

Modified Severity of Dyspepsia Assessment pain scale: a new tool for measuring upper abdominal pain in osteoarthritis patients taking NSAIDs

Jennifer Welle¹
John Fort²
Joseph Crawley³
Byron Cryer⁴
Rene Dickerhoof¹
Michelle P Turner¹
Kimberly L Miller¹

¹ICON Clinical Research, The Embarcadero, San Francisco, CA, USA;
²POZEN Inc, Raleigh Road, Chapel Hill, NC, USA; ³AstraZeneca LP, Concord Pike, Wilmington, DE, USA;
⁴UT Southwestern Medical Center, Harry Hines Boulevard, Dallas, TX, USA

Background: This study evaluated the electronically administered modified Severity of Dyspepsia Assessment (mSODA) pain scale, a six-item measure of upper abdominal pain intensity, for daily use in osteoarthritis patients taking nonsteroidal anti-inflammatory drugs.

Methods: Once the mSODA pain scale was isolated, cognitive debriefing interviews (n = 30) were used to examine its appropriateness in the target population. Following administration of the instrument in two Phase III pivotal trials, the data were analyzed to examine reliability, validity, responsiveness, and the minimal important difference.

Results: Using a subset of trial data (n = 90 patients), the mSODA pain scale proved to be a unidimensional, highly internally consistent instrument ($\alpha = 0.93$) with good test-retest reliability (intraclass correlation coefficient 0.77). Construct validity was established via moderate correlations with other similar patient-reported outcomes. Additionally, known-groups validity demonstrated that the mSODA pain scale could distinguish between subjects who did and did not report gastrointestinal symptoms and antacid use (both *P* values ≤ 0.05). The mSODA pain scale was also responsive to change in heartburn at weeks 6 and 12 (Guyatt's statistic = 1.7 and 2.6, respectively), and the minimal important difference obtained via $\frac{1}{2}$ SD was 5.7 (range 2–47).

Conclusion: This research suggests that the mSODA pain scale is both feasible and valid for assessing dyspepsia in patients taking nonsteroidal anti-inflammatory drugs for relief of symptoms of osteoarthritis.

Keywords: mSODA pain scale, questionnaire, antacid, patients, heartburn

Introduction

Patient-reported outcomes (PROs) have become increasingly important tools for understanding how various conditions affect patients. Both the US Food and Drug Administration and the Committee for Medicinal Products for Human Use of the European Medicines Agency emphasize the value of PRO measures in identifying and quantifying the patient's perspective on their disease and its treatments. Additionally, patient-reported symptoms are of critical interest in the evaluation of osteoarthritis therapies, such as nonsteroidal anti-inflammatory drugs (NSAIDs), because of the subjective nature of both efficacy (joint pain and mobility) and safety (upper abdominal pain and dyspepsia) outcomes.

The purpose of this study was to evaluate the electronically administered modified Severity of Dyspepsia Assessment (mSODA) pain scale, a PRO for daily use in osteoarthritis patients taking NSAIDs. To do this, a previously published questionnaire was identified in the existing literature, and cognitive debriefing interviews were used to further validate the instrument's content. Following content validation,

Correspondence: Rene Dickerhoof
ICON Clinical Research, 188 The Embarcadero, San Francisco, CA, 94105, USA
Tel +1 415 371 2123
Fax +1 415 8560840
Email rene.dickerhoof@iconplc.com

the psychometric properties of the modified questionnaire were assessed and are reported here.

Methods

PRO search

To identify a suitable PRO questionnaire to assess dyspepsia in two planned clinical trials, a detailed literature search was conducted in PubMed. The search was restricted to English-only articles published in the past 10 years. These PRO questionnaires were evaluated against a set of pre-specified criteria. Specifically, the PRO had to be designed for English-speaking adults (>18 years), developed for self-administration, and indicated for use as an evaluative instrument (ie, diagnostic tools were excluded). Finally, the PRO questionnaire had to provide evidence of previous validation, which was loosely defined as having any available psychometric data.

Cognitive debriefing interviews

A convenience sample of patients mirroring the clinical trial target population was recruited from Chicago, IL, New York, NY, and San Francisco, CA, using Craigslist.org. The protocol and informed consent form used to conduct the interviews were approved by the Copernicus Group independent review board. Eligible participants needed to be diagnosed with osteoarthritis (as indicated by self-report), be <50 years old, able to speak and read English, able to travel to a local interview location, and currently taking NSAIDs. Participants who reported having rheumatoid arthritis, gout/pseudogout, fibromyalgia syndrome, severe erosive esophagitis, Zollinger Ellison syndrome, and/or peptic ulcer disease were excluded.

