Aurora Health Care®

# Feasibility and Validity of Asking Patients to Define Individual Levels of Meaningful Change on Patient-Reported Outcomes

Salene M.W. Jones
Yuxian Du
Ari Bell-Brown
Kaylin Bolt
Joseph M. Unger

@JPCRR

# Feasibility and Validity of Asking Patients to Define Individual Levels of Meaningful Change on Patient-Reported Outcomes

Salene M.W. Jones, PhD,[1] Yuxian Du, PhD,[1,2] Ari Bell-Brown, MPH,[1] Kaylin Bolt, MPH, MEd, MSW,[3] Joseph M. Unger, PhD[1]

[1]Fred Hutchinson Cancer Research Center, Seattle, WA; [2]Data Generation and Observational Studies, Bayer HealthCare LLC, Whippany, NJ; [3]independent scholar

| | |
|---|---|
| **Purpose** | Patient-reported outcomes (PROs) are frequently used in clinical care to monitor treatment response. However, most guidelines on PRO use treat all patients the same. This study tested the feasibility and validity of a method for determining individually meaningful change in PRO measures. |
| **Methods** | Participants (n=398) completed 12 pain and distress questions to define individually meaningful change. This mixed-methods study used both quantitative and qualitative analyses, including descriptive statistics, inferential statistics, and content analysis. |
| **Results** | Two-thirds (67%) of the sample reported at least one medical condition, including depression and back pain. Most participants (70%–90%) were able to answer the questions as intended. Participants varied widely in the amount of change they considered meaningful (coefficients of variation: 40%–99%). Higher symptom levels were associated with larger amounts of change considered meaningful and with greater likelihood of answering questions as intended. Participants reported a variety of reasons for why they considered an amount of change in pain or distress meaningful. The hypothetical nature of the questions and the need to reference previous questions was found to be confusing. |
| **Conclusions** | Asking patients to define an individual level for meaningful change on PROs was feasible and valid. Having patients define their own goals on PROs for treatment of pain or distress could make treatment more patient-centered. (*J Patient Cent Res Rev.* 2020;7:239-248.) |
| **Keywords** | patient-reported outcomes; minimal clinically important difference; treatment response; mixed methods; qualitative; quantitative; pain; distress |

Patient-reported outcomes (PROs) are questionnaire-based measures of a patient's perception of their current health state.[1] PROs are frequently used in measurement-based care (MBC), an approach in which patients are monitored to ensure treatments are working or if treatment needs to be changed.[2-5] Determining the amount of change needed to classify a treatment as working or not working in MBC has been challenging. Previous work has focused on defining the minimally important difference (MID), which is the smallest amount of change between two groups necessary to consider a treatment successful.[6-9] Other work has focused on defining statistically significant or noticeable change at the individual, rather than group, level.[10,11] However, these approaches differ from the original definition of the MID, which is the smallest amount of change that is meaningful to the individual patient and would warrant a change in treatment.[12]

The need for some measure of what is meaningful to each individual patient, rather than changes between groups or statistical significance, has been acknowledged.[13-18] Personalizing the items on PROs makes them more sensitive to treatment effects,[19] suggesting that personalizing the amount of change

Correspondence: Salene M.W. Jones, PhD,
Fred Hutchinson Cancer Research Center, 1100 Fairview
Ave. N., Seattle, WA 98109 (smjones3@fredhutch.org)

considered meaningful could also improve treatment and MBC. Patients differ from each other in what they consider a meaningful change on a PRO,[20] yet the only current options for determining whether patients improved a meaningful amount use group-based MID to assess individual patients (eg, a 5-point change is considered meaningful for all) or to look for statistically significant change that might not be meaningful. Recent methods of determining a group-based MID have involved asking patients and providers to review case reports and judge whether one is worse than the other or, for patients, whether the case had worse symptoms than their own,[20,21] showing that patients can report directly what they consider meaningful. However, these methods were cognitively burdensome and did not reference changes in the patient's own health state. Rather than creating an omnibus MID that is applied to all patients equally, methods for determining what is individually meaningful to each patient are needed.

This study tested the feasibility of a new method of determining individually meaningful change on PROs using a prospective approach. We chose a prospective approach, as previous attempts at retrospective approaches have had problems.[22,23] Previous attempts to create measures of PRO change that is meaningful to the individual patient have involved significant clinician effort and input.[24,25] Our proposed method, simply asking patients to define for themselves what is meaningful, would be less burdensome and, surprisingly, has not been tested yet. If proved feasible and valid, it could help make MBC more patient-centered by using the amount of change meaningful to the individual patient as the marker of whether a treatment is working or not.

