

## Research article

## Measuring outcomes of treatment for patients with episodic abdominal pains; conclusions from the EPISOD study

Valerie Durkalski-Mauldin<sup>\*1</sup>, Qi Pauls<sup>1</sup>, Peter Cotton<sup>2</sup>

1. Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA

2. Digestive Disease Center, Medical University of South Carolina, Charleston, SC, USA

<sup>\*</sup>Corresponding author: Valerie Durkalski-Mauldin, Department of Public Health Sciences, Medical University of South Carolina, 135 Cannon St, Ste 301 Charleston, SC 29425; Tel: 843-876-1911; E-mail: durkalsv@musc.edu

Received: June 07, 2019; Accepted: July 16, 2019; Published: July 18, 2019

**Abstract**

**Objective:** Measuring outcomes, that are relevant to both the patient and clinician, in patients with intermittent attacks of abdominal pain, remains challenging. The aim of this study was to examine several definitions of 'success' for patients undergoing treatment for suspected Sphincter of Oddi dysfunction and identify valid and reliable outcome measures for future research studies. **Methods:** The recently completed EPISOD trial incorporated several patient reported outcome measures to determine improvement in patients' pain and disability over time. The trial's primary outcome was dichotomized as 'success/failure' based on the days of disability due to episodic abdominal pain using the RAPID, a 90-day patient recall instrument. Additional measures included the SF-36, frequency and intensity of pain, and the Patient Global Impression of Change (PGIC), which were collected periodically during the long term follow up period. Correlations between the different instruments were calculated accounting for repeated measures within patients. Agreement using a dichotomized definition of 'success' was also examined. **Results:** There was a moderate negative correlation of the RAPID score with the specific SF-36 physical domain scores of bodily pain, physical functioning and role limitation. Negative correlations were expected since a higher RAPID score indicates greater role impairment and a higher SF-36 score indicates higher levels of physical functioning. Aggregated one month assessments of pain frequency correlated well with the results of the 90 day recall ( $r=0.84$ ). When comparing the dichotomized definitions of 'success', the RAPID and PGIC had a high percentage of agreement (72%) and a moderate kappa coefficient of 0.44 (0.23, 0.65). **Conclusion:** These results support the validity of the 90-day recall of the RAPID instrument and a dichotomized definition of success based on disability days in patients with intermittent attacks of abdominal pain. We recommend using the objective RAPID disability score and the subjective PGIC instrument in future studies.

**Keywords:** abdominal pain, sphincter of Oddi dysfunction, post-cholecystectomy pain, pain disability

**Introduction**

Measuring pain and its response to treatment is an important, but challenging, aspect of clinical practice and related research. There are many well-validated instruments for measurement of chronic daily pain, where several use a simple visual analogue scale [1]. However, the pains caused by gallstones and other biliary pathologies, such as sphincter of Oddi dysfunction (SOD) are episodic and vary in frequency among and within individuals. Pain episodes can be severe, often interfere with ability to function in primary roles, and can have a signifi-

cant impact on quality of life. Endoscopic sphincterotomy has become a popular treatment, with uncertain benefit and significant risk. The authors recently completed a multicenter, blinded, randomized, sham-controlled clinical trial (EPISOD - Evaluating Predictors and Interventions in Sphincter of Oddi Dysfunction; Trial registration: NCT 00688662) that examined the efficacy of sphincterotomy [2]. A major challenge in planning the study was to measure the patient's pain and disability, and its response to treatment [3].

Past research studies in SOD have not used structured measures of pain to assess impact of treatments, but rather have relied on subjective global end-points such as “improved” or “very much improved”, without details of the methods used to actually measure the outcomes [4,5]. One of the largest studies of endoscopic treatment claimed success simply because most of the patients did not return to the treating center for further intervention [6].

