Bluett *et al.* (2014)	36.6	75.9	24.1	PTSD	IV + IS (TA)	Yes	None	IFA (SR)
Boulougouris *et al.* (1977)	30.3	33.3	0.0	OCD	IV + IS (TA)	Yes	Clearly evident	IFA (P)
Busscher *et al.* (2013)	38.4	56.0	36.7	Aviophobia	IV (TA)	NA	Possible	IFA (P)
Foa *et al.* (1983)	34.0	50.0	51.2	OCD	IV + IS (TA)	Yes	Clearly evident	IFA (SR) WSH (SR) BSH (SR)
Foa *et al.* (1995)	39.4	100	NA	PTSD	IS (TA)	No	None	IFA (ER, SR)
Freyth (2009)	38.9	62.5	25.0	ASD	IS (TA)	Yes	Possible	IFA (P, SR)
Gallagher and Resick (2012)	32.0	100	27.3	PTSD	IV + IS (TA)	Yes	Possible	BSH (SR)
Konig *et al.* (2014)	21.5	72.0	21.1	PTSD	WED (NTA)	No	Possible	IFA (P)
Kozak *et al.* (1988)	34.4	57.1	17.6	OCD	IV + IS (TA)	Yes	Clearly evident	IFA (P, SR) BSH (P, SR)
Mathews and Shaw (1973)	NA	100	NA	Simple phobia (spiders)	IS (NTA)	No	Possible	IFA (SR) WSH (SR)
Muehlberger *et al.* (2005)	43.9	86.7	25.0	Aviophobia	VR (TA)	Yes	None	IFA (P, SR) BSH (P, SR)
Orenstein and Carr (1975)	NA	100	0.0	Simple phobia (rats)	IS (NTA)	No	None	IFA (P, SR) WSH (P)
Pitman *et al.* (1996a)	42.5	0.0	42.9	PTSD	IS (TA)	Yes	Clearly evident	IFA (P) WSH (P) BSH (P)
Rauch *et al.* (2004)	31.0	100	1.4	PTSD	IV + IS (TA)	Yes	Possible	BSH (SR)
Sloan *et al.* (2005)	19.0	74.7	2.5	PTSD	WED (NTA)	No	Clearly evident	BSH (SR)
Stern and Marks (1973)	36.0	56.3	NA	Agoraphobia	IV + IS (TA)	No	Clearly evident	IFA (P)
Telch *et al.* (2000)	17.9	85.2	10.0	Claustrophobia	IV (NTA)	No	Possible	IFA (P)
Van Minnen and Hagenaars (2002)	35.1	50.0	24.4	PTSD	IS (TA)	Yes	Possible	WSH (SR) BSH (SR)
Watson and Marks (1971)	30.3	NA	0.0	Agoraphobia/simple phobia	IS (TA)	No	Clearly evident	IFA (P)
Woodward *et al.* (1997)	NA	0.0	25.0	PTSD	IS (TA)	Yes	Possible	IFA (P)

Note. Only the studies by Busscher *et al.* (2013), Watson and Marks (1971) and Woodward *et al.* (1997) applied a group therapy setting, all other studies employed individual therapy settings. The seventh column gives information on whether the exposure treatment was amended by any additional treatment elements or not. The last column provides information concerning the process measures for which correlations were extracted from each study. *Abbreviations*: PV = process variables. clin = clinical. sub = subclinical. IV = *in vivo*. IS = *in sensu*. WED = written emotional disclosure. VR = virtual reality. TA = therapist-assisted. NTA = non-therapist-assisted. expo. = exposure. ORB = outcome reporting bias. P = physiological. SR = self-report. ER = external rating; NA = missing value.

instance, referred to the case that 'IFA' was measured during confrontation with any non-phobic stimulus material, or measured directly at the beginning of the exposure session, since fear levels measured right at the start of the exposure situation can be assumed to reflect anticipatory anxiety rather than actual fear activation. The latter criterion was based on the argument that raw post-scores do not reflect an adequate measure of outcome, unless there is some form of statistical control applied to the pre-scores, whose variance would otherwise be ignored.

Outcome Reporting Bias

In seven studies, ORB was clearly evident, and nine studies at least gave rise to some suspicion (labelled *possible*). Only five studies did not raise any suspicion concerning reporting bias (labelled *none*). Taken together, 35.3% of all studies reporting IFA were affected by clearly evident ORB, as well as 40% of all studies on WSH and 50% of those reporting BSH

Meta-Analysis

As described in the method section, we computed separate meta-analyses for each of the subsets resulting from the various combinations of process variables, process measures and outcome measures. All meta-analyses were based on the original data set with all individual endpoints, so that they all reflect models based on RVE. Table 2 displays the results from these meta-analyses using the metric of Fisher's z transformed values. As can be seen from the results displayed, none of the weighted mean effects in the IFA subsets was significantly different from zero. For BSH, there emerged a significant weighted mean effect size of medium size for the case of non-physiological process measures, whereas the corresponding effect size for physiological process measures assumed a similar value but failed to reach significance due to the small number of endpoints. The weighted mean effect sizes for WSH resemble those for BSH; however, statistical power was too low to yield significant results. Figures 2 to 6 display the corresponding forest plots.

With regard to heterogeneity, it should be noted that, taking all relevant statistics (Q, τ^2, I^2) together, there was strong evidence in favor of a distribution of effect sizes (i.e., large variation between studies) within the IFA subsets, whereas the WSH and BSH subsets were marked by only little between-study variation. Thus, a random-effects model seemed appropriate.

Meta-Regression Concerning Type of Process Measure

Type of process measure could be subject to hypothesis testing via meta-regression (RVE) only concerning the three subsets IFA-symptoms, IFA-situations and BSH-symptoms. In the WSH subsets case, the small number of studies and endpoints precluded any reliable meta-regression. For the case of BSH-symptoms, the effect was virtually not evident ($b = -0.009$, $p = 0.966$). The results for both IFA-symptoms ($b = 0.254$, $p = 0.332$) and IFA-situations ($b = 0.865$, $p = 0.152$) pointed in the direction that physiological process measures lead to higher correlations between IFA and treatment outcome as opposed to non-physiological ones, although the results

Table 2. Results from the random-effects meta-analyses on the three process variables (RVE)

