## Supplementary File 6 – Measurement properties of commonly used outcome measures in FND treatment studies

| Outcome domain | Outcome measure | Measurement properties | Data extracted |
|---|---|---|---|
| **Symptom change** | **Clinical Global Impression – Improvement scale (CGI-I) (patient-rated)** | **Validity** | ***Demartini et al. 2014***<br>Convergent validity: change in HoNOS scores significantly predicted CGI-I (patient) ratings of general health (Wald (1)=4.52, p=0.033) |
| | | **Reliability** | No data available |
| | | **Responsiveness** | ***Demartini et al. 2014***<br>• Post-treatment:<br>  o 75% patients rated their presenting symptom as 'better' or 'much better'<br>  o 66.2% rated general health 'better' or 'much better'<br>• 12m: 66.6% (general health) and 63.9% (presenting symptom) were 'better' or 'much better'<br><br>***Demartini et al. 2019b***<br>• Improvement reported by 66.7% of sample at 12wk and 77.8% at 24wk<br>• No patient worse at either time point<br><br>***Dreissen et al. 2019***<br>• CGI-I scores improved ('minimally improved, 'much improved' or 'very much improved') in 48% of the BoNT group, vs 52.2% in placebo group<br>• Difference not significant (OR = -0.042, CI= −0.300, 0.225, p=1.00)<br><br>***Garcin et al. 2013***<br>• Follow-up (average 20m): 71% reported improvement (CGI-I rating of 'minimally improved', 'much improved' or 'very much improved')<br>• Median CGI-I score=2 ('much improved') (range 1–6)<br><br>***Garcin et al. 2017****:*<br>• After 2 TMS sessions, 60% of patients reported being 'much improved' or 'very much improved' (CGI-I scores of 1 or 2)<br>• 1y: 56% of patients still 'much' or 'very much improved' |

| | | |
|---|---|---|
| | | ***Nielsen et al. 2015***<br>• 96% of patients reported improvement at end of treatment (64% 'much improved' or 'very much improved')<br>• 85% reported improvement at 3m (55% 'much improved' or 'very much improved')<br>• None of the patients were worse at the end of treatment<br><br>***Nielsen et al. 2017b***<br>• 6m: 72% of physiotherapy arm reported a 'good' outcome (CGI-I ratings of 'much improved' or 'improved'), vs 18% controls<br>• 3% physiotherapy arm reported worsening of symptoms from start therapy to 6m, vs 32% controls<br><br>***Sharpe et al. 2011 (CGI-I global / presenting symptom):***<br>• Global:<br>  • Post-treatment (3m) (GSH vs TAU) (OR=2.4, CI=1.2,4.7, p=0.016)<br>  • Participants in GSH arm reported greater improvement (30% GSH arm 'better' or 'much better' vs 17% for TAU)<br>  • Number needed to treat=8<br>  • Not significant at 6m, GSH=38%, TAU=27% (OR=1.5, CI=0.7,2.8, p=0.27)<br>• Presenting symptom:<br>  • 3m: GSH arm reported greater improvement (OR=2.33, CI=1.19, 4.56, p=0.014); 38% GSH arm 'better' or 'much better' vs 29% for TAU (3m)<br>  • Remained significant at 6m (OR=3.2, CI=1.3, 4.5, p=0.014); 47% GSH arm 'better' or 'much better' vs 30% for TAU<br><br>***Taib et al. 2019***<br>• Scores at all timepoints were between 2-3 ('minimally' - 'much improved') in the active arm, and between (3-4) 'unchanged' - 'minimally improved' in the control arm (except the final assessment)<br>• Mean (SD) scores:<br>  o 1m: active=2.3 (1.1), control=3.3 (1.2)<br>  o 2m: active=2.5 (1.3), control=3.6 (1.5)<br>  o 6m: active=2.2 (1.0), control= 3.6 (2.1)<br>  o 12m: active=2.6 (1.5), control=2.8 (1.6) |
| **CGI-I (clinician-rated)** | **Validity** | ***Conwill et al. 2014***<br>Convergent validity: No significant associations between SF-36, HADS anxiety/depression scores, and the CGI-I (clinician) categorical variables of 'improved' or 'not improved' |

| | | |
|---|---|---|
| | **Reliability** | ***Dreissen et al. 2019***<br>Inter-rater reliability: CGI-I (clinician) and CGI-S had an average (SD) weighted kappa=0.65 (0.16) |
| | **Responsiveness** | ***Conwill et al. 2014***<br>• Mean (SD) post-treatment CGI-I score=2.4 (1.1), indicating ('minimally' – 'much improved')<br>• No significant difference between seizures and other FND groups<br><br>***Dreissen et al. 2019***<br>• 16/25 (64.0%) of the BoNT arm rated as improved (clinician-rated symptoms as 'minimally improved', 'much improved' or 'very much improved'), vs 56.5% in placebo arm<br>• Difference not significant (OR=0.075, CI = −0.189, 0.327, p=0.77)<br><br>***Hubschmid et al. 2015***<br>• Poor outcome (defined by a score of ≥4) significantly less likely in the treatment group compared to TAU (p=0.02)<br>• Percentage participants rating >4:<br>o Baseline: TAU=63.6, treatment=60<br>o End of therapy (2m): TAU=30, treatment= 66.7<br>o 6m: TAU=44.4, treatment= 50<br>o 12m: TAU=37.5, treatment=28.6<br><br>***Khattak et al. 2006***<br>• No significant between group difference in CGI-I scores (global) at baseline (mean difference=0.00, CI = -0.32, 0.318, p=0.100)<br>• Active arm had significantly lower scores (i.e., more improved) at end of treatment (mean difference = -1.32, CI = -1.75, -1.1, p<0.001) and follow up (mean difference = -2.89, CI = -3.4, -2.4, p<.001)<br>• Mean (SD) CGI-I scores:<br>o Baseline: treatment arm=4.30 (0.91), control=4.30 (0.68)<br>o End treatment: treatment arm=3.04 (0.83), control=4.46 (0.813)<br>o Follow-up: treatment arm=1.58 (0.87), control=4.48 (1.49)<br><br>***Kompoliti et al. 2014:***<br>• Significant main effect of time, showing similar improvement in CGI-I scores (FMD) across arms (3m, 6m) (p=0.04)<br>• No significant between group difference in improvement at 3m (p=0.46) or 6m (p=0.15)<br>• No significant between arm difference in proportion of patients with positive outcome (i.e., CGI-I score of <4, at least minimally improved) (p=0.40) |

| | | |
|---|---|---|
| | | • Average (range) CGI-I scores by group:<br>  ○ 3m: Immediate=3.0 (range 1e6), Delayed 4.0 (range 1-7)<br>  ○ 6m: immediate=4.0 (range 1e5), delayed=2.0 (range 1-4)<br><br>***LaFrance et al. 2010***<br>• No significant between group difference in CGI-I scores (p>0.05)<br>• Mean (SD) CGI-I scores (end of treatment): treatment arm=3.5 (1.6), placebo=2.9 (1.4)<br><br>***LaFrance et al. 2014***<br>• Treatment arms: sertraline (n=9), CBT (n=9), CBT+sertraline (n=9), TAU (n=7)<br>• Significant decrease in scores over time for:<br>  ○ CBT (mean difference = -2.3, CI = -4.1, -0.5, ES (d) = -1.2, p<0.05)<br>  ○ CBT+sertraline (mean difference = -1.9, CI = -3.1, -0.6, ES (d) = -1.2, p<0.01)<br>• No significant change for:<br>  ○ TAU (mean difference=0.1, CI = -1.1, 1.4, ES(d)=0.1, p>0.05)<br>  ○ sertraline only (mean difference = -1.4, CI = -3.0, 0.1, ES(d)=0.7, p>0.05)<br><br>***Taib et al. 2019***<br>• Scores at all time points were between 2-3 ('minimally' - 'much improved') in the active arm, and between (3-4) 'unchanged' - 'minimally improved' in the control arm<br>• Mean (SD) scores:<br>  • 1m: active=2.1 (0.9), control= 3.1 (1.0)<br>  • 2m: active=2.1 (0.9), control=3.6 (1.0)<br>  • 6m: active=2.1 (0.9), control=3.6 (1.5)<br>  • 12m: active=2.2 (1.2), control=3.4 (1.6)<br><br>***Voon & Lang (2005)***<br>• 80% of primary FMD patients were rated as having improved (both motor and global)<br>• Mean (SD) CGI-I (clinician) scores:<br>  ○ Motor symptoms: 2.3 (1.4) ('minimally improved' – 'much improved')<br>  ○ Global: 2.4 (1.5) ('minimally improved' – 'much improved') |
| **Seizure frequency** | **Validity** | ***De Barros et al. 2018***<br>Convergent validity - significant positive correlations at 8w between seizure frequency and:<br>  ○ depressive symptoms (HAM-D) (*r*=3.4; p=0.02)<br>  ○ anxiety symptoms (HAM-A) (*r*=2.8; p=0.03)<br>  ○ alexithymia scores (TAS) (*r*=3.1; p=0.02) |

| | | |
|---|---|---|
| | | ***Kuyk et al. 2008***<br>Convergent validity - at 6m, seizure frequency (weekly) showed significant negative correlation with SF-36 'energy vitality' (r = -0.56; p = 0.025) |
| | **Reliability** | No data available |
| | **Responsiveness** | ***Barry et al. 2008 (daily seizure log, monthly seizure count)***<br>6/7 patients (86%) reported a decrease in seizure frequency during treatment<br><br>***Bullock et al. 2015 (weekly seizure log/count)***<br>• 9/17(53%) treatment completers reported at least a 50% reduction in seizure frequency (vs baseline)<br>• Significant reduction in mean seizure frequency observed (66% decrease) (seizure frequency ratio=0.34, CI=0.19, 62, p=0.002).<br>• Mean seizure frequency/week: baseline=13.8, end=4.7<br><br>***Chen et al. (2014) (seizure log)***<br>No significant between arm difference in reported seizure frequency at 3m (p=0.359) or 6m (p=0.394)<br><br>***Conwill et al. 2014 (monthly seizure count)***<br>• Mean (SD) seizure frequency improved between start-end of treatment<br>• Start=13.8 (12.6)/month; end=12.3 (13.7)/month<br><br>***Cope et al. 2017 (monthly seizure count)***<br>12/16 patients (75%) reported reduction in seizure frequency during treatment<br><br>***De Barros et al. 2018 (weekly seizure diaries/chart review, weekly seizure count)***<br>Significant reduction in mean seizure frequency in treatment group at 8w (p=0.02), relative to controls<br><br>***De Oliveira Santos et al. 2014***<br>51.4% of sample (n=19) reported reduced seizure frequency by end of treatment<br><br>***Drane et al. 2016 (monthly seizure count)***<br>• Significant reduction in seizure frequency at 8wks (vs baseline) for inpatient psychiatric consultation arms but not for TAU<br>• Mean (SD) seizure frequency: |

- Structured inpatient feedback: baseline=2.9 (0.9), 8wk=1.7 (0.5), p=0.005
- Structured ongoing feedback: baseline=2.9 (0.9), 8wk=1.7 (=0.6), p=0.001)
- TAU: baseline=3.2 (1.1), 8wk=2.5 (1.0) (p=ns)

***Goldstein et al. 2004 (seizure log, monthly seizure count)***

- Significant decrease in reported seizure frequency after 12 CBT sessions, maintained at 6m (p<0.01)
- Mean (SD) seizure frequency/month: pre-treatment=18.22 (43.70), end treatment=2.88 (4.73), 6m=2.59 (4.14)
- ITT analysis also statistically significant (p<0.01)
- 13/16 treatment completers (81.25%) reported >50% reduction in seizure frequency (pre-treatment-follow-up)