We considered using daily diaries for the EPISOD study, but concluded that a diary was too cumbersome for the planned one year follow-up. For these reasons, we developed the RAPID (Recurrent Abdominal Pain Intensity and Disability) instrument to measure disability within the previous 90 days due to abdominal pain [7]. The RAPID closely mirrors a validated instrument used in migraine research and specialty care, the Migraine Disability Assessment (MIDAS) questionnaire [8]. RAPID is a patient-completed instrument that asks 5 questions about the effect of the abdominal pain episodes on the ability of patients to function in home/school, work and play over the prior 90 days. A total score is derived that ranges from 0 to 270 days (Supplement Table). Two supplementary questions are asked regarding the frequency and intensity of pain over the prior 90 days. Details of the rationale and the results of initial validation studies have been published [2,7]. Previous studies showed that the RAPID score had very good to excellent test-retest reliability of 0.81 and 0.95 [7]. The goal of the present study was to examine multiple measures of pain and disability in this patient population, and how they correlate with one another using the data from the recently completed EPISOD trial.

## Methods

The EPISOD study was a multicenter, randomized, blinded, sham-controlled clinical trial designed to assess the efficacy of endoscopic sphincterotomy for the treatment of patients suffering from suspected SOD. Nine U.S. centers participated in the trial that randomized a total of 214 patients to either endoscopic sphincterotomy or sham treatment (irrespective of the results of sphincter manometry) [2]. Local institutional review board approval was obtained at all participating sites and written informed consent was obtained from the patients prior to study enrollment. Patients were initially followed quarterly for one year by telephone from the central site, and completed the RAPID and SF-36 quality of life instruments. Participants also completed questions each month on a one-month recall of the frequency (days of abdominal pain) and pain intensity (scale 1-10) in the previous month. The primary study outcome was dichotomized as success or failure. Success was defined as patients having a RAPID score of <6 days at months 9 and 12 post-procedure, without re-intervention and without use of narcotic analgesics.

A supplemental grant award allowed the extension of

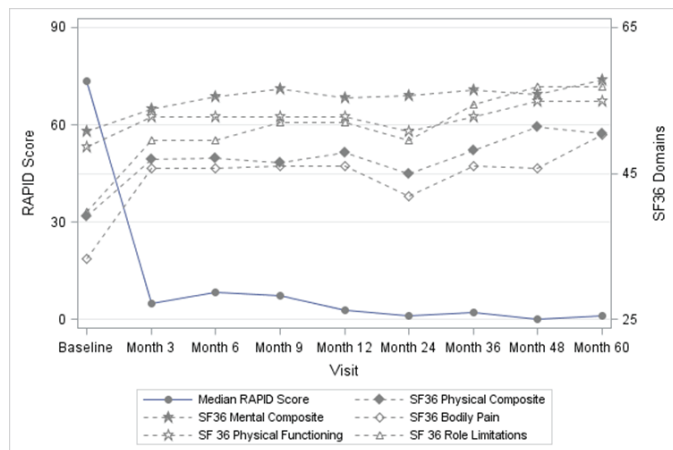
the follow-up period to a maximum of 5 years, and consenting patients continued to complete the RAPID every 6 months and SF-36 questionnaires every 12 months. They were asked also for their overall assessment of the outcome of their treatment every 6 months from Months 42 through 60, using the Patient Global Impression of Change (PGIC) instrument [10]. The PGIC asks the patient, since the initial randomized treatment, whether their overall status is: Very Much Improved; Much Improved; Minimally Improved; No Change, Minimally Worse, Much Worse or Very Much Worse. Success was defined as a patient response of ‘Much Improved’ or ‘Very Much Improved’ with no narcotic use at the last time of response and no re-intervention. The success criterion for RAPID in the long-term follow phase was a score of <6 days without narcotic use on the last recorded visit and no re-intervention.

