Taxometric analysis of biceps femoris EMG following electrocutaneous stimulation over the sural nerve: Determining the latent structure of the nociceptive flexion reflex (NFR)

Jamie L. Rhudy, Bradley A. Green, Randolph C. Arnau, Christopher R. France

Department of Psychology, University of Tulsa, USA

Department of Psychology, University of Southern Mississippi, USA

Department of Psychology, Ohio University, USA

ABSTRACT

The nociceptive flexion reflex (NFR) is a polysynaptic withdrawal reflex typically assessed from biceps femoris electromyogram (EMG) following noxious stimulation of the ipsilateral sural nerve. Electrophysiological evidence suggests the reflex is elicited following the activation of small diameter A-delta afferents. As a result, the NFR is assumed to be a categorically distinct construct that emerges from EMG activity only following nociceptor activation. Despite the widespread use of the NFR in pain research, there has been little attempt to verify the latent structure of the NFR. The present study used “coherent cut kinetics” taxometric analyses to examine whether the latent structure of biceps femoris EMG reflects the taxonic structure that would be predicted from electrophysiological evidence. To achieve this end, preliminary analyses first compared different methods of scoring NFR magnitude. Results suggested the presence of a taxon in the covariance of biceps femoris EMG and stimulus intensity that is likely to be the NFR. Furthermore, preliminary analyses suggested the best method of scoring NFR magnitude was using Cohen’s d. Implications of these results are discussed.

When small diameter afferents (A-delta fibers) of a limb are excited, a nociceptive impulse is relayed to the spinal cord that elicits a protective withdrawal reflex in the ipsilateral limb. This nociceptive flexion reflex (NFR) is polysynaptic and functions to minimize tissue damage. Although the NFR can be elicited in any limb, experimental paradigms typically record the NFR from the leg’s biceps femoris muscle electromyogram (EMG) following electrocutaneous stimulation of the ipsilateral sural nerve.

Research has shown that the stimulus intensity that reliably elicits the NFR (NFR threshold) significantly correlates with subjective pain threshold (Chan and Dallaire, 1989; Guieu et al., 1992; Willer, 1977; Willer et al., 1979) and the magnitude of the NFR correlates with pain intensity (Chan and Dallaire, 1989; Rhudy et al., 2005; Willer et al., 1979). Thus, the NFR can be used in two ways. NFR threshold is used as an objective measure of nociceptive threshold, and NFR magnitude is used to assess changes in nociceptive responding following repeated administrations of a constant-level suprathreshold stimulus. For these reasons, the NFR has emerged as an important tool for research on nociceptive responding in humans (Sandrini et al., 2005; Skljarevski and Ramadan, 2002).

The presence and size of the NFR is assessed directly from biceps femoris EMG (see Fig. 1). However, biceps femoris EMG following electrocutaneous stimulation is not solely influenced by A-delta fiber activation. Electrical stimulation activates non-nociceptive A-beta fibers that can result in early (≤70 ms post-stimulation) biceps femoris activity (RII reflex) (Sandrini et al., 2005; Skljarevski and Ramadan, 2002). Furthermore, late activity (>150 ms post-stimulation) can be associated with startle and/or voluntary movements (Dowman, 1992). To avoid contamination from these non-nociceptive responses, the NFR (RII reflex component) can be assessed from a 90–150 ms post-stimulation interval (Edwards et al., 2001; France et al., 2002a,b, 2005; Rhudy et al., 2005).

To accurately index the NFR, biceps femoris activity in the 90–150 ms post-stimulation interval should be relatively unaffected by stimulus intensities too low to activate nociceptive fibers. For stimulus intensities that activate nociceptive fibers, however, significant biceps femoris activity should emerge, and the activity level should correlate with stimulus intensity (and pain ratings) within individuals (Fig. 1). Thus, the latent structure of the NFR (as assessed from biceps femoris EMG) should be taxonic, i.e., the NFR should be a categorically distinct construct in biceps femoris EMG that occurs only at stimulus intensities high enough to activate nociceptors (and elicit a withdrawal reflex). Nonetheless, within the NFR taxon, biceps femoris EMG should be continuous and covary with the level of nociceptive input...
Although electrophysiological studies have led to this hypothesis (Chan and Dallaire, 1989; Guieu et al., 1992; Rhudy et al., 2005; Willer, 1977; Willer et al., 1979), the latent structure of NFR has not been tested empirically. Verifying the NFR’s taxonic nature is important for its validity as a measure of nociceptive responding.

Statistical methods developed by Paul Meehl et al. (Meehl and Yonce, 1994, 1996; Meehl, 1995) and recently extended by others (Ruscio and Ruscio, 2004) are believed to reveal naturally-occurring, non-arbitrary groups, or dimensions from manifest indicators of a construct. These taxometric procedures collectively referred to as the “coherent cut kinetics method” are well-suited to empirically test the latent structure of NFR (Waller and Meehl, 1998). Unlike the older “formal-numeric” taxometric methods (e.g., cluster analysis) using algorithms designed to maximize between-group differences and minimize within-group differences, coherent cut kinetics taxometric procedures do not arbitrarily impose any particular structure on the data (Gangestad and Snyder, 1985; Waller and Meehl, 1998).