At the beginning of each interview, participants first completed the mSODA pain scale questionnaire on a personal digital assistant device. A semistructured interview guide was used to ask questions about the instrument items, the mode of administration, and the recall period. Specifically, participants were asked to describe their understanding of abdominal pain and discomfort. Each interview lasted between 20 and 30 minutes and was digitally recorded and transcribed verbatim for coding purposes. Upon completion of the interviews, two independent reviewers coded each transcript and assessed saturation.

Psychometric testing procedures

The mSODA pain scale was implemented into a clinical research program assessing the efficacy and safety of a combination drug (esomeprazole [a proton pump inhibitor] + naproxen

[an NSAID]) for treatment of osteoarthritis of the knee. These trials and their populations are described elsewhere (Protocols PN400-307 [NCT00664560] and PN400-309 [NCT00665431]).^{1,2}

Patients completed daily assessments of the mSODA pain scale, Heartburn Symptom Severity Rating (HSSR), and a question regarding antacid use on the personal digital assistant device. The Multi-Dimensional Health Assessment Questionnaire (MDHAQ) Global Status (GS) Scale, and GI Symptoms report were completed at preset time points in the study (see Figure 1 for questionnaire assessment schedule).³ The HSSR is a one-item assessment of the severity of patients' self-reported heartburn (ie, "Over the past 24 hours, please rate your heartburn symptoms"). Antacid use was measured by asking, "How many antacids did you take in the past 24 hours?" (responses dichotomized into 'none'/'any'). The MDHAQ GS Scale stated, "Considering all the ways in which illness and health conditions may affect you at this time, please indicate below how you are doing". Finally, the MDHAQ GI Symptom report asked patients to indicate whether they experienced "heartburn or stomach gas", "stomach pain or cramps", "nausea", and/or "vomiting" in the last month.

Psychometric statistical testing methods

The scale structure of the mSODA pain scale was confirmed by conducting cluster analysis using the SAS VARCLUS procedure.⁴ VARCLUS uses iterative splitting and factor analytic methods to divide a group of variables into discrete (ie, nonoverlapping) subgroups or clusters that are relatively highly correlated and represented by a single given domain. The initial clustering was based on the correlation matrix

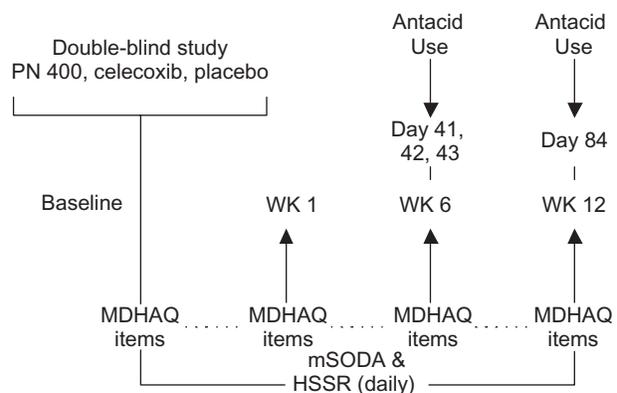


Figure 1 Assessment schedule for the modified Severity of Dyspepsia Assessment pain scale and validation items.

Abbreviations: mSODA, modified Severity of Dyspepsia Assessment pain scale; HSSR, Heartburn Symptom Severity Rating; WK, week; MDHAQ, Multi-Dimensional Health Assessment Questionnaire.

produced and the default stopping rule (ie, clusters with a second eigenvalue below 1 were not further subdivided).

Variability was assessed by examining the instrument's mean and standard deviation (SD) as well as the minimum, median, and maximum values obtained from patient assessments (ie, from day 0 through day 84). Floor and ceiling effects were also examined, which determine whether a disproportionately large percentage of responders were prone to providing either the lowest or highest values. Unless otherwise specified in the methods below, data were pooled across study evaluations to include as much variation in patient health status during the study as possible.

Cronbach's alpha demonstrates the extent to which items within an instrument assess the same construct (ie, internal consistency).⁵ Test-retest reliability measures the degree to which a questionnaire yields stable scores in patients over a short period of time (ie, when no clinical change or intervention is anticipated), and can be estimated with the intraclass correlation coefficient.⁶ Given that clinical change was anticipated in participants, a stable cohort of patients reporting no change in heartburn from baseline (ie, day 0) to day 7 was used to assess this type of reliability.

Convergent validity measures the extent to which an instrument performs as intended when compared with other related constructs. Convergent validity was assessed by examining Pearson correlation coefficients between the mSODA pain scale and the HSSR and MDHAQ-GS scales. It was hypothesized that both the HSSR and MDHAQ-GS scales would be moderately correlated with the mSODA pain scale. Specifically, patients who reported greater heartburn and poorer global health were also hypothesized to report increased dyspepsia.

