


Fibromyalgia and Nociceptive Flexion Reflex (NFR) Threshold: A Systematic Review, Meta-Analysis, and Identification of a Possible Source of Heterogeneity

Mohammadreza Amiri ¹

Jamie Rhudy ²

Kei Masani ^{1,3}

Dinesh Kumbhare ^{1,3}

¹KITE Research Institute, University Health Network, Toronto, ON, M5G 2A2, Canada; ²College of Arts & Sciences Psychology, University of Tulsa, Tulsa, OK, 74104, USA; ³Department of Medicine, Division of Physical Medicine and Rehabilitation, University of Toronto, Toronto, ON, M5S, Canada

Background: The nociceptive flexion reflex is a physiological, polysynaptic reflex and refers to the level that an appropriate withdrawal response activates when a painful stimulus is detected. The nociceptive flexion reflex threshold (NFRthr) is defined as the lowest noxious stimulation intensity required to trigger a reflex motor response. Despite wide utilization and reports of the NFRthr, there has been no consensus on a standard and/or best method in assessment of the NFRthr.

Objective: To systematically review the literature that compared the NFRthr between individuals with fibromyalgia (FM) and healthy controls; and to identify a source of heterogeneity in these trials.

Methods: Employing the Cochrane methodology, we systematically searched Ovid MEDLINE, Embase, Cochrane Clinical Answers, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, and PsycINFO for clinical case-controlled trials assessing the NFRthr in individuals with and without fibromyalgia from inception to July 2019. Selected articles were passed for data extraction and meta-analyses. We utilized the random-effects model for meta-analysis assuming the true effect size may vary between studies. The sample sizes as a possible source of heterogeneity in multiple meta-regressions were investigated. This systematic review and meta-analysis were registered in PROSPERO before data extraction.

Results: Nine studies met our criteria and were included in the meta-analysis. Methodologies and settings varied between studies, eg, stimulation intensity, duration, and the current increments. Only two articles comprehensively described and reported details about electromyogram amplification, latency, and sampling rate. Evidence from 423 patients with fibromyalgia and 326 healthy individuals suggested that there may not be a meaningful decreased NFRthr in patients (overall mean difference = -3.16; 95% CI: -6.82 to 0.50; $Z = 1.69$; $P = 0.09$). Published effect sizes were not homogenous ($I^2 = 0.91$, $\tau^2 = 25.04$, $\chi^2 = 91.22$, $df = 8$, $P < 0.00001$). The multiple meta-regression analyses indicated that total and female sample sizes might be the main sources of heterogeneity for the effect sizes $SS_{total} = -0.0570$, $P = 0.040$; $SS_{female} = -0.0569$; $P = 0.047$.

Conclusion: Evidence suggests that the nociceptive flexion reflex threshold may not be different between patients with fibromyalgia and healthy controls. A unified and rigorous methodology and sample size calculation (probably sex specific investigation) is required for the assessment of nociceptive flexion reflex threshold in patients with fibromyalgia.

Keywords: nociceptive flexion reflex threshold, NFR, fibromyalgia, systematic review and meta-analysis, meta-regression, heterogeneity, sex differences

Correspondence: Dinesh Kumbhare
KITE Research Institute, University Health Network, 550 University Ave, Toronto, ON, M5G 2A2, Canada
Tel +1 416 597-3422 x 4520
Email dinesh.kumbhare@uhn.ca

Summary

Systematic review and meta-analysis suggest that the NFRthr is not different between fibromyalgia and healthy controls. A rigorous methodology and sample size calculation is required.

Introduction

Fibromyalgia (FM) is usually accompanied by chronic widespread pain, fatigue, tenderness, sleep disturbances, and psychological distress.¹⁻⁴ Achieving accurate diagnosis and prognosis is critical since treatment is multimodal and expensive.⁵ The mechanisms responsible for the pain associated with FM include peripheral nociceptive input and spinal cord hyperexcitability.⁶⁻⁸ To evaluate chronic pain, efforts have been made to develop a measure of spinal nociceptive processes compared to relying on subjective patient self-reporting.^{9,10} For that, the nociceptive flexion reflex threshold (NFRthr) has gained particular attention in recent years¹⁰ to uncover the pain mechanisms behind the central sensitization and spinal cord hyperexcitability⁷ that was originally described as a direct measure of spinal cord segmental excitability and neuroplasticity.⁹ This threshold may play an important role in a diagnostic categorization that provides a better response to treatment.¹¹

The NFRthr is a physiological, polysynaptic reflex and refers to the level that an appropriate withdrawal response activates when a painful stimulus is detected.¹⁰ The NFRthr is defined as the lowest noxious stimulation intensity required to trigger a reflex motor response.^{12,13}

Despite wide utilization and reports of the NFRthr, there has been no consensus on a standard and/or best method in assessment of the NFRthr.¹⁴ However, a brief method¹⁵ and a possible solution has been suggested to standardize the methodology for the evaluation of NFRthr in different populations.¹⁴ In this procedure, the accurate and reliable criteria for defining NFRthr have been suggested to be the standardized peak (NFR Interval Peak *z* score) and mean (NFR Interval *z* score) biceps femoris electromyogram (EMG) activity.¹⁴ Nonetheless, it is known that human characteristics such as sex/gender can influence the magnitude of the threshold.¹⁶ For instance, females have significantly lower NFRthr compared to males^{16,17} which was discussed as a potential source of heterogeneity in trials investigating the NFRthr since the higher population of females in some trials in FM patients seemed to influence