The present study analyzed biceps femoris EMG data following electrodermal stimulation of the sural nerve in a procedure to assess pain tolerance in 50 healthy young adults. The primary goal was to determine whether empirical evidence could be obtained for the presence of a NFR taxon as would be expected if the reflex is a categorically distinct construct that emerges following nociceptor activation. To achieve this goal, we also conducted analyses to identify an optimum method of scoring NFR so as to maximize covariation with subjective pain ratings (i.e., a reflex magnitude scoring strategy that would explain the greatest variance in subjective ratings of stimulation intensity).

1. Methods

1.1. Participants

Participants were a random subsample of 50 healthy young adults (25 men and 25 women) who participated in a larger study examining...
opiate blockade on nociceptive responding (France et al., 2005). Participants in the larger study were tested on two separate days (i.e., consuming placebo pill or 50 mg of naltrexone). For the present study, only data collected on the placebo day were considered. One participant was excluded from analysis because biceps femoris EMG was too noisy to score (see Biceps Femoris Magnitude Scoring section). Participants provided informed consent after they were provided a full description of the study. Participants were provided an honorarium of $20 per hour of testing. All procedures were approved by the university’s IRB.

1.2. Electric stimulation and biceps femoris recording

Before testing, electrode sites were cleaned and abraded with Omni Prep electrode paste to achieve impedances of less than 10 kΩ. To record biceps femoris activity associated with the nociceptive flexion reflex (NFR), an electromyographic (EMG) electrode was secured over the muscle of the left leg, 10 cm superior to the popliteal fossa. A reference electrode was attached over the lateral epicondyle of the femur. EMG activity was amplified using a DelSys, Bagnoli-2 differential amplifier and the signal was recorded and processed using a CED Micro1401 analog-to-digital converter and Spike2 software. To elicit biceps femoris activity associated with the NFR, electrocutaneous stimulation was applied over the retromalleolar pathway of the sural nerve of the left leg using a Nicolet bar electrode (anode inferior) and a Digitimer D57A constant-current stimulator. Each stimulation trial consisted of a train of five 1 ms rectangular pulses with a 3 ms interpulse interval (total duration=17 ms). The NFR is typically defined by biceps femoris EMG activity in the 90 to 150 ms post-stimulation interval relative to a 60 ms pre-stimulation baseline (~65 to ~5 ms) interval (France and Suchowiecki, 2001; France et al., 2002a, b; Page and France, 1997). Use of the 90–150 ms interval avoids possible contamination by low-threshold cutaneous flexor reflex (RIF) which can precede 90 ms and by startle reactions and voluntary movements that can begin as early as 150 ms post-stimulation (Downman, 1991, 1992). EMG was sampled at 2000 Hz and recorded 400 ms prior to and 1600 ms after each stimulation.

1.3. Procedures

Participants were scheduled for two laboratory sessions, approximately 24 h apart (± 1 h). They were instructed to refrain from caffeine, nicotine, alcohol, and strenuous exercise for at least four hours and from analgesic medication for 24 h prior to testing. Participants were seated in a Hi-Seat rehabilitation chair (model 2000) with the left leg rest adjusted to maintain knee flexion at 60±5° from horizontal. Once the leg rest was adjusted, participants received a series of four electrocutaneous stimulations of increasing intensity (0, 2, 4, and 6 mA) to acclimatize them to the stimulation. NFR threshold was then determined three times, with a 5-minute rest period between assessments. Data collected during these NFR threshold assessments are not central to the current analyses and are not discussed further. The interested reader is referred to the following source for additional details (French et al., 2005).

Upon completion of threshold assessments, electrocutaneous pain tolerance level was measured. Specifically, sural nerve stimulation trials were delivered using a train of five 1 ms rectangular pulses with a 3 ms interpulse interval (total duration=17 ms). Following each train, participants rated the perceived intensity of the stimulation using a Verbal Rating Scale (VRS) with anchors of 1 (sensory threshold), 25 (uncomfortable), 50 (painful), 75 (very painful), and 100 (maximum tolerable). Stimulation intensity began at 0 mA and increased in 2 mA steps until a maximum stimulation intensity of 40 mA was reached or the participant reported that they had reached their tolerance threshold (100 on the VRS). EMG data recorded during this series of increasingly intense stimulations was used in the present study. This testing procedure generated a data file for every electrocutaneous stimulation trial. Each file included rectified biceps femoris EMG and a monitor signal for electrical stimuli. A total of 795 files were collected from the original 50 participants.