We examined feasibility both in a general sample and also in people reporting pain, distress, back pain, or depression. As participants with higher symptom levels have more room to change, we examined validity of the new method through association of the amount of change considered meaningful with current symptom level.

## METHODS
### Participants and Procedures
Participants were recruited from Prolific Academic, a crowdsourcing website that helps connect potential study participants with online research studies. Participants had to be 18 years of age or older and able to read and write English. Once the study was posted on Prolific Academic, potential participants either checked the website periodically or received notification that a study was available. They then went to the website and completed the survey. Participants received $8.00 after survey completion. All participants provided informed consent, and the study was reviewed and approved by the institutional review board (#8703).

### Measure
The numerical rating scale (0=no pain or distress, 10=worst pain or distress) was used because of its ubiquity and likely familiarity to all patients.[26] Patients were first asked to rate their current levels of pain or distress. A set of 5 questions (Table 1) was drafted to ask participants about the amount of change considered meaningful, either in general or in response to treatment. Each set of 5 questions was asked per construct (pain and distress), along with the numerical rating scale for that construct, resulting in a total of 12 questions. The research team, colleagues, and a paid patient consultant reviewed the questions and made revisions. After each of the 10 drafted questions, participants were asked to provide a brief explanation of the reasons for each answer.

### Quantitative Analyses
We examined two quantitative outcomes per question: paradoxical vs logical responding; and amount of change. Paradoxical vs logical responding was considered a measure of feasibility as it showed how many participants could answer the questions as intended. Individual meaningful change for pain or distress worsening was defined as the difference between current pain or distress and the level indicating worsening. For pain or distress worsening, an answer was defined as paradoxical if the person marked a level of pain or distress at or below their current level (ie, the difference was negative or zero). Individual meaningful change for pain or distress improving was defined as the difference between current pain or distress and the level indicating improvement. For pain or distress improving, an answer was defined as paradoxical if the person marked a level of pain or distress at or above their current level (ie, the difference was negative or zero).

We also calculated the difference in the level at which one would want treatment and the levels at which one would consider a treatment working or successful. For

**Table 1.** Survey Questions for Determining Individually Meaningful Change*

| |
|---|
| Q1. In the past 7 days, how would you rate your [pain/distress] on average? |
| Q2. Given your current [pain/distress] level, what level on this question would mean your pain was getting worse? (Pain/Distress Worsen) |
| Q3. Given your current [pain/distress] level, what level on this question would mean your pain was getting better? (Pain/Distress Improve) |
| Q4. On this 0 to 10 scale, at what level would you want treatment for your [pain/distress]? |
| If you were experiencing a high level of pain and received treatment for [pain/distress] … |
| Q5. … what level on this question would mean this treatment was working? (Pain/Distress Treatment Work) |
| Q6. … what level would your [pain/distress] have to be to consider the treatment a success? (Pain/Distress Treatment Success) |

*Questions 2 and 3 were used in conjunction with Question 1 to determine meaningful level of change. Questions 4 and 5 were used in conjunction with Question 6 to determine a different metric of meaningful change.*

a treatment working or being a success, an answer was defined as paradoxical if the person marked an answer at or above the level they marked for wanting treatment (ie, treatment is working if pain or distress had increased). For amount of change, we examined the difference between their numerical rating scale answer for current pain or distress and the amount of change for worsening or improving symptoms. We also examined the difference between the level at which one wanted treatment and the change constituting treatment working or successful.

To examine feasibility, we calculated the frequency of paradoxical responding by each question and for the following subgroups: participants reporting back pain; participants reporting depression; participants experiencing pain in the past 7 days; and participants experiencing distress in the past 7 days. Feasibility was determined by the proportion of respondents answering questions as intended. To assess validity, we also used logistic and linear regressions to test the association of current level of pain/distress to the level at which one would want treatment, with paradoxical responding and amount-of-change considered meaningful.

Analyses controlled for demographic variables, and demographics were included based on significant associations with the outcomes. First, each demographic variable was tested for a univariate association with the 8 paradoxical responding outcomes and 8 amount-of-change outcomes. Demographic variables with P-values of <0.25 in the univariate analyses were then entered into a multivariate regression. Lastly, the final multivariate regression was run using only significant variables from the previous multivariate regression. The following demographic variables were tested: age, gender, income, race/ethnicity, education, living in the United States, hypertension diagnosis, depression diagnosis, and back pain diagnosis.