***Goldstein et al. 2010 (seizure log, monthly seizure count)***

- No significant between arm difference at start of treatment
- Significantly lower seizure frequency at end of treatment in treatment group (vs controls), ES=0.75
- Non-significant trend for lower seizure frequency at follow-up, ES=0.42
- Median (IQR) seizure frequency/month:
  - Start treatment: treatment arm=12.0 (22.50), control=33 8.00 (29.25)
  - End treatment: treatment arm=2.0 (6.00), control=6.75 (38.63)
  - 6m: treatment arm=1.5 (8.00), control=5.00 (24.00)

***Hovorka et al. 2007 (seizure logs)***

- At 2y, 19/56 (33.9%) patients showed reduction in seizure frequency by >50%
- 18/56 (32.1%) patients did not respond to treatment

***Khattak et al. 2006 (daily seizure log)***

- At baseline, no significant between arm difference in seizure frequency observed (p=0.112)
- Significant difference in frequency between treatment and control arms at end of treatment and follow-up (p<0.001)

***Kuyk et al. 2008 (nursing staff/patient seizure logs, weekly seizure count)***

- Mean seizure frequency/week significantly declined between start treatment, end treatment and 6m (p<.05)
- Mean (SD) seizure frequency/week: start=6.6 (9.8), end=3.0 (4.7), 6m=0.9 (1.8)

***LaFrance et al. 2009 (daily seizure log, weekly seizure count)***

- 16/21 participants reported a 50% reduction in seizure frequency
- In treatment completers, seizure frequency decreased during treatment
- Median seizure frequency/week: pre-treatment=8, end=0
- Mean (SD) frequency/week: start=17.2 (23.2), end=7.1 (14.6), p=0.001

***LaFrance et al. 2010 (daily seizure log, biweekly seizure count):***

- Treatment arm showed a significant 45% reduction in biweekly seizure frequency during treatment (baseline-end), from 22.24 -12.18 (ratio 0.55, CI=0.32, 0.93, p=0.03)
- Control arm showed 8% increase in biweekly seizure frequency (not significant) from 13.38-14.38 (ratio 1.08, CI=0.65, 1.77, p=0.78)
- In patients with seizure rates of >1 at baseline, 8/17 in treatment arm reported >50% reduction in seizure rates by end of treatment (vs 3/16 in placebo arm) (ITT rates: 47.1% vs 18.8%, p=0.18)
- Number needed to treat=3.53
- Mean (SD) seizure frequency by group (raw):
    - Baseline**:** treatment arm=11.3 (12.1), placebo=19.9 (43.5)
    - Week 2: treatment arm=8.9 (8.5), placebo=17.8 (37.7)
    - Week 4: treatment arm=10.3 (10.6), placebo=16.1 (31.5)
    - Week 6: treatment arm=10.9 (16.4), placebo=13.1 (31.5)
    - Week 8: treatment arm=12.1 (17.4), placebo=12.1 (24.4)
    - Week 10: treatment arm=11.7 (12.4), placebo=18.7 (30.7)
    - Week 12: treatment arm=11.6 (14.0), placebo=11.7 (20.3)

***LaFrance et al. 2014 (daily seizure log, monthly seizure count)***

- Significant reductions (baseline-end treatment) observed in (monthly) seizure frequency for CBT and CBT+sertraline arms, but not sertraline only or TAU
- CBT arm: 51.4% fewer seizures (post/pre-treatment ratio=0.49, CI=0.28, 0.84, p=0.01)
- CBT+sertraline: 59.3% fewer seizures (post/pre-treatment ratio=0.41, CI=0.21, 0.79, p=0.008)
- Sertraline only: 26.5% fewer seizures (post/pre-treatment ratio=0.74, CI=0.52, 1.03, p=0.08)
- TAU: 33.8% fewer seizures (post/pre-treatment ratio=0.67, CI=0.37, 1.21, p=0.19)

***Mayor et al. 2010 (monthly seizure count)***

- 19 patients (40.4%) reported a >50% reduction in seizure frequency at follow-up (vs baseline)
- Median seizure frequency/month: baseline=6, follow-up=1 (p<0.007)

***Mayor et al. 2013 (monthly seizure count)***

- Median (IQR) seizure frequency in treatment completers: baseline=8/month

(IQR=2–15), follow-up=8 (IQR=0–16)
- 3 patients (23%) had >50% reduction in seizure frequency
- 5 patients unchanged
- 1 patient reported worsening

***Metin et al. 2013 (weekly seizure log, weekly seizure count)***
- Significant reduction in seizure frequency between start of treatment and 12m (p<0.0001)
- At 12m, 100% of patients reported >50% reduction in seizure frequency
- Median (range) seizure frequency: pre-treatment=24 (2–60), 1m=2 (0–19), 2m=2 (0–18), 3m=2 (1–43), 4m=2 (0–10), 6m=1 (0–30), 9m=0 (0–14), 12m=0 (0–6)

***Myers et al. 2017 (daily seizure count)***
- Significant reduction in seizure frequency between baseline and end of treatment (p=0.001)
- Reduction maintained from the end of treatment to follow-up (p =0.285)
- Seizure frequency at follow-up was significantly lower than at baseline (p=0.028)

***Oto et al. 2010 (monthly seizure count)***
- Median seizure frequency/month reduced for both study groups
- Median (range frequency/month:
    - immediate AED withdrawal: baseline=20 (range 5–720), 9m=2 (0–290), 18m=1 (0–6)
    - delayed AED withdrawal: baseline=12 (6–120), 9m=6 (0–100), 18m=4 (0–32)
- Decrease in seizure frequency from baseline to 9m was statistically significant (p=0.028) for the immediate withdrawal arm (but not delayed, p=0.415)

***Pintor et al. 2010 (seizure records, 15 days/monthly, 15-day seizure count)***
- Significant reduction in seizure frequency between baseline and follow-up (p<0.001)
- Mean (SD) scores (and p-values for comparisons to baseline scores):
    - baseline=7.78 (8.2)
    - 1m=4.62 (6.1) (p=0.017)
    - 2m=3.60 (6.5) (p=0.01)
    - 3m=3.10 (5.0) (p=0.009)
    - 4m=2.25 (2.8) (p=0.001)
    - 5m=1.62 (2.3) (p<0.001)

***Thompson et al. 2013 (weekly seizure count)***
No significant change in seizure frequency in either group (data not provided)

| | | | |
|---|---|---|---|
| | | | ***Tolchin et al. 2019 (interview, weekly seizure count)***<br>• Significant between arm difference in seizure frequency change (p=0.034, Cohen's d=0.59 (CI=0.04, 1.14)<br>• Mean (SD) change in seizure frequency: control=34.8% (89.7%), MI arm=76.2% (39.2%)<br><br>***Wiseman et al. 2016 (monthly seizure count)***<br>• End of treatment: 12/18 patients (67%) reported decline in seizure frequency (seizure free/fewer seizures) relative to start of treatment<br>• Reduction in seizure frequency (median: start=8/month, end=3/month) not significant<br><br>***Zaroff et al. 2004 (structured study-specific questionnaire, daily/weekly seizure frequency)***<br>• 4/7 patients reported no change in seizure frequency between pre- and post-treatment (3 reported seizure freedom pre-treatment so no change expected)<br>• 2/7 reported a reduction in seizure frequency<br>• 1/7 reported an increase in seizure frequency |
| | **Seizure freedom** | **Validity** | ***Kuyk et al. 2008***<br>Convergent validity: at 6m, seizure free patients reported significantly greater improvement on: SF-36 (mental health, energy vitality, pain), BDI, and STAI, relative to patients who were not seizure free (p<0.05) |
| | | **Reliability** | No data available |
| | | **Responsiveness** | ***Ataoglu et al. 2003***<br>• 14/15 (93.3%) treatment completers (paradoxical intention) showed improvement by the end of treatment (i.e., no symptoms reported in previous 2 weeks)<br>• 9/15 (60%) patients in diazepam arm showed improvement<br>• The recovery rate was significantly greater in the PI arm than the diazepam arm (p=0.034)<br><br>***Barry et al. 2008***<br>4/7 patients (57%) reported seizure freedom<br><br>***Bullock et al. 2015***<br>6/17 participants (35%) reported seizure cessation during treatment |

***Cope et al. 2017***
- 7/18 (38.9%) of patients reported seizure freedom at end of treatment
- Significant increase from pre- to post-treatment (2/18, 11.1%, p=0.032)

***De Oliveira Santos et al. 2014***
9.7% of sample (n=11) reported seizure freedom

***Duncan et al. 2016***
At 6m, 43/81 patients (53.1%) reported 2-month seizure freedom

***Goldstein et al. 2004***
- 4/16 participants (25%) reported seizure cessation at 6m, since end of treatment
- 7 participants (43.75%) reported seizure cessation in the previous month (at 6m follow-up)

***Goldstein et al. 2010***
- Non-significant trend for the treatment arm to be more likely to report seizure freedom at 6m (OR=3.125, CI=0.852, 11.468, p=0.086)
- Absolute risk reduction=19.5%
- Number needed to treat (for seizure cessation in 1 patient)=5.13

***Hovorka et al. 2007***
At 2y, 16/56 (28.6%) patients were seizure free for at least 12 months

***Kuyk et al. 2008***
Proportion of patients reporting seizure freedom: start=7.7%, end=27.3%, 6m=44%

***LaFrance et al. 2010***
11/17 treatment completers (65%) reported seizure cessation by end of treatment

***LaFrance et al. 2010***
- Six participants in treatment arm reported seizure freedom, vs 1 in placebo arm
- ITT rates: 35.3% vs 6.3%, p=0.08

***LaFrance et al. 2014***
- Seizure freedom rates at end of study by group (n):
    - CBT= 3
    - CBT+sertraline=5
    - sertraline=1

| | | |
|---|---|---|
| | | ○ TAU= 1<br>• Odds of seizure cessation=6.2 greater for CBT arm, vs patients not receiving CBT (p=0.06)<br><br>***Mayor et al. 2010***<br>• 12/47 patients (25.5%) reported seizure freedom (>1 month) at end treatment<br>• At follow-up (median 50m after baseline) 12/47 (25.5%) patients reported seizure freedom<br><br>***Mayor et al. 2013***<br>4/13 patients (31%) seizure free at follow-up<br><br>***Metin et al. 2013***<br>6/9 (67%) of patients were seizure free at 12m<br><br>***Myers et al. 2017***<br>13/16 treatment completers (81%) reported seizure freedom by end of treatment<br><br>***Oto et al. 2010***<br>• At 9m, proportions of seizure free patients were similar in the immediate and delayed AED arms<br>• Number of seizure free patients (%) at 9m: immediate withdrawal=3/14 (21%), delayed=3/11 (27%)<br>• At 18m, twice as many patients in the immediate arm were seizure free, relative to the delayed arm (p=0.173)<br>• Number of seizure free patients (%) at 18m: immediate withdrawal=7/14 (50%), delayed=3/11 (27%)<br><br>***Tolchin et al. 2019***<br>• 3 control participants (10.7%) reported seizure freedom in the 4 weeks before follow-up, vs 8 in the MI arm (30.8%)<br>• This was not a significant difference (p=0.095)<br><br>***Wiseman et al. 2016***<br>6/18 patients (33%) reported seizure freedom for the past month, at end of treatment |
| **Psychogenic Movement Disorders Rating Scale (PMDRS)** | **Validity** | ***Taib et al. 2019***<br>Convergent validity:<br>• SF-36 physical role domain significantly negatively correlated with PMDRS total scores ($r = -0.77$, p=0.0002)<br>• Significant positive correlation between PMDRS and CGI-severity scores ($r$=0.88, p=0.001) |