1.4. Biceps femoris magnitude scoring

In a prior report using the present dataset (Rhudy and France, 2007), different scoring methods were compared to identify an optimum empirical criterion to detect the presence/absence of the NFR (i.e., had the highest sensitivity and specificity for detecting trials that contained a reflex). The present report extends these analyses by addressing a related but independent question. Specifically, what is the optimum empirical criterion to measure NFR magnitude that is most related to pain ratings (i.e., explains the most variance in subjective ratings of stimulation intensity). A program to score biceps femoris activity was constructed by the first author using LabVIEW (National Instruments, Austin, TX). The program provided a graphic display of the rectified biceps femoris waveform and the electrocutaneous stimulation pulses. The program clearly identified for the operator the ~65 to ~5 ms pre-stimulation baseline interval as well as the 90–150 ms interval used to define the NFR. Two experts (one from Ohio University and one from The University of Tulsa) familiar with NFR methodology and recording, independently scored each waveform and determined whether there was noise in the pre-stimulation baseline that precluded accurate scoring of the biceps femoris EMG. A total of 40 waveforms (5.0%) were identified by at least one expert as having significant noise during the pre-stimulation baseline. These waveforms (which included all trials from one participant) were excluded from subsequent analyses. Analyses included a final set of 755 (95% of original) waveforms from 49 participants.

The program scored NFR magnitude from biceps femoris EMG in several ways (see Table 1). The NFR Interval peak was defined as the highest point (in µV) within the EMG curve between 90–150 ms post-stimulation, NFR interval mean was computed as the average level of EMG activity (in µV) within the same 90–150 ms window. NFR Interval AUC, or area under the curve, was computed as the sum of EMG activity (in µV) between 90–150 ms post-stimulation. Each of these criteria was calculated relative to EMG activity in the ~65 to ~5 ms pre-stimulation baseline interval: Baseline Adjusted NFR Interval Peak, Baseline Adjusted NFR Interval Mean, Baseline Adjusted NFR Interval AUC. Standardized versions of the peak and mean activity were also examined (by dividing by the variability at baseline), as was the NFR Interval Cohen's d (which standardizes based on variability from baseline and NFR interval).

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFR magnitude scoring methods</td>
</tr>
<tr>
<td>Scoring method</td>
</tr>
<tr>
<td>Baseline adjusted NFR interval peak</td>
</tr>
<tr>
<td>NFR interval peak</td>
</tr>
<tr>
<td>NFR interval peak z score</td>
</tr>
<tr>
<td>Baseline adjusted NFR interval mean</td>
</tr>
<tr>
<td>NFR interval mean</td>
</tr>
<tr>
<td>NFR interval mean z score</td>
</tr>
<tr>
<td>Baseline adjusted NFR interval AUC</td>
</tr>
<tr>
<td>NFR interval Cohen's d'</td>
</tr>
</tbody>
</table>

Note: NFR = nociceptive flexion reflex, AUC = area under the curve.

* Refers to standardized criteria. z scores are standardized because the standard deviation of baseline activity is used in the denominator, plus placing the variable in standard deviation units. The d' score is standardized because the standard deviation of baseline activity and NFR activity is used in the denominator.
1.5. Preliminary data analysis: comparison of NFR magnitude scoring methods

Multilevel linear modeling was used to determine which scoring method best predicted trial-by-trial pain ratings (Tabachnick and Fidell, 2007). Two-level models predicted VRS ratings from different NFR magnitude scores (separate analyses were conducted for each scoring method). First-level units were NFR magnitudes and second-level units were the 49 participants. Initially, an intercept only model (no NFR magnitude predictor) was conducted. Next, models were conducted with the NFR magnitude score entered as a predictor. The null model likelihood ratio test ($\chi^2$) was used to determine whether the addition of the predictor suggested a better fitting model than the intercept only model (p < .05 suggests a better fitting model). Models with different predictors (non-nested models) were compared using AIC (smaller is better criterion) to determine which scoring method performed best (Hox, 2002). The eta-squared ($\eta^2$) effect size was calculated by subtracting the residual variance in the predictor model from the residual variance in the intercept only model and dividing by the residual variance in the intercept only model. This was conducted to determine the proportion of variance in pain ratings explained by NFR magnitude.

1.6. Data analysis: taxometric analyses

All taxometric analyses were conducted using a set of computer programs written by John Ruscio that have been shown to produce results that are valid and reliable (Ruscio et al., 2006, 2007). An added benefit is that the Ruscio programs have data simulation programs built in to most of the taxometric methods. The programs developed by Ruscio take the distributional properties of data being analyzed and create simulated taxonic and dimensional datasets with the same distributional properties for comparison analyses. The simulated datasets are run through the taxonic programs along with the research data, so the output results from the taxonic and dimensional simulations can be compared to each other, as well as to the output results from the research data. The simulations are compared to each other to determine whether the data are appropriate for taxometric analysis. If they are, then the simulated dimensional and taxonic results should be clearly distinguishable. The results of the simulations are compared to the results of the research data to aid in interpretation by determining whether the research data results fit better with a simulated taxonic or dimensional result.

There were only two indicators of interest for the present study: a) amount of electrical stimulation of the sural nerve, measured in milliamperes (Stimulation Intensity); and b) biceps femoris EMG activity (NFR Magnitude). Having only two indicators limits the types of taxometric procedures to MAMBAC and MAXSLOPE, because MAXCOV, MAXEIG and L-Mode require at least 3 indicators.