Known-groups validity confirms whether an instrument can appropriately distinguish between two or more groups "known" to differ on a given outcome. Known-groups validity was evaluated by examining mean mSODA pain scale scores between patients who did and did not report upper gastrointestinal symptoms on the MDHAQ and antacid use on days 41, 42 (mid trial), 43, and 84 (trial end). The analyses at day 41 and day 43 were conducted to account for variation in patient antacid use that may occur on a day-to-day basis. It was predicted that patients who reported upper gastrointestinal symptom(s) or antacid use would report greater dyspepsia than patients who did not.

Responsiveness measures the degree to which a questionnaire detects change in individuals known to change clinically on the construct of interest (or a related construct). Responsiveness was evaluated using Guyatt's statistic,

which represents the ratio of change occurring in the cohort of individuals known to improve clinically to the SD of change occurring in the cohort of individuals known to remain stable.⁷ Responsiveness of the mSODA was examined at week 6 (ie, average of days 36–42 minus baseline mSODA pain scale scores) and week 12 (ie, average of days 78–84 minus baseline mSODA pain scale scores). Change in HSSR was used to define clinically improved, stable, or worsened cohorts (ie, week 6 and week 12 HSSR average scores minus baseline values, respectively). To do this, cohorts were categorized as "improved" (change in heartburn from baseline to follow-up ≤ -1), "stable" (change > -1 but < 1), and "worsened" (change ≥ 1). Patients who reported improved, little to no change, and worsened heartburn were expected to report improved, little to no change, and worsened dyspepsia, respectively.

When using PROs to evaluate a treatment benefit, the FDA recommends the sponsor establish an a priori responder definition.⁸ To determine the appropriate change in mSODA pain scale scores constituting a potential responder definition, one distribution-based method was examined, ie, a calculation of $\frac{1}{2}$ SD. The $\frac{1}{2}$ SD was calculated by multiplying 0.5 by the SD of the day 0 mSODA pain scale scores. This approach has been suggested as a good starting point for interpreting a clinically important change when no other supplemental information is available (eg, when no anchor-based approach is possible). Further, this approach provides a conservative estimate of the minimal important difference that is likely to be a lower bound.⁹

Results

PRO selection and modification

One-hundred and eighty-nine abstracts were identified referencing 26 dyspepsia-focused PRO instruments. Eighteen instruments met the initial inclusion criteria and were investigated further to identify those that contained an abdominal pain-specific scale (ie, single items did not meet the criteria) and those with demonstrated psychometric properties in an osteoarthritis/NSAID-user population. Only one instrument met these inclusion/exclusion criteria, ie, the pain intensity subscale of the SODA.

The SODA pain intensity scale was initially developed to assess the predominant feature of dyspepsia, ie, upper abdominal pain.¹⁰ The scale was established by adapting items from pain-related studies.¹⁰ Confirmatory factor analysis and various other psychometric tests (eg, effective measurement range) were used to reduce the final scale to six items.¹⁰ Although not originally developed in this population, the

pain intensity scale has been validated in and administered to patient taking NSAIDs to treat symptoms of arthritis.^{11,12}

Once the SODA pain intensity scale was selected for inclusion in the clinical program, two significant modifications were made to the scale and the new name was applied (ie, the modified SODA pain scale, ie, mSODA pain scale). Because electronic data collection has been shown to result in greater compliance and reporting accuracy, the mode of administration was updated from a paper-and-pencil to electronic format.¹³ Further, because one-day recall periods for pain have been demonstrated to be more accurate than one-week recall, the instrument's time frame was changed from the "past seven days" to the "past 24 hours".¹⁴

Cognitive debriefing interviews

Content validity was assessed in a convenience sample of 30 individuals (>50 years of age) with osteoarthritis and self-reporting NSAID use. Most participants were female (n = 18, 60%), Caucasian (n = 27, 90%), and diagnosed with osteoarthritis within the past three years (n = 19, 63%). Participants were also well-educated (at least a college degree; n = 23, 77%), fairly affluent (\$75,000+ per year; n = 18, 60%), and taking a variety of oral pain medications to treat osteoarthritis symptoms (aspirin [n = 2, 7%], ibuprofen [n = 8, 27%], COX-2 inhibitors [n = 7, 23%], and naproxen [n = 13, 43%]).