## Qualitative Analyses

Responses to the questions asking for the reasons were coded by members of the research team using qualitative content analysis. The codebook was developed inductively; the first coder read all the statements of all the surveys and developed a preliminary codebook based on the codes that emerged. The first coder then trained the second coder on the codebook and the two independently applied the codes to the first 25 records. The two coders discussed major discrepancies and applied agreed-upon revisions to the codebook. They then independently coded the next 75 records (records 26 to 100) and again discussed major discrepancies, resulting in a further refined and consolidated codebook. The first 25 and next 75 records were chosen because this allowed enough variability to inform changes to the codebook but also enough additional records to test the revisions. The two coders then independently coded the remaining 298 records.

A third coder used the final codebook to reconcile the two independent sets of codes for all 398 records.

## RESULTS

We originally recruited 400 participants, but 2 did not successfully answer 3 or more of the 4 attention check questions, resulting in a final sample of 398. Consistent with previous studies using Prolific samples, the study

sample was predominantly college-educated and from Europe (Table 2). Two-thirds (67%) of the sample reported at least one medical condition, including depression, high blood pressure, back pain, and arthritis. Means for what participants considered meaningful are reported in Table 3. The standard deviations for meaningful PRO change were relatively large (coefficients of variation ranging from 40% to 99% in the total sample), indicating participants had differing views of what constituted a meaningful change.

## Paradoxical Responding

The frequency of paradoxical responding to the individually meaningful change questions is reported in Table 3. The percentage of participants reporting a paradoxical response (answering the questions not as intended), including missing responses, ranged from 10% to 30%. Questions about individually meaningful change for pain or distress worsening or what level would mean the treatment was working tended to have higher levels of paradoxical responding. Most participants (76%) were able to answer 6 to 8 of the questions as intended. Rates of paradoxical responding were consistent across patient groups.

Demographic variables associated with nonparadoxical responding are reported in Table 4. People with higher education were more likely to answer questions as intended when asked about what constituted a worsening of symptoms than people with a high school diploma or less. As was expected, people reporting more pain or distress were more likely to answer questions as intended compared to people reporting less pain or distress. The two exceptions were: higher initial pain level was found to be associated with more paradoxical responding for pain treatment working; and higher initial pain levels was associated with higher likelihood of paradoxical response on treatment being a success.

## Level of Individually Meaningful Change

Results of the regressions comparing demographics, symptom levels, and levels at which one wanted treatment are reported in Table 5. Higher current pain or distress levels were associated with larger amounts of improvement (higher pain level, larger decrease in pain to be considered an improvement) but smaller amounts for worsening of pain or distress (higher pain level, smaller increase in pain to be considered worsening). People with a higher level of pain or distress at which they wanted treatment tended to want larger decreases in pain or distress to consider treatment working or successful. Few demographic variables were associated with level of change considered meaningful, but being non-Hispanic was associated with considering smaller changes in distress meaningful compared to Hispanic participants.

## Qualitative Results

Results from the qualitative analysis of reasons participants considered certain amounts meaningful showed 4 general categories of reasons: amount; context outside of pain or distress; context specific to pain or distress; and confusion about answering the questions. Each theme had 4 to 9 subthemes (child codes). Two of the subthemes from context outside of pain or distress were further divided into 3 or 4 grandchild codes (see

**Table 2.** Demographics of the Study Sample (N=398)

| Characteristic | n (%) or mean (SD) |
|---|---|
| Age in years, mean (SD) | 33.10 (12.07) |
| Gender, n (%) | |
| Female | 234 (58.8%) |
| Other gender | 3 (0.8%) |
| Declined to answer | 3 (0.8%) |
| Male | 158 (39.7%) |
| Race/Ethnicity,* n (%) | |
| White | 372 (93.5%) |
| Hispanic | 12 (3.0%) |
| Asian | 19 (4.8%) |
| Black | 18 (4.5%) |
| Multiracial | 16 (4.0%) |
| Yearly income in U.S. dollars, mean (SD) | $31,689 ($34,654) |
| Education, n (%) | |
| High school diploma, GED or lower | 88 (22.1%) |
| Some college or associate's degree | 141 (35.4%) |
| Bachelor's degree | 116 (29.1%) |
| Graduate degree | 53 (13.3%) |
| Marital status, n (%) | |
| Married | 152 (38.2%) |
| Long-term relationship | 125 (31.4%) |
| Single | 121 (30.4%) |
| Living in United States, n (%) | 70 (17.6%) |