Taxometric analyses require relatively large datasets to produce reliable results. A minimum of 300 subjects is recommended for data with ideal distribution and validity characteristics (i.e., base rate near .500, little within-group indicator correlation, indicators that separate groups by 2.00 SDs). However, 600 or more subjects has been the typical recommendation, the norm in the literature, and the minimum demonstrated to produce interpretable results with data distributions that are not ideal (Meehl and Yonce, 1994; Ruscio et al., 2006). The present study included only 49 participants, but each participant provided multiple data points that corresponded to each level of electrical stimulation. The number of trials varied across participants as a function of pain tolerance. The potential taxon of interest was conjectured to be the NFR; therefore, stimulation intensity (in mA) and biceps femoris EMG activity in the 90–150 ms interval (NFR magnitude) constitute the data distributions of interest. A given participant could provide as many as 21 data points (0 to 40 mA in 2 mA increments) with biceps femoris EMG for each trial. The resulting dataset comprised 755 data points, more than enough for taxometric analysis. The NFR magnitude score to be used to calculate biceps femoris EMG activity was determined by preliminary multi-level modeling analyses.

1.6.1. MAMBAC

MAMBAC is an acronym for Mean Above Minus mean Below A sliding Cut. Briefly, when data are gathered on any two independent indicators of a construct a MAMBAC analysis can be conducted if at least one of the indicators is continuously distributed. The indicator variables are related as an input, continuous variable ($x$) and a corresponding output variable ($y$). If both variables are continuously distributed the analysis can be bi-directional, with each indicator serving as an input variable and an output variable (Meehl and Yonce, 1994). The data are divided by a sliding cut made at successive intervals along the input variable, which corresponds to the x-axis. The mean value for the output variable, which corresponds to the y-axis, is calculated for the data points below the x-cut and this value is subtracted from the mean value of the output variable for the data points above the x-cut: Difference ($d_i$) = Mean of $y$ Above Cut – Mean of $y$ Below Cut.

This difference-of-means operation is conducted at each cut point and the resulting values ($d_i$) are plotted. If the construct being investigated is dimensional, the greatest difference between the means will be at the cuts made at the extreme high and low values of the input variable. This will generate a MAMBAC plot with a dished, or concave, shape. If the construct being investigated is taxonic (categorical) then the greatest difference between the means will occur at the point of the cut on the input variable that most accurately separates the mixed sample into the constituents of taxon members and non-taxon (or complement) members. The taxonic condition will generate a convex MAMBAC plot with a peak near the cut point that best separates the two classes.

1.6.2. MAXSLOPE

MAXSLOPE analysis (Grove and Meehl, 1993; Ruscio et al., 2006) requires two continuously distributed indicators. A scatterplot is created by assigning one indicator to the x-axis, and the other to the y-axis of a graph. After the data points are plotted a smoothed local regression curved is fitted to the data. If the construct being examined is continuous, the data points should form a single cloud, and the resulting regression line should be relatively straight, and similar to a linear regression line produced by two positively correlated variables. If the construct is taxonic (categorical), the data points should create two data clouds. The first cloud should be roughly located in the lower left area of the graph, being produced by subjects (or in the present

<table>
<thead>
<tr>
<th>Predictor</th>
<th>z Log likelihood</th>
<th>AIC model</th>
<th>Null model likelihood ratio test outcome</th>
<th>Residual variance</th>
<th>$\eta^2$ squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (intercept only)</td>
<td>7316.00</td>
<td>917.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline adjusted NFR interval peak</td>
<td>6793.48</td>
<td>6805.48</td>
<td>p &lt; .05</td>
<td>363.09</td>
<td>.60</td>
</tr>
<tr>
<td>Baseline adjusted NFR interval peak</td>
<td>7093.42</td>
<td>7101.42</td>
<td>p &lt; .05</td>
<td>645.86</td>
<td>.30</td>
</tr>
<tr>
<td>Baseline adjusted NFR interval mean</td>
<td>6704.50</td>
<td>6716.50</td>
<td>p &lt; .05</td>
<td>314.14</td>
<td>.66</td>
</tr>
<tr>
<td>Baseline adjusted NFR z score</td>
<td>6726.26</td>
<td>6738.26</td>
<td>p &lt; .05</td>
<td>326.81</td>
<td>.64</td>
</tr>
<tr>
<td>Baseline adjusted NFR interval AUC</td>
<td>6705.19</td>
<td>6717.19</td>
<td>p &lt; .05</td>
<td>314.54</td>
<td>.66</td>
</tr>
<tr>
<td>Cohen’s d</td>
<td>6542.97</td>
<td>6554.97</td>
<td>p &lt; .05</td>
<td>279.53</td>
<td>.70</td>
</tr>
</tbody>
</table>
study, trials) with low values on both indicators (complement group).

The second cloud should be located roughly in the upper right area of the graph, being produced by subjects/trials with high values on both variables. It is not uncommon for the taxon and complement data clouds to overlap such that upon visual inspection the scatterplot would appear to comprise only one group of data points related by a dimensional construct. In other words, taxonic grouping may not be visually obvious. If the data are taxonic, a local regression line for the data should be relatively flat toward both extremes, where the data points are more purely taxon or purely complement, because the indicators are all largely absent in the complement and largely present in the taxon. That is, they do not correlate within groups, or correlate less than in the mixed sample of taxon and complement members. The regression line should slope upward in the middle region where the two groups are more mixed. The local regression line for a taxonic construct more or less approximates an ogive, or stretched "S" shape, depending on the validity of the indicators and the nuisance covariance (or correlation) between the indicators within the groups.