Although four participants (14%) reported difficulty using the stylus or their finger to advance the personal digital assistant screens, only one participant reported physical difficulty handling the device/stylus. Several participants (n = 5, 17%) reported difficulty reading the screen or suggested increasing font size; however, this problem was remedied when reading glasses were worn. When the intent to administer the mSODA pain scale on a daily basis in a subsequent study was discussed, no participants reported unwillingness to do so. Finally, when asked whether they would prefer to complete questions using another format (eg, paper-and-pencil), the majority of participants (n = 16, 55%) selected the personal digital assistant as their first choice.

As expected, pain (n = 23, 79%) was the most commonly reported symptom, followed by discomfort (n = 18, 62%) and ache (n = 8, 28%). Other less commonly reported experiences included bowel symptoms (n = 12, 41%), acid regurgitation/indigestion (n = 12, 41%), bloating (n = 11, 38%), gas (n = 10, 34%), and nausea/queasiness (n = 10, 34%). Additionally reported terminology included words such as uneasy, upset, unpleasant, unsettled, uncomfortable, bother, distress, irritation, and problems (n = 18, 62%).

Interviewers also probed participants to describe their abdominal discomfort using various temporal reference points. When asked to compare symptoms "right now" to those experienced in the "the past 24 hours", reports were less severe for the former versus the latter. However, when distinguishing between "the last 24 hours" and "yesterday", no distinction in terms was identified by participants. Several participants (n = 12, 41%) reported some difficulty recalling their abdominal discomfort or stomach ache using "last week" as a reference point. Saturation of coding concepts was reached after the 11th interview, and all identified themes were reported by at least two participants at the 17th interview.

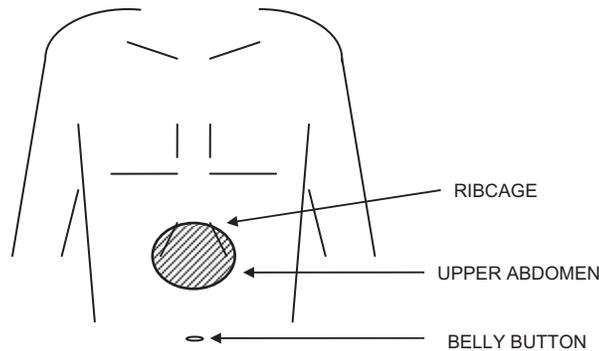
Based upon patient feedback from the cognitive debriefing interviews, modifications were made to the mSODA pain scale instrument. These changes included: adding instructions to the first question regarding use of the personal digital assistant device; changing the item's language to reference abdominal discomfort (or stomach ache) rather than pain, a term which was often confused for osteoarthritis pain rather than upper abdominal pain; simplifying the response options for the third question (eg, from, "horrible" and "excruciating" to "severe pain" and "very severe pain"); and, finally, inserting a diagram of the upper abdomen into the instructions. Figure 2 presents the finalized mSODA pain scale administered in the PN400 trials along with the instrument's scoring instructions. Pending use of the instrument's recoding algorithm, scores on the instrument range from 2 to 47, with higher scores indicating increasing pain.

Validation study

The mSODA pain scale and other questionnaires were administered to 1230 patients participating in the PN400-307 and PN400-309 trials. However, because the validation study was conducted within the context of the clinical trials, only a subset of trial participants could be allotted to the evaluation of the mSODA pain scale's psychometric properties. This analytic population consisted of 90 randomly selected participants. Forty-five subjects were randomly selected from each trial and from each of three treatment groups (ie, PN400, celecoxib, and placebo; all n = 15). The demographic characteristics for all patients and within each trial are presented in Table 1. The distribution of these characteristics was similar in both trials. Table 2 presents the means and SDs for the mSODA pain scale and other validation items at each time point (ie, baseline, week 6, and week 12). The mSODA pain scale completion rates ranged from 90 patients

Instructions: There are six questions in this questionnaire. Please read and answer each question carefully. We are interested in your responses to the questions; there are no right or wrong answers. The questions that follow ask about your abdominal discomfort (or stomach ache). You may find that some of the questions are very similar to each other, but by answering all of them you can help us understand your pain better.

When answering these questions, please think about the pain or discomfort in the region of your abdomen below your rib cage and above your belly button, as illustrated in the diagram here:



1. Please enter the number between 0 and 100 that best describes your abdominal discomfort (or stomach ache) during the PAST 24 HOURS, ON AVERAGE. A zero '0' would mean "No pain or discomfort", and a one hundred '100' would mean "Pain as bad as it could be".



(Tap on the triangles to select a number from 0 to 100)

2. During the PAST 24 HOURS, ON AVERAGE, how intense was your abdominal discomfort (or stomach ache) rated on a 0–10 scale where '0' is "No discomfort" and '10' is "Discomfort as bad as it could be"?