*Participants were able to select more than one race/ethnicity option, including multiracial.*

*SD, standard deviation.*

**Table 3.** Distribution of Meaningful Change and Paradoxical Responding* for the Total Sample and Subsamples With Back Pain, Any Current Pain, Depression, and Any Current Distress

| | Total sample (N=398) | | | People with back pain (n=153) | | | People with pain in past 7 days (n=330) | | | People with depression (n=178) | | | People with distress in past 7 days (n=329) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean (SD) | CoV | Para | Mean (SD) | CoV | Para | Mean (SD) | CoV | Para | Mean (SD) | CoV | Para | Mean (SD) | CoV | Para |
| **Pain level** | | | | | | | | | | | | | | | |
| Improvement | 1.42 (1.40) | 99% | 14% | 2.05 (1.48) | 72% | 11% | 1.84 (1.33) | 72% | 13% | 1.90 (1.40) | 74% | 14% | 1.82 (1.30) | 71% | 16% |
| Worsening | 1.39 (1.31) | 94% | 27% | 1.73 (1.01) | 58% | 13% | 1.77 (0.97) | 55% | 19% | 1.80 (1.07) | 59% | 14% | 1.82 (1.11) | 61% | 17% |
| Treatment working | 2.25 (2.05) | 91% | 25% | 3.31 (1.88) | 57% | 24% | 3.17 (1.81) | 57% | 26% | 3.24 (1.66) | 51% | 28% | 3.18 (1.83) | 58% | 30% |
| Treatment successful | 3.23 (2.49) | 77% | 16% | 4.27 (2.28) | 53% | 16% | 4.18 (2.09) | 50% | 18% | 4.15 (1.99) | 48% | 21% | 4.11 (2.08) | 51% | 21% |
| Level want treatment for pain | 5.13 (2.70) | 53% | --- | 5.82 (2.52) | 43% | --- | 5.48 (2.52) | 46% | --- | 5.51 (2.59) | 47% | --- | 5.32 (2.65) | 50% | --- |
| **Distress level** | | | | | | | | | | | | | | | |
| Improvement | 1.64 (1.57) | 96% | 12% | 2.32 (1.57) | 68% | 11% | 2.13 (1.44) | 68% | 11% | 2.30 (1.59) | 69% | 10% | 2.13 (1.46) | 69% | 8% |
| Worsening | 1.60 (1.28) | 80% | 20% | 1.93 (1.18) | 61% | 15% | 1.93 (1.16) | 60% | 16% | 1.92 (1.16) | 60% | 10% | 1.84 (1.00) | 54% | 14% |
| Treatment working | 2.78 (2.21) | 79% | 25% | 3.48 (1.97) | 57% | 27% | 3.57 (1.93) | 54% | 25% | 3.43 (1.81) | 53% | 23% | 3.59 (1.85) | 52% | 23% |
| Treatment successful | 3.91 (2.73) | 70% | 21% | 4.85 (2.41) | 50% | 23% | 4.82 (2.18) | 45% | 24% | 4.88 (2.13) | 44% | 21% | 4.88 (2.21) | 45% | 21% |
| Level want treatment for distress | 5.90 (2.87) | 49% | --- | 6.12 (2.87) | 47% | --- | 5.96 (2.85) | 48% | --- | 6.48 (2.58) | 40% | --- | 6.21 (2.72) | 44% | --- |

*The percentage of respondents with a paradoxical response included those who did not answer.

CoV, coefficient of variation; Para, paradoxical response; SD, standard deviation.

Table 6 for example quotes and number of participants reporting each category).

***Amount:*** Most participants reported that some change was meaningful to them and should be taken seriously even if the change was small. Other participants cited only a large change would be meaningful either because their pain or distress fluctuated or because side effects of treatment could only be justified by a large change. For improvements and considering a treatment successful or working, some participants specifically said they wanted no pain or distress. Others cited a set level of pain or distress as constituting worsening or improvement, including that some pain or distress may be inevitable given their situation.

***Context Outside of Pain or Distress:*** Participants mentioned activities in their daily lives, such as exercising too much or losing a job, as a reference for how they answered at the time of the survey. Others cited specific medical conditions like migraines or anxiety. Participants also mentioned specific treatments they would consider, had used previously, or were already taking. Some comments about treatment were more general and did not cite a specific treatment. Several participants stated they did not want treatment, usually in response to the treatment-specific personal MID questions. Reasons for not wanting treatment included a desire to manage their pain or distress on their own, such as not wanting to take available resources from others or thinking doctors could not help. Other reasons for not wanting treatment were drawbacks of treatment, a general belief that treatments don't work, and a belief that treatments do not address the root cause of pain or distress. When discussing reasons for not wanting treatment, participants referenced specific treatments like medication.