| | | |
|---|---|---|
| | | • Negative correlation (12m) between PMDRS total or tremor scores and general health domain of SF-36 (PMDRS, $r = -0.67$, p=0.02; tremor, $r = -0.53$, p=0.05) |
| | **Reliability** | ***Dreissen et al. 2019***<br>Inter-rater reliability: average ICC=0.76, SD=0.11 |
| | **Responsiveness** | ***Dallocchio et al. 2010***<br>• Total PMDRS scores improved by 70% (range 48–100%) compared to baseline, in 10/16 patients (p=0.014)<br>• Function scores improved significantly(p=.043)<br>• Mean (SD) scores:<br>  ○ PMDRS total: start=69.3 (41.3), end=28.2 (19.9)<br>  ○ PMDRS function: start=9.8 (7.8), end=4 (5.7)<br><br>***Dallocchio et al 2016***<br>• Significant improvement over time for both active arms (p<0.001) but not for controls (p>0.05)<br>• No significant differences in improvement between the two active arms (p>0.05)<br>• PMDRS (mean (SD)) scores:<br>  ○ CBT: baseline=71.5 (21.4), end=33.2 (30.2) (p<0.001)<br>  ○ CBT+physical activity: baseline=76.7 (16.6), end=38.8 (18.1) (p<0.001)<br>  ○ TAU: baseline=72.4 (22.3), end=69.8 (20.8)<br><br>***Demartini et al. 2019b:***<br>• Significant improvement for PMDRS scale total score over time (start treatment, 12, 24 wks): p<0.001<br>• Significant differences between start and both follow-ups (p<0.001), but no difference between 12 and 24 wks (p=ns)<br>• Mean (SD): start=22.6 (9.6), 12 wk=12.8 (7.9), 24 wk=10.4 (6.5)<br><br>***Dreissen et al. 2019***<br>• No significant between arm differences in change scores<br>• Median difference = −3.0, CI = −2.0, 4.0, p=0.438<br><br>***Espay et al. 2014***<br>• Tremor improved across all participants<br>• PMDRS tremor scores: baseline=22.2 (13.4), end=4.3 (5.5), p=0.0019<br>• Three subjects reported tremor freedom<br>• Improvement maintained for 1wk to 6m |

- At 6m, four patients had experienced relapse whereas six remained markedly improved, including one free of tremor

***Ferrara et al. 2011***

- Significant improvement in total PMDRS scores at 6-7m follow-up (p=0.02)
- 5 patients experienced a >50% improvement during initial TENS session

***Hinson et al. 2006***

- Significant improvement in total PMDRS scores from baseline to post-treatment
- Mean (SD) PMDRS total scores: baseline=71.2 (42.5), post-treatment=29.0 (20.6) (p=0.0195)
- PMDRS function score declined significantly
- Mean (SD) PMDRS function scores: baseline=7.4 (6.1), post-treatment=2.1 (3.3) (p=0.0142)
- 7 patients improved in PMDRS scores (75%)

***Kompoliti et al. 2014***

No significant between arm differences in PMDRS scores for baseline-3m or 3m-6m comparisons (minimum p=0.28)

***Taib et al. 2019***

- Mean PMDRS scores (1m) decreased significantly in the active arm (p=0.0004) but not in the control rTMS arm
- Mean (SD) change in PMDRS scores: active = -11.7 (5.5), control = -4.9 (5.0)
- Mean Tremor scores (1m) decreased significantly in the active arm (p=0.0001) but not in the control arm
- Mean (SD) change in PMDRS scores: active = -5.9 (2.4), control = −3.0 (3.2)

***Vizcarra et al. 2019***

- All patients showed improvement (reductions) in PMDRS scores at follow-up
- PMDRS scores improved significantly (relative to baseline) only in the Placebo+CBT arm (mean change=9.0, CI = -16.5, 1.5, p=0.02)
- Mean (SD) PMDRS scores:
    - BoNT+CBT: baseline=21.3(13.8), follow-up: 12(10.3), change= −9.3 (CI = −19.9, 1.3), p=0.07
    - Placebo+CBT: baseline=15.3(9.6), follow-up= 6.3(9.9), change = −9.0 (CI = −16.5, −1.5), p=0.02

| | | | |
|---|---|---|---|
| **Other physical symptoms** | **SCL-90 (somatisation subscale)** | **Validity** | No data available |
| | | **Reliability** | No data available |
| | | **Responsiveness** | ***Barry et al. 2008***<br>Reduction in SCL-90 Somatisation scores observed during treatment (ES=0.31)<br><br>***LaFrance et al. 2009***<br>• Significant reduction in somatisation scores from baseline to end of treatment (p=.027)<br>• Mean (SD) scores: baseline=94.3 (77.6), 1m=87.7 (68.2), end=62.4 (52.8)<br><br>***LaFrance et al. 2010***<br>• No significant between arm difference in change scores (baseline-end of treatment), after adjusting for baseline scores (p>0.05)<br>• Mean (SD) scores by group:<br>o Baseline: placebo=109.4 (70.9), treatment arm=91.4 (77.2)<br>o End: placebo=84.9 (73.3), treatment arm=78.9 (67.2)<br><br>***LaFrance et al. 2014***<br>• Significant reduction in scores over time for CBT arm (mean difference=46.9, CI = -74.7, -19.0, ES(d) = -1.6, p<0.001).<br>• No significant change for:<br>o CBT+sertraline: mean difference = -24.1, CI = -87.2,39.0, ES(d) = -0.3, p >0.05<br>o Sertraline only: mean difference = -14.4, CI = -45.9,17.0, ES(d) = -0.4<br>o TAU: mean difference=23.7, CI = -43.0,90.3, ES(d)=0.4, p>0.05<br><br>***Pleizier et al. 2017***<br>• No significant between arm differences in mean change scores<br>• Mean (SD) SCL-90 somatisation scores:<br>o Neurologist management: baseline=28.24 (8.21), 12m=24.77 (7.59), change = −3.48 (7.41)<br>o GP management: baseline=28.39 (8.62), 12m=24.92 (9.25), change = −3.47 (8.40)<br>• Mean treatment change difference = −0.01 (CI = −2.25, 2.24), p=0.95 |
| | **Patient Health Questionnaire-15 (PHQ-15)** | **Validity** | No data available |
| | | **Reliability** | No data available |

| | | | |
|---|---|---|---|
| | | **Responsiveness** | ***Dallocchio et al. 2016***<br>• Significant improvement over time for both active arms (p<0.001) but not for controls (p>0.05)<br>• No significant differences between the two active arms (p>0.05)<br>• Mean (SD) scores:<br>  ○ CBT: baseline=19.8 (4.4), end=8.7 (7.1) (p<0.001)<br>  ○ CBT+physical activity: baseline=20.9 (3.3), end=10.6 (3.8) (p<0.001)<br>  ○ TAU: baseline=19.3 (5.6), end=21.1 (6.4)<br><br>***DeMartini et al. 2014:***<br>• Significant improvement in PHQ-15 scores between start/end treatment (p<0.001, Cohen's d=0.53)<br>• Mean (SD): start=12.6 (5.5), end=9.8 (5.1)<br><br>***Reuber et al. 2007***<br>• Significant decline in the mean number of physical symptoms between pre- and post-treatment (p=0.008)<br>• Mean (SD) scores: pre-treatment=13.8 (5.5), post-treatment=11.6 (6.2), 6m=11.0 (6.5)<br><br>***Sharpe et al. 2011 (PHQ-13, minus sexual/menstrual items):***<br>• GSH arm reported fewer physical symptoms than TAU at post-treatment (3m) (AMD = -1.0, CI = -1.7, -0.2, p=0.009).<br>• Mean (SD) scores: GSH=6.2 (3.3), TAU=7.0 (3.0)<br><br>***Williams et al. 2018***<br>• PHQ-15 scores declined between pre- and post-treatment, but this was not significant following Holm–Bonferroni correction (p=0.03, CI=0.35, 3.43, critical p-value= =0.025)<br>• Mean (SD) scores: pre-treatment=14.05 (5.35), post-treatment=12.14 (6.32) |
| **Psychological symptoms** | **Hospital Anxiety and Depression Scale (HADS)** | **Validity** | ***Conwill et al. 2014***<br>Convergent validity: No significant associations between HADS anxiety/depression scores and the CGI-I (clinician) categorical variables of 'improved' or 'not improved' |
| | | **Reliability** | No data available |
| | | **Responsiveness** | ***Conwill et al. 2014***<br>• HADS anxiety and depression scores improved start-end treatment, but not significantly (p>0.05)<br>• Mean (SD) scores:<br>  ○ HADS anxiety (n = 16): start=8.6 (6.0), end=7.8 (5.0), p=0.34 |

- HADS depression (n = 16): start=9.1 (5.3), end=8.4 (5.0), p=0.46

***DeMartini et al. 2014***

- Significant improvements in HADS scores between start/end treatment (p=0.004, Cohen's d=0.26)
- Mean (SD): start=15.8 (8.5), end=13.3 (8.2)

***Goldstein et al. 2004***

- HADS anxiety and depression scores decreased significantly between pre- and post-treatment (p<.05) and the improvements were sustained to 6m follow-up
- Mean (SD) HADS anxiety scores: pre-treatment=10.06 (5.62), end=7.81 (5.52), 6m=8.13 (6.71)
- Mean (SD) HADS depression scores: pre-treatment= 6.75 (3.55), end=4.63 (4.22), 6m=4.63 (5.08)

***Goldstein et al. 2010***

- No significant group x time interactions or main effects of group or time, for either HADS anxiety or depression scores
- Mean (SD) anxiety scores:
  - Start of treatment: treatment arm=8.83 (4.95), TAU=9.02 (4.82)
  - End of treatment: treatment arm=7.93 (3.58), TAU=8.79 (4.77)
  - Follow-up: treatment arm=7.15 (5.16), TAU=8.79 (5.22)
- Mean (SD) depression scores:
  - Start of treatment: treatment arm=6.74 (4.05), TAU=7.88 (5.07)
  - End of treatment: treatment arm=6.20 (4.08), TAU=7.04 (4.93)
  - Follow-up: treatment arm=5.69 (5.34), TAU=7.38 (5.21)

***Khattak et al. 2006***

- No significant between arm difference in HADS anxiety scores at baseline (mean difference = -0.5, CI = -1.93, 0.93, p=0.490)
- Significantly lower anxiety scores in the treatment arm (vs controls) at end of treatment (mean difference= -2.28, CI = (-3.666, -0.894), p=0.002) and follow-up (mean difference= -6.01, CI = -7.519, 4.493, p<0.001) (scores lower in treatment arm)
- Mean (SD) HADS anxiety scores by group:
  - Baseline: treatment arm=9.60 (3.69), control=10.1 (3.52)
  - End treatment: treatment arm=7.42 (3.41), control=9.7 (3.57)
  - Follow-up: treatment arm=3.21 (2.82), control=9.21 (4.33)