3. During the PAST 24 HOURS, ON AVERAGE, how intense was your abdominal discomfort (or stomach ache)? Select the phrase that best describes your pain.

- 0 No pain
- 1 Mild pain
- 2 Moderate pain
- 3 Severe pain
- 4 Very severe pain

4. During the PAST 24 HOURS, ON AVERAGE, how intense was your abdominal discomfort (or stomach ache)? Make a mark on the line that shows how intense your pain was. A mark ALL THE WAY TO THE LEFT would indicate "No pain". A mark ALL THE WAY TO THE RIGHT would indicate "Pain as bad as it could be".



5. During the PAST 24 HOURS, ON AVERAGE, how intense was your abdominal discomfort (or stomach ache)? Select the phrase that best describes your pain.

- 0 No pain
- 1 Some pain
- 2 Considerable pain
- 3 Pain which could not be more severe

6. During the PAST 24 HOURS, how intense was your WORST abdominal discomfort (or stomach ache) rated on a 0–10 scale where '0' is "No discomfort" and '10' is "Discomfort as bad as it could be"?



Figure 2 Modified Severity of Dyspepsia Assessment pain scale. Copyright © 2011, AstraZeneca.

at baseline to 65 at week 12 (Mean = 78; including day 0) yielding a retention rate of 72%. As expected, the six mSODA pain scale items formed a coherent single factor for upper abdominal pain (84% of variation explained, second eigenvalue <1.0).

Across all daily observations provided within and between patients (n = 5639), mSODA pain scale scores spanned the entire range of values (ie, 2–47) demonstrating that the scale covers a range of dyspepsia experiences during the entire study period. The mean and median scores were 8.6 ± 9.2 and 2.0, respectively. Given that the median mSODA score was 2 (ie, the lowest possible mSODA score indicating no reported dyspepsia), we also examined mSODA scores in only those patients who reported dyspepsia (ie, scores >2; n = 3369 observations). The mean and median mSODA pain scale scores in this population were 12.0 ± 9.8 and 14.0, respectively.

Internal consistency reliability of the mSODA pain scale was high ($\alpha = 0.93$). Test-retest reliability was also good in a cohort of patients with no change in heartburn severity from day 0 to day 7 (intraclass correlation coefficient 0.77, n = 44).

As hypothesized, the mSODA pain scale was positively correlated with both the HSSR ($r = 0.55$) and the MDHAQ-GS ($r = 0.34$). Further, and also as predicted, Table 3 shows that subjects who reported gastrointestinal symptoms and who took antacids reported more upper abdominal pain than those who did not (all *P* values < 0.05).

The cohort of patients who improved on the HSSR showed the greatest decrease in mSODA pain scale scores

at weeks 6 and 12. Additionally, the stable and worsened cohorts reported only small decreases in mSODA pain scale scores. A Guyatt's statistic of 1.7 and 2.6 was found at weeks 6 and 12, respectively (see Table 4, Guyatt's statistic ≥ 1.0 suggests high responsiveness).

Finally, the distribution-based $\frac{1}{2}$ SD result, which suggests merely a starting point for assessing meaningful change on the mSODA pain scale, was 5.7 (range 2–47).

Discussion

In general, the mSODA pain scale was a feasible and psychometrically sound tool. Specifically, cognitive debriefing interviews, although acquired via a convenience sample, established that the 24-hour recall period and electronic mode of administration were appropriate in the target population. Most participants preferred to complete the mSODA pain scale on the personal digital assistant and reported less difficulty recalling pain in the past 24 hours versus in the last week. Additionally, patients confirmed the importance of assessing NSAID-induced upper abdominal pain by self-describing this symptom as the most prominent feature.

Given the similarity of the instrument items, it was not surprising that the mSODA pain scale proved to be a unidimensional, highly internally consistent instrument. Indeed, an identical alpha was found for the SODA pain intensity scale, demonstrating that changes made to the instrument (both content and administration mode) did not affect this psychometric property. The mSODA pain scale also demonstrated good test-retest reliability. Unfortunately, no data currently