Subset										95% CI						
PV	OM	PM	k	n	b_0	SE	t	df	$p(t)$	LL	UL	τ^2	I^2	Q	df^*	$p(Q)$
IFA	sympt.	any	12	39	0.16	0.12	1.30	9.95	0.22	−0.11	0.43	0.11	74.17	42.90	11.08	<0.01
	sympt.	phys	9	22	0.26	0.14	1.82	7.42	0.11	−0.07	0.60	0.15	74.87	32.15	8.08	<0.01
	sympt.	non-phys	7	17	−0.05	0.12	−0.43	5.03	0.69	−0.35	0.25	0.05	51.38	12.49	6.07	0.05
	sit.	any	6	18	0.24	0.27	0.87	4.96	0.42	−0.47	0.94	0.52	88.77	45.41	5.10	<0.01
	sit.	phys	5	15	0.43	0.27	1.62	3.97	0.18	−0.31	1.17	0.61	89.48	39.09	4.11	<0.01
WSH	sympt.	any	3	4	0.37	0.10	3.65	1.84	0.08	−0.10	0.85	0.00	0.00	1.81	1.81	0.36
	sit.	any	2	3	0.54	0.09	6.06	1.00	0.10	−0.59	1.68	0.00	0.00	0.18	0.18	0.16
BSH	sympt.	any	8	22	0.35	0.05	6.95	4.42	**0.00**	0.22	0.49	0.00	0.00	6.27	6.27	0.42
	sympt.	phys	3	10	0.31	0.17	0.19	1.99	0.20	−0.40	1.03	0.10	55.83	4.73	2.09	0.10
	sympt.	non-phys	7	12	0.34	0.05	6.79	3.92	**0.00**	0.20	0.48	0.00	0.00	3.48	3.48	0.40

Note. Subsets were omitted if only data from one study were available. All values are given in the metric of Fisher's z transformed values. I^2 is a percentage. *Abbreviations:* sympt. = symptoms. sit. = situations. PV = process variable. OM = outcome measure. PM = process measure. k = number of independent studies. n = number of endpoints. b_0 = intercept of model (weighted mean effect size). SE = standard error. df = degrees of freedom (numbers represent small-sample corrected, Satterthwaite approximated values). CI = confidence interval. LL = lower limit. UL = upper limit. df^* = degrees of freedom for Q and I^2 (also representing small-sample corrected, approximate values). $p(Q) = p$ value for the chi-square test on whether the amount of variation between studies is significant.

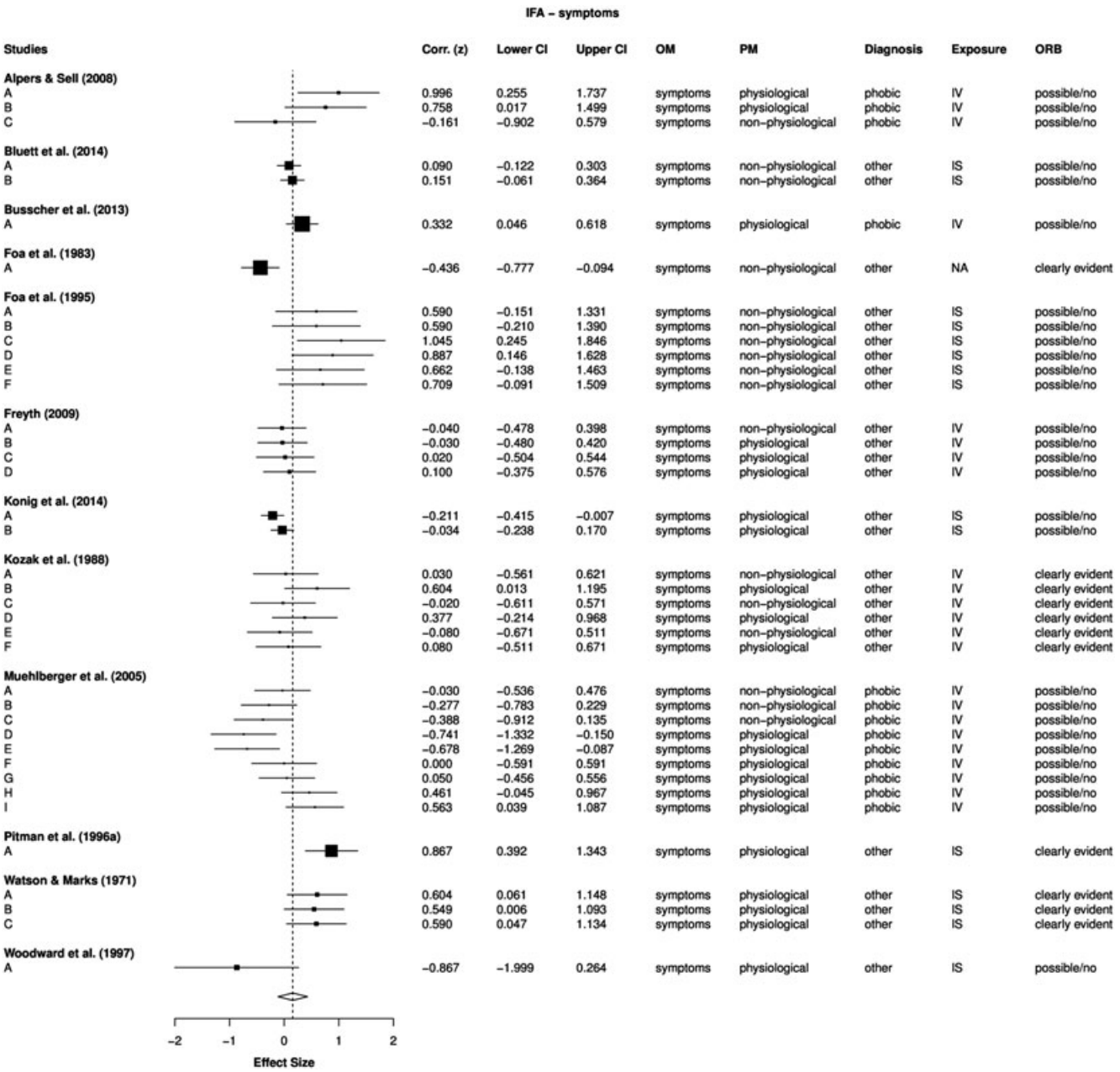

Figure 2. Forest plot for RVE meta-analysis of IFA-symptoms (Fisher's *z* transformed correlations). The uppercase letters below the study names indicate the various endpoints extracted from each study. Abbreviations: CI: confidence interval, OM: type of outcome measure, PM: type of process measure, ORB: outcome reporting bias, IV: *in vivo*, IS: *in sensu*

did not reach significance. It is, however, worth noting that while for IFA-symptoms the results appear very heterogeneous, the comparably large effect for IFA-situations can be predominantly traced back to the high relative weight of the study by Mathews and Shaw (1973), as demonstrated by the corresponding forest plot (Figure 3).