***Context Specific to Pain or Distress:*** Several participants referenced their

**Table 4.** Logistic Regression Results for Answering the Question as Intended*

| | Pain improve | Pain worsen | Distress improve | Distress worsen | Pain treatment work | Pain treatment success | Distress treatment work | Distress treatment success |
|---|---|---|---|---|---|---|---|---|
| Long-term relation (not married) vs never married | 2.513‡ (0.968) | | | | | | | |
| Married vs never married | 0.900 (0.336) | | | | | | | |
| Previously married, now single vs never married | 0.541 (0.355) | | | | | | | |
| Declined marital status | 0.767 (1.447) | | | | | | | |
| Some college vs ≤high school | | 1.867§ (0.674) | 1.369 (0.514) | | | | | |
| Bachelor vs ≤high school | | 1.763 (0.642) | 2.701‡ (1.161) | | | | | |
| Master/doctorate vs ≤high school | | 2.897‡ (1.463) | 8.575† (6.757) | | | | | |
| Initial pain | 1.905† (0.200) | | | | 0.751† (0.062) | 0.787† (0.068) | | |
| Pain level when treatment wanted | 1.222† (0.072) | 1.195† (0.062) | | | 2.327† (0.214) | 2.243† (0.200) | | |
| Initial distress | | | 2.361† (0.264) | | | | | |
| Distress level when treatment wanted | | 1.211† (0.059) | 1.174† (0.072) | 1.303† (0.066) | | | 1.916† (0.147) | 1.816† (0.127) |
| Constant | 0.272† (0.099) | 0.275† (0.113) | 0.161† (0.066) | 0.705 (0.276) | 0.134† (0.052) | 0.269† (0.088) | 0.133† (0.054) | 0.219† (0.077) |
| Observations | n=355 | n=338 | n=344 | n=348 | n=348 | n=366 | n=333 | n=343 |

*Odds ratios are reported with standard errors in parentheses.

†$P<0.01$; ‡$P<0.05$; §$P<0.1$.

previous experiences with pain or distress, such as a previous medical diagnosis or injury. In some cases, participants cited pain or distress as variable and that this was why they considered only larger changes meaningful or potentially longer lasting. Similarly, other participants cited pain or distress as stable and that they considered even small changes meaningful. Other participants referenced the type or location of their pain, such as a sprained ankle. Some participants used the frequency of the pain or distress as a benchmark for what they considered meaningful improvement, worsening, or needing treatment. Other participants used whether pain or distress was noticeable or interfering with their function to determine meaningful change. Innate tolerance for pain or distress was also cited. Some participants noted that their pain or distress was already at the top or bottom of the scale.

*Confusion:* The qualitative analysis revealed several ways the participants found the questions confusing. Some did not understand the question asking for their reasons, even if they may have understood the personal MID questions. A few gave contradictory reasons. Others specifically contradicted the level of pain or distress they had reported, such as saying they had no pain or distress but marking a 4. Several participants did not understand the hypothetical, prospective nature of the questions. A few participants reversed the scales, assuming 10 was no pain or distress. A few participants also misinterpreted the questions as asking about