the effect size. Lim et al⁷ synthesized the available information from the trials evaluating the NFRthr trials until 2010. For the FM subgroup, a random-effects model was correctly employed as the trials' methodologies to evaluate the NFRthr differed in stimuli parameters. The studies published up to 2010 showed a significantly lower standardized mean difference in NFRthr among patients suffering from FM compared to healthy controls (-0.63 ; 95% confidence interval [CI]: -0.93 to -0.34 ; $P < 0.0001$) However, the authors noted that the presence of an efficacious diffuse noxious inhibitory control (DNIC) in some participants may result in the non-significant effect sizes.⁷ The highly effective descending DNIC modulation of noxious input in some patients may impose irresponsiveness to the noxious stimulus with a reflex activation of the bicep femoris (BF) muscle.⁷ In addition, it was shown in a systematic review that the population of males in a study may exert great DNIC effects compared to females.¹⁸ Moreover, the experimental methodology and the measurement modes of the effect size may affect the sex differences in the DNIC.¹⁸ For that reason, inclusion of unbalanced sample size of males and females with effective descending DNIC may introduce a biased effect size in the NFRthr difference between FM and healthy individuals.^{18,19} Nonetheless, we noted that in the Lim et al's systematic review,⁷ the only FM trial in which the NFRthr deficit did not exist, the number of female participants was comparatively lower²⁰ than the rest of trials in which effect sizes were consistent.^{9,21,22} This mixture of findings may imply that different structures of sample sizes including unbalanced number of females and males might be a source of heterogeneity to the estimation of effect sizes (ie, NFRthr mean difference). Accordingly, using meta-analysis and multiple meta-regression we planned to investigate whether the total, female, and male sample sizes are influential in determination of NFRthr in FM patients. This effort is due to the ultimate concern that is the results obtained from an objective biomarker should be without characteristics that impose any source of heterogeneity or at least minimizing such risks. Therefore, to investigate possible source of heterogeneity and eliminate or minimize them in future research we aimed to synthesize the available information on the evidence for the NFRthr in FM patients and provide evidence that sex-specific sample size may act as a possible source of bias. A secondary objective was to report a meta-analysis on the available evidence on the NFRthr in FM patients compared to healthy controls.

Methodology

Study Registration

The review protocol of this study was registered in PROSPERO (CRD42019140354). Registration of systematic reviews and meta-analyses ensures adherence to initial goals, improves research quality and avoids data dredging.²³

Systematic Search Strategy

Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses, the PRISMA Statement,²⁴ we systematically searched Ovid MEDLINE, Embase, Cochrane Clinical Answers, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, and PsycINFO using methodology outlined by the Cochrane collaboration²⁵ from inception to July 2019. We used the following search terms: nocicept* (flex* OR withdraw*) adj7 reflex; NFR; flex* withdraw* reflex; withdrawal reflex; RIII reflex; (spinal OR central) hyperexcitability; pain; Fibromyalg*; widespread pain; chronic adj3 (widespread OR body wide) pain. The only limits applied to our search strategy were “human” and “English”. We also included abstracts if they reported the complete methodologies taken and the NFRthr comparison within fibromyalgia or chronic widespread pain conditions to control subjects.

Inclusion and Exclusion Criteria

The scope of this review was to target the conducted clinical controlled human trials that compared the NFRthr between patients suffering from fibromyalgia or chronic widespread pain and healthy individuals that shared similar demographic characteristics. Studies that only reported patients' feedback about pain or lacked comparators and not report NFRthr results were not included in the analysis. Studies were included if the methodology and results were comprehensively reported, ie, sample sizes for healthy controls and patients, NFRthr mean (standard deviation) or median (interquartile range or quartiles), device settings and methods used for the NFRthr measurement.

Selection Agreement and Disagreement Settlement

All the articles found by the search were reviewed and assessed for eligibility by both authors (MA and DK)

independently. The only disagreement²⁶ was resolved by an experienced colleague.

Data Extraction

The following data were extracted from each study: study characteristics such as sample size; duration of symptoms; testing procedure including stimulation and recording; and, reflex threshold in milliamperes (mA: reported as mean and standard deviation or median and interquartile range). Data were extracted from each of the articles by one of the authors and confirmed by another author. Furthermore, where the NFRthr values were presented in plots, values were extracted with the best precision possible and were confirmed by two authors (MA and DK) and then included in the meta-analysis. Also, when only mean NFRthr and the probability of difference between FM and healthy controls were reported we utilized the Cochrane additional calculators to calculate the best estimates of the standard deviations for each group.²⁷ Finally, if mean and standard error was reported in the studies the proper conversion formulae were used to calculate the standard deviation.

Source of Bias

We used the contour-enhanced funnel plot that displays areas of statistical significance for the effect sizes of trials.²⁸ We followed the recommendation that discourages the use of scales for quality assessment since the number (produced by the scale) does not reflect the specific weakness or strength of the trial.²⁵ Using the effect sizes of trials and the inversed standard errors (ie, precision), the statistical significance (eg, $P < 0.01$, $P < 0.05$, $P < 0.1$) of any point on a funnel plot can be calculated and categorizes studies based on their precision. Funnel plot asymmetry has long been equated with publication bias.²⁹ However, the contour-enhanced funnel plot is a generic means of displaying small-study effects – a tendency for the intervention effects estimated in smaller studies to differ from those estimated in larger studies.³⁰ Thus, in a contour-enhanced funnel plot, we evaluated the presence of potential source of heterogeneity among the included studies. Peters et al²⁸ suggest that

If studies appear to be missing in areas of statistical non-significance, then this adds credence to the possibility that the asymmetry is due to publication bias. Conversely, if the supposed missing studies are in areas of higher statistical significance, this would suggest the cause of the

asymmetry may be more likely to be due to factors other than publication bias, such as variable study quality.

Meta-Analysis

Pooled results using formal meta-analysis techniques using RevMan 5.3 (The Cochrane Collaboration, 2006) and Stata (Release 2014, Stata Corp, College Station, TX) were analyzed comprehensively. In order to generate a point estimate for the effect and to incorporate the variability of the reflex thresholds described in the literature, we calculated mean differences and 95% confidence intervals.³¹ Following Borenstein et al,³¹ due to variability of the stimuli employed for the evaluation of the NFRthr we employed a random-effects model. Studies lacking important pertinent information, eg, mean and standard deviation of NFRthr in FM patients and/or in healthy controls, were handled as follows: i) If the median and range (interquartile range or first-quartile and third-quartile) were mentioned: we used an update and robust methodology to calculate the required statistics;³² ii) If the required statistics could not be extracted, obtained from contacting the authors, estimated from figures, appendices, etc, then the study was determined to not satisfactorily fit the inclusion criteria and therefore was excluded.

Following the registered protocol for this study, we did not investigate any other trial characteristics that are deemed to act as a possible source for heterogeneity. Nevertheless, for the meta-regression analysis we used the metareg package in Stata³³ utilizing a random-effects model due to versatile methodologies incorporated to estimate the NFRthr in the trials. The random-effects model assumes the within-study and between-studies effects have normal distributions (Ch. 6).³⁴ In multiple meta-regression models, the trials' effect sizes (difference in mean of NFR threshold estimated between FM patients and healthy individual) was regressed against the total, female, or male sample sizes. For the metareg package, the following options were chosen: i) restricted maximum likelihood (REML) for the estimation of τ^2 ;^{17,36} and, ii) standard errors, p-values, and confidence intervals for coefficients were estimated with modifications suggested by Knapp and Hartung.³⁵

Results

Systematic Literature Search and Selected Studies

Using the designed search domains, we systematically searched the selected electronic databases ([Appendix 1](#)).