Table 1 Demographic characteristics of study population

	All subjects (n = 90)	Trial 1 (n = 45)	Trial 2 (n = 45)
Age (years)			
Mean \pm standard deviation	62.3 \pm 8.5	62.6 \pm 9.0	62.1 \pm 8.0
Minimum, median, maximum	50.0, 61.0, 90.0	50.0, 61.0, 90.0	50.0, 61.0, 80.0
<65	53 (58.9%)	26 (57.8%)	27 (60.0%)
≥ 65	37 (41.1%)	19 (42.2%)	18 (40.0%)
Female, n (%)	53 (58.9%)	25 (55.6%)	28 (62.2%)
Race, n (%)			
White	71 (78.9%)	34 (75.6%)	37 (82.2%)
Black	14 (15.6%)	9 (20.0%)	5 (11.1%)
Asian	5 (5.6%)	2 (4.4%)	3 (6.7%)
Hispanic or Latino ethnicity, n (%)	6 (6.7%)	2 (4.4%)	4 (8.9%)
Education, n (%)			
Less than high school graduate	8 (9.2%)	5 (11.6%)	3 (6.8%)
High school graduate/GED	27 (31.0%)	15 (34.9%)	12 (27.3%)
Some college	25 (28.7%)	10 (23.3%)	15 (34.1%)
Bachelor's degree	17 (19.5%)	8 (18.6%)	9 (20.5%)
Graduate degree	10 (11.5%)	5 (11.6%)	5 (11.4%)

Table 2 Means and standard deviations across assessments for the modified Severity of Dyspepsia Assessment pain scale and validation items

Measure ^a	n	M ± SD
mSODA pain scale ^b		
Day 0	90	13.4 ± 11.4
Week 6 (days 36–42)	71	8.3 ± 8.6
Week 12 (days 78–84)	62	7.3 ± 8.0
HSSR ^c		
Day 0	90	0.61 ± 0.82
Week 6 (days 36–42)	71	0.36 ± 0.50
Week 12 (days 78–84)	62	0.33 ± 0.49
MDHAQ-GS scale ^d		
Day 0	88	6.2 ± 2.3
Week 6 (days 36–42)	49	4.2 ± 2.5
Week 12 (days 78–84)	50	3.7 ± 2.5
MDHAQ-GI Symptoms ^e		
Day 0	88	40% ± 49%
Week 6 (days 36–42)	49	47% ± 50%
Week 12 (days 78–84)	50	42% ± 50%

Notes: ^aNonmissing week 6 and week 12 responses are averaged across the associated seven-day period; ^bmSODA pain scale (see scoring instructions; range 2–47); ^cHSSR, 0–3 point scale (none, mild, moderate, severe); ^dMDHAQ-GS, 0–10 point scale (very well, very poorly); ^eMDHAQ-GI (any “Yes” response at a given assessment to any of the following: heartburn or stomach gas; stomach pain or cramps; nausea; or vomiting).

Abbreviations: HSSR, Heartburn Symptom Severity Rating; MDHAQ-GS, Multi-dimensional Health Assessment Questionnaire-Global Status Scale; MDHAQ-GI, Multi-dimensional Health Assessment Questionnaire-Gastrointestinal Symptom Scale; mSODA, modified Severity of Dyspepsia Assessment; M, mean; SD, standard deviation.

exist to compare this psychometric property across modes of administration (ie, paper versus electronic delivery). Finally, the instrument also showed sufficient variability. That is, change in mSODA pain scale scores spanned the entire range of values, and did not demonstrate floor or ceiling effects in the population of participants who reported at least some dyspepsia at baseline.

The construct validity of the mSODA pain scale was verified by examining the instrument’s correlation with assessments of heartburn and global health. The moderate correlation between the heartburn assessment and the mSODA pain scale suggests that heartburn is a related but distinct concept from dyspepsia. Further, the correlation between global health and upper abdominal pain indicates that increases in upper abdominal pain correspond to reported reductions in global perceptions of health. Finally, increased reporting of upper abdominal pain was correlated with antacid use and the reporting of gastrointestinal symptoms.

The mSODA pain scale also proved to be responsive to changes in heartburn severity when examined at both week 6 and week 12 suggesting that it is an appropriate tool for

Table 3 Modified Severity of Dyspepsia Assessment construct validity (known-groups)

	n	mSODA M ± SD	P value
MDHAQ-GI symptoms			
None	53	9.1 ± 10.0	0.039
Any	35	19.0 ± 10.4	
Use of antacids* at day 41			
None	54	6.4 ± 8.0	0.026
Any	5	15.0 ± 8.2	
Use of antacids at day 42			
None	51	8.0 ± 8.1	0.005
Any	13	15.5 ± 8.8	
Use of antacids at day 43			
None	50	6.6 ± 7.8	0.027
Any	13	12.2 ± 8.7	
Use of antacids at day 84			
None	41	6.9 ± 8.1	0.005
Any	6	17.5 ± 9.4	

Note: *Maalox[®] regular strength tablet antacids were the only antacids allowed for use during the trials. Use of other proton pump inhibitors or H₂ antagonists were prohibited.