- No significant between arm difference in HADS depression scores at baseline (mean difference=0.76, CI=-1.04, 2.56 p=0.41) or at end of treatment (mean difference= -1.48, CI = -3.21, -2.5, p=0.09)
- Depression scores significantly lower in treatment arm (vs controls) at follow-up (mean difference = -5.38, CI = -7.56, -3.81, p<0.001)
- Mean (SD) HADS depression scores:
    - Baseline: treatment arm= 13.5 (4.6), control= 12.8 (4.4)
    - End treatment: treatment arm=10.9 (4.25), control= 12.3 (4.47)
    - Follow-up: treatment arm= 5.6 (3.97), control= 11.3 (4.97)

***Nielsen et al. 2015***

- No significant change in HADS anxiety scores during treatment (mean (SD)): start=7.1 (4.4), end=6.1 (4.7), follow-up=6.9 (4.6) (p=.114)
- No significant change in HADS depression scores during treatment (mean (SD)): start=6.0 (3.9), end=5.3 (4.0), follow-up=6.0 (4.6) (p=.96)

***Nielsen et al. 2017b***

- No significant between arm differences at 6m follow-up for HADS anxiety and depression scores (adjusting for baseline differences)
- Anxiety mean scores (SD): physiotherapy=6.9 (4.8), control arm=7.9 (5.6) (coefficient = −0.1, CI = −2.1, 2.0, Cohen's d = 0.02)
- Depression mean scores (SD): physiotherapy arm=5.2 (3.9), control arm=8.4 (5.0) (coefficient = −1.4, CI = −3.2, 0.5, Cohen's d=0.30)

***Oto et al. 2010***

- No significant between arm differences in change scores over time
- Anxiety mean (SD) scores:
    - Baseline: immediate=10.9 (6.4), delayed=9.9 (3.4)
    - 9m: immediate=8.7 (4.4), delayed=8.1 (4.3)
    - 18m: immediate=10.0 (3.2), delayed=9.5 (3.4)
- Depression mean (SD) scores:
    - Baseline: immediate=9.0 (5.5), delayed=6.2 (4.0)
    - 9m: immediate=6.2 (4.8), delayed=7.5 (3.5)
    - 12m: immediate=6.9 (3.5), delayed=7.0 (3.8)

***Pintor et al. 2010***

- Both HADS depression (p=0.008) and anxiety (p = .001) scores improved significantly between baseline and follow-up assessments
- Mean (SD) HADS depression scores (plus significance values for comparison to baseline, where available): baseline=8.23 (5.4), 1m=7.34 (4.9), 2m=5.97 (4.4), 3m=5.38 (4.9), 4m=5.07 (5.3) (p=0.02), 5m=4.85 (4.8) (p=0.008)
- Mean (SD) HADS anxiety scores (plus significance values for comparison to baseline, where available): baseline=11.18 (4.3), 1m=10.16 (5.6), 2m=8.97 (4.9) (p=0.04), 3m=8.77 (6.2), 4m=8.50 (6.0) (p=0.02), 5m=7.79 (5.5) (p<0.001)

***Pleizier et al. 2017***

- No significant between arm differences in mean change scores
- Mean (SD) HADS anxiety scores:
    - Neurologist management: baseline=7.79 (4.68), 12m=6.84 (4.17), change = −0.95 (3.82)
    - GP management: baseline=8.46 (4.67), 12m=7.31 (4.33), change = −1.15 (3.26)
- HADS anxiety treatment change difference=0.20 (CI = −0.81, 1.21), p=0.90
- Mean (SD) HADS depression scores:
    - Neurologist management: baseline= 6.62 (4.95), 12m= 5.65 (4.88), change = −0.97 (4.13)
    - GP management: baseline=6.81 (4.34), 12m=5.55 (4.74), change = −1.26 (4.04)
- HADS depression treatment change difference= 0.29 (CI = −0.88, 1.45), p=0.69

***Sharpe et al. 2011:***

- No significant differences between treatment arms (GSH, TAU) post-treatment (3m):
    - depression (AMD=-1.1, CI=-2.4, 0.1, p=0.075).
    - anxiety (AMD=-0.8, CI=-1.8, 0.3, p=0.153)
- Mean (SD) subscale scores (3m)
    - depression: GSH=6 (4.9), TAU=7.3 (4.1)
    - anxiety: GSH=6.6 (3.9), TAU=8.2 (4.9)

***Taib et al. 2019***

- No significant change over time in either group
- Mean (SD) anxiety scores:
    - Baseline: active=9.6 (4.1), control=7.6 (3.6)
    - 1m: active=8.0 (5.0), control=7.1 (2.9)
    - 6m: active=7.8 (4.2), control=8.1 (3.3)
    - 12m: active=7.1 (3.0), control=7.1 (1.7)
- Mean (SD) depression scores:
    - Baseline: active=4.8 (3.0), control=8.1 (4.9)

| | | |
|---|---|---|
| | | ○ 1m: active=5.3 (2.7), control=6.2 (3.0)<br>○ 6m: active=5.5 (3.2), control=6.6 (4.8)<br>○ 12m: active=6.2 (4.1), control=6.0 (3.8) |
| **Beck Depression Inventory (BDI) / BDI-II** | **Validity** | No data available |
| | **Reliability** | No data available |
| | **Responsiveness** | ***Barry et al. 2008 (BDI)***<br>• A significant reduction in scores (improvement) reported during treatment<br>• Mean (SD) scores: pre-treatment=16.6 (10.1), post-treatment=13.3 (7.9) (p<0.01)<br><br>***Drane et al., 2016 (BDI-II)***<br>• Significant decrease in BDI-II scores for one intervention arm (baseline=8wk), but not in TAU or a second intervention arm<br>• Structured ongoing feedback arm (mean (SD)): baseline=23.7 (13.0), 8wk=15.5 (9.6), ES ($n^2$)=0.64, p<0.001)<br>• Structured feedback (mean (SD)): baseline=22.7 (11.0), 8wk=18.8 (9.7), ES ($n^2$)=0.24, p=NS<br>• TAU (mean (SD)): baseline= 25.9 (14.5), 8wk= 25.1 (14.0), ES ($n^2$)=0.10, p=NS<br><br>***Dreissen et al. 2019 (BDI)***<br>• No significant between arm difference in change scores<br>• Median difference = −1.0, CI = −3.0, 3.0, p=0.802)<br><br>***Hubschmid et al. 2015 (BDI)***<br>• Likelihood of BDI score ≥20 (cut-off for moderate depression) significantly lower in treatment group (p<0.05) vs TAU (<0.05)<br>• Percentage participants with scores >20 by group:<br>○ Baseline: TAU=10, treatment=40<br>○ End of therapy (2m): TAU=22.2, treatment= 28.6<br>○ 6m: TAU=22.22, treatment=42.9<br>○ 12m: TAU=28.6, treatment=28.6<br><br>***Kuyk et al. 2008 (BDI)***<br>• Mean BDI scores were in the 'mild-moderate' range at start of treatment, reducing to the 'minimal' range at end of treatment and 6m<br>• Mean (SD): start=19.7 (9.4), end=11.5 (10.9), 6m=9.2 (7.5) |

| | | | |
|---|---|---|---|
| | | | ***LaFrance et al. 2009 (BDI-II)***<br>• Significant reduction in scores from baseline to end of treatment (p=0.015)<br>• Mean (SD) scores: baseline=19.1 (15), 1m=18.5 (22.4), end=10.7 (7.8)<br><br>***LaFrance et al. 2010 (BDI-II)***<br>• No significant between arm difference in change scores (baseline-end of treatment), when adjusting for baseline differences (p>0.05)<br>• Mean (SD) BDI-II scores by group:<br>• Baseline: placebo=22.1 (13.9), treatment arm=16.7 (13.0)<br>• End: placebo=11.7 (11.5), treatment arm=11.7 (11.5)<br><br>***LaFrance et al. 2014 (BDI-II)***<br>• Significant reduction in scores over time for CBT arm (mean difference = -9.0, CI = -15.8, -2.2, ES(d) = -1.2, p<0.01)<br>• No significant change for:<br>o CBT+sertraline: mean difference = -1.0, CI = -11.4, 9.4, ES(d) = -0.1, p>0.05<br>o Sertraline only: mean difference = -8.6, CI = -16.4, -0.7, ES(d) = -0.8, p>0.05<br>o TAU: mean difference = -2.0, CI = -20.2, 16.2, ES(d) = -0.1, p>0.05<br><br>***Myers et al. 2017 (BDI-II)***<br>• Significant improvement in BDI-II scores between baseline and end treatment (p<0.0001)<br>• Mean (SD) scores: baseline=27.00 (8.5), end=13.44 (7.94) |
| | **Beck Anxiety Inventory (BAI)** | **Validity** | No data available |
| | | **Reliability** | No data available |
| | | **Responsiveness** | ***Dallocchio et al. 2010***<br>• Significant pre-post treatment improvement (p=0.034)<br>• Mean (SD) scores: start=27.71 (13.2), end=19.28 (9.5)<br><br>***Dallochio et al. 2016***<br>• Significant improvement over time for both active arms (p<0.001) but not for controls (p>0.05)<br>• No significant differences between two active treatment arms (p>0.05)<br>• Mean (SD) BAI scores: |

| | | |
|---|---|---|
| | | ○ CBT arm: baseline=27.6 (6.8), end=18.6 (6.5) (p<0.001)<br>○ CBT+physical activity: baseline=27.5 (6.6), end=15.2 (4.1) (p<0.001)<br>○ TAU=baseline=28.2 (5.9), end=26.9 (6.2)<br><br>***Dreissen et al. 2019***<br>• No significant between arm difference in change scores<br>• Median change scores=1.0, CI = −5.0 to 1.0, p=0.213<br><br>***Hinson et al. 2006***<br>• Significant improvement in BAI scores observed (p=0.002)<br>• Mean (SD): pre-treatment=19.7 (10.2), post-treatment=4.9 (2.4)<br><br>***Kompoliti et al. 2014***<br>• Significant change over time (p<0.0005), but no significant difference between groups (immediate vs delayed treatment) (p=0.07)<br>• Mean (SD) BAI scores by group:<br>○ Baseline: immediate=21.1 (7.3), delayed=11.9 (8.0)<br>○ 3m: immediate=5.0 (10.0), delayed=10.5 (6.7)<br>○ 6m: immediate=10.0 (5.9), control=7.9 (8.0)<br><br>***LaFrance et al. 2014***<br>• Significant reduction in scores over time for CBT arm (mean difference=13.3, CI = -21.7, -5.0, ES(d) = -1.7, p<0.001)<br>• No significant change for:<br>○ CBT+sertraline: mean difference = -10.7, CI = -27.9,6.5, ES(d) = -0.7, p>0.05<br>○ Sertraline only: mean difference =1.5, CI = -9.7,12.7, ES(d)=0.1, p>0.05<br>○ TAU: mean difference = -5.4, CI = -19.9,9.1, ES(d) = -0.5, p>0.05 |
| **Hamilton Rating Scale – Anxiety (HAM-A)** | **Validity** | ***De Barros et al. 2018***<br>Convergent validity - significant positive correlations at 8w between seizure frequency and HAM-A scores (r = 2.8; p=0.03)<br><br>***Demartini et al. 2019b:***<br>Known groups validity: mean (SD) baseline scores: FMS=11.78 (7.95), HC=4.14 (5.7), p=0.048 |
| | **Reliability** | No data available |
| | **Responsiveness** | ***Ataoglu et al. 2003***<br>• HAM-A scores decreased significantly by end of treatment in both PI (p=0.0007) and diazepam arms (p=0.0012)<br>• The decrease was significantly greater in the PI arm vs diazepam (p=0.015) |