**Table 5.** Linear Regression Results for Magnitude of Personal Minimally Important Difference*

| | Pain improve | Pain worsen | Distress improve | Distress worsen | Pain treatment work | Pain treatment success | Distress treatment work | Distress treatment success |
|---|---|---|---|---|---|---|---|---|
| Income | | | | | | | | |
| $100~$999 vs $0~$99 | | | | | | -0.799 (0,701) | | |
| $1k~$9999 vs $0~$99 | | | | | | -0.457 (0.371) | | |
| $10k~$19,999 vs $0~$99 | | | | | | -0.351 (0.323) | | |
| $20k~$49,999 vs $0~$99 | | | | | | -0.868† (0.294) | | |
| $50k~$99,999 vs $0~$99 | | | | | | -0.833‡ (0.348) | | |
| $100k~$199,999 vs $0~$99 | | | | | | -0.652 (0.515) | | |
| ≥$200k vs $0~$99 | | | | | | -0.919 (1.064) | | |
| Non-Hispanic vs Hispanic | | | -1.172† (0.410) | | | | -1.699† (0.536) | -1.190‡ (0.548) |
| Declined ethnicity | | | 2.217‡ (0.938) | | | | 0.481 (1.581) | 0.252 (1.620) |
| Initial pain (0=none, 10=most) | 0.342† (0.032) | -0.065‡ (0.031) | | | | | | |
| Pain level when treatment wanted | | | | | 0.588† (0.050) | 0.733† (0.046) | | |
| Initial distress (0=none, 10=most) | | | 0.303† (0.031) | -0.118† (0.025) | | | -0.107† (0.039) | -0.155† (0.039) |
| Distress level (0=none, 10=most) when treatment wanted | | | | | | | 0.622† (0.052) | 0.810† (0.049) |
| Male vs female | | | | | | | | -0.488‡ (0.193) |
| Other gender (≤10) | | | | | | | | 0.459 (0.887) |
| Declined gender (≤10) | | | | | | | | 0.786 (1.527) |
| Constant | 0.628† (0.132) | 2.036† (0.117) | 1.858† (0.444) | 2.360† (0.110) | -0.578§ (0.322) | 0.252 (0.384) | 1.269‡ (0.639) | 1.231§ (0.661) |
| Observations | n=280 | n=288 | n=287 | n=313 | n=253 | n=283 | n=262 | n=277 |

*Unstandardized coefficients are reported with standard errors in parentheses. People with paradoxical values were excluded from these regressions.

†P<0.01; ‡P<0.05; §P<0.1.

amount of change not level of pain or distress. Several participants had difficulty connecting the MID questions to their current pain or distress level or the level at which they would want treatment. In some cases, participants connected the level for treatment working or success to their current level of distress or pain instead of the level at which they would want treatment.

## DISCUSSION

This study tested the feasibility and validity of asking people to define what they considered meaningful change on pain and distress numerical rating scales. Overall, this simple method appeared feasible, as most respondents answered all or nearly all the questions as intended. Higher levels of current pain or distress

**Table 6.** Qualitative Codes for Reasons Participants Considered Specific Amounts of Change Meaningful

| Parent | Child | Grandchild |
|---|---|---|
| Amount | Some change (292); *"Any reduction would be positive"* | |
| | Large amount (191); *"A change of 2 could be within normal ups and downs. A change of 3 would make me worry"* | |
| | No distress or pain (221); *"If the treatment is successful, I should be experiencing minimal to no pain"* | |
| | Set level including some is normal or to be expected (310); *"I wouldn't expect it to be 0 because we all get distressed but as long as it's manageable"* | |
| Context, outside pain or distress | Activities, outside events (86); *"I'm quite anxious about my job and family life at the minute"* | |
| | Medical diagnosis (70); *"I have MS [multiple sclerosis] so my pain threshold varies"* | |
| | Medical treatment | Specific treatment (74); *"At this level I would require Advil or possibly something stronger depending on how long the pain lasted"* |
| | | Want treatment, nonspecific (150); *"I need treatment at this level"* |
| | | Already taking treatment (36); *"Tablets are working so it's great"* |
| | Don't want treatment | Manage on own (120); *"I can manage the pain on my own"* |
| | | Treatment drawbacks (30); *"Medications make me sick, so it needs to be a lot of distress before I will resort to it"* |
| | | Treatments don't work (38); *"I take ibuprofen and paracetamol at this level, but they don't really help as I can't sleep"* |
| | | Treatments don't address root cause (13); *"Generally my distress is sourced in something reasonable so cannot be completely removed, but a 4 would mean it can be managed without wasteful amounts of distress"* |
| Context, pain- or distress-specific | History of pain or distress (previous experience) (111); *"I have a chronic low-level pain at around a 3 level, so back to this base level [for me]"* | |
| | Type or location of pain (81); *"I suffer with back pain"* | |
| | Pain or distress is variable (82); *"My distress is up and down"* | |
| | Pain or distress is stable (54); *"Pain is fairly consistent on a daily basis"* | |
| | Frequency of pain or distress (81); *"Lower score represents less frequent pain"* | |
| | Disposition/tolerance (63); *"Have a low pain tolerance"* | |
| | At 0 or 10 (106); *"I'm not experiencing pain right now"* | |
| | Noticeable (93); *"Noticeable enough to be problematic"* | |
| | Interference/function (235); *"That's the level at which the pain becomes easily manageable"* | |
| Confusion | Don't understand (115); *"I was unsure what qualifies as 'high,' so 4 is a good safe number to pick because it's definitely on the lesser side"* | |
| | Didn't understand the reasons question (29); *"Climate change"* | |
| | Contradictory reasons (7); *"I feel the pain I experience is not bad enough to be treated, but is also something that should be treated"* | |
| | Contradicts numerical rating scale (47); *"I am currently not experiencing pain"* [marked 3 for numerical rating scale] | |
| | Didn't understand hypothetical (142); *"My pain is getting better"* [in response to the question about worsening pain] | |
| | Interpreted higher as better (instead of worse) (73); *"It would improve"* [in response to distress treatment working question, participant marked 8 but marked 4 for wanting treatment for distress] | |
| | Interpreted as change not level (3); *"That's the same rate of the previous question. [Because] it's a balance... 5 for getting worse, 5 for getting better"* | |
| | Didn't connect to correct numerical rating scale (146); *"4 is considerably greater than 8"* [in response to the question about distress treatment success, but participant marked 10 for wanting treatment and 8 for current distress] | |