Briefly, we obtained a total of 73 articles from electronic databases that satisfied our search strategy. After removing duplicates ($n = 39$), a total of 34 articles were selected for comprehensive screening. The authors applied the inclusion and exclusion criteria after which 9 scientific trials were passed to perform full-text comprehensive information extraction. [Figure 1](#) contains the PRISMA flowchart. The systematic literature search was performed by MA with the help of our librarian, Ms J. Babineau, who is a professional Information Specialist and has received training from the Cochrane Collaboration methodology and double-checked, reviewed and confirmed by author DK.

Risk of Bias

In [Figure 2](#), from each included study the precision (1/standard errors) is plotted against the effect size (ie, difference in means). The standard errors on a reversed scale place the more statistically powered studies towards the top of the plot (Ch. 10).²⁵ Results from the contour-funnel precision plot is as the following: i) effect sizes are non-significant ($P > 0.1$): Four trials were present in this zone that indicates that their effect sizes were not different from zero; ii) effect sizes are marginally non-significant ($0.05 < P < 0.1$): none of the studies fell inside this area; iii) effect sizes are different from zero ($0.01 < P < 0.05$): none of the trials' effect sizes are located inside this category; iv) effect sizes are highly statistically different than zero ($P < 0.01$): Five trials (55%) had highly significant effect sizes from which 4 trials (80%) indicated deficits in NFRthr in FM patients compared to healthy controls and 1 trial (20%) found positive difference in NFRthr means in FM patients and healthy individuals. The spread of the trials in this figure and the absence of symmetry between the dots implies the possibility of bias among trials' effect sizes.

Meta-Analysis

The sample size, mean, and standard deviation from the FM patients and healthy controls were extracted ([Appendix 2](#)) and were meta-analyzed in the forest plot illustrated in [Figure 3](#). The plot revealed interesting results. First, among the included studies four trials (44%) reported significantly lower NFRthr among FM patients,^{9,21,36,37} one study (11%) found significantly higher NFRthr,³⁹ and four trials (44%) reported non-significant differences between the FM patients and the healthy controls.^{20,22,26,28} Secondly, there was a large heterogeneity between the studies ($I^2=91\%;\tau^2=25.04$;

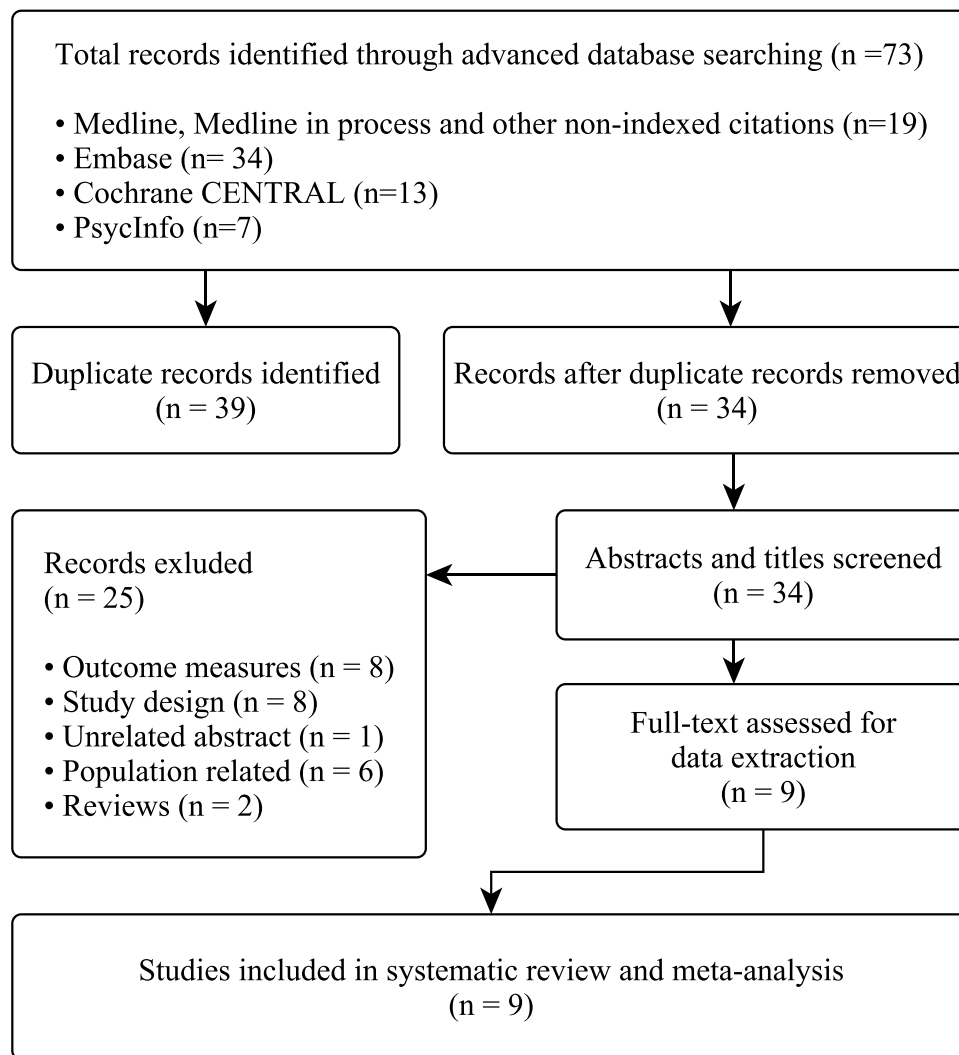


Figure 1 PRISMA flowchart of study selection.

Notes: PRISMA figure adapted from Liberati A, Altman D, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of Clinical Epidemiology*. 2009;62(10):e1-e34. Creative Commons.

$\chi^2=91.22$, $df=8$; $P<0.00001$). Finally, there was a statistically non-significant reduction in NFRthr among FM patients compared to the healthy controls (mean difference = -3.16 ; 95%CI : -6.82 to 0.50 ; $Z = 1.69$; $P=0.09$). The versatile findings of the studies may imply the possibility of the existence of source of bias among these trials. This was further investigated using meta-regression to test our hypotheses.