| | | |
|---|---|---|
| | | • Mean (SD) HAM-A scores:<br>  ○ Pre-treatment: diazepam=25.60 (4.27), PI arm=27.6 (5.00) (p=0.280)<br>  ○ Post-treatment: diazepam=18.2 (3.47), PI=14.47 (5.36)<br>  ○ Mean difference: diazepam=7.27 (4.56), PI=13.13 (5.67) (p=0.015)<br><br>***De Barros et al. 2018***<br>• Significant improvement in physical symptom dimension of HAM-A (p=0.02) between pre- and post-treatment<br>• Mean (SD) scores:<br>  ○ Anxious cognition: pre-treatment=17 (9), post-treatment=15 (5), p=0.59, Cohen's d=0.22<br>  ○ Anxious mood: pre-treatment=10 (5), post-treatment=10 (3), p=0.45, Cohen's d=0.00<br>  ○ Physical symptoms: pre-treatment=6 (5), post-treatment=5 (3), p=0.02, Cohen's d=0.20<br><br>***Demartini et al. 2019b:***<br>• Non-significant trend for improvement over time (start treatment, 12, 24 wk) (p=0.06)<br>• Mean (SD): start=9.5 (6.7), 12 wk=9.8 (8.1), 24 wk= 7.9 (6.6)<br><br>***Pintor et al. 2010***<br>• HAM-A scores significantly decreased between baseline and follow-up assessments (p=0.001)<br>• Mean (SD) HAM-A scores (plus significance values where available): baseline=26.22 (11.8), 1m= 26.60 (12.1), 2m=23.61 (9.6), 3m=22.67 (15.0) (p=0.02), 4m=17.69 (14.1) (p=0.006), 5m=15.16 (12.4) (p=0.001)<br><br>***Vizcarra et al. 2019***<br>• No significant between arm differences reported<br>• Mean (SD) HAM-A scores:<br>  ○ BoNT+CBT: baseline=19.8(10.5), follow-up=13.5(9.7), change = −6.3 (CI = −12.6, −0.1), p=0.05<br>  ○ Placebo+CBT: baseline=20(5.4), follow-up=14.5(13.7), change = −5.5 (CI = −24.8, 13.8), p=0.47 |
| **Hamilton Rating Scale – Depression (HAM-D)** | **Validity** | ***De Barros et al. 2018***<br>Convergent validity - significant positive correlations at 8w between seizure frequency and HAM-D scores (r = 3.4; p=0.02)<br><br>***Demartini et al. 2019b:***<br>Known groups validity: mean (SD) scores at baseline: FMS=14.7 (9.3), HC=5.1 (7.03), p=0.041 |

| | | |
|---|---|---|
| | **Reliability** | No data available |
| | **Responsiveness** | ***Dallocchio et al. 2010***<br>• HAM-D scores improved significantly from start to end of treatment (p<0.028)<br>• Mean (SD) scores: start=13.3 (6.7), end=7.5 (3.3)<br><br>***Dallocchio et al. 2016***<br>• Significant improvement in HAM-D scores over time in both active arms (p<0.001) but not for controls (p>0.05)<br>• No significant differences between the two active groups (p>0.05)<br>• Mean (SD) HAM-D scores:<br>  ○ CBT: baseline=14.9 (3.4), end=7.6 (3.5) )(p<0.001)<br>  ○ CBT+physical activity: baseline=12.6 (3.9), end=7.1 (3.1) (p<0.001)<br>  ○ TAU: baseline=13.6 (3.4), end=12.9 (3.2)<br><br>***De Barros et al. 2018***<br>• Significant improvement in HAM-D scores observed between pre- and post-treatment (p<0.0001, Cohen's d=0.71)<br>• Mean (SD) scores: pre-treatment=18(7), post-treatment=13 (4)<br><br>***Demartini et al. 2019b:***<br>• No significant difference in scores between start/end (p=0.155)<br>• Mean (SD): start=9.0 (6.7), 12 wk=10.2 (8.9), 24 wk=8.4 (7.4)<br><br>***Hinson et al. 2006***<br>• Significant improvement in HAM-D scores was reported (p=0.009)<br>• Mean (SD): pre-treatment=14.8 (7.1), post-treatment=3.9 (2.1)<br><br>***Kompoliti et al. 2014***<br>• Significant main effect of time (p=0.001), but no significant between arm difference in HAM-D scores (p=0.719).<br>• Mean (SD) HAM-D scores by group:<br>  ○ Baseline: immediate=13.3 (7.5), delayed=8.6 (6.4)<br>  ○ 3m: immediate=9.6 (7.1), delayed=5.8 (5.7)<br>  ○ 6m: immediate=8.6 (8.8), delayed=2.8 (2.9)<br><br>***LaFrance et al. 2009***<br>• HAM-D scores did not change significantly over treatment (p=0.055) (baseline-end treatment) |

- Mean (SD) scores: baseline=14.6 (7.3), 1m=10.4 (7), end=11.6 (7.2)

***LaFrance et al. 2010***

- No significant between arm difference in change scores (baseline-end of treatment), when adjusting for baseline scores (p>0.05)
- Mean (SD) HAM-D scores by group:
    - Baseline: placebo=16.8 (8.8), treatment arm=17.8 (21.0)
    - End: placebo=13.3 (8.4), treatment arm=11.6 (9.0)

***LaFrance et al. 2014***

- Significant reduction in scores over time for CBT arm (mean difference = -6.0, CI = -9.6, -2.4, ES(d) = -1.8, p<0.001)
- No significant change for:
    - CBT+sertraline: mean difference = -4.3, CI = -9.5,0.9, ES(d) = -0.8, p>0.05
    - Sertraline only: mean difference=0.1, CI = -5.2,5.5, ES(d)=0.0, p>0.05
    - TAU: mean difference=1.1, CI = -7.5,9.8, ES(d) = 0.1, p>0.05

***Pintor et al. 2010***

- Scores showed significant improvement between baseline and follow-ups (p<0.001)
- Mean (SD) HAM-D scores (plus significance values where available): baseline=21.21 (8.5), 1m= 19.26 (8.0) (p=0.009), 2m=15.53 (8.1) (p=0.02), 3m=13.36 (10.2) (p=0.02), 4m=11.97 (10.6) (p=0.001), 5m=10.2 (7.6) (p<0.001)

***Taib et al. 2019***

- No significant change over time in either group
- Mean (SD) scores:
    - Baseline: active=13.8 (5.5), control=14.1 (8.0)
    - 1m: active=12.0 (6.6), control=12.2 (5.9)
    - 6m: active=9.8 (5.1), control=11.2 (5.4)
    - 12m: active=11.6 (7.0), control=13.1 (4.2)

***Vizcarra et al. 2019***

- After sensitivity analysis, only the Placebo+CBT arm showed a significant improvement in HAM-D scores (mean change = −13.2, CI = - 23.8, −2.6, p = 0.02)
- Mean (SD) raw HAM-D scores:
    - BoNT+CBT: baseline=17.2(7.2), follow-up=10.3(10.1), change = −6.8 (CI = −18.2, 4.5), p=0.19

| | | | |
|---|---|---|---|
| | | | ○ Placebo+CBT: baseline=16.3 (7.1), follow-up=6.8 (9.1), change = −9.5 (CI = −23.5, 4.5), p=0.09 |
| **Quality of life / disability** | **QoLIE (10, 31)** | **Validity** | No data available |
| | | **Reliability** | No data available |
| | | **Responsiveness** | ***Drane et al. 2016 (QoLIE-10)***<br>• Significant improvement in scores (baseline-8wk) in both intervention arms but not in TAU arm<br>• Mean (SD) scores:<br>○ Structured feedback arm: baseline=19.4 (5.9), 8wk=33.6 (5.5), ES ($n^2$)=0.80, p<0.001)<br>○ Structured ongoing feedback arm: baseline=21.7 (6.8), 8wk= 34.7 (3.8), ES ($n^2$)=0.87, p<0.001<br>○ TAU: baseline=34.3 (6.6), 8wk= 31.0 (5.8), ES ($n^2$)=0.29, p=ns<br><br>***LaFrance et al. 2009 (QoLIE-31)***<br>• Significant improvement in scores from baseline to end of treatment (p=.049)<br>• Mean (SD) scores: baseline=46.7 (24.0), 1m=53.3 (20.7), end=62.8 (19.4)<br><br>***LaFrance et al. 2010 (QoLIE-31)***<br>• No significant between arm difference in change scores (baseline-end of treatment), after adjusting for baseline scores (p>0.05)<br>• Mean (SD) scores by group:<br>○ Baseline: placebo=38.2 (19.0), treatment arm=48.4 (20.7)<br>○ End: placebo=46.9 (24.0), treatment arm=56.7 (25.1)<br><br>***LaFrance et al. 2014 (QoLIE-31)***<br>• Significant improvement in scores in CBT arm (mean difference=20.9, CI=10.4,31.4, ES(d)=1.8, p<0.001)<br>• No significant change in scores for:<br>○ CBT+sertraline: mean difference=11.0, CI = -3.6,25.5, ES(d)=0.6, p>0.05<br>○ Sertraline only: mean difference=7.3, CI = -0.5,15.1, ES(d)=0.7, p>0.05<br>○ TAU: mean difference=9.7, CI = -15.4,34.8, ES(d)=0.4, p>0.05<br><br>***Thompson et al. 2013 (QoLIE-31)***<br>• No significant between arm difference observed for: overall score, seizure worry, emotional well-being, energy/fatigue, cognitive, and social life at baseline or follow-up (descriptive data not provided) |

| | | |
|---|---|---|
| | | • At follow-up, the treatment arm (brief educational intervention) had higher overall quality of life scores (58) relative to controls (44), but not significantly (p=0.12)<br>• Treatment group had a reduced medication effect score over the control group (43 vs. 57), but not significantly (p=0.09)<br><br>***Tolchin et al. 2019 (QoLIE-10)***<br>Mean (SD) improvement scores: controls=1.8 (7.9), MI arm=7.2 (10.0) (p=0.047)<br>Cohen's *d*=0.60, CI=0.01, 1.12)<br><br>***Zaroff et al. 2004 (QoLIE-31)***<br>• A non-significant trend indicated improvement on total QoLIE-31 scores<br>• (p=0.07)<br>• QoLIE-31 Overall scores (mean (SD)): pre-treatment=58.27 (17.62), post-treatment=67.68 (24.32) |
| **Short Form Health Survey (SF-36 / SF-12)** | **Validity** | ***Conwill et al. 2014***<br>Convergent validity: No significant associations between SF-36 scores and the CGI-I (clinician) categorical variables of 'improved' or 'not improved'<br><br>***Kuyk et al. 2008***<br>Convergent validity: significantly greater improvements on SF-36 domains: mental health, energy-vitality and pain (p<0.05) for seizure free vs non seizure-free patients at 6m follow-up<br><br>***Williams et al. 2018***<br>Convergent validity: significant positive correlation between SF-36 mental health component and EPS-25 post-treatment change scores (*r*=0.634, p <0.008) |
| | **Reliability** | No data available |
| | **Responsiveness** | ***Aybek et al. 2013 (SF-36)***<br>• Joint consultation arm had superior SF-36 scores for several domains (relative to TAU): physical functioning (p=0.021), social functioning (p=0.033), general health (p=0.006), role limitation due to physical symptoms (p=0.017)<br>• No significant between arm differences for mental health: (p=0.09), role limitation due to emotional symptoms (p=0.9), energy/vitality (p=0.26) or pain (p=0.1)<br>• Mean (SD) SF-36 scores:<br>  ○ Physical functioning: joint=83 (28), TAU=57 (27), p=0.021<br>  ○ Role physical: joint=67 (39), TAU=28 (26), p=0.017<br>  ○ Social functioning: joint=74 (35), TAU=48 (25), p=0.033<br>  ○ Bodily pain: joint=72 (37), TAU=47 (36), p=0.10) |