Reporting Bias

The subset structure of the data implied that the only case for which a funnel plot made sense was the IFA subset dealing with symptom-based outcome measures. For all other possible combinations, fewer than 10 studies were available, so the corresponding funnel plots were

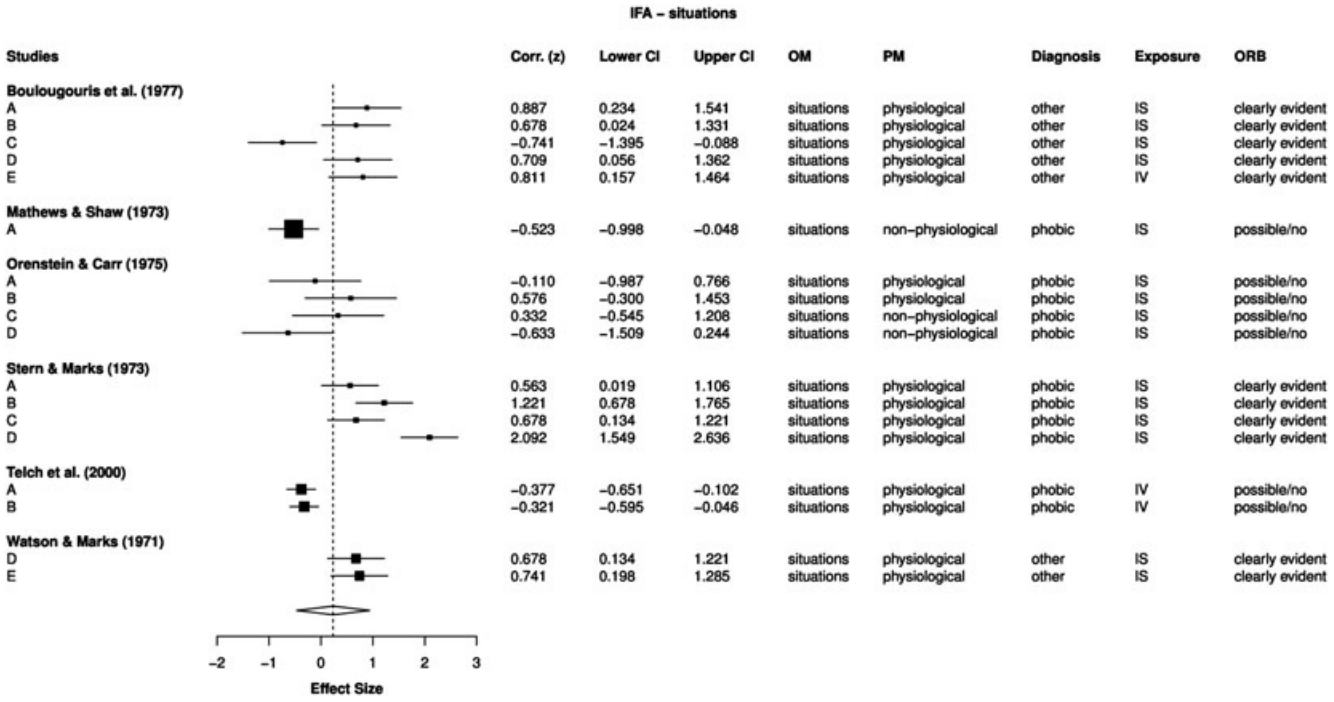

Figure 3. Forest plot for RVE meta-analysis of IFA-situations (Fisher's *z* transformed correlations). For further annotations, please see the legend below Figure 2

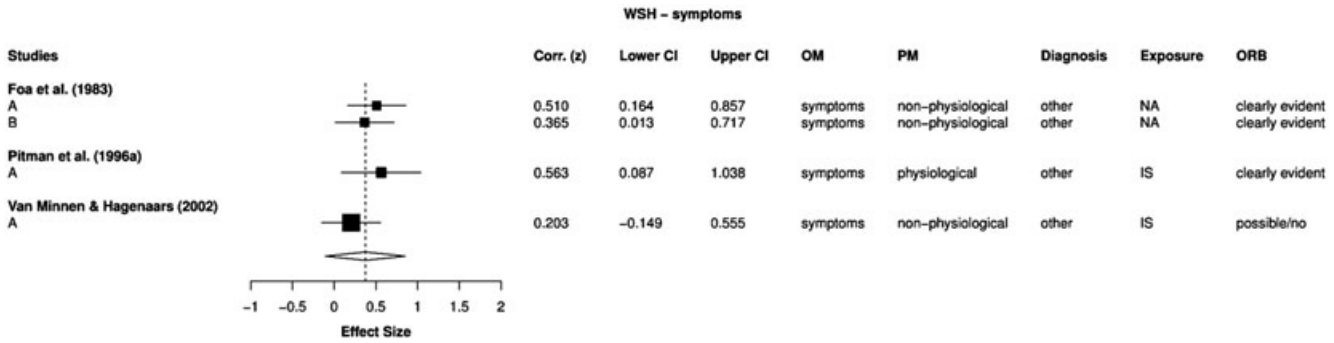

Figure 4. Forest plot for RVE meta-analysis of WSH-symptoms (Fisher's *z* transformed correlations). For further annotations, please see the legend below Figure 2

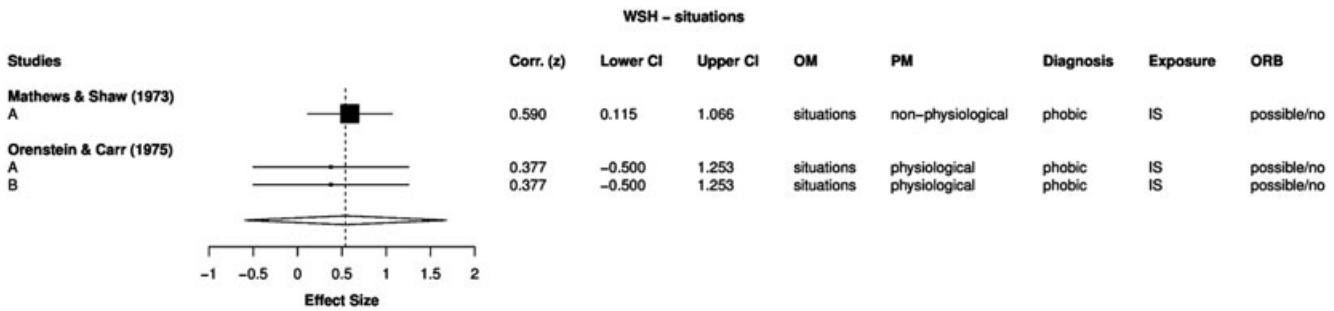

Figure 5. Forest plot for RVE meta-analysis of WSH-situations (Fisher's *z* transformed correlations). For further annotations, please see the legend below Figure 2