1.6.3. Multiple hurdles consistency testing

Coherent cut taxometric procedures employ no significance tests to assess the likelihood that a given finding is spurious. Instead, multiple-hurdles consistency testing is used to allow any apparent result the opportunity to replicate, or fail to do so (Meheül, 1995; Ruscio et al., 2006; Waller and Meehl, 1998). The most common form of consistency testing is the use of multiple taxometric methods. We used the MAMBAC procedure to determine whether there was a discontinuity in the relationship between the indicators, marking a region where Stimulation Intensity and NFR Magnitude are more strongly correlated than at either extreme. Again, such a finding would suggest that the relationship between Stimulation Intensity and NFR Magnitude resembles a step function, with low to moderate correlation at lower level, low to moderate correlation at high levels, and relatively greater correlation across the entire range of the indicators. We used the MAXSLOPE procedure as a consistency test to determine whether our findings, be they taxonic or dimensional, were replicable via multiple taxometric procedures, and therefore likely to be legitimate. MAXSLOPE determines whether there was a discontinuity in the relationship between the indicators, marking a region where Stimulation Intensity and NFR Magnitude are more strongly correlated than at either extreme. Again, such a finding would suggest that the relationship between Stimulation Intensity and NFR Magnitude resembles a step function, with low to moderate correlation at lower level, low to moderate correlation at high levels, and a narrow range in between where both indicators increase abruptly.

2. Results

2.1. Comparison of NFR magnitude scoring methods

Results from multilevel models are reported in Table 2. The addition of each NFR magnitude predictor improved model fit over the intercept only model (as indicated by the null model likelihood ratio test significance). Moreover, effect sizes (η²) suggested most scoring methods (except NFR Interval Peak z score) explained 60% or greater of the variance in stimulus intensity ratings. Examination of AIC suggested NFR Cohen’s d was the best method of scoring NFR magnitude and correlated very highly with intensity ratings (explained 70% of the variance). In addition, this scoring method resulted in data that were more normally distributed than other scoring methods, eliminating the need to transform. Thus, NFR Interval Cohen’s d was used to score NFR Magnitude for taxometric analyses.

2.2. Taxometric analyses

2.2.1. MAMBAC

The parameters for the analysis were to make 30 evenly spaced cuts along the input indicator, starting after the 25th case from the beginning and ending before the 25th case from the end, producing 30 d₀ points on each output MAMBAC graph. When cases in a data set are sorted along an indicator, several cases may have the same value for the sorting indicator, and therefore their order in the sorted distribution will be arbitrary. As a result, the cases with the same value, that fall above or below a cut in the distribution are also arbitrary, adding some chance to the mean values calculated above and below the cut. Arbitrary ordering is especially problematic when there is a limited set of values a case may take for the sorting indicator, as was the case with the Stimulation Intensity mA indicator. Reliability

Table 3

<table>
<thead>
<tr>
<th>Indicator</th>
<th>MAMBAC</th>
<th>MAXSLOPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full data set/minus outliers</td>
<td>Full data set/minus outliers</td>
</tr>
</tbody>
</table>

Note: Results are presented from primary analyses on all of the data (Full Data Set), as well as the post-hoc analysis that eliminated persons who reflexed below 10 mA (Minus Outliers).

<table>
<thead>
<tr>
<th>Indicator</th>
<th>MAMBAC</th>
<th>MAXSLOPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Stimulation intensity 3.06/2.67 3.18/3.06
NFR interval Cohen’s d 1.76/2.82 1.53/1.87

<table>
<thead>
<tr>
<th>Indicator covariance in terms of correlation across entire sample, as well as within taxon and complement groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full sample</td>
</tr>
<tr>
<td>Taxon</td>
</tr>
<tr>
<td>Complement</td>
</tr>
</tbody>
</table>

Fig. 2. The pair of graphs on the left side of the page are the individual MAMBAC graphs created by entering Stimulus Intensity mA as the input, with NFR Interval Cohen’s d as the output (first graph), then entering NFR Interval Cohen’s d as the input with Stimulus Intensity mA as the output. (second graph). The graph on the right side of the page was created by averaging the two individual MAMBAC graphs on the left side of the page.
and interpretability of the output graphs were enhanced by conducting 25 replications of the analysis, resorting cases with identical input values before cutting into successive intervals for each replication, then averaging the output values for the graph. Multiple replications of this sort reduce the potentially misleading effects of arbitrarily assigning cases with the same input value to the cut intervals (Ruscio et al., 2006). MAMBAC validity and indicator correlations data are summarized in Table 3. Validity of indicators for taxometric analysis is reported in standard deviation units (Cohen's $d$) separating the mean of the taxon group from the mean of the complement group, if in fact there are groups (if the latent structure is dimensional then there is only one group). Separation of at least 1.25 SD is desirable, and separation of 2.00 SD or more is optimal. Validity for the Stimulation Intensity mA indicator was excellent at $d=3.06$. The NFR Interval Cohen's $d$ (NFR magnitude) indicator had good validity at $d=1.70$. The correlation coefficients for MAMBAC demonstrate moderate nuisance covariance in both the taxon (.33) and complement (.47) groups. However, the MAMBAC procedure is adequately robust to moderate nuisance covariance. The correlation of the combined sample (.70) is much greater than that of either the taxon or complement, as should be the case when the latent structure is taxonic.