- Mental health: joint=59 (23), TAU=44 (13), p=0.09
- Role emotional: joint=67 (45), TAU=64 (46), p =0.9
- Energy /Vitality: joint=46 (24), TAU=35 (18), p=0.26
- General health: joint=56 (27), TAU=25(16), p=0.006

***Conwill et al. 2014 (SF-36)***

- Total SF-36 scores improved non-significantly between start-end treatment
- Mean (SD) total SF-36 scores: start=37.4 (24.7), end=41.6 (20.9), p=0.22
- SF-36 domains showing significant improvements:
  - Emotional well-being (mean (SD)): start=43.1 (26.2), end=57.4 (26.2), p=0.04
  - Role limitation - emotional problems (mean (SD)): start=0.0 (75.0), end=4.76 (100.0), p=0.04
- Non-significant improvements for:
  - Physical functioning (mean (SD)): start=45.1 (36.5), end=45.4 (36.8), p=0.92
  - Bodily pain (mean (SD)): start= 40.3 (32.2), end=43.9 (29.6), p=0.44
  - Energy/fatigue (mean (SD)): start=25.5 (24.7), end=30.0 (21.8), p=0.21
  - Role limitation due to physical problems (mean (SD)): start=0.0 (75.0), end=4.76 (100.0), p=0.04
  - General health (mean (SD)): start=26.5 (50.0), end=27.5 (36.3), p=0.96
- Social functioning worsened but not significantly (mean (SD): start=47.3 (33.0), end=41.1 (32.0), p=0.13)

***De Barros et al. 2018 (SF-36)***

- Significant improvement in quality of life between pre- and post-treatment, in the following domains (mean (SD)):
  - functional capacity: pre-treatment=63 (24), post-treatment=73 (20), p=0.003, d=0.41
  - physical: pre-treatment=53 (33), post-treatment=67 (34), p=0.03, d=0.42
  - emotional: pre-treatment=49 (34), post-treatment=63 (30), p=0.02, d=0.41
- No significant differences for:
  - pain: pre-treatment=64 (27), post-treatment=69 (27), p=0.45, d=0.18
  - energy/vitality: pre-treatment=42 (12), post-treatment=41 (13), p=0.67, d=0.08
  - general well-being: pre-treatment=43 (14), post-treatment=40 (13), p=0.32, d=0.21
  - social: pre-treatment=70 (26), post-treatment=69 (24), p=0.65, d=0.03
  - mental health: pre-treatment=39 (15), post-treatment=37 (13), p=0.73, d=0.13

***Demartini et al. 2019b (SF-36)***

- Significant improvement over time (start treatment, 12, 24 wk) for several domains (mean (SD)):
  - general health: start=35.4 (29.3), 12wk=51.6 (21.4), 24wk=53.8 (25.6) (p=0.001)

- vitality: start=32.0 (30.4), 12wk=53.6 (27.1), 24wk=58 (28.1) (p=0.008)
- social functioning: start=33.7 (37.5), 12wk=56.5 (25), 24wk=60.8 (29.7) (p=0.003)
- mental health: start=33.9 (25.1), 12wk=58.3 (21.8), 24wk=66.4 (23.2) (p=0.001)

- No significant improvement for: physical function, physical role, pain, and emotional role (Ps>0.05)

***Dreissen et al. 2019 (SF-36)***

- No between arm differences in change scores on either physical or mental components
- Physical: median difference in change scores=0.1, CI = −4.2, 4.2, p=0.964
- Mental: median difference in change scores = −0.8, CI = −3.6, 5.4, p=0.768

***Hubschmid et al. 2015 (SF-36)***

- Treatment arm had superior scores for the mental health component relative to the TAU arm (p<0.05)
- Mean (SE) mental health component scores by group:
  - Baseline: TAU=64 (6.17), treatment=54.93 (6.3)
  - End of therapy (2m): 62.2 (9.48), treatment=61.71 (9.67)
  - 6m: TAU=63.11 (8.19), treatment=65.14 (8.67)
  - 12m: 61.71 (8.5), treatment=69.71 (7.75)
- No significant differences for: physical functioning, emotional and physical limitation, energy-vitality, social functioning, pain and general health domains (p>0.05)

***Jordbru et al. 2013 (SF-12)***

- SF-12 physical component: mean difference between treatment vs no treatment=11.7 (p<0.001, CI=7.2, 16.1)
- SF-12 mental component: mean difference between treatment vs no treatment=6.9 units (p<0.01, CI=2.1–11.8)
- SF-12 Physical (score range 0–100)
  - Intervention arm: baseline=25.7 (8.0), start treatment=37.2 (10.8), end treatment=35.5 (11.5), 1m=35.5 (11.5), 1y=28.6 (10.2)
  - Waiting list: baseline=28.3 (8.6), start treatment=27.3 (8.1), end treatment=36.6 (13.9), 1m=40.1 (14.2), 1y=44.5 (13.7)
- SF-12 Mental (score range 0–100)
  - Intervention arm: baseline=47.3 (14.3), start treatment=54.9 (9.0), end treatment=51.6 (10.7), 1m=51.6 (10.7), 1y=49.3 (13.6)
  - Waiting list: baseline=42.9 (12.9), start treatment=45.8 (13.5), end treatment=54.3 (10.4), 1m=54.8 (9.8), 1y=52.1 (9.1)

***Kuyk et al. 2008 (SF-36)***

- No significant improvement reported between pre- and post-treatment
- At 6m follow-up, improvement observed (relative to start of treatment) for 'role limitation due to emotional problems' domain (p=0.03)
- Mean SF-36 domain scores:
    - Physical functioning: start=78.2, end=80.8, 6m=76.3
    - Role limitation due to physical problems: start=47.2, end=66.7
    - Role limitation due to emotional problems: start=55.6, end=81.5, 6m=84.4
    - Social functioning: start=55.6, end=60.8, 6m=68.9
    - Mental health: start=62.5, end=69.3, 6m=74.1
    - Energy vitality: start=56.8, end=59.5, 6m=54.7
    - Pain: start=58.5, end=68.4, 6m=65.2
    - General health perception: start=54.9, end=65.2, 6m=60.
    - Change in health: start=53.9, end=63.2, 6m=65

***Mayor et al. 2013 (SF-36)***

- Median (IQR) scores:
    - Mental health component: baseline=28.6 (20.2-46), follow-up=44.6 (21-51.5)
    - Physical health component: baseline=38.6 (31.2-47.6), follow-up=44.8 (31.6-53.9)

***McWhirter et al. 2016 (SF-12)***

- At 3m, there were no significant differences on SF-12 physical component (p=0.17) or mental health component scores (p=0.91)
- Mean (SD scores):
    - Physical: pre-treatment=34.9 (11.4), 3m=29.2 (12.8)
    - Mental: pre-treatment=45.3 (11.2), 3m=44.8 (10.5)

***Metin et al. 2013 (SF-36)***

- Significant improvement in the mental health component (p=0.03), between start and end of treatment
- No significant change in the other subscales

***Nielsen et al. 2015 (SF-36)***

- Significant improvements on physical function and physical role domains
- Physical function (mean (SD)): start=26.1 (25.6), 3m=35.0 (27.1), p=0.001, r=0.32; Physical role (mean (SD)): start=12.8 (26.5), 3m=27.7 (31.4), p=0.001, r=0.31.

- No significant change was found for the other domains (p>0.05)

***Nielsen et al. 2017b (SF-36)***

- Physiotherapy arm had significantly higher scores on several domains at 6m (adjusting for baseline scores):
    - Physical function (mean (SD)): physiotherapy arm=51.9 (27.2), control arm=23.2 (21.3), coefficient=19.8, CI=10.2,29.5, p<0.001, Cohen's d = 0.70
    - Physical role (mean (SD)): physiotherapy arm=47 (30.3), control arm=26.8 (22.5), coefficient=13.0, CI=0.8, 25.2, p=0.037, Cohen's d=0.46
    - Social function (mean (SD)): physiotherapy arm=56.9 (30.2), control arm=37.0 (25.1), coefficient=17.1, CI=5.0, 29.2, p=0.007, Cohen's d=0.58
- No significant between arm differences for:
    - Bodily pain (mean (SD)): physiotherapy arm=47.4 (33.1), control arm=33.9 (27.4), coefficient=3.6, CI = −8.0, 15.3, Cohen's d = 0.12
    - General health (mean (SD)): physiotherapy arm=54.1 (28.3), control arm=39.6 (22.6), coefficient=9.0, CI = −0.1, 18.2, Cohen's d = 0.34
    - Vitality (mean (SD)): physiotherapy arm=39.2 (27.3), control arm=28.3 (20.2), coefficient= 6.2, CI = −3.6, 15.9, Cohen's d = 0.25
    - Role emotional (mean (SD)): physiotherapy arm=68.7 (34.5), control arm=62.5 (35.4), coefficient=0.1, CI = −15.1, 15.4, Cohen's d = 0.00
    - Mental health (mean (SD)): physiotherapy arm=67.9 (23.8), control arm=59.3 (25.2), coefficient=3.4, CI = −6.4, 13.2, Cohen's d = 0.14

***Pleizier et al. 2017 (SF-36)***

- Both treatment arms showed improvements on almost all SF-36 subscales
- No significant between arm differences observed in mean change scores, except for 'emotional role functioning' and 'mental component' scores
- When adjusting for baseline scores and stratification variables, there were no significant treatment effects on SF-36 scores at 12m
- Mean (SD) scores on SF-36 domains:
    - Physical functioning: neurologist baseline=53.78 (26.49), neurologist 12m=62.61 (30.41) neurologist change=8.83 (27.12); GP baseline=56.58 (25.78), GP 12m=64.42 (28.39), GP change=7.84 (22.01), treatment difference=0.99 (CI = −6.05, 8.03), p=0.99
    - Physical role functioning: neurologist baseline=19.33 (32.97), neurologist 12m=35.57 (41.90), neurologist change=16.24 (47.74); GP baseline=15.43 (28.89), GP 12m=37.23 (39.61), GP change=21.81 (44.49), treatment difference = −5.57 (CI = −18.75, 7.61), p=0.63