In **Tables 1** and **2**, the details about the procedure of measuring the NFRthr including patient positions, electrical stimulation details, study conclusion, electromyography (EMG), latency, sampling rate, and determinant of NFR threshold of the trials are summarized. For the site of the electrical stimulation, all the studies used the sural nerve and recording over the biceps femoris. We

noticed that the stimulation intensity, duration, and the increments were variable between the studies for example, one study reported 20 ms trains of 6 rectangular pulses with 1 mA increments,²⁰ another utilized five 1 ms rectangular shocks delivered at 2 mA and variable milliamperere increments,²² and Desmeules et al²¹ incorporated single rectangular impulses of a half a second and variable increments. The increments were either variable,^{21,22,36,38} 1 mA,^{9,20,37,39} and 4 mA.²⁶ All patients were relaxed in a sitting position or a supine position where the knee angle was variable between the studies. The EMG amplification and filtering were only reported in three studies^{9,38,39} from which the actual configuration is only reported by Rhudy et al³⁹ and Umeda et al.³⁸ From these two tables the versatile

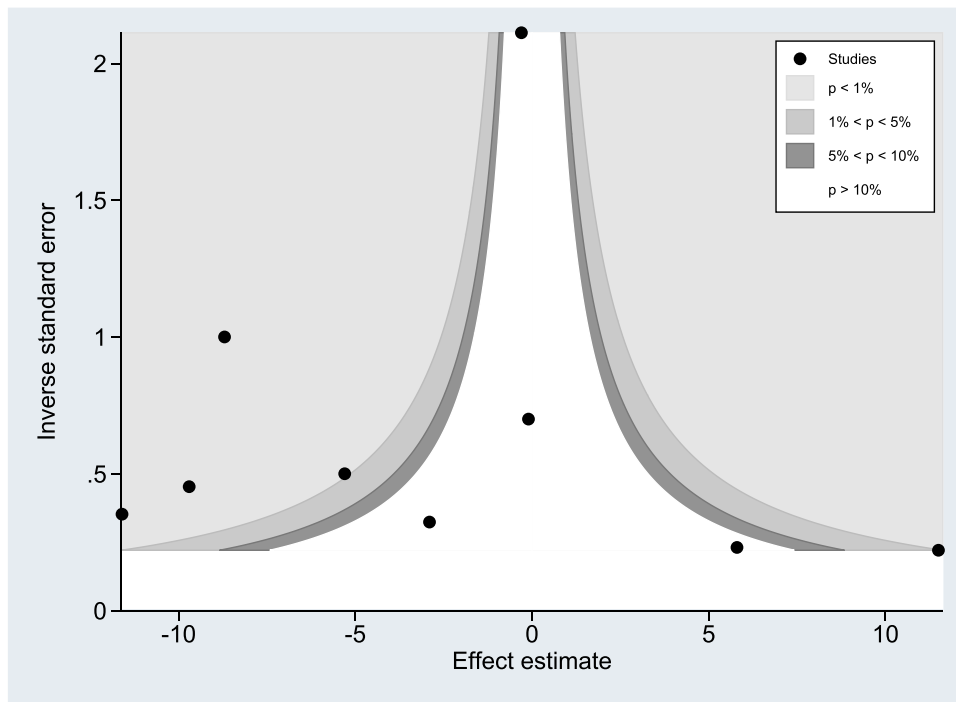


Figure 2 Contour-funnel plot.

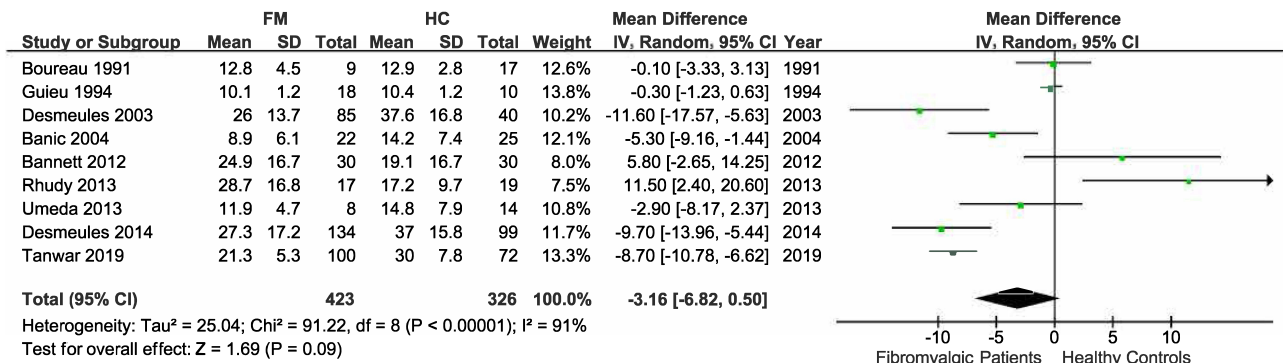


Figure 3 Forest plot.

methodologies used for evaluation of the NFRthr are highly obvious and might be a source of effect size bias.

Utilizing the REML meta-regression methodology, the total, female, and male sample sizes were meta-regressed in separate models against the effect sizes (Table 3). The analyses indicated that the total and female sample sizes significantly influenced the effect size while the sample size of males did not show such significance even though the slope of its meta-regression line is steeper compared to the ones representing the total or female sample size (Figure 4). Interpreting the coefficients, of all other study characteristics held unchanged. A ten-unit increment in the

number of total sample size in a trial the effect size is expected to decrease by 0.5 unit ($\beta_{\text{totalsamplesize}} = -0.057; P = 0.040$) and very similar results for findings from the female sample size ($\beta_{\text{femalesamplesize}} = -0.057; P = 0.047$). Nevertheless, the sample size of males did not depict any significant influence on the magnitude of the effect size ($\beta_{\text{malesamplesize}} = -0.327; P = 0.404$) even though the magnitude of the slope is larger. Figure 4 is a visualization of the trends of the effect sizes against total, female, and male sample sizes. Note here that the total and female sample sizes were able to explain about 62.15% and 65.15% of the

Table 1 Study Characteristics of Patient Position, Electrical Stimulation Details, and Authors' Conclusion