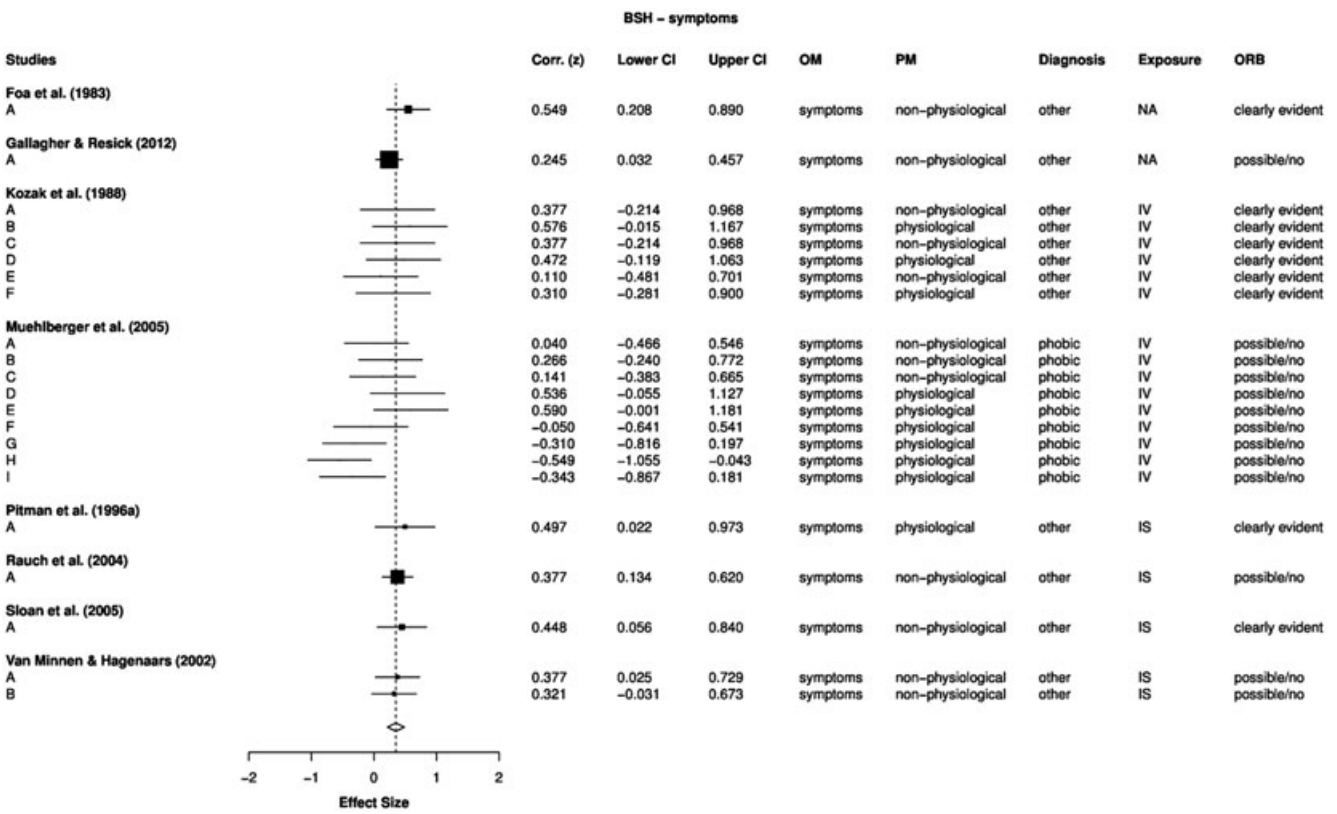

Figure 6. Forest plot for RVE meta-analysis of BSH-symptoms (Fisher's *z* transformed correlations). For further annotations, please see the legend below Figure 2

Table 3. Meta-regression concerning the influence of ORB (RVE)

Subset											95% CI			
PV	OM	PM	*k*	*n*	Coefficient	*b*	*SE*	*t*	*df*	*p*(*t*)	LL	UL	τ^2	I^2
IFA	sympt.	any	12	39	Intercept	0.11	0.11	0.97	6.17	0.37	−0.16	0.38	0.13	76.24
					Clearly evident	0.15	0.34	0.45	6.02	0.67	−0.67	0.98	—	—
	sympt.	phys	9	22	Intercept	0.09	0.15	0.59	4.19	0.58	−0.32	0.49	0.09	64.76
					Clearly evident	0.53	0.21	2.50	3.92	0.07	−0.06	1.12	—	—
	sympt.	non-phys	7	17	Intercept	0.05	0.09	0.58	2.68	0.61	−0.27	0.38	0.02	27.54
					Clearly evident	−0.36	0.21	−1.69	1.90	0.24	−1.33	0.61	—	—
	sit.	any	6	18	Intercept	−0.33	0.12	−2.63	1.83	0.13	−0.91	0.26	0.16	67.71
					Clearly evident	1.11	0.23	4.83	3.82	**0.01**	0.46	1.76	—	—
	sit.	phys	5	15	Intercept	−0.12	0.28	−0.43	1.00	0.74	−3.73	3.48	0.30	74.85
					Clearly evident	0.90	0.34	2.62	2.27	0.11	−0.42	2.23	—	—
BSH	sympt.	any	8	22	Intercept	0.29	0.05	5.95	2.06	0.03	0.09	0.50	0.00	0.00
					Clearly evident	0.20	0.06	3.18	4.25	**0.03**	0.03	0.36	—	—
	sympt.	phys	3	10	Intercept	−0.02	0.10	−0.21	1.00	0.87	−1.29	1.24	0.10	52.14
					Clearly evident	0.50	0.10	4.88	1.00	0.13	−0.80	1.79	—	—
	sympt.	non-phys	7	12	Intercept	0.30	0.05	6.37	2.10	0.02	0.11	0.49	0.00	0.00
					Clearly evident	0.17	0.08	2.23	2.76	0.12	−0.09	0.43	—	—

Note. The moderator variable *ORB* was dummy-coded so that *intercept* reflects the weighted mean effect size for the reference category (i.e., *possible/no*). All values are reported in the metric of Fisher's *z* transformed values. I^2 is a percentage. *Abbreviations*: sympt.: symptoms; sit: situations; phys: physiological; non-phys: non-physiological; PV: process variable; OM: outcome measure; PM: process measure; *k*: number of independent studies; *n*: number of endpoints; *b*: coefficient of model; *SE*: standard error; *df*: degrees of freedom (numbers represent small-sample corrected, Satterthwaite approximated values); *p*(t): p value; CI: confidence interval; LL: lower limit; UL: upper limit.

not computed. The funnel plot for the aforementioned subset seemed sufficiently symmetric; however, the Egger regression test bordered on significance ($z = 1.88$, $p = 0.06$). Further investigation of reporting bias was thus indicated.