- Bodily pain: neurologist baseline=37.75 (27.05), neurologist 12m=49.55 (27.64), neurologist change=11.79 (24.20); GP baseline=40.74 (26.37), GP 12m=50.14 (25.40), GP change=9.39 (23.56), treatment difference=2.40 (CI = −4.42, 9.22), p=0.71
- Social functioning: neurologist baseline=45.36 (30.10), neurologist 12m=60.82 (26.18), neurologist change=15.46 (30.29); GP baseline=47.50 (30.43), GP 12m=60.53 (28.38), GP change=13.03 (32.10), treatment difference=2.44 (CI = −6.44, 11.32), p=0.76
- Mental health: neurologist baseline=62.27 (21.93), neurologist 12m=67.88 (19.74), neurologist change=5.61 (18.01); GP baseline=60.99 (22.54), GP 12m=68.67 (20.36), GP change=7.68 (18.31), treatment difference = −2.07 (CI = −7.27, 3.13), p=0.51
- Emotional role functioning: neurologist baseline=65.46 (44.50), neurologist 12m=63.23 (43.96), neurologist change = −2.23 (56.30); GP baseline=49.10 (44.93), GP 12m=66.31 (42.70), GP change=17.20 (43.58), treatment difference = −19.44 (−33.82, −5.06), p=0.16
- Vitality: neurologist baseline=42.42 (20.71), neurologist 12m=48.92 (21.06), neurologist change=6.49 (20.91); GP baseline=41.47 (20.92), GP 12m=48.92 (22.19), GP change=7.46 (21.02), treatment difference = −0.96 (CI = −6.96, 5.04), p=0.83
- General health: neurologist baseline=46.65 (20.65), neurologist 12m=50.33 (21.31), neurologist change=3.68 (18.85), GP baseline=47.47 (22.39), GP 12m=51.90 (24.17), GP change=4.43 (18.53), treatment difference = −0.75 (CI = −6.10, 4.60), p=0.68
- Physical component: neurologist baseline=32.78 (10.06), neurologist 12m=37.82 (12.42), neurologist change=5.04 (10.35); GP baseline=34.79 (10.69), baseline 12m=38.08 (11.93), GP change=3.29 (9.15), treatment difference=1.75 (CI = −1.08, 4.58), p=0.40
- Mental component: neurologist baseline=43.21 (13.21), neurologist 12m=44.71 (12.50), neurologist change=1.51 (13.16); GP baseline=40.26 (12.96), GP 12m=45.55 (11.30), GP change=5.29 (12.00), treatment difference = −3.79 (CI = −7.43, −0.15), p=0.17

***Reuber et al. 2007 (SF-36)***

- SF-36 scores significantly improved between pre- and post-treatment (p<0.001)
- SF-36 Completers: pre-treatment=77.0 (17.1), post-treatment=89.3 (25.0), 6m=86.5 (26.9)
- Significant improvements of mean scores were reported for:
    - General health pre-treatment=12.0 (3.6), post-treatment=13.2 (4.9), 6m=0.3 (p=0.024)
    - Vitality: pre-treatment=9.2 (3.7), post-treatment=11.9 (5.5), 6m=0.7 (p<0.001)
    - Social functioning: pre-treatment=5.3 (2.3), post-treatment=6.8 (3.0), 6m=0.7 (p<0.001)
    - Role limitation–emotional: pre-treatment=4.1 (1.3), post-treatment=4.5 (1.3), 6m=0.4 (p=0.013)

- Mental health: pre-treatment=17.3 (5.8), post-treatment=19.9 (6.3), 6m=0.5 (p=0.001)

- No significant improvements for:
  - Physical functioning: pre-treatment=18.2 (6.5), post-treatment=18.7 (6.2), 6m = −0.1 (ns)
  - Role limitation–physical: pre-treatment=5.0 (1.5), post-treatment=5.4 (1.7), 6m = −0.3 (ns)
  - Bodily pain: pre-treatment=5.9 (2.7), post-treatment=6.0 (2.8), 6m=0.04 (ns)

***Sharpe et al. 2011 (SF-12):***

- No significant between arm difference post-treatment (3m) for physical function scores (AMD=4, CI = -4, 12.0, p=0.347).
- Mean (SD) scores at 3m: GSH=60 (39), TAU=50 (40)
- Physical function scores superior in GSH group at 6m (AMD=11.05, CI = 3.03, 19.06, p=0.007).
- Mean (SD) scores at 6m: GSH=68 (36), TAU=50 (41) (represents a clinically significant difference)

***Taib et al. 2019 (SF-36)***

- Mean vitality domain scores improved significantly (1m) relative to baseline in both groups (p=0.02); however, this was not sustained at 6m and 12m
- Mean (SD) vitality scores:
  - Baseline: active=42.6 (17.8), control=41.8 (22.2)
  - 1m: active=54.6 (21.3), control=64.4 (25.7)
  - 6m: active=40.1 (14.8), control=40.1 (21.1)
  - 12m: active=48.4 (17.8), control=47.5 (19.5)
- Mean general health domain scores improved in both groups at 12m (p<0.01)
- Mean (SD) general health scores:
  - Baseline: active=47.3 (11.9), control=50.4 (19.3)
  - 1m: active=50.0 (12.6), control=58.3 (20.4)
  - 6m: active=53.1 (14.2), control=53.2 (24.2)
  - 12m: active=61.4 (16.3), control=61.5 (19.9)

***Williams et al. 2018 (SF-36)***

- SF-36 scores were significantly greater for both the mental and physical health components post-treatment (vs pre-treatment) (p=0.02, ES ($\eta 2p$)=0.14)
- Mean (SD) physical health scores: pre-treatment=36.24 (11.45), post-treatment=38.10 (11.95)

| | | | |
|---|---|---|---|
| | | | • Mean (SD) mental health scores: pre-treatment=40.10 (10.11), post-treatment=42.31 (11.12) |
| **General functioning** | **Work and Social Adjustment Scale (WSAS)** | **Validity** | No data available |
| | | **Reliability** | No data available |
| | | **Responsiveness** | ***Chen et al. 2014***<br>• No significant between arm difference in baseline WSAS scores (p=0.629)<br>• Mean (SEM) baseline scores: treatment arm=23.05 (1.54), control=24.17 (1.69)<br>• At 3m, treatment arm showed reduced scores (vs control arm)<br>• Mean (SEM) 3m scores: treatment arm=18.4 (1.91), control=25.52 (1.95), p=0.013<br>• At 6m, treatment arm continued to score lower than controls<br>• Mean (SEM) 6m scores: treatment arm=18.75 (1.85); control=24.86 (2.15), p=0.038<br><br>***Cope et al. 2017***<br>• No significant pre-post treatment difference in scores (p=0.105, ES(d)=0.30)<br>• Mean (SD) scores: pre-treatment=16.37 (11.40), post-treatment=13.00 (10.38)<br><br>***Goldstein et al. 2004***<br>• Significant reduction (improvement) in WSAS scores observed post-treatment, sustained to 6m (p<0.01)<br>• Mean (SD) scores: pre-treatment=18.88 (11.19), end=11.38 (11.16), 6m=11.13 (11.11)<br><br>***Goldstein et al. 2010***<br>• No significant group x time interaction (p=0.120); therefore, no benefit for treatment arm over TAU<br>• Mean (SD) scores:<br>  ○ Start of treatment: CBT=19.73 (8.07), TAU=22.62 (8.88)<br>  ○ End of treatment: CBT=12.97 (9.62), TAU=18.99 (10.75)<br>  ○ Follow-up: CBT=11.81 (11.05), TAU=19.44 (12.75)<br><br>***Mayor et al. 2013***<br>Median (IQR) scores: baseline=26 (7.5-34), follow-up=13 (16-32.5)<br><br>***Nielsen et al. 2015***<br>• WSAS scores improved significantly over time, start/end treatment and 3m follow-up (p<0.01)<br>• Differences significant from end-3m (p=.015) and from start-3m (p<0.001) |

| | | | |
|---|---|---|---|
| | | | • Mean (SD) scores: start=24.2 (8.0), end=23.0 (7.6), 3m=21.0 (7.2)<br><br>***Nielsen et al. 2017b***<br>• No significant between arms difference at 6m follow-up (adjusting for baseline differences)<br>• Mean (SD) scores: physiotherapy arm=20.2 (10.5), control arm=26.9 (10.2), coefficient = −4.2, CI = −8.4,0.1, Cohen's d = 0.39<br><br>***Wiseman et al. 2016***<br>• No significant change in WSAS scores between pre- and post-treatment<br>• Median (IQR): pre-treatment=26 (17.9), post-treatment=20.5 (14) p=0.122) |
| | **Global Assessment of Functioning (GAF)** | **Validity** | No data available |
| | | **Reliability** | No data available |
| | | **Responsiveness** | ***Hinson et al. 2006***<br>• Significant improvement in GAF scores observed (p=0.0083)<br>• Mean (SD) scores: pre-treatment=62.3 (6.1), post-treatment=69.4 (9.2)<br><br>***Kozlowska et al. 2018***<br>Mean GAF scores for PNES group=41 (median=43, range 11-65)<br>Mean GAF score increased from 41 to 67 (p<0.001) at 12m follow-up<br><br>***LaFrance et al. 2009***<br>Significant improvement in scores from baseline to end of treatment (p=0.0005)<br>Mean (SD) scores: baseline=50.1 (7.7), 1m=54.1 (11.7), end=59 (12.5)<br><br>***LaFrance et al. 2010***<br>• No significant between arm difference in change scores (baseline-end of treatment), after adjusting for baseline scores (p>0.05)<br>• Mean (SD) GAF scores by group:<br>  ○ Baseline: placebo=49.1 (7.1), treatment arm=53.3 (10.3)<br>  ○ End: placebo=52.0 (7.9), treatment arm=56.8 (11.0)<br><br>***LaFrance et al. 2014***<br>• Significant improvements in scores over time for:<br>  ○ CBT: mean difference = 19.9, CI=9.6,30.1, ES(d)=1.8, p<0.001 |

| | | | |
|---|---|---|---|
| | | | ○ CBT+sertraline: mean difference= 9.0, CI=3.1,14.9, ES(d)=1.2, p<0.01<br>• No significant change for:<br>○ Sertraline only: mean difference=7.0, CI = -1.0,15.0, ES(d)=0.7, p>0.05<br>○ TAU: mean difference=1.7, CI = -8.0,11.4, ES(d)=0.2, p>0.05 |
| **Health economics / cost utility** | **Quality adjusted life years (QALYs)** | **Validity** | No data available |
| | | **Reliability** | No data available |
| | | **Responsiveness** | ***Nielsen et al. 2015 (EQ-5D-5L)***<br>• Utility scores increased significantly between start treatment and 3m follow-up<br>• Mean utility scores: start=0.35 (CI=0.27, 0.43), 3m=0.47 (CI=0.39, 0.55)<br>• Gain=0.125 (CI=0.19, 0.06).<br><br>***Nielsen et al. 2017b (EQ-5D-5L)***<br>• Mean QALYs (6m): physiotherapy arm=0.34 (CI=0.31, 0.37), control arm=0.26 (CI=0.22, 0.30)<br>• Mean gain (QALYs/patient)=0.08 (CI=0.03, 0.13)<br>• Mean cost: physiotherapy=£1200/patient, control=£233/patient<br>• Mean cost/QALY=£12,087<br><br>***Reuber et al. 2007 (SF-36)***<br>• Mean SF-6D index: baseline=0.53, post-treatment=0.57<br>• Mean difference=0.04, t=−2.088, p=0.042).<br>• Mean cost/QALY=£5,328 |
| | **Healthcare resource use** | **Validity** | No data available |
| | | **Reliability** | No data available |
| | | **Responsiveness** | ***Aybek et al. 2013***<br>Significantly fewer patients from joint consultation arm sought further medical advice, relative to TAU (16 vs. 73%, p<0.01)<br><br>***Chen et al. 2014***<br>• No significant between arm difference during follow-up period in:<br>○ ED visits (p=0.184)<br>○ Initiation of counselling/psychotherapy (p=0.655)<br>○ Initiation of new psychotropic medications (p=0.523) |

***Drane et al. 2016***
No significant between arm difference in healthcare utilization for uncontrolled events at 8wk

***Goldstein et al. 2010***
- Number of ED visits:
  - tended to be lower in the CBT arm relative to controls at 6m (group x time interaction, Wald(1)=2.73, p=0.098)
  - significant reduction over time across groups (p=0.036)
  - no significant main effect of group (p=0.897)
- Number of ED visits by ambulance:
  - significantly declined over time (across groups) (p=0.001)
  - no significant main effect of group, and no group x time interaction
- No significant differences for: GP visits, prescribed medication, AEDs