*Parentheses indicate number of participants reporting that code. Italicized phrases are example quotes with explanations in square brackets.*

and higher levels at which one wanted treatment were associated with larger amounts of change considered meaningful, suggesting this method was valid, as this would be expected because these participants had more range on the rating scale to change. Paradoxical responding occurred most often for whether a treatment would be working, but also for whether a treatment was a success. Participants reported a wide range of what changes in pain and distress they considered meaningful and reported a diverse set of reasons for what they considered a meaningful change in pain or distress. Participants seemed to be thinking both in level and amount of change in pain or distress. Overall, the quantitative and qualitative results suggest that simply considering the same amount of change as similarly important for all patients is likely not ideal.

Results showed a variety of reasons for what people considered meaningful change, consistent with previous research.[20,21] Coefficients of variation were large, and the qualitative analyses identified a large variety of reasons people considered an amount of change meaningful. Interference of pain or distress in function was cited often, but participants also mentioned frequency of symptoms, medical conditions, and the variability of their pain or distress. While some participants considered any change important, others only valued large changes in distress or pain. Our results support continued research into ways of determining what is individually meaningful change on a PRO within MBC.

The qualitative and quantitative results also suggest several potential revisions to reduce paradoxical responding. Revisions should emphasize the prospective nature of the questions and eliminate the need for respondents to reference a previous answer. The qualitative data showed some participants would not want treatment even if needed, either because they were worried about side effects or did not want to prevent others from getting care. Therefore, further use of the question asking participants to define at what level they would want treatment is discouraged and a different anchor should be devised. As the questions about treatment working had more paradoxical responding than treatment being a success, the wording about treatment working needs further research to ensure patients understand the question.

## Limitations

The limitations of the study should be noted. This was a convenience sample. Respondents may have been savvier with survey questions due to being recruited through a crowdsourcing website. However, Prolific Academic has been used in hundreds of research studies.[27,28] Our use of attention check questions meant careless responding was unlikely, and the Prolific platform prevents people from completing a survey more than once. The sample also was not selected for previous experience with pain or distress, although we did stratify feasibility numbers by pain, distress, back pain, and depression. While the use of the numerical rating scale was warranted for this first study, these results might not translate to multi-item, scored PROs.

## CONCLUSIONS

A surprisingly simple method of asking patients to define what is meaningful change on a patient-reported outcome could be feasible for measurement-based care. The method also showed initial validity. Using a patient's own definition of meaningful change on a PRO may help improve the effectiveness of MBC. Future research should examine revisions to this method to reduce paradoxical responding as well as the utility of using individually meaningful change in MBC. Future research is also needed to determine whether this approach could be adapted for clinical trials.

### Patient-Friendly Recap

- Patient-reported outcomes (PROs) measure patients' perceptions of their health and are used to track how patients are responding to treatment.

- Current use of PROs in clinical care assumes the same amount of change in symptoms must be similarly meaningful to all patients.

- The authors tested a new model that seeks to personalize each patient's definition of meaningful change (ie, effective treatment).

- This method of asking patients to define the amount of symptom change they would find personally meaningful to them proved feasible and valid.

## Acknowledgments

## Author Contributions

## Conflicts of Interest

None.

## References

1. U.S. Food and Drug Administration. *Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims.* Available at https://www.fda.gov/media/77832/download. Rockville, MD: U.S. Department of Health and Human Services, 2009.