Study	Patient Position	Site of ES	Stimulation Parameters and Intensity	Interval Between Pulses or Trains	Increments	Authors' Conclusion
Banic (2004) ⁹	Sitting; Leg rest was placed under the knee to obtain 30° semi-flexion	Just distal to the right lateral malleolus (innervation area of the SN)	A 25 ms, five times, 1 ms, square-wave impulse	5 ms	1 mA	FM<HC (significant)
Boureau (1991) ²⁰	Sitting in a comfortable armchair to obtain good muscular relaxation	SN behind lateral malleolus	20 ms train of 6 rectangular pulses (1 ms duration each)	2.8 ms	1st session: 10 consecutive ascending series of 1 mA (method of limits); 2nd session (for patients only): 5 fixed intensity intervals (method of constant stimuli)	FM<HC (non-significant)
Desmeules (2003) ²¹	Rested comfortably in a supine position to obtain muscular relaxation	SN in Retromalleolar track	Single rectangular impulses of 0.5 ms	Variable: 6–10 sec	Variable: 1–100 mA	FM<HC (significant)
Guieu (1994) ²²	Sitting in an armchair in a semi-reclining position	SN	Series of five 1 ms rectangular shocks delivered at 2 ms intervals with a total duration of 13 ms	2 ms	Variable: Current of ES is not reported. 3 “increasing-decreasing” intensity series at a frequency of 4/min (each intensity level repeated 6 times)	FM<HC (non-significant)
Bennett (2012) ²⁶	Sitting	Volar forearm, mid trapezius, mid gluteal area and mid-point of anterior thigh with average of 3 reading at each site	NR	NR	4 mA	FM>HC (non-significant)
Desmeules (2014) ³⁶	NR	SN in Retromalleolar track	Variable intensities between 1–100 mA	Variable: 6–10 sec	Variable: 1–100 mA	FM<HC (significant)
Tanwar (2019) ³⁷	Supine position rested for 30 min prior to recording session	SN in the Retromalleolar region	Train of five rectangular pulses of 1 ms duration at 200 Hz.	Variable: 5–15 sec	Steps of 5 V until the response was recorded; Strength was between 0.1–50 V randomly at 5–15 s to avoid habituation	FM<HC (significant)

(Continued)

Table I (Continued).

Study	Patient Position	Site of ES	Stimulation Parameters and Intensity	Interval Between Pulses or Trains	Increments	Authors' Conclusion
Umeda (2013) ³⁸	Sit upright in an experimental chair; hip joint flexed about 90° and the right ankle positioned on a footrest with the right knee flexed 60° from horizontal	Retromalleolar pathway of the SN	Trains of five 1-ms rectangular pulses with a 3-ms interpulse interval	3 ms	4-2-1 mA; Series of 1 mA test stimuli until Max. 40 mA reached	FM<HC (non-significant)
Rhudy (2013) ³⁹	Seated in a reclining chair; knee angle at 160°	Left leg over Retromalleolar pathway of the SN	Each ES was a train of 1-ms square wave pulses delivered at 250 Hz	8 ms	1 mA; Max. 50 mA ensuring safety.	FM>HC (significant)

Abbreviations: ES, electrical stimulation; EMG, electromyography; NFR, nociceptive flexion reflex; NR, not reported; SN, sural nerve; BF, biceps femoris.

between-study effect size variance, respectively. In addition, we noted that the percentage of residual variation due to heterogeneity ($I^2_{residual}$) is considerably lower compared to other models (ie, Model I compared to Model III and Model IV) at 67.13% and 66.69% which indicate a lower heterogeneity. Interestingly, the likelihood-ratio test of $\tau^2 = 0$ is non-significant (total sample size: $P = 0.0869$; female sample size: $P = 0.0829$) that indicates the chance alone does not explain the heterogeneity among trials included and the total and female sample sizes may have accounted for it sufficiently. Another point from this analysis is the Adjusted R^2 for Model III (-9.93%) that includes male sample size as the explanatory variable. The negative value indicates that this model is not superior to a constant-only model with respect to the between-study variance (τ^2). When Model I (constant-only) is juxtaposed with Model III (constant + male sample size), one would notice the magnitude of τ^2 increases from 38.4 (Model I) to 42.2 (Model III) that results in -9.9% ($1 - \frac{42.2}{38.4}$) indicating including the male sample size increases between-study variation.

Discussion

In this study, utilizing contour-enhanced funnel and forest plots it was concluded that there exists at least one possible source of heterogeneity. Following the protocol, in multiple meta-regressions we found the importance of

sample size – as a possible source of heterogeneity – in determination of the effect size (ie, difference in means between the NFRthr in FM patients and healthy controls) found in the studies.

Different methodological procedures were utilized in the evaluation of the NFRthr in trials included in this review. These dissimilarities might be the potential source of the asymmetry in the contour-enhanced funnel plot. Judging from the plot,¹⁵ as a modern tool to pinpoint source of heterogeneity among studies, we found that there exists at least a possible source of heterogeneity that would increase heterogeneity among studies. Our results indicated that the total sample size and the sample size of female participants were responsible for about 67% of the variability of the effect sizes. However, determining the significance of the male sample size is difficult since the number of male participants is considerably lower compared to female counterparts. Lim et al⁷ discussed that null effects (ie, specifically in studies that included fibromyalgic patients) might be due to efficacious diffuse noxious inhibitory control (DNIC) that are exerted more in males than females.^{13,40} The DNIC refers to the inhibition of the NFR as a result of a heterotopic noxious stimulation that is delivered concurrent or before the NFR is tested.¹⁹ Lim et al⁷ noted that trials focusing on FM and reported null effects had more male participants compared to other research about headache, whiplash, or knee pain. This finding might be due to the larger magnitude of DNIC

Table 2 Study Characteristics for Electromyography (EMG), Latency, Sampling Rate, and NFR Definition

Study	Site of EMG	EMG Amplification	Filter	NFR Post-Stimulus Latency	Sampling Rate	NFR Definition
Banic (2004) ⁹	Middle of the BF and the rectus femoris muscles	Amplified but no details were reported	1.5–150 Hz by a single channel EMG – EEG amplifier	NR	NR	A reflex with an amplitude exceeding $20\mu V$ for at least 10 ms in the 70–150 ms post-stimulation interval was detected
Boureau (1991) ²⁰	Ipsilateral BF muscle using a pair of surface electrodes on the degreased skin overlying the muscle	NR	NR	90–180 ms	NR	When an amplitude greater than $50\mu V$ was observed and calculated based on RIII frequency of occurrence
Desmeules (2003) ²¹	Pair of surface electrodes placed over the tendon of the ipsilateral BF	NR	NR	NR	NR	Identified as a multiphasic signal appearing at least 90 ms but less than 250 ms after each stimulation. Present when the corrected computed surface was $> 0.5 \mu V/ms$
Guieu (1994) ²²	Opposite the distal third of the short part of the ipsilateral femoral biceps: calf, the insertion points of the anserine bursa site, knee	NR	NR	NR	NR	Definition of NFR was not reported. NFR threshold: Mean minimum stimulation intensity level required to trigger a reflex motor response in 100% of the trials
Bennett (2012) ²⁶	BF	NR		NR	NR	Definition of NFR was not reported. when the NFR interval mean minus baseline mean divided by the baseline SD exceeded 1.5 (NFRz was used)
Desmeules (2014) ³⁶	Ipsilateral BF	NR	NR	90–250 ms	NR	A multiphasic signal appearing and considered present when the corrected computed surface was $>0.5 \mu V/ms$
Tanwar (2019) ³⁷	BF	NR	NR	80–180 ms	NR	Definition of NFR was not reported. Average of the voltage which elicited the reflex during the increasing sequence and the current strength at which the reflex could no longer be elicited during the decreasing sequence
Umeda (2013) ³⁸	BF at half the distance between the ischial tuberosity and the lateral tibial condyle, with a reference electrode on the ipsilateral patella	$\times 1000$	13–1000 Hz	90–150 ms	2000 Hz	Definition of NFR was not reported. Stimulus intensity was increased in 4 mA increments until an NFR response was detected using an online script written with software.