MAMBAC analysis produced a clearly taxonic result. The individual graphs and averaged graph are clearly taxonic (Fig. 2). Simulated taxonic and dimensional datasets (10 of each) were created, with distribution and validity characteristics similar to the research data. The simulated datasets were run to create simulated output curves (taxonic and dimensional), which are combined to estimate a range within which each point of a curve would be likely to fall, depending on whether it is taxonic or dimensional. As can be seen in Fig. 3, the averaged MAMBAC output from the study data (leftmost graph) is congruent with the simulated taxonic output (middle graph), and incongruent with the simulated dimensional output (rightmost graph). Fig. 4 takes the MAMBAC graphs produced by the study data and superimposes them over the areas covered by the taxonic and dimensional simulations, demonstrating that the graph produced by the study data fits a taxonic simulation, and not a dimensional simulation. The Ruscio program for MAMBAC analysis calculates a comparison curve fit index (CCFI), which provides numerical estimate, ranging from .00 to 1.00, of relative fit of the study data to either the simulated taxonic or simulated dimensional curves. CCFI values close to .00 support a non-taxonic structure. Values close to 1.00 support a taxonic structure. Values near .50 are interpreted as ambiguous (Ruscio et al., 2006). The data from the present MAMBAC analysis produced a CCFI of .72, supporting a taxonic structure.

The analysis produced an averaged base rate estimate of $\text{.479 (SD=.018)}$, estimating that the complement consisted of 362 data points, and the taxon consisted of 393 data points. To convert the base rate estimate into the practically meaningful metrics of the indicators we use the base rate, or percentage for the complement group (362 is 47.9% of 755) and find the next highest percentage value in the distribution of that indicator in the data set. For Stimulus Intensity in mA the next value in the distribution above 47.9% is 14 mA at 50.9%, suggesting that most subjects have begun to demonstrate a reflex at 14 mA. The NFR magnitude value just above 47.9% in the distribution was NFR Interval Cohen's $d=.453$ at 48.1%, suggesting that any reflex producing a NFR Interval Cohen's $d$ value of .453 or greater is likely to be a genuine reflex response.

2.2.2. MAXSLOPE

The MAXSLOPE results also suggested a taxonic structure upon visual inspection, although the graphs are not as clear as the MAMBAC graphs (Fig. 5). When the input indicator was Stimulation Intensity the
3. Discussion

The results of MAMBAC and MAXSLOPE analyses suggest a taxonic structure for biceps femoris EMG activity in response to increasing electrical stimulation of the sural nerve. The base rate estimates from the two procedures are not greatly divergent, further supporting a taxonic structure for biceps femoris EMG activity, as the base rate estimates of a class or type (taxon) should be similar when calculated by different valid methods (Waller and Ross, 1997). The relatively high nuisance covariance estimated for the complement group by both procedures is congruent with a construct in which indicators of the construct display a moderate linear relationship across a lower portion of the domain of either indicator. The change to a lower level of nuisance covariance in the taxon group demonstrates a discernibly different relationship between those indicators across the higher portion of the domain of either indicator, in which higher stimulation elicits proportionally less NFR response than was evident in the complement group (i.e., NFR approaches a plateau).

Together, these results provide statistical evidence for the presence of a categorically distinct response (taxon) in the covariance of biceps femoris EMG and stimulus intensity. Although it is impossible to determine from the present analyses whether the taxon corresponds to the NFR, this is the most plausible explanation. Indeed, evidence suggested that the taxon existed above stimulation intensities of 14 mA. This is consistent with prior research that suggests the NFR (when elicited using similar equipment and procedures) does occur near this intensity (France et al., 2005). Furthermore, our use of the 90–150 ms post-stimulus interval reduces the probability that non-nociceptive confounds contaminated the EMG signal, such as A-beta fiber activation (RII reflex), startle effects, or voluntary movements (Downman, 1992; Sandrini et al., 2005).

These findings provide additional validity for the use of the NFR as a measure of nociceptive responding. The covariation between stimulus intensity and biceps femoris EMG suggested the presence of a taxon with a cut point at an intensity near NFR threshold. This is consistent with electrophysiological evidence suggesting the NFR emerges only from stimulus intensities that activate A-delta fibers, a categorically distinct class of neurons (Hugon, 1973; Wiesenfeld-Hallin et al., 1984; Willer, 1977; Willer et al., 1978). Thus, the latent structure of the NFR matches its neurophysiological underpinnings.

Our observations that NFR magnitude (NFR Interval Cohen's d) covaried with stimulus intensity in both the taxon and complement groups is worthy of comment. This is inconsistent with our prediction that biceps femoris EMG would only covary with stimulus intensity after the reflex was observed (in the taxon group). However, there is reason to believe the covariance within the complement group may be an artifact, at least in part, of the combination of within-subject and between-subject nature of our dataset.