2. Scott K, Lewis CC. Using measurement-based care to enhance any treatment. *Cogn Behav Pract.* 2015;22:49-59. CrossRef

3. Guo T, Xiang YT, Xiao L, et al. Measurement-based care versus standard care for major depression: a randomized controlled trial with blind raters. *Am J Psychiatry.* 2015;172:1004-13. CrossRef

4. Basch E, Deal AM, Dueck AC, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *JAMA.* 2017;318:197-8. CrossRef

5. Basch E, Deal AM, Kris MG, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. *J Clin Oncol.* 2016;34:557-65. CrossRef

6. Streiner DL, Norman GR. Health Measurement Scales: *A Practical Guide to Their Development and Use, Fourth Edition.* Oxford, United Kingdom: Oxford University Press, 2008.

7. Cella D, Nichol MB, Eton D, Nelson JB, Mulani P. Estimating clinically meaningful changes for the Functional Assessment of Cancer Therapy – Prostate: results from a clinical trial of patients with metastatic hormone-refractory prostate cancer. *Value Health.* 2009;12:124-9. CrossRef

8. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol.* 1999;52:861-73. CrossRef

9. Morgan EM, Mara CA, Huang B, et al. Establishing clinical meaning and defining important differences for Patient-Reported Outcomes Measurement Information System (PROMIS®) measures in juvenile idiopathic arthritis using standard setting with patients, parents, and providers. *Qual Life Res.* 2017;26:565-86. CrossRef

10. Weinfurt KP. Clarifying the meaning of clinically meaningful benefit in clinical research: noticeable change vs valuable change. *JAMA.* 2019 Dec 2 [Epub ahead of print]. CrossRef

11. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol.* 1991;59:12-9. CrossRef

12. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989;10:407-15. CrossRef

13. King MT, Dueck AC, Revicki DA. Can methods developed for interpreting group-level patient-reported outcome data be applied to individual patient management? *Med Care.* 2019;57 Suppl 5 Suppl 1:S38-45. CrossRef

14. Moinpour CM, Donaldson GW, Davis KM, et al. The challenge of measuring intra-individual change in fatigue during cancer treatment. *Qual Life Res.* 2017;26:259-71. CrossRef

15. Donaldson G. Patient-reported outcomes and the mandate of measurement. *Qual Life Res.* 2008;17:1303-13. CrossRef

16. Jones SMW, Crane PK, Simon GE. A comparison of individual change using Item Response Theory and sum scoring on the Patient Health Questionnaire-9: implications for measurement-based care. *Ann Depress Anxiety.* 2019;6(1):1098.

17. Cook CE. Clinimetrics Corner: The minimal clinically important change score (MCID): a necessary pretense. *J Man Manip Ther.* 2008;16(4):E82-3. CrossRef

18. Gatchel RJ, Mayer TG. Testing minimal clinically important difference: additional comments and scientific reality testing. *Spine J.* 2010;10:330-2. CrossRef

19. Lindhiem O, Bennett CB, Orimoto TE, Kolko DJ. A meta-analysis of personalized treatment goals in psychotherapy: a preliminary report and call for more studies. *Clin Psychol (New York).* 2016;23:165-76. CrossRef

20. Cook KF, Kallen MA, Coon CD, Victorson D, Miller DM. Idio Scale Judgment: evaluation of a new method for estimating responder thresholds. *Qual Life Res.* 2017;26:2961-71. CrossRef

21. Thissen D, Liu Y, Magnus B, et al. Estimating minimally important difference (MID) in PROMIS pediatric measures using the scale-judgment method. *Qual Life Res.* 2016;25:13-23. CrossRef

22. Schmitt J, Abbott JH. Global ratings of change do not accurately reflect functional change over time in clinical practice. *J Orthop Sports Phys Ther.* 2015;45:106-11, D1-3. CrossRef

23. Garrison C, Cook C. Clinimetrics Corner: The Global Rating of Change Score (GRoC) poorly correlates with functional measures and is not temporally stable. *J Man Manip Ther.* 2012;20:178-81. CrossRef

24. Wolpe J. *The Practice of Behavior Therapy.* New York, NY: Pergamon Press, 1969.

25. Kiresuk TJ, Sherman RE. Goal attainment scaling: a general method for evaluating comprehensive community mental health programs. *Community Ment Health J.* 1968;4:443-53. CrossRef

26. Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA. Studies with pain rating scales. *Ann Rheum Dis.* 1978;37:378-81. CrossRef

27. Palan S, Schitter C. Prolific.ac — a subject pool for online experiments. *J Behav Exp Finance.* 2018;17:22-7. CrossRef

28. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the Turk: alternative platforms for crowdsourcing behavioral research. *J Exp Soc Psychol.* 2017;70:153-63. CrossRef