(Continued)

Table 2 (Continued).

Study	Site of EMG	EMG Amplification	Filter	NFR Post-Stimulus Latency	Sampling Rate	NFR Definition
Rhudy (2013) ³⁹	BF EMG recorded from 2 active Ag-AgCl electrodes placed 10 cm superior to the popliteal fossa	x10,000	Bandpass filtered (10–300 Hz) online	90–150 ms	1000 Hz	EMG activity in the 90–150 ms post-stimulus interval that is greater than the 60ms pre-stimulus baseline interval by at least 1 SD

Abbreviations: ES, electrical stimulation; EMG, electromyography; NFR, nociceptive flexion reflex; NR, not reported; SN, sural nerve; BF, biceps femoris.

Table 3 Meta-Regression of the Effect Sizes on the Sample Sizes of Females and Males with Knapp-Hartung Modification (Total Number of Studies Included in Each Model: n = 9)

Effect Size (Difference in Mean)	Model			
	I	II	III	IV
Constant	-2.971 (p = 0.244)	1.105 (p = 0.678)	-0.257 (p = 0.949)	1.556 (p = 0.570)
SS _{female}		-0.057 (p = 0.047)		
SS _{male}			-0.327 (p = 0.404)	
SS _{Total}				-0.057 (p = 0.040)
REML estimate of between-study variance (τ^2)	38.4	14.54	42.22	13.37
Test for residual between-study variance (of $\tau^2=0$) $Q_{residual}$	98.35 (p < 0.0001)	21.3 (p = 0.0034)	85.72 (p < 0.0001)	21.02 (p = 0.0037)
Likelihood-ratio test of $\tau^2 = 0$ (prob.)	71.03 (p < 0.0001)	1.92 (p = 0.0829)	62.02 (p < 0.0001)	1.85 (p = 0.0869)
% residual variation due to heterogeneity ($I^2_{residual}$)	91.87%	67.13%	91.83%	66.69%
Proportion of between-study variance explained (Adj. R^2)	NA	62.15%	-9.93%	65.15%

Notes: Model I: Constant only; Model II: Constant + Female Sample Size; Model III: Constant + Male Sample Size; Model IV: Constant + Total Sample Size. Figures in parentheses are probabilities.

Abbreviation: SS, sample size.

that in males than females and, therefore, an unbalanced sample size between males and females may impose deviation from the actual effect size. On a side note, while running meta-regressions we speculated whether the ratios of male or female to total sample size could be a significant determinant of the effect size of which none was meaningfully significant.

In addition, the previous studies have shown that females have lower NFRthr than males.¹⁷ Therefore, an unequally distributed sample of males and females in trials, due to different neurophysiological characteristics, may not find generalizable findings applicable to both sexes. Our findings emphasize the importance of sample size in the studies reporting the NFRthr among FM patients. Further investigation, however, is necessary to

illustrate whether the sample size of females or males or the total sample size is important in regard to the magnitude of the effect size in trials comparing the NFRthr in the FM patients and healthy populations. Considering the abovementioned facts, one would speculate the low number of males in the included trials may attenuate the ability to find a reliable relationship between the male sample size and the effect size. In contrast, one may reasonably and logically conclude that the total sample size matters significantly in determination of the magnitude of the effect size. Another possible reason behind heterogeneity among studies is the sampling variation,⁴¹ eg, equal sample size distribution, age, and body mass index. For instance, from available literature in this context, we noticed that only trials conducted by Rhudy et al³⁹ (FM = 18; healthy

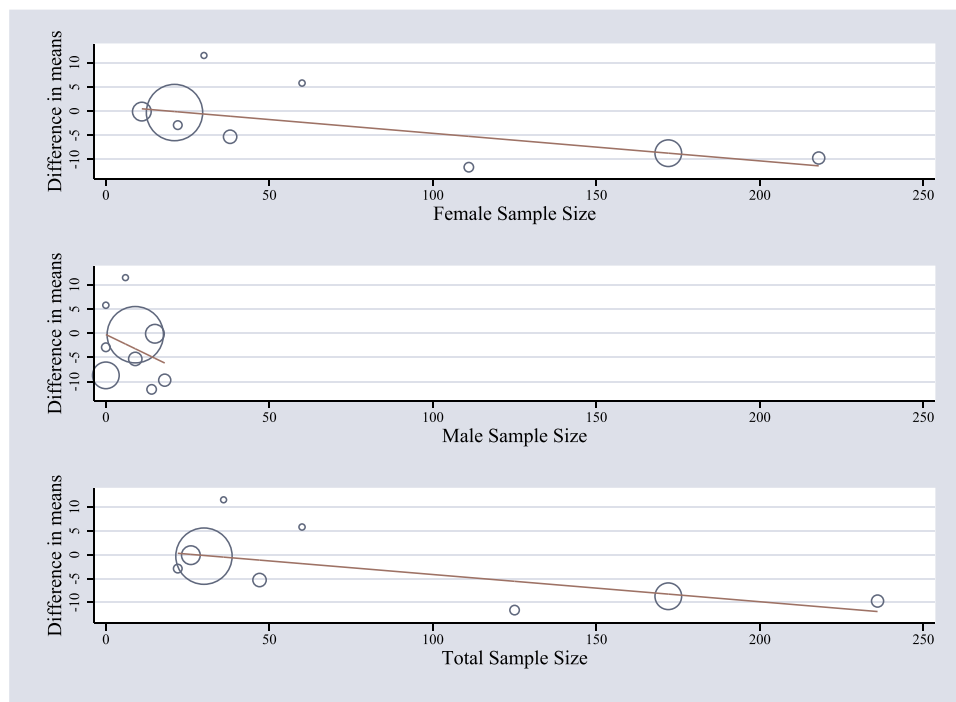


Figure 4 Meta-regression plots of effect sizes on total, male, and female sample sizes.