External validity of the RAPID score (disability due to pain over the past 90 days) was examined in relation to the four-week recall of the physical domain scores of the SF-36 during the entire follow up period of EPISOD. Correlations were calculated to measure the strength of the association between the RAPID score and the SF-36 composite physical score and the domain specific scores of bodily pain, physical functioning and physical role limitation. The RAPID score and the RAPID frequency of pain question also were correlated with the monthly abdominal pain assessment (i.e., days of abdominal pain in the previous month). The monthly abdominal pain days were summed over a consecutive 3-month period and compared to the corresponding 90-day RAPID (i.e., Month 3 RAPID and Months 1-3 monthly pain days, Month 6 RAPID and Months 4-6 monthly pain days). Mean and median values of each measurement were examined by visit. All correlations were adjusted for repeated measures (corresponding baseline through all follow up visits) using a linear mixed effects model and fitted using SAS PROC MIXED [9]. For this analysis, data collected at a particular visit were excluded if either relevant measure was missing. The EPISOD study involved active- and sham-treated patients. All were included in the analysis, since treatment was not expected to impact the correlation between measurements within a study participant. The distribution of long-term outcomes as measured by RAPID and SF36 during the long-term follow up phase were compared with the corresponding PGIC response as well as the agreement between the primary definition of success. Two alternative definitions of a 50% change from baseline and a 75% change from baseline in the final RAPID score were examined. All authors had access to the study data and reviewed and approved the final manuscript. All analyses were conducted using SAS Version 9.3 or higher (SAS, Cary, North Carolina).

## Results

The EPISOD trial randomized 214 participants. Figure

1 shows the distribution of RAPID scores and Sf-36 domain scores by visit. The average number of days with disability due to abdominal pain (RAPID score) at the baseline visit was 84.5 (SD: 58.1; median: 73.5). At month 3 the mean RAPID score was 31.6 (SD: 52.4; median: 5) and remained at this level through Month 12 where it was 32.8 (SD: 56.5; median: 3). The baseline visit SF-36 physical composite score was 38.7 (SD: 7.9; median: 39.2), compared to a healthy population score of 50 (SD: 10). The score improved to 45 (SD: 8; median: 47) by Month 3 and was maintained at this level during follow up with a slight improvement at Month 12 with a mean score of 45.9 (SD: 9.6; median: 48). The additional SF-36 domain scores showed similar patterns.



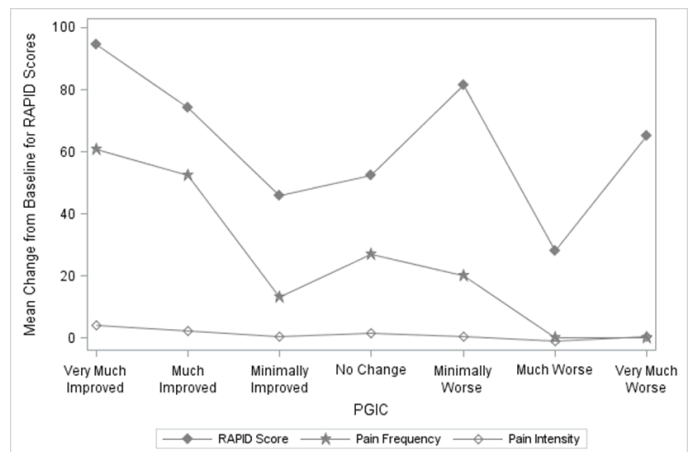
**Figure 1.** Median Instrument Scores over Visits

The correlation accounting for repeated measurements across visits between the RAPID score and the SF-36 physical composite score was  $-0.54$ . A negative correlation coefficient was expected since a higher RAPID score indicates greater role impairment and a higher SF-36 score indicates higher levels of physical functioning. The correlation of the RAPID score with specific SF-36 domain measures were  $-0.55$  for bodily pain,  $-0.41$  for the physical functioning, and  $-0.60$  for the physical role limitations (Table 1). The correlation between the RAPID score and the patient reported monthly abdominal pain question (i.e., days of abdominal pain in the previous month) was  $0.54$ . When comparing the specific RAPID question on frequency of pain in the past 90 days to the monthly abdominal pain question summed over a 3-month consecutive period, the correlation was  $0.84$ .