***Hubschmid et al. 2015***
- Significantly fewer inpatient hospital days after treatment in the treatment arm vs TAU (p<0.05)
- Mean (SE) number of inpatient hospital days by group:
  - Baseline: TAU=8.6 (2.36), treatment=10.1 (2.85)
  - End of therapy (2m): TAU=13.4 (6.31), treatment=20.56 (8.05)
  - 6m: TAU=10.33 (10.09), 0 (0)
  - 12m: TAU=17.25 (11.23), 0 (0)
- No difference in overall physical outpatient visits or number of ED visits (ps>0.05)

***LaFrance et al. 2014***
Significantly fewer ED visits reported in CBT arm (during trial vs baseline) (p<0.001)

***Mayor et al. 2010***
- Significant reduction in health care utilization between baseline and follow-up
- Median health care utilization category changed from 3 (5–10 contacts in 3 months) to 2 (1–4 contacts) (p<0.039)
- Number of health care contacts (3 months) dropped by a mean of 1.7 between baseline (6.2) and follow-up (4.5)
- Minimum cost per health care contact=£36 ($60)
- Savings calculated for outpatient health care costs=£244.80 ($408)

***Mayor et al. 2013***

| | | | |
|---|---|---|---|
| | | | • Number of healthcare contacts (previous 3 months) (median (IQR)): baseline=7 (1.5-10), follow-up=6 (4-12)<br>• ED visits (previous 3 months) (% attended): baseline=35%, follow-up=10%<br>• Psychological treatment (previous 3 months) (% attended): baseline=0%, follow-up=15%<br><br>***Oto et al. 2010***<br>• Emergency healthcare service use (PNES related) reduced by 9m: 1/14 (0.07%) 0f immediate arm, vs 3/11 (27%) of the delayed arm (p=0.183)<br>• Use of rescue medication reduced by 9m: immediate arm=0/14 (0%), delayed=4/11 (36%) (p=0.026)<br>• Emergency healthcare service use (non-PNES related) reduced to 0% in both arms<br><br>***Thompson et al. 2013***<br>• At 6w, 9/9 (100%) of patients in the treatment arm sought contact with a mental health professional<br>• There was a significant between arm difference (p=0.003)<br><br>***Tolchin et al. 2019***<br>• Monthly ED visits:<br>  ○ Controls=increase of 0.06 (SD=0.47) visits/month<br>  ○ MI arm=decrease of 0.15 (SD=0.76)<br>• Difference not significant (p=0.23) |
| **Illness beliefs** | **Illness Perceptions Questionnaire (IPQ) / Brief IPQ / IPQ – Revised** | **Validity** | ***Demartini et al. 2014 (IPQ-R):***<br>Predictive validity: start of treatment IPQ-R scores did not significantly predict CGI-I (patient) outcome (p=0.77) and only 3.8 % of the variance in outcome scores could be explained (Nagelkerke R2) |
| | | **Reliability** | No data available |
| | | **Responsiveness** | ***Cope et al. 2017 (B-IPQ)***<br>• Significant pre-post treatment reduction (improvement) in B-IBQ scores (p<0.001, d=0.54)<br>• Mean (SD) scores: pre-treatment=6.20 (2.23), post-treatment=5.00 (2.13)<br>• Significant pre-post treatment differences for specific items (p<0.026):<br>  ○ Beliefs about permanence of illness<br>  ○ Level of concern about illness<br>  ○ Understanding of illness |

***DeMartini et al. 2014 (IPQ-R):***

- Significant change between start/end treatment on several subscales:
  - timeline acute/chronic (p<0.001, Cohen's d=0.51)
  - illness coherence (p<0.001, d=0.84)
  - emotional representations (p=0.009, d=0.32)
  - consequences (p=0.02, d=0.26)
- Mean (SD) scores:
  - timeline acute/chronic: start=18.4 (4.4), end=16.2 (4.4)
  - illness coherence: start=13.5 (5), end=17.5 (4.6)
  - emotional representations: start=20.5 (5.6), end=18.7 (5.5)
  - consequences: start=22.7 (4.9), end=21.5 (4.4)

***Goldstein et al. 2004 (IPQ)***

- Significant post-treatment decrease in:
  - beliefs that physical factors were important causes of the illness (p<.05)
  - self-perceived negative consequences of illness, which was maintained at follow-up (p<.01)
- Significant post-treatment increase in:
  - beliefs in the possibility of a control/cure, maintained at follow-up (p<.01).
- No significant post-treatment change in:
  - attributions to emotional causes
  - beliefs about permanence of illness ('timeline')
- Mean (SD) scores:
  - Physical causes: pre-treatment=1.91 (0.58), end=1.56 (0.49), 6m=1.55 (0.56)
  - Emotional causes: pre-treatment=2.91 (0.91), end=3.23 (0.99), 6m=3.01 (0.96)
  - Timeline: pre-treatment=2.77 (0.45), end=2.77 (0.51), 6m=2.54 (0.72)
  - Consequences: pre-treatment=3.71 (0.59), end=3.32 (0.59), 6m=3.40 (0.60)
  - Control/Cure: pre-treatment=3.19 (0.52), end=3.86 (0.74), 6m=3.67 (0.83)

***Nielsen et al. 2017b (B-IPQ)***

- Physiotherapy group had lower composite B-IPQ scores (fewer maladaptive illness beliefs), relative to controls, at 6m (adjusting for baseline scores)
- Mean (SD) scores: physiotherapy arm=39.4 (16.1), control arm=51.0 (13.0), coefficient = −8.0, CI = −14.4, −1.6, p=0.015, Cohen's d = 0.51

***Oto et al. 2010 (IPQ)***

- Number of reported symptoms reduced in both arms (immediate/delayed AED withdrawal)
- Mean (SD) number of symptoms:

| | | |
|---|---|---|
| | | ○ Baseline: immediate=8.3 (2.2), delayed=7.3 (2.5)<br>○ 9m: immediate=7.4 (4.3), delayed=8.0 (2.4)<br>○ 18m: immediate=6.6 (4.2), delayed=5.2 (3.5)<br>• At 18m, more patients attributed the disorder to psychological causes (p=0.005)<br>• Mean (SD) scores – attribution to psychological causes:<br>○ Baseline: immediate=8.0 (3.4), delayed=8.0 (2.6)<br>○ 9m: immediate=10.0 (2.4), delayed=9.0 (3.4)<br>○ 18m=11.2 (1.8), delayed=8.3 (2.8)<br><br>***Sharpe et al. 2011 (IPQ):***<br>• Significant between arm differences for health anxiety items post-treatment (3m), both significantly lower in GSH arm vs TAU, but not at 6m (p>0.05)<br>• Concern that there is something seriously wrong with your body (3m) (OR=3.3, CI=1.3, 8.6, p=0.012)<br>• Worry a lot about your health (3m) (OR=3.4, CI=1.4, 8.6, p=0.009).<br>• No significant between arm differences for beliefs about health (symptoms are permanent / symptoms are a mystery) at 3m (p=0.231, p=0.059 respectively)<br>• At 6m, beliefs in symptom permanence reduced in GSH group relative to TAU (OR = -0.5, CI = -0.9, -0.1, p=0.024)<br><br>***Williams et al. 2018 (B-IPQ)***<br>• Significant improvement in B-IPQ scores (p=0.01, CI=2.57, 12.39, critical *p*-value=0.016)<br>• Mean (SD) scores: pre-treatment=55.51 (11.84), post-treatment=48.83 (15.79) |
| **Study specific questions / data** | **Validity** | ***Duncan et al. 2016 (data taken from psychologists' records/assessments)***<br>• Belief that the seizures could be helped (OR=3.86, CI=1.60, 9.35), p=0.003) and the patient feeling as though he/she had some control over his/her seizures (OR=3.30, CI=1.20, 9.09, p=0.021), and a (psychologist-rated) internal locus of control (OR=7.46, CI=2.62, 21.28), p<0.001) predicted seizure freedom (for 2 months) at 6m<br>• The best predictor was (psychologist-rated) presence of internal locus of control, which predicted outcome in 70.5% of cases |
| | **Reliability** | N/A |
| | **Responsiveness** | ***Chen et al. 2014***<br>• No significant differences in illness beliefs at baseline (p>0.09)<br>• Treatment completers were more likely to agree with the statements below:<br>○ 'My attacks do not bother me as much anymore'(3m/6m, p<0.001)<br>○ 'I am able to carry on with most daily activities despite my attacks' (3m, p=0.037, 6m p=0.021) |

- ‘I am able to avoid triggers to my attacks’ (3m, p=0.016, not sustained at 6m)
- ‘I have some control over my attacks’ (3m, p=0.006, not sustained at 6m)

***Cope et al. 2017***

- A significant improvement in perceived understanding reported post-treatment (p=0.004)
- Patients also less likely to disagree with seizures ‘not affecting my life’ or ‘not
- scaring me’, but differences non-significant (p>0.05)
- N (%) of participants agreeing with the following:
  - ‘Accept PNES diagnosis’: pre-treatment=7 (38.9), post-treatment=11 (61.1), p=0.180
  - ‘Understand my PNES’: pre-treatment=6 (31.6), post-treatment=14 (73.7), p=0.004
  - ‘Despite my PNES, I am able to carry on most of my essential daily activities: pre-treatment=6 (33.3), post-treatment=8 (44.4), p=0.250
- N (%) of participants disagreeing with the following:
  - ‘PNES does not affect my life’: pre-treatment=12 (70.6), post-treatment=7 (41.2), p=0.070
  - ‘PNES does not scare me’: pre-treatment=12 (63.2), post-treatment=8 (42.1), p=0.125

***Drane et al. 2016***

- Number of patients agreeing with the statement: ‘I understand the diagnosis’:
  - Discharge: TAU=6, Structured feedback=7, Structured ongoing feedback=9
  - 8 weeks: TAU=8, Structured feedback=8, Structured ongoing feedback=12

**Key:** AED=antiepileptic drugs; AMD=adjusted mean difference; BAI=Beck Anxiety Inventory; BDI=Beck Depression Inventory; BoNT=botulinum toxin; CBT=cognitive behavioural therapy; CGI=Clinical Global Impression; CI=confidence interval; ED=emergency department; EPS-25=Emotional Processing Scale-25; ES=effect size; FMS=functional motor symptoms; FND=functional neurological disorder; GAF=gloval assessment of functioning; GP=general practitioner; GSH=guided self-help; HADS=Hospital Anxietty & Depression Scale; HAM-A=Hamilton rating Scale for Anxiety; HAM-D=Hamilton Rating Scale for Depression; HC=healthy controls; HoNoS=Health of the Nation Outcomes Scale; ICC=intraclass correlation; IQR=interquartile range; IPQ=Illness Perceptions Questionnaire; IRR=inter-rater reliability; ITT=intention-to-treat; m=month; MI=motivational interviewing; ns=non-significant; OR=odds ratio; PHQ=Patient Health Questionnaire; PI=paradoxical intention; PMD=psychogenic movement disorder; PMDRS=Psychogenic Movement Disorder Rating Scale; QALY=quality adjusted life years; QoLIE=Quality of Life in Epilepsy; SCL=Symptom Checklist; SD=standard deviation; SEM=standard error of mean; SF=Short Form Health Survey; SMD=standardised mean difference; STAI=State Trait Anxiety Inventory; TAS=Toronto Alexithymia Scale; TAU=treatment as usual; TMS=transcranial magnetic stimulation; wk=week; WSAS=Work & Social Adjustment Scale; y=year