control = 19) and Bennett et al²⁶ (FM = 30; healthy control = 30) included equal sample sizes for both FM patients and healthy controls. However, except for the three studies that included only females^{26,37,38} the rest did not include equally distributed sample sizes of males and females. Thompson and Higgins⁴¹ emphasize the complexity of interpretation of the coefficients of covariates such as ratio of each sex or other characteristics included in the meta-regression as possible source of heterogeneity. It is a valid point since meta-regression analysis allocates more weight to larger studies and therefore their effect sizes are more powerful in pulling the overall influence of one characteristic on the meta-regression outcome.⁴² However, unlike larger studies, smaller studies that are conducted and analyzed with less methodological rigor⁴³ (and lower quality trials) tend to show larger effect sizes.⁴⁴ For this reason, they may overly influence the meta-regression results. Thus, the size of a trial and its methodological robustness would impact the magnitude and direction of the effects found. Second, it is wise to allow for the residual heterogeneity among intervention effects not modeled by the explanatory variables. This gives rise to the term “random-effects meta-regression” since the extra variability is incorporated in the same way as in a random-effects meta-analysis.⁴² Also, the meta-regression technique does not assess the within studies

variability but examines it across trials. Thus, for a within-study versatility, more research is needed. Nonetheless, we intended to investigate the across trials’ importance of sex-specific sample sizes. Considering the physiologic differences between sexes, the NFR is, in fact, different between females and males^{15,17,18} and logically a biased sample including more of one of the sexes may deviate the effect size. The regression coefficient obtained from a meta-regression analysis will describe how the outcome variable (the intervention effect) changes with a unit increase in the explanatory variable (the potential effect modifier). The statistical significance of the regression coefficient is a test of whether there is a linear relationship between intervention effect and the explanatory variable.^{31,41,42} Our findings are a statistical indication of such relationship and a possible source of heterogeneity. Note here that we do not claim a within-study comparison of effect sizes.

Based on the characteristics extracted from the trials, one might consider other factors that would possibly increase heterogeneity among the trials. We noted that the trials varied in the site of electrical stimulation, stimulus intensity ranges, increment in the electric current, EMG amplification and filters. Previous research by Rhudy and France¹⁵ reported that larger stimulus increments (4 mA) to assess peak 1 NFR tended to result in higher NFRthr estimates, whereas smaller stimulus

increments (2 mA) tended to result in lower NFRthr estimates. Considering that their findings were from healthy controls, one might question if such stimulation procedures might result in different NFRthr in fibromyalgia patients. In addition, Lim et al⁷ found that inter-pulse duration was a possible source of heterogeneity with their meta-regression for NFRthr in patients and healthy controls. Nonetheless, only two trials found higher NFRthr in FM patients compared to healthy controls. Rhudy et al³⁹ found that with or without inclusion of the five FM patients whose NFRthr was not achieved before meeting the 50 mA maximum stimulation intensity, the threshold was “surprisingly” still higher compared to healthy controls. In addition, Bennett et al²⁶ found increased but statistically non-significant NFRthr among all female patients with fibromyalgia in their study compared to healthy individuals. The information reported in this publication is limited and therefore discussing these results further is difficult. However, we noted that the average age for FM patients was higher than the median of the trials included.

We acknowledge that this research has several limitations. First, the trials assessing the NFRthr have investigated several other chronic pain conditions such as whiplash, headache, osteoarthritis, etc.⁷ Further investigations are needed to find out whether female sample size can universally explain the variation in the effect sizes of these chronic pain conditions. Second, aside from the inter-pulse range⁷ and female sample size there may be other sources of bias in the trials investigated that were out of the scope of this study. Other methodology characteristics such as the stimulation, current intensity, amplifier filters, etc. might account for sources of bias as we noticed a considerable extent of variability on parameters incorporated to evaluate the NFRthr in the included case-control trials. Finally, the limited number of the included studies might be a statistical obstacle to mark the total or female sample size as a truly robust source of heterogeneity. Further research may be required to investigate this concern.

Conclusion

Our study echoes two very important needs in future trials investigating NFRthr in FM patients: i) need for a standardized protocol using clinically available instruments to be used consistently to evaluate the NFRthr in patients suffering from FM; ii) need for a robust calculation for statistically powerful equally distributed sex-specific sample sizes for both cases and healthy control

groups. However, this recommendation may not be feasible in the “real world” as FM is more prevalent in females compared to males. Therefore, finding an equally distributed sample including balanced sample sizes for both males and females may impose tremendous difficulties on researchers. This way, at least, one possible source of heterogeneity that causes heterogeneity would be eliminated and the effect size might be more accurate.

Funding

Authors received no source of funding for this research.

Disclosure

The authors reported no conflicts of interest for this work.

References

1. Clauw DJ. Fibromyalgia and related conditions. In: *Proceedings of the Mayo Clinic Proceedings*. Vol. 90. Elsevier; 2015:680–692.
2. Rudin NJ. Fibromyalgia. In: *Pain*. Springer; 2019:693–697.
3. Wolfe F. Fibromyalgia. *Rheum Dis Clin North Am*. 1990;16(3):681–698.
4. Wolfe F, Ross K, Anderson J, Russell IJ, Hebert L. The prevalence and characteristics of fibromyalgia in the general population. *Arthritis Rheum*. 1995;38(1):19–28. doi:10.1002/art.1780380104
5. Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC. The biopsychosocial approach to chronic pain: scientific advances and future directions. *Psychol Bull*. 2007;133(4):581–624. doi:10.1037/0033-2909.133.4.581
6. Borsook D, Youssef AM, Simons L, Elman I, Eccleston C. When pain gets stuck: the evolution of pain chronification and treatment resistance. *Pain*. 2018;159(12):2421–2436.
7. Lim EC, Sterling M, Stone A, Vicenzino B. Central hyperexcitability as measured with nociceptive flexor reflex threshold in chronic musculoskeletal pain: a systematic review. *Pain*. 2011;152(8):1811–1820. doi:10.1016/j.pain.2011.03.033
8. Nicholas MK. Why do some people develop chronic, treatment-resistant pain and not others? *Pain*. 2018;159(12):2419–2420. doi:10.1097/j.pain.0000000000001404
9. Banic B, Petersen-Felix S, Andersen OK, et al. Evidence for spinal cord hypersensitivity in chronic pain after whiplash injury and in fibromyalgia. *Pain*. 2004;107(1–2):7–15. doi:10.1016/j.pain.2003.05.001
10. Skljarevski V, Ramadan NM. The nociceptive flexion reflex in humans – review article. *Pain*. 2002;96(1–2):3–8. doi:10.1016/S0304-3959(02)00018-0
11. Smith SM, Dworkin RH, Turk DC, et al. The potential role of sensory testing, skin biopsy, and functional brain imaging as biomarkers in chronic pain clinical trials: IMMPACT considerations. *J Pain*. 2017;18(7):757–777.
12. Leroux A, Belanger M, Boucher JP. Pain effect on monosynaptic and polysynaptic reflex inhibition. *Arch Phys Med Rehabil*. 1995;76(6):576–582. doi:10.1016/S0003-9993(95)80514-1
13. Peters ML, Schmidt AJ, Van den Hout MA, Koopmans R, Sluiter ME. Chronic back pain, acute postoperative pain and the activation of diffuse noxious inhibitory controls (DNIC). *Pain*. 1992;50(2):177–187. doi:10.1016/0304-3959(92)90159-9
14. Rhudy JL, France CR. Defining the nociceptive flexion reflex (NFR) threshold in human participants: a comparison of different scoring criteria. *Pain*. 2007;128(3):244–253. doi:10.1016/j.pain.2006.09.024