During the long-term follow up period, the distribution of the RAPID change from baseline scores as well as the change in pain frequency and intensity correlated well with the PGIC responses, indicating a decrease in these outcome for subjects with reports of improvement (Figure 2a). For the two cases which reported 'Very Much Worse', this was from the same subject at Months 54 and 60. Despite having a decrease in pain-related disability by almost 70 days, this subject had a re-intervention at Month 24 and reported no change in the frequency and a decrease in intensity by 1 grade. For the seven cases in the

'Minimally Worse' response category, six were from unique subjects. Two of the subjects did not have a re-intervention. One of these subjects had a baseline RAPID score of 13, which decreased to 0 days of disability by Month 54. That small change corresponded with a large decrease in the frequency (35 day difference) and intensity (decrease by 2 grades) of pain. The other subject reported 98 days of disability at baseline, which reduced to 8 days at Month 60 with no change in frequency or intensity of pain attacks. The change from baseline in the SF-36 overall physical and mental composite scores as well as the individual pain domain all coincided with higher post baseline scores (better physical/mental status) for subjects with reports of improvement (Figure 2b).

Examining the most appropriate definition of 'success' based on the final visit included comparing the PGIC success definition of 'Very Much Improved' and 'Much Improved' to the RAPID success definition as well as a change from baseline in the RAPID score of 50% and 75%. These definitions incorporate the requirement of no re-interventions and no narcotic use since the last contact period. Based on Table 2, the percent agreement by definition was 72%, 73% and 78%. The corresponding kappa values were 0.44 (95% confidence interval: 0.23, 0.65), 0.44 (0.21, 0.66), and 0.55 (0.34, 0.76).



**Figure 2a.** Mean Change from Baseline in RAPID Score, Pain Frequency and Intensity

## Discussion

The assessment of benefit is central to establishing the value of a treatment. This is especially difficult when dealing with a subjective symptom end-point such as pain, since that may be affected by many factors. There are many well-validated instruments for measurement of pain when it occurs every day, where several use a simple visual analogue scale [1]. The RAPID instrument was developed to measure disability from abdominal pains that occur only in unpredictable intermittent attacks (Supplementary Table 1). We used data from the EPISOD trial to assess the external validity of the RAPID among patients suffering intermittent abdominal pain and examined other measurements of patient pain and disability.

The RAPID score (90 day recall of days of disability

**Table 1.** Correlation Between Total RAPID Score and SF-36 Aggregate T-Score and Individual Physical Domains

| SF36 Domains   | Specific Question  | Correlation with RAPID Score |
|--|--|------------------------------|
| <b>Physical Composite</b>  | -  | -0.54                        |
| <b>Bodily Pain</b>   | How much bodily pain have you had in the past 4 weeks?<br>During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?   | -0.55                        |
| <b>Physical Functioning</b><br>The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much? Please check the circle that comes the closest to the way you have been feeling. | Vigorous activities such as running, lifting heavy objects, participating in strenuous sports:<br>Moderate activities such as moving a table, pushing a vacuum cleaner, bowling, or playing golf:<br>Lifting or carrying groceries:<br>Climbing several flights of stairs:<br>Climbing one flight of stairs:<br>Bending, kneeling, or stooping | -0.41                        |
| <b>Physical Role Limitation</b><br>During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?   | Cut down on the amount of time you spent on work or other Activities:<br>Accomplished less than you would like:<br>Were limited in the kind of work or other activities:<br>Had difficulty performing the work or other activities (for example, it took extra effort):  | -0.60                        |