Abbreviations: MDHAQ-GI, Multi-dimensional Health Assessment Questionnaire-Gastrointestinal Symptom Scale; mSODA, modified Severity of Dyspepsia Assessment; M, mean; SD, standard deviation.

detecting change in dyspeptic pain over time. The ½ SD, a commonly used and conservative estimate of minimal important difference, was similar to that value found for the original SODA pain intensity scale.⁸ This minimal important difference can be interpreted to mean that six-week change in a given patient’s mSODA pain scale score would need to be ≥5.7 (range 2–47) to be considered a minimal important or detectable difference by patients (ie, for group comparisons).

Although the mSODA pain scale appears to be valid for assessing upper abdominal pain in osteoarthritis patients taking NSAIDs, these results are not without caveats. To begin, the cognitive debriefing interviews were based on a convenience sample of highly educated, affluent, and urban-dwelling individuals. Although the recall period and electronic mode of administration were preferred in this population, it is unclear whether all trial participants would provide the same (or similar) feedback.

Additionally, although participants indicated that “pain”, “discomfort”, and “ache” were the fundamental features of dyspepsia, it is possible that this language was encouraged by prior completion of the questionnaire, which regularly referenced these terms. It is also true that the current study isolated NSAID-induced pain intensity specifically in the upper abdomen, although pain could arguably be a multidimensional concept that is applicable to the lower gastrointestinal tract as well.

Table 4 Modified Severity of Dyspepsia Assessment responsiveness at week 6 and week 12

	Heartburn severity						Guyatt's statistic
	Clinically improved		Clinically stable		Clinically worsened		
	≤ -1		$> -1 < 1$		≥ -1		
	n	Mean \pm SD	n	Mean \pm SD	n	Mean \pm SD	
Week 6 change	16	-13.0 \pm 9.6	49	-3.1 \pm 7.8	6	-1.6 \pm 14.7	-1.7
Week 12 change	16	-15.1 \pm 8.9	39	-2.0 \pm 5.8	7	-3.8 \pm 14.9	-2.6

Note: *Guyatt's statistic is defined as the ratio of the average mSODA change score (at each time point) in the improved cohort to the SD of the mSODA change score in the stable cohort.

Abbreviations: mSODA, modified Severity of Dyspepsia Assessment; M, mean; SD, standard deviation.

Another potential concern pertains to the reliability of the instrument. Although the mSODA pain scale was highly internally consistent, it could be argued that the items are, to a fault, "too" consistent. Indeed, future research may consider potential item redundancies that could lead to a more simplified form of the measure. Additionally, the assessment of test-retest reliability was problematic in the clinical trials as only one baseline (ie, preintervention) assessment of dyspepsia was available. Given this, a cohort of the trial patients with stable heartburn reports was used to examine reliability of the mSODA over time. However, it must be noted that heartburn is only moderately related to dyspepsia and is, therefore, a proxy cohort at best.

Another limitation involves interpretation of the relationship between heartburn and dyspepsia. As previously discussed, the medical literature has attempted to separate heartburn from dyspepsia to distinguish the latter from gastroesophageal reflux disease.¹⁵ Additionally, although several studies examining patient feedback, including the current examination, have identified pain as the definitive feature of dyspepsia, reports of heartburn and other symptomatology are not uncommon.¹⁶⁻¹⁸ Given this, it remains arguable whether dyspepsia is a symptom complex that includes some aspects of heartburn or whether dyspepsia and heartburn are a related but distinct phenomenon entirely. In either case, using change in heartburn to assess test-retest reliability and responsiveness of the mSODA pain scale is less than ideal.

An additional limitation concerns the instruments used to validate the mSODA pain scale. To begin, the MDHAQ-GI item references a one-month retrospection period, whereas the mSODA pain scale encapsulates a 24-hour period. Further, the antacid item and the HSSR are not previously validated instruments. Thus, it is unclear whether these comparisons are valid construct representations. Finally, these instruments/items, like the mSODA pain scale, provide only patient-reported data rather than a clinical assessment with which to compare/contrast results.

In addition to the aforementioned issues, the most conservative estimate of meaningful change obtained in this study for group comparisons (ie, 5.7) should be considered merely a starting point for interpreting the change in mSODA pain scale scores. Anchor-based methods are necessary to evaluate fully the relevance of change to patients and clinicians.

Conclusion

In summary, the mSODA pain scale proved to be a feasible and appropriate questionnaire for assessing dyspepsia in osteoarthritis patients taking NSAIDs. Initial results suggest that the content validity, scale dimensionality, variability, internal consistency, test-retest reliability, construct validity, responsiveness, and minimal important difference are all reasonably sound. Although these results are encouraging, further validation is encouraged to confirm the accuracy of these psychometric properties in the target population.