15. Rhudy JL, France CR. Reliability and validity of a brief method to assess nociceptive flexion reflex (NFR) threshold. *J Pain*. 2011;12(7):782–791. doi:10.1016/j.jpain.2011.01.002
16. Racine M, Tousignant-Laflamme Y, Kloda LA, Dion D, Dupuis G, Choinière M. A systematic literature review of 10 years of research on sex/gender and experimental pain perception—part 1: are there really differences between women and men? *Pain*. 2012;153(3):602–618. doi:10.1016/j.pain.2011.11.025
17. Mylius V, Kunz M, Schepelmann K, Lautenbacher S. Sex differences in nociceptive withdrawal reflex and pain perception. *Somatosens Mot Res*. 2005;22(3):207–211. doi:10.1080/08990220500262414
18. Popescu A, LeResche L, Truelove EL, Drangsholt MT. Gender differences in pain modulation by diffuse noxious inhibitory controls: a systematic review. *Pain*. 2010;150(2):309–318. doi:10.1016/j.pain.2010.05.013
19. Baad-Hansen L, Poulsen HF, Jensen HM, Svensson P. Lack of sex differences in modulation of experimental intraoral pain by diffuse noxious inhibitory controls (DNIC). *Pain*. 2005;116(3):359–365. doi:10.1016/j.pain.2005.05.006
20. Boureau F, Luu M, Doubreire JF. Study of experimental pain measures and nociceptive reflex in chronic pain patients and normal subjects. *Pain*. 1991;44(2):131–138. doi:10.1016/0304-3959(91)90126-1
21. Desmeules JA, Cedraschi C, Rapiti E, et al. Neurophysiologic evidence for a central sensitization in patients with fibromyalgia. *Arthritis Rheum*. 2003;48(5):1420–1429. doi:10.1002/art.10893
22. Guieu R, Serratrice G, Pouget J. Counter irritation test in primary fibromyalgia. *Clin Rheumatol*. 1994;13(4):605–610. doi:10.1007/BF02243002
23. Sideri S, Papageorgiou SN, Eliades T. Registration in the international prospective register of systematic reviews (PROSPERO) of systematic review protocols was associated with increased review quality. *J Clin Epidemiol*. 2018;100:103–110. doi:10.1016/j.jclinepi.2018.01.003
24. Moher D, Liberati A, Tetzlaff J, Altman DG, Prisma Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097
25. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. 2008.
26. Bennett RM, Jones KD, Hoffman J. A comparison of the nociceptive flexion reflex, pressure algometry and summated widespread pain in the diagnosis of fibromyalgia. *Arthritis Rheum*. 2012;10:S350.
27. Cochrane. RevMan calculator. Cochrane training. Vol. 2019. 2019.
28. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *J Clin Epidemiol*. 2008;61(10):991–996. doi:10.1016/j.jclinepi.2007.11.010
29. Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011;343(jul22 1):d4002. doi:10.1136/bmj.d4002
30. Sterne JA, Harbord RM. Funnel plots in meta-analysis. *Stata J*. 2004;4(2):127–141. doi:10.1177/1536867X0400400204
31. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. West Sussex, England: John Wiley & Sons Ltd.; 2009.
32. Luo D, Wan X, Liu J, Tong T. Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. *Stat Methods Med Res*. 2018;27(6):1785–1805. doi:10.1177/0962280216669183
33. Harbord RM, Higgins JPT. Meta-regression in stata. *Stata J*. 2008;8(4):493–519. doi:10.1177/1536867X0800800403
34. Sterne JA. *Meta-Analysis in Stata: An Updated Collection from the Stata Journal*. StataCorp LP; 2009.
35. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Stat Med*. 2003;22(17):2693–2710. doi:10.1002/sim.1482
36. Desmeules J, Chabert J, Rebsamen M, et al. Central pain sensitization, COMT Val158Met polymorphism, and emotional factors in fibromyalgia. *J Pain*. 2014;15(2):129–135. doi:10.1016/j.jpain.2013.10.004
37. Tanwar S, Mattoo B, Kumar U, Bhatia R. Can aberrant spinal nociception be a marker of chronicity of pain in fibromyalgia syndrome? *J Clin Neurosci*. 2019;65:17–22. doi:10.1016/j.jocn.2019.04.029
38. Umeda M, Corbin LW, Maluf KS. Preliminary investigation of absent nociceptive flexion reflex responses among more symptomatic women with fibromyalgia syndrome. *Rheumatol Int*. 2013;33(9):2365–2372. doi:10.1007/s00296-013-2725-0
39. Rhudy JL, DelVentura JL, Terry EL, et al. Emotional modulation of pain and spinal nociception in fibromyalgia. *Pain*. 2013;154(7):1045–1056. doi:10.1016/j.pain.2013.03.025
40. Arendt-Nielsen L, Sluka KA, Nie HL. Experimental muscle pain impairs descending inhibition. *Pain*. 2008;140(3):465–471. doi:10.1016/j.pain.2008.09.027
41. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med*. 2002;21(11):1559–1573. doi:10.1002/sim.1187
42. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med*. 1999;18(20):2693–2708. doi:10.1002/(SICI)1097-0258(19991030)18:20<2693::AID-SIM235>3.0.CO;2-V
43. Egger M, Juni P, Bartlett C, Hohenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical Study. *Health Technol Assess*. 2003;7(1):1–76. doi:10.3310/hta7010
44. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273(5):408–412. doi:10.1001/jama.1995.03520290060030