Contrary to publication bias, ORB constituted a considerable issue in this meta-analysis. The results of the corresponding meta-regression are displayed in Table 3. Obviously, the influence of this moderator variable on the effect size was only significant in one IFA subset and one BSH subset (even though the effect indeed bordered on significance in two further IFA subsets). In both cases, the direction of the effect emerged as suggested by Furukawa *et al.* (2007): Clearly evident ORB was shown to be associated with significantly higher weighted mean effect sizes. However, in the BSH subset, the weighted mean effect size was still significant after subtracting the ORB effect, which was not the case in the IFA subset. In terms of WSH, statistical power was too low to perform the meta-regression concerning ORB.

The Copas and Jackson (2004) worst case estimate for the bias introduced by ORB corroborates the findings from Table 3: For IFA, the possibility of substantial downward ORB could not be excluded, implying that the weighted mean effect sizes for the IFA subsets may indeed reflect an overestimation. For BSH and WSH, however, none of the adjusted pooled correlations (based on the aggregated data) failed to be significant after taking into account the size of the bias. Thus, sensitivity analysis with the Copas selection model was in line with the other two methods.

Sensitivity Analysis

As planned, sensitivity analyses were performed on the 10 random-effects models presented in Table 2. In none of all the cases, any substantial influence of ρ was evident, suggesting that all estimates are fairly robust.

DISCUSSION

The main goal of our study was to estimate the association between the process variables IFA, BSH and WSH on the one hand and the outcome of exposure to situations provoking fear or anxiety on the other hand. We found evidence that BSH is moderately associated with treatment outcome measured by standardized clinical scales, whereas for the physiological measures, this association missed statistical significance due to the small sample size. With respect to WSH, the strength of association was also moderate to high, but due to the small number of studies investigating this association we only found a trend towards statistical significance. Finally, there was no evidence for a significant association between IFA and outcome of exposure. Of note, however, there was clear evidence that selective reporting is associated with the weighted mean effect sizes for BSH and IFA. Therefore, single results that show strong associations between IFA, BSH and outcome of exposure should be regarded with caution since they probably reflect overestimations.

Another important result of this meta-analysis is that, as anticipated, there was a large variation between studies concerning the strength of association between IFA and treatment outcome. However, only very little of this variation could eventually be explained by the moderator variables we examined, which can in part be traced back to the small number of studies and endpoints that kept statistical power small. Nevertheless, it is noteworthy that in terms of IFA, the results of the meta-regression suggested that the correlation with outcome of exposure is higher for physiological compared with non-physiological process measures, even though this effect did not reach statistical significance owing to the lack of power. Yet, it should be emphasized that this finding is in line with the predictions from EPT, which regards physiological measures as more valid indicators of the fear network's activation than non-physiological ones. As the lack of statistical power precluded the detection of any further moderators that may help to explain the remaining variation, further research is needed to shed light on this.

Thus, taking all results together, our meta-analysis has provided rather little evidence in support of the predictions derived from EPT. First, IFA as the alleged precondition for the success of exposure therapy (i.e., the element that is most specific to EPT) was not shown to be significantly associated with outcome, at least not in a linear way. Taking into account the significant impact of ORB, it is likely to assume that the association we found is even overestimated. With regard to BSH and WSH, we did find some evidence for an association with treatment outcome. However, whereas the results concerning WSH should be treated with maximal caution because of limited power, those for BSH were shown to be biassed by selective reporting and therefore are probably overestimated. Moreover, the association we found with regard to habituation could also be explained as a side-effect that occurs when symptoms decrease within or between sessions due to mechanisms other than emotional processing—which is why some alternative theoretical considerations seem appropriate.

Following cognitive theory (e.g., Clark, 1999), exposure should aim at correcting patients' faulty beliefs concerning the phobic situation or the phobic stimulus, such as the expectation that something terrible will happen when being confronted with it. In contrast to the habituation rationale, thus, the primary goal is not to experience fear reduction (unless the critical belief is that

fear will never decline), but rather to stay in the feared situation as long as necessary to test whether such expectations actually come true. This perspective draws on classical cognitive theories stating that the relevant intervening variable between stimulus and response is a cognition in the form of an appraisal or an interpretation (e.g., Beck & Clark, 1988), which can be directly tested by means of such a re-designed exposure task that is generally referred to as *behavioural experiment* (e.g., Clark, 1999). Following cognitive theories, it is not important to experience a high level of fear, but to refrain from safety strategies that prevent the testing of faulty beliefs about a feared outcome (Salkovskis, Clark, Hackmann, Wells, & Gelder, 1999). If relevant faulty beliefs are challenged, the patient is likely to experience a considerable amount of fear during the exposure task (or *behavioural experiment*). However, it is possible that a medium level of arousal is likely to be the best condition for new learning because very high levels of arousal are supposed to interfere with new learning (Baldi & Bucherelli, 2005), suggesting a curvilinear rather than a linear statistical association between fear activation and outcome—as it has been found for the treatment of depression (Carryer & Greenberg, 2010). Given that this is the case, it would make sense that we did not find any evidence in favour of a (linear) correlation between IFA and outcome.

Finally, our findings are also in line with the perspective of inhibitory learning. In accordance with compelling evidence demonstrating that extinction actually reflects the formation of newly learned associations, which compete with the original ones (rather than the definite elimination of these, e.g., Bouton, 2002), Foa and McNally (1996) suggested that instead of weakening old associations within the fear structure, exposure therapy involves the formation of 'new structures that override the influence of pathological ones' (p. 339). This perspective was enlarged by Craske *et al.* (2008), who emphasize that exposure should aim at 'developing competing, non-threat associations […] and ways of enhancing the accessibility and retrievability of those associations' (p. 21.), which they argue can be best achieved by varying emotional and situational contexts in which a person enters exposure situations during treatment—rather than working towards quick reductions of fear. Craske, Treanor, Conway, Zbozinek, and Vervliet (2014) furthermore highlight expectancy violation and removal of safety signals as key ingredients of effective exposure. The inhibitory learning perspective can therefore be reconciled with the result of our meta-analysis that, regarding the size of the correlation coefficient, WSH and BSH may be beneficial, but certainly not sufficient conditions for exposure to be successful.