Our data was made up of 755 trials from 49 different individuals. There is considerable between-subject variability in the stimulus graph approximated the ogive shape characteristic of a taxon, though the line is not as flat as one would desire in the lower (complement) range, tending to slope gradually up until flattening in the upper (taxon) range. When the input was NFR Interval Cohen's d, the plot slopes consistently up through the lower range, much like the line of a linear regression of two related variables, but then plateaus, flattening out for the remainder of the range of the input indicator. This formation is congruent with a taxonic structure with high nuisance covariance (within-group indicator correlation) in the complement group (Ruscio et al., 2006), which is in fact what the numeric output indicates. Validity for the indicators in MAXSLOPE was again excellent for Stimulus Intensity mA (d = 3.18) and good for NFR Interval Cohen's d (d = 1.53). However, nuisance covariance, though improved in the taxon group (r = .16), was much higher in the complement group (r = .60), which would account for failure to produce a definitive taxonic MAXSLOPE plot in the complement range. MAXSLOPE indicator validity and nuisance covariance are summarized in Table 3. The first plotted curve, using Stimulation Intensity as the input, did not produce a deflection point adequate to estimate a base rate. The second plotted curve produced a base rate estimate of .540. The MAXSLOPE program does not generate and analyze simulations of taxonic and dimensional data. The base rate estimate from MAMBAC is the better estimate, as the MAMBAC analysis was able to produce valid estimates using either indicator as the input, and as the MAMBAC results appeared to be less affected by nuisance covariance in the complement group.

The high nuisance covariance in the complement was conjectured to be due to outlier cases of participants who began responding to electrical stimulation at unusually low levels (10 mA or less). All participants for whom a reflex began at 10 mA or less were removed from the data set and the procedures were repeated. As would be expected, nuisance covariance was reduced for MAMBAC (complement = .32; taxon = .20), but in MAXSLOPE nuisance covariance became more evenly balanced across the groups, decreasing in the complement and increasing in the taxon (complement = .47; taxon = .38) while the correlation across the entire sample increased to .74. As can be seen in Table 3, indicator validity remained excellent for Stimulus Intensity and improved for NFR interval Cohen's d. These results confirm that nuisance covariance in the complement group was largely due to some participants responding at unusually low levels of stimulation.

Fig. 5. MAXSLOPE graphs, first with Stimulus Intensity mA as the input and NFR Interval Cohen's d as the output (left side), then with NFR Interval Cohen's d as the input and Stimulus Intensity mA as the output (right side).
intensity that elicits a NFR (Sandrini et al., 2005; Skljarevski and Ramadan, 2002), which is one of the reasons that the NFR is of interest to pain researchers. Ostensibly, these individual differences in NFR threshold are due to endogenous factors (e.g., descending inhibition) that alter the threshold of spinal nociceptive processes. The impact of these individual differences in NFR threshold on our results is that persons who reflex early (e.g., 8 mA) are combined with those who reflexed later (e.g., 22 mA). Thus, the break in the covariance between biceps femoris EMG and stimulus intensity that separates the complement group (pre-reflex EMG) from the taxon group (NFR) should be different for different individuals (as a function of their NFR threshold). When the data are combined from all participants and analyzed for the latent structure, the complement group derived from the total sample will have data (trials) from persons whose reflex threshold was lower than the cut point in the total sample (<14 mA). Likewise, the taxon group derived from the total sample will have data (trials) from persons whose reflex threshold was higher than the cut point in the total sample (>14 mA). As a result, this could inflate the covariance in the complement group and reduce the covariance in the taxon group. In addition, problematic nuisance covariance in either group can influence base rate estimation. In MAMBAC, excess nuisance covariance in the complement can inflate the estimated size of the complement, therefore reducing the estimated base rate of the taxon. Although the effects of nuisance covariance on MAXSLOPE have not been fully documented, a similar bias in base rate estimate is likely to occur. As the nuisance covariance pulls lower members of the taxon group into the complement group, leaving the highest scoring taxon members, the covariance within the taxon group should decrease. In fact, MAXSLOPE results from the full sample demonstrated problematic nuisance covariance in the complement (r = .60) with very low covariance in the taxon (r = .16).

To determine whether these issues could have influenced our results, we eliminated data from persons with a low NFR threshold (<10 mA) and re-analyzed the data. These post-hoc analyses lent partial support for this hypothesis. Specifically, the covariance in the complement group decreased in both the MAMBAC and MAXSLOPE analyses after these data were omitted (Table 3). Interestingly, the post-hoc MAXSLOPE analysis suggested the covariance in the taxon group increased as though persons with a lower threshold contaminated the taxon group (perhaps due to an asymptote of reflex activity for persons who reflex at low intensities). However, the covariance in the taxon group was not increased in the post-hoc MAMBAC analysis. This discrepancy between analyses may have been due to differences in base rate estimates for the taxon group. The nuisance covariance was more evenly balanced in the taxon and complement groups in the MAMBAC results from the full sample. MAMBAC nuisance covariance estimates from the reduced sample were not expected to change in a similar fashion to MAXSLOPE, and in fact, covariance in the taxon decreased, as the base rate estimate of the taxon decreased. Again, MAMBAC and MAXSLOPE are mathematically independent, and are expected to produce similar, not identical results.