**Table 2.** Agreement in Definitions of Success

| Agreement |                | RAPID <6 |         | 50% Change |         | 75% Change |         |
|-----------|----------------|----------|---------|------------|---------|------------|---------|
|           |                | Success  | Failure | Success    | Failure | Success    | Failure |
| PGIC      | <b>Total</b>   | 34       | 31      | 34         | 31      | 34         | 31      |
|           | <b>Failure</b> | 21       | 5       | 21         | 5       | 21         | 5       |
|           | <b>Success</b> | 13       | 26      | 13         | 26      | 13         | 26      |

due to pain) was compared with various external measures of general physical health and well-being from the SF36 (4week recall as well as overall health in past year) and a monthly recall of days of abdominal pain. The modest correlations between the RAPID score and the SF-36 physical composite and domain specific scores were somewhat expected as the two instruments do not measure the exact same spectrum. The SF-36 physical domain offers a broad measure of functional health and

well-being that can be influenced by the overall health status of the patient. The RAPID score is more narrowly focused on impaired role function attributable to abdominal pain episodes. As such, the SF-36 physical domain score is subject to more external influences than is the RAPID score. This was evident in the stronger correlation between the RAPID score and the more specific physical role limitation domain score of the SF36 as these two scores represent impairment to role function and focus on what a patient cannot do due to their health status. The monthly abdominal pain assessment was a measurement of the frequency of pain days and expressed as a summation of the total pain days in a three month recall period. Similar to each of the SF-36 domain scores, the abdominal pain score was moderately correlated with the RAPID Score. The validity of a 90-day recall for pain days was confirmed by the strong correlation between the overall frequency of pain reported from the RAPID and the self-reported total pain days from the monthly assessments.

The PGIC instrument allows patients to grade their subjective assessment of benefit. We showed that the responses moderately correlated with the change in days of disability, and change from baseline, as measured by the RAPID. However, it also revealed that the frequency and intensity of pain play a role in the relationship between disability and overall impression of benefit. This was a similar finding when examining the agreement between the RAPID and PGIC definitions of success, controlling for re-interventions and narcotic use. When using a strict definition of success (RAPID score <6), patients tended to report much or very much improved more often, even when the RAPID score was greater than 5 days. When the RAPID success criteria were relaxed to a 50% or 75% decrease from baseline, the disagreement between the two measures was reduced but not eliminated. Although a patient may have low disability days or a large decrease in disability, they do not always associate this with being improved.

These results highlight the complexity in defining clinically relevant outcomes when studying interventions to relieve intermittent pain. The choice of outcomes that measure the disability due to pain versus the pain itself need to be considered as the tolerability threshold for not working or participating in daily activities due to pain will vary among the population. Overall, there is utility of the RAPID instrument for measuring the disability due to pain as well as the intensity and frequency of the pain in patients with intermittent abdominal pains in the context of suspected SOD (patients with pain following cholecystectomy). It should be suitable for use in patients with gallstones, other episodic painful abdominal conditions, and indeed for diseases in other systems.

Since this work was completed, there have been publications reviewing published patient-reported outcomes for GI diseases (but not mentioning RAPID) [11], and reporting the "Development of the NIH Patient-Reported

Outcomes Measurement Information System (PROMIS) Gastrointestinal Symptom Scales" [12]. The latter includes a measurement tool for "belly pain" which asks patients about their pain over only the prior 7 days. Whilst this might be sufficient for patients with some functional digestive disorders, it would not be for the biliary pains, which can be much less frequent. The current and previous studies confirm the reliability of the 90 day recall for the RAPID instrument, which makes it easy to apply in practice.

A question that needs more study and discussion is what decrease in the RAPID score should define "successful" treatment. That likely will depend on the patient population, the severity of the symptoms and the perceived likely burdens and risks of the treatment. The EPISOD study required a decrease to a score of <6 days of disability (which turned out to be a reduction from baseline of roughly 94%). We had planned a stringent criterion because of the risks involved in the treatment, but, in retrospect, that may have been too high a hurdle. However, that did not impact the results of the trial since the active and sham arms fared the same, even when we explored the use of lesser criteria, i.e., reduction in RAPID of 50% and 75%. In a comparable situation, patients undergoing cholecystectomy for "gall bladder dyskinesia", we found significant variations in patient's expectations (13). Only 50% stated that their pain would have to be removed completely for them to judge the treatment as successful. Perhaps not surprisingly, the others felt that partial relief would be worthwhile, assuming no serious adverse events.