Overall, our results do not specifically support EPT. First, IFA as the precondition for the success of exposure therapy (i.e., the element that is most specific to EPT) was not significantly associated with the outcome. For BSH and WSH, we found some evidence for an association with treatment outcome, but (apart from the statistical limitations described above) the question whether this fear reduction can be traced back to habituation (which would support EPT) or to cognitive processes, as described above, remains unresolved and requires further experimental research. For instance, an experimental design that allows either habituation or cognitive processes to be optimized could help to shed light on the question which process is more strongly linked to outcome of exposure therapy.

Our study has a number of strengths. First and most importantly, to our knowledge, this is the first systematic quantitative analysis of the association between IFA, WSH and BSH on the one hand and treatment outcome on the other hand. Especially when regarding the comparably small sample sizes of some of the studies included, a meta-analysis renders an enormous increase of information as compared with reviewing single studies because it maximizes statistical power and can therefore shed light on the nature of a statistical effect that largely differs across studies. Second, the fact that we found a large effect of reporting bias in the case of the association between IFA and the success of exposure therapy underlines the importance of a systematic quantitative analysis before drawing conclusions regarding the best way of delivering exposure therapy.

However, a number of limitations need to be taken into account when interpreting our findings. First and most importantly, our conclusions are limited by the comparably small number of studies that entered the meta-analysis, which is especially true of studies reporting correlations between WSH and outcome. It should therefore be emphasized that in order to further evaluate the model there is an urgent need for further empirical research on the association between outcome of exposure-based interventions and process variables such as those suggested by EPT.

Second, the studies included in our meta-analysis dealt with a comparably small range of different disorders, comprising specific phobia, agoraphobia, OCD, trauma-related disorders and related subclinical phenomena. Moreover, trauma-related disorders and OCD are in fact no longer subsumed under the diagnostic group of anxiety disorders in DSM-5 (APA, 2013), but it seemed reasonable to include these disorders in our meta-analysis as EPT also substantially refers to them (Foa and Kozak, 1986).

Third, our results concerning the moderator analysis are limited by the fact that due to a lack of power we only were able to analyse a subset of the initially coded moderators. Specifically, we could not analyse the impact of the *purity of exposure* and the *duration of treatment*. This implies that we cannot be sure whether the heterogeneous results

can, for instance, be traced back to differences concerning the type of protocol used or the rationale according to which exposure was delivered. Neither can we be certain about the role that the type of diagnosis or a possible interaction between diagnosis and treatment rationale play. Future research should therefore address the question to what extent correlations between the process variables of EPT and treatment outcome are affected by those parameters. More studies on the mechanisms underlying exposure-based interventions are still needed.

CONCLUSIONS

The most striking result of this meta-analysis is that, contrary to the predictions of EPT (Foa & Kozak, 1986), it did not find any systematic association between IFA and success of exposure therapy, while finding questionable evidence for the beneficial role of WSH and BSH. Moreover, the associations between IFA and BSH on the one hand and outcome of exposure therapy on the other hand were shown to be influenced by ORB evident in the studies included, which highlights the caution that seems appropriate with regard to drawing conclusions concerning the best way of applying exposure therapy. Thus, we conclude that the premises of EPT have no sufficient empirical foundation to draw final recommendations, since our results are equally in line with other perspectives on the working mechanisms of exposure therapy, such as the inhibitory learning perspective and behavioural testing approaches to exposure. Therefore, more attention should be paid to these approaches in the future. In sum, further experimental research is needed before final recommendations can be provided on the best ways to deliver exposure therapy under specific circumstances.

REFERENCES

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders, 4th edition, text revision* (4th ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders, fifth edition* (5th ed.). Washington, DC: American Psychiatric.

Baldi, E., & Bucherelli, C. (2005). The inverted "U-shaped" dose–effect relationships in learning and memory: Modulation of arousal and consolidation. *Nonlinearity in Biology, Toxicology, Medicine*, *3*, 9–21. DOI:10.2201/nonlin.003.01.002.

Beck, A. T., & Clark, D. A. (1988). Anxiety and depression: An information processing perspective. *Anxiety Research*, *1*, 23–36. DOI:10.1080/10615808808248218.

Bisson, J. I., Ehlers, A., Matthews, R., Pilling, S., Richards, D., & Turner, S. (2007). Psychological treatments for chronic post-traumatic stress disorder: Systematic review and meta-analysis. *The British Journal of Psychiatry, 190*, 97–104. DOI:10.1192/bjp.bp.106.021402.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry, 52*, 976–986. DOI:10.1016/S0006-3223(02)01546-9.

Carpenter, J. R., Schwarzer, G., Rücker, G., & Künstler, R. (2009). Empirical evaluation showed that the Copas selection model provided a useful summary in 80% of meta-analyses. *Journal of Clinical Epidemiology, 62*, 624–631. DOI:10.1016/j.jclinepi.2008.12.002.

Carryer, J. R., & Greenberg, L. S. (2010). Optimal levels of emotional arousal in experiential therapy of depression. *Journal of Consulting and Clinical Psychology, 78*, 190–199. DOI:10.1037/a0018401.

Clark, D. M. (1999). Anxiety disorders: Why they persist and how to treat them. *Behaviour Research and Therapy, 37*, S5–S27. DOI:10.1016/S0005-7967(99)00048-0.

Copas, J. (1999). What works?: Selectivity models and meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *162*, 95–109. DOI:10.1111/1467-985X.00123.

Copas, J., & Jackson, D. (2004). A bound for publication bias based on the fraction of unpublished studies. *Biometrics*, *60* (1), 146–153. DOI:10.1111/j.0006-341X.2004.00161.x.

Craske, M. G., Kircanski, K., Zelikowsky, M., Mystkowski, J., Chowdhury, N., & Baker, A. (2008). Optimizing inhibitory learning during exposure therapy. *Behaviour Research and Therapy, 46*, 5–27. DOI:10.1016/j.brat.2007.10.003.

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy, 58*, 10–23. DOI:10.1016/j.brat.2014.04.006.

Crits-Christoph, P., Gibbons, M. B. C., & Mukherjeed, D. (2013). Psychotherapy process-outcome research. In Lambert, M. J. (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed. pp. 298–340). Hoboken, New Jersey: John Wiley & Sons.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*, 629–634. DOI:10.1136/bmj.315.7109.629.

Emmelkamp, P., & Ehring, T. (2014). *The Wiley handbook of anxiety disorders* (1st ed.). Chichester, West Sussex: John Wiley & Sons.