Disclosure

This research was funded by AstraZeneca LP and POZEN Inc. JC is an employee of AstraZeneca LP and JF is an employee of POZEN Inc. BC has current consulting relationships with AstraZeneca LP, Horizon Therapeutics, McNeil Consumer Products, NiCox Inc, PLx Pharma, Pfizer Inc, and POZEN Inc, and has received current research grants from PLx Pharma and Pfizer Inc. RD, KM, MPT, and JW have consulting relationships with AstraZeneca LP and POZEN Inc.

References

1. Pozen, Inc. Study Evaluating the Efficacy of PN 400 BID and Celebrex QD in Patients With Osteoarthritis of the Knee. ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). 2000-[2011 April 06]. Available from: <http://clinicaltrials.gov/ct2/show/NCT00664560?term=NCT00664560&rank=1>. Accessed 5 April 2011.
2. Pozen, Inc. Study Evaluating the Efficacy of PN 400 BID and Celecoxib QD in Patients With Osteoarthritis of the Knee. ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). 2000-[2011 April 06]. Available from: <http://clinicaltrials.gov/ct2/show/NCT00665431>. Accessed 5 April 2011.

3. Pincus T, Yazici Y, Bergman M. Development of a multi-dimensional health assessment questionnaire (MDHAQ) for the infrastructure of standard clinical care. *Clin Exp Rheumatol*. 2005;23 Suppl 39:S19–S28.
4. SAS System for Windows [Computer program]. Cary, NC: SAS Institute Inc; 2010.
5. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16:297–233.
6. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–428.
7. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis*. 1987;40:171–178.
8. US Food and Drug Administration. Guidance for Industry Patient-Reported Outcome Measures: Use in Medicinal Product Development to Support Labeling Claims. Rockville, MD: US Department of Health and Human Services; 2009.
9. Sloan JA. Assessing the minimally clinically significant difference: Scientific considerations, challenges and solutions. *COPD*. 2005;2:57–62.
10. Rabeneck L, Cook KF, Wristers K, Soucek J, Menke T, Wray NP. SODA (severity of dyspepsia assessment): A new effective outcome measure for dyspepsia-related health. *J Clin Epidemiol*. 2001;54:755–765.
11. Rabeneck L, Wristers K, Goldstein JL, Eisen G, Dedhiya SD, Burke TA. Reliability, validity, and responsiveness of severity of dyspepsia assessment (SODA) in a randomized clinical trial of a COX-2-specific inhibitor and traditional NSAID therapy. *Am J Gastroenterol*. 2002;97:32–39.
12. Rabeneck L, Goldstein JL, Vu A, Mayne TJ, Rublee DA. Valdecoxib is associated with improved dyspepsia-related health compared with nonspecific NSAIDs in patients with osteoarthritis or rheumatoid arthritis. *Am J Gastroenterol*. 2005;100:1043–1050.
13. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient non-compliance with paper diaries. *BMJ*. 2002;324:1193–1194.
14. Broderick JE, Schwartz JE, Vikingstad G, Pribbernow M, Grossman S, Stone AA. The accuracy of pain and fatigue items across different reporting periods. *Pain*. 2008;139:146–157.
15. Talley NJ. American Gastroenterological Association medical position statement: Evaluation of dyspepsia. *Gastroenterology*. 2005;129:1753–1755.
16. Veldhuyzen van Zanten S. Can severity of symptoms be used as an outcome measure in trials of non-ulcer dyspepsia and helicobacter pylori associated gastritis? *J Clin Epidemiol*. 1993;46:273–279.
17. Meineche-Schmidt V. Classification of dyspepsia and response to treatment with proton-pump inhibitors. *Aliment Pharmacol Ther*. 2004;20:1171–1179.
18. Meineche-Schmidt V. Empiric treatment with high and standard dose of omeprazole in general practice: two-week randomized placebo-controlled trial and 12-month follow-up of health-care consumption. *Am J Gastroenterol*. 2004;99:1050–1058.

Patient Related Outcome Measures

Publish your work in this journal

Patient Related Outcome Measures is an international, peer-reviewed, open access journal focusing on treatment outcomes specifically relevant to patients. All aspects of patient care are addressed within the journal and practitioners from all disciplines are invited to submit their work as well as healthcare researchers and patient support groups. Areas covered will

Submit your manuscript here: <http://www.dovepress.com/patient-related-outcome-measures-journal>

include: Quality of life scores; Patient satisfaction audits; Treatment outcomes that focus on the patient; Research into improving patient outcomes; Hypotheses of interventions to improve outcomes; Short communications that illustrate improved outcomes; Case reports or series that show an improved patient experience; Patient journey descriptions or research.

Dovepress