## Conclusion

The objective measure of success, the RAPID score, correlated well with the patient's subjective assessment using the PGIC instrument. We recommend that both be used in future studies of treatment for episodic abdominal pains.

## Acknowledgements and Disclosures

This work was supported by National Institutes of Diabetes, and Digestive and Kidney Diseases (NIDDK, grant U01 DK074739). Durkalski-Mauldin and Pauls have no potential conflicts; Cotton consults for Olympus America and Cook Medical, and receives royalties from Cook Medical for devices not used in the study. We would like to thank the EPISOD study team for their efforts in the planning and conduct of the trial and the collection of the data.

## Clinical Trial Registration

This study is registered at <https://clinicaltrials.gov/ct2/show/NCT00688662?term=EPISOD&rank=1>. The registration identification number is NCT00688662.

## References

1. Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*. 2005;113(1-2):9-19.
2. Cotton PB, Durkalski V, Romagnuolo J, et al. Effect of endoscopic sphincterotomy for suspected sphincter of Oddi dysfunction on pain-related disability following cholecystectomy – the EPISOD randomized clinical trial. *JAMA* 2014; 311 (20): 2101-2109.
3. Cotton PB, Durkalski V, Orrell KB, et al. Challenges in planning and initiating a randomized clinical study of sphincter of Oddi dysfunction. *Gastrointest Endosc*. 2010;72(5):986-991.
4. Petersen BT. Sphincter of Oddi dysfunction, part 2: Evidence-based review of the presentations, with "objective" pancreatic findings (types I and II) and of presumptive type III. *Gastrointest Endosc*. 2004;59(6):670-87.
5. Petersen BT. An evidence-based review of sphincter of Oddi dysfunction: part I, presentations with "objective" biliary findings (types I and II). *Gastrointest Endosc*. 2004;59(4):525-34.
6. Park SH, Watkins JL, Fogel EL, et al. Long-term outcome of endoscopic dual pancreatobiliary sphincterotomy in patients with manometry-documented sphincter of Oddi dysfunction and normal pancreatogram. *Gastrointest Endosc*. 2003;57(4):483-91.
7. Durkalski V, Stewart W, MacDougall P, et al. Measuring episodic abdominal pain and disability in suspected sphincter of Oddi dysfunction. *World J Gastroenterol*. 2010;16(35):4416-4421.
8. Stewart W, Lipton R, Kolodner K, et al. Validity of the Migraine Disability Assessment (MIDAS) score in comparison to a diary-based measure in a population sample of migraine sufferers. *Pain*. 2000; 88:41-42.
9. Hamlett A, Ryan L, Wolfingner R. On the use of PROC MIXED to Estimate Correlation in the Presence of Repeated Measures. SUGI 29 Statistics and Data Analysis. Paper 198-29. Available from: URL: [www2.sas.com/proceedings/sugi29/198-29.pdf](http://www2.sas.com/proceedings/sugi29/198-29.pdf).
10. Farrar JT, Young JP, LaMoreaux L, et al. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain*. 2001; 94: 149-158.
11. Khanna P, Agarwal N, Khanna D, et al. Development of an Online Library of Patient-Reported Outcome Measures in Gastroenterology: The GI-PRO Database. *Am J Gastroenterol*. 2014: 109:234-248.
12. Spiegel BMR, Hays RD, Bolus R, et al. Development of the NIH Patient-Related Outcomes Measurement Information System (PROMIS) Gastrointestinal Symptom Scales. *Am J Gastroenterol* 2014;109: 1804-1814.
13. Suarez AL, Kutlu O, Cunningham SC, et al. Tu1517 How Much Pain Relief Do Patients Expect After Cholecystectomy? *Gastroenterology*. 2016;150:S1257.

To cite this article: Durkalski-Mauldin V, Pauls Q, Cotton P. Measuring outcomes of treatment for patients with episodic abdominal pains; conclusions from the EPISOD study. *British Journal of Gastroenterology*. 2019; 1:1.

© Durkalski-Mauldin V, et al. 2019.