Fisher, Z., & Tipton, E. (2014). robumeta: Robust variance meta-regression. R package version 1.1. Retrieved from: http://CRAN.R-project.org/package=robumeta

Fisher, Z., & Tipton, E. (n.d). robumeta: An R-package for robust variance estimation in meta-analysis. Retrieved from: http://blogs.cuit.columbia.edu/let2119/files/2013/03/Fisher-and-Tipton-.pdf

Foa, E. B., Huppert, J. D., & Cahill, S. P. (2006). Emotional processing theory: An update. In Rothbaum, B. O., & Rothbaum, B. O. (Ed.), *Pathological anxiety: Emotional processing in etiology and treatment* (ed. pp. 3–24). New York, NY, US: Guilford Press.

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin*, *99*, 20–35. DOI:10.1037/0033-2909.99.1.20.

Foa, E. B., & McNally, R. J. (1996). Mechanisms of change in exposure therapy. In Rapee, M. (Ed.), *Current controversies in the anxiety disorders* (ed. pp. 329–343). New York: The Guilford Press.

Furukawa, T. A., Watanabe, N., Omori, I. M., Montori, V. M., & Guyatt, G. H. (2007). Association between unreported outcomes and effect size estimates in Cochrane meta-analyses. *Jama*, *297*, 465–470. DOI:10.1001/jama.297.5.468-b.

Gerew, A. B., Romney, D. M., & Leboeuf, A. (1989). Synchrony and desynchrony in high and low arousal subjects undergoing therapeutic exposure. *Journal of Behavior Therapy and Experimental Psychiatry, 20*, 41–48. DOI:10.1016/0005-7916(89)90006-2.

Grey, S. J., Rachman, S., & Sartory, G. (1981). Return of fear: The role of inhibition. *Behaviour Research and Therapy, 19*, 135–143. DOI:10.1016/0005-7967(81)90037-1.

Harbord, R. M., Egger, M., & Sterne, J. A. (2006). A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine*, *25*, 3443–3457. DOI:10.1002/sim.2380.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*, 39–65. DOI:10.1002/jrsm.5.

Hofmann, S. G., & Smits, J. A. J. (2008). Cognitive-behavioral therapy for adult anxiety disorders: A meta-analysis of randomized placebo-controlled trials. *Journal of Clinical Psychiatry*, *69*, 621–632. DOI:10.4088/JCP.v69n0415.

Jaycox, L. H., Foa, E. B., & Morral, A. R. (1998). Influence of emotional engagement and habituation on exposure therapy for PTSD. *Journal of Consulting and Clinical Psychology, 66*, 185–192.

Kamphuis, J. H., & Telch, M. J. (2000). Effects of distraction and guided threat reappraisal on fear reduction during exposure-based treatments for specific fears. *Behaviour Research and Therapy, 38*, 1163–1181. DOI:10.1016/S0005-7967(99)00147-3.

Kirkham, J. J., Dwan, K. M., Altman, D. G., Gamble, C., Dodd, S., Smyth, R., et al. (2010). The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ [British Medical Journal], 340*, 637–640. DOI:10.1136/bmj.c365.

Lang, P. J. (1977). Imagery in therapy: An information processing analysis of fear. *Behavior Therapy, 8*, 862–886. DOI:10.1016/S0005-7894(77)80157-3.

Lang, P. J. (1979). A bio-informational theory of emotional imagery. *Psychophysiology, 16*, 495–512. DOI:10.1111/j.1469-8986.1979.tb01511.x.

Leitenberg, H., Agras, S., Butz, R., & Wincze, J. (1971). Relationship between heart rate and behavioral change during the treatment of phobias. *Journal of Abnormal Psychology, 78*, 59–68. DOI:10.1037/h0031478.

Mitte, K. (2005). A meta-analysis of the efficacy of psycho- and pharmacotherapy in panic disorder with and without agoraphobia. *Journal of Affective Disorders*, *88*, 27–45. DOI:10.1016/j.jad.2005.05.003.

Nesse, R. M., Curtis, G. C., Thyer, B. A., McCann, D. S., Huber-Smith, M. J., & Knopf, R. F. (1985). Endocrine and cardiovascular responses during phobic anxiety. *Psychosomatic Medicine*, *47*(4), 320–332. DOI:10.1097/00006842-198507000-00002.

Powers, M. B., Halpern, J. M., Ferenschak, M. P., Gillihan, S. J., & Foa, E. B. (2010). A meta-analytic review of prolonged exposure for posttraumatic stress disorder. *Clinical Psychology Review, 30*, 635–641. DOI:10.1016/j.cpr.2010.04.007.

Salkovskis, P. M., Clark, D. M., Hackmann, A., Wells, A., & Gelder, M. G. (1999). An experimental investigation of the role of safety-seeking behaviours in the maintenance of panic disorder with agoraphobia. *Behaviour Research And Therapy, 37*, 559–574. DOI:10.1016/S0005-7967(98)00153-3.

Schwarzer, G., Carpenter, J., & Rücker, G. (2010). Empirical evaluation suggests Copas selection model preferable to trim-and-fill method for selection bias in meta-analysis. *Journal of Clinical Epidemiology, 63*, 282–288. DOI:10.1016/j.jclinepi.2009.05.008.

Schwarzer, G., Carpenter, J., & Rücker, G. (2014). metasens: Advanced statistical methods to model and adjust for bias in meta-analysis. R package version 0.1-0. Retrieved from http://CRAN.R-project.org/package=metasens

Sterne, J. A. C., Egger, M., & Moher, D. (2008). Addressing reporting biases. In Higgins, J. P. T., & Green, S. (Ed.), *Cochrane handbook for systematic reviews of interventions* (1st ed. pp. 297–333). Chichester, England: John Wiley & Sons.

Sterne, J. A., Sutton, A. J., Ioannidis, J., Terrin, N., Jones, D. R., Lau, J., et al. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, 343. DOI:10.1136/bmj.d4002.

Telch, M. J., Cobb, A. R., & Lancaster, C. L. (2013). Agoraphobia. In Hofmann, S. G., Dozois, D. A., Rief, W., & Smits, J. J. (Ed.), *The Wiley handbook of cognitive behavioral therapy (Vol. 3)* (ed. pp. 941–978). Hoboken, NJ: Wiley-Blackwell.

Telch, M. J., Valentiner, D. P., Ilai, D., Petruzzi, D., & Hehmsoth, M. (2000). The facilitative effects of heart-rate feedback in the emotional processing of claustrophobic fear. *Behaviour Research and Therapy, 38*, 373–387. DOI:10.1016/S0005-7967(99)00038-8.

Tipton, E. (2014). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*. DOI:10.1037/met0000011.supp.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48.