While these post-hoc analyses suggest the covariance in the complement group may be an artifact of our dataset, it does not rule out the possibility that that biceps femoris EMG does show some covariation with stimulus intensity in the complement group (pre-reflex EMG). One way to address this issue in a future study would be to collect 600+ trials from each participant and analyze the latent structure of each person's data separately. That would eliminate any artifacts due to analyzing trials from several individuals. While a study of this nature is possible, there are limitations inherent to this approach as well. Exposure to so many stimuli could activate endogenous mechanisms (e.g., habituation, sensitization) that can alter afferent neural transmission, spinal nociceptive processes, and/or the efferent motor response. Therefore, the covariance between stimulus intensity and biceps femoris EMG could change as a result of the extended testing session. Such changes in covariance could impact the results of taxometric analyses. Nonetheless, it may be important to consider this approach to assessing the latent structure of the NFR.

As we demonstrated in the current study, NFR magnitude covaries with pain rating. For this reason, experimental paradigms that deliver a constant-level suprathreshold stimulus are used to study within-subject changes in NFR magnitude. This procedure assumes that changes in the magnitude of biceps femoris EMG activity reflect differences in noceptive responding. In light of the present findings, until the reason for the covariance in the complement group can be determined it is recommended that researchers interested in measuring within-subject changes in spinal noception begin by ensuring that they accurately identify NFR threshold. Then, a constant intensity stimulus must be chosen to reliably elicit an NFR, but also to minimize floor and ceiling effects in NFR magnitude measurement (e.g., 120% of NFR threshold). In sum, because biceps femoris EMG also covaries with stimulus intensity in the complement group, studies that measure changes in NFR magnitude should use a threshold detection criterion that optimizes the specificity of NFR detection to ensure that changes in biceps femoris EMG are related to the NFR. To aid researchers in this endeavor, Rhudy and France (2007) compared different NFR threshold scoring criteria using ROC analyses to derived the best scoring methods and cutoffs for NFR detection. Graphs depicting the relationship between cutoffs, sensitivity, and specificity are presented to facilitate cutoff decision-making. Cutoffs should be chosen that maximize specificity, such that observed changes in NFR magnitude are related to noiception, but not so high as to impose a ceiling on measurement. Once the NFR is correctly detected (NFR threshold is determined), the next step is to determine the suprathreshold stimulus level. Several studies have used a stimulus set at 120% NFR threshold (Danziger et al., 1998; France and Suchowiecki, 2001; Rhudy et al., 2005; Skljarevski and Ramadan, 2002); but, it may be important for future studies to compare different suprathreshold stimulus criteria.

3.1. NFR magnitude scoring

To our knowledge, this is the first study to compare methods of scoring NFR magnitude. To achieve this end, we decided a priori to base our decision on the method that best predicted subjective ratings of the stimulus (pain ratings). Multilevel analyses suggested the model with NFR Interval Cohen’s d had the best fit and explained 70% of the variance in ratings. Because NFR Interval Cohen’s d incorporates baseline variability and NFR Interval variability into its formula, it also had the best distributional properties (most normal distribution) which facilitate the use of parametric statistics. As a result, we concluded that NFR Interval Cohen’s d was the best method of calculating NFR magnitude of those that were tested.

Multilevel modeling was chosen to analyze the relationship between NFR magnitude and pain ratings because it provides a means to control for the non-independence of observations assessed within-subject. In the present study this analytic procedure was used for the sole purpose of determining the best scoring method for NFR magnitude to use in taxometric analyses. However, the use of multilevel analysis to model the relationship between NFR Interval Cohen’s d and pain ratings has advantages that could be used in future studies. For example, researchers could use this procedure to study factors that could alter the relationship between subjective pain ratings and the objectively-determined NFR. Indeed, between-subjects manipulations (psychological, pharmacological, or surgical) could be modeled to determine whether those manipulations impact the degree of NFR–pain covariance. While NFR magnitude and pain covary under most circumstances, there are pharmacological, psychological, and physical conditions under which the two can be dissociated (Andersen et al., 1995; Arendt-Nielsen et al., 1995; Danziger et al., 1998; Petersen-Felix et al., 1995; Rhudy et al., 2006). The multilevel analysis using NFR Interval Cohen’s d could be used to...
model this divergence, potentially providing important insights into chronic conditions in which a dissociation between the NFR and pain perception has been noted (e.g., chronic pain, patello-femoral dysfunction) (Skljarevski and Ramadan, 2002). Moreover, individual differences in the NFR–pain relationship could be modeled by allowing the slope of the relationship to vary between-subjects. Differences in the slope could be predicted from individual differences variables (e.g., catastrophizing, sex, affectivity, hormone levels). Thus, we believe that modeling the relationship between NFR interval Cohen’s d and pain ratings may provide a valuable tool for future research.

4. Summary

In sum, the present study provided empirical evidence that the NFR is a taxon, validating previous notions about its latent structure. Additionally, methods of scoring NFR magnitude were compared and it was determined that using the Cohen’s d scoring method covaried most with pain ratings. We believe these findings have important implications for researchers interested in using the NFR as a research tool for studying nociceptive processing.

References