Wolpe, J. (1969). *The practice of behavior therapy.* New York: Pergamon Press.

APPENDIX A:

Search term used for the keyword-based literature search

('exposure therapy' OR 'exposure treatment' OR 'exposure session' OR 'fear modification' OR flooding OR '*in vivo* exposure' OR 'imaginal exposure'
OR
(('behavior therapy' OR 'behaviour therapy') AND ('anxiety disorder' OR phobia OR 'generalized anxiety disorder' OR GAD OR OCD OR 'obsessive–compulsive disorder' OR 'posttraumatic stress disorder' OR PTSD OR panic OR agoraphobia)))
AND
(activation OR process OR habituation OR 'anxiety during treatment' OR 'fear during treatment' OR 'initial fear' OR 'initial anxiety' OR arousal OR within-session OR between-session OR 'heart rate' OR 'skin conductance' OR 'emotional processing' OR 'process variable' OR 'process measure')

APPENDIX B:

References of the studies included in the meta-analysis

Alpers, G. W., & Sell, R. (2008). And yet they correlate: Psychophysiological activation predicts self-report outcomes of exposure therapy in claustrophobia. *Journal of Anxiety Disorders*, *22*, 1101–1109. doi:10.1016/j.janxdis.2007.11.009

Bluett, E. J., Zoellner, L. A., & Feeny, N. C. (2014). Does change in distress matter? Mechanisms of change in prolonged exposure for PTSD. *Journal of Behavior Therapy and Experimental Psychiatry, 45*, 97–104. doi:10.1016/j.jbtep.2013.09.003

Boulougouris, J. C., Rabavilas, A. D., & Stefanis, C. (1977). Psychophysiological responses in obsessive–compulsive patients. *Behaviour Research and Therapy, 15*, 221–230. doi:10.1016/0005-7967(77)90019-5

Busscher, B., Spinhoven, P., van Gerwen, L. J., & de Geus, E. J. C. (2013). Anxiety sensitivity moderates the relationship of changes in physiological arousal with flight anxiety during in vivo exposure therapy. *Behaviour Research and Therapy, 51*, 98–105. doi:10.1016/j.brat.2012.10.009

Foa, E. B., Grayson, J. B., Steketee, G. S., Doppelt, H. G., Turner, R. M., & Latimer, P. R. (1983). Success and failure in the behavioral treatment of obsessive–compulsives. *Journal of Consulting and Clinical Psychology, 51*, 287–297. doi:10.1037/0022-006X.51.2.287

Foa, E. B., Riggs, D. S., Massie, E. D., & Yarczower, M. (1995). The impact of fear activation and anger on the efficacy of exposure treatment for posttraumatic stress disorder. *Behavior Therapy, 26*, 487–499. doi:10.1016/S0005-7894(05)80096-6

Gallagher, M. W., & Resick, P. A. (2012). Mechanisms of change in cognitive processing therapy and prolonged exposure therapy for PTSD: Preliminary evidence for the differential effects of hopelessness and habituation. *Cognitive Therapy and Research, 36*, 750–755. doi:10.1007/s10608-011-9423-6

Konig, A., Eonta, A., Dyal, S. R., & Vrana, S. R. (2014). Enhancing the benefits of written emotional disclosure through response training. *Behavior Therapy, 45*, 344–357. doi:10.1016/j.beth.2013.12.006

Kozak, M. J., Foa, E. B., & Steketee, G. (1988). Process and outcome of exposure treatment with obsessive–compulsives: Psychophysiological indicators of emotional processing. *Behavior Therapy, 19*, 157–169. doi:10.1016/S0005-7894(88)80039-X

Mathews, A., & Shaw, P. (1973). Emotional arousal and persuasion effects in flooding. *Behaviour Research and Therapy, 11*, 587–598. doi:10.1016/0005-7967(73)90117-4

Orenstein, H., & Carr, J. (1975). Implosion therapy by tape-recording. *Behaviour Research and Therapy, 13*, 172–182. doi:10.1016/0005-7967(75)90013-3

Muehlberger, A., Petrusek, S., Herrmann, M. J., & Pauli, P. (2005). Biocyberpsychologie: Subjektive und physiologische Reaktionen von Flugphobikern und Gesunden bei Exposition mit virtuellen Flügen [Biocyberpsychology: Subjective and physiological responses of flight phobics and healthy controls during exposure with virtual reality flights]. *Zeitschrift Für Klinische Psychologie Und Psychotherapie: Forschung Und Praxis, 34*, 133–143. doi:10.1026/1616-3443.34.2.133

Pitman, R. K., Orr, S. P., Altman, B., & Longpre, R. E. (1996). Emotional processing and outcome of imaginal flooding therapy in Vietnam Veterans with chronic posttraumatic stress disorder. *Comprehensive Psychiatry, 37*, 409–418. doi:10.1016/S0010-440X(96)90024-3

Rauch, S. A. M., Foa, E. B., Furr, J. M., & Filip, J. C. (2004). Imagery vividness and perceived anxious arousal in prolonged exposure treatment for PTSD. *Journal of Traumatic Stress, 17*, 461–465. doi:10.1007/s10960-004-5794-8

Sloan, D. M., Marx, B. P., & Epstein, E. M. (2005). Further examination of the exposure model underlying the efficacy of written emotional disclosure. *Journal of Consulting and Clinical Psychology, 73*, 549–554. doi:10.1037/0022-006X.73.3.549

Stern, R., & Marks, I. (1973). Brief and prolonged flooding: A comparison in agoraphobic patients. *Archives of General Psychiatry, 28*, 270–276.

Telch, M. J., Valentiner, D. P., Ilai, D., Petruzzi, D., & Hehmsoth, M. (2000). The facilitative effects of heart-rate feedback in the emotional processing of claustrophobic fear. *Behaviour Research and Therapy, 38*, 373–387. doi:10.1016/S0005-7967(99)00038-8

Van Minnen, A., & Hagenaars, M. (2002). Fear activation and habituation patterns as early process predictors of response to prolonged exposure treatment in PTSD. *Journal of Traumatic Stress, 15*, 359–367. doi:10.1023/A:1020177023209

Watson, J. P., & Marks, I. M. (1971). Relevant and irrelevant fear in flooding—A crossover study of phobic patients. *Behavior Therapy, 2*, 275–293.

Woodward, S. H., Drescher, K. D., Murphy, R. T., Ruzek, J. I., Foy, D. W., Arsenault, N. J., & Gusman, F. D. (1997). Heart rate during group flooding therapy for PTSD. *Integrative Physiological and Behavioral Science, 32*, 19–30.

