World Journal of Gastroenterology
Reply to reviewer's comments

Thank you very much for reviewing our manuscript.
We have read the comments carefully and have made the corrections required.

To editor in chief

Thank you very much for your gentle comments.

I made the title of the manuscript shorter, and I will resubmit my manuscript according to your instructions.

To reviewer

(1) - The Authors found a major number of patients with ERD than NERD and this is the opposite of what happens in western countries. Is there an explanation for this difference?

Thank you very much for your valuable comment. Indeed, the proportion of ERD patients were somehow larger than that of NERD patients in this study against to previous reports. We have no potent explanation for this issue. Since as many as 29 institutions participated in the present study, we believe that our results also reveal a certain aspects of GERD patients in Japan.

(2) - The rate of response in terms of GERD symptoms is very high in this study, even though the overlapping with dyspeptic symptoms is also frequent and present in about two thirds of patients. As it is well known from many clinical studies that the concomitant presence of GERD with dyspepsia or irritable bowel symptoms is predictive of a poor response of reflux to PPIs, how can the Authors explain the good results they obtained?

Thank you very much for your relevant comment. Surely, as you mentioned, that NERD and concomitant dyspeptic symptoms were the significant predictive factors of refractory GERD to PPIs therapy in our previous study too (reference 23). Comparatively high proportion of ERD patients in our study may, in part, explain the good response to PPIs.

(3) - Can this simple and effective questionnaire be translated and applied in western countries or its use is limited to Asiatic populations?

The GERD-TEST has developed for diagnosing and for evaluating the therapeutic efficacy of GERD. The English version of the GERD-TEST is available and is attached to this manuscript as Table 1. The authors hope that the GERD-TEST will be utilized for GERD patients in the western countries as well as eastern countries.

(4) - This reviewer suggests to add to the reference list a relevant paper by Savarino E et al (GUT 2009; 58:1185-1191) showing the frequent overlap of functional dyspepsia with functional heartburn.

Thank you very much for precious comment. I quoted the relevant paper by Savarino E et al as reference 12.

(5) - The discussion could be reduced in length by one third

Thank you very much for your valuable comment. I have deleted the redundant sentences or even paragraphs, and then, made the discussion as short as possible.

The deleted sentences or paragraph, and added sentences is marked as red letter below.

DISCUSSION

Approximately 85% of reports from GERD patients recruited under the Montreal

definition were diagnosed as having GERD based on the results of the GERD-TEST, providing evidence in support of the diagnostic usefulness of the GERD-TEST. The Cronbach's α for GERD-SS, FD-SS, and DS-SS in the GERD-TEST ranged from 0.75 to 0.82, indicating a superior internal consistency and high reliability. Significant correlations were observed between symptom or living status items/subscales of the GERD-TEST and the PCS or MCS of the SF-8, ~~as well as between the symptom items/subscales and the living status items/subscales of the GERD-TEST,~~ demonstrating a good convergent validity. Both GERD and FD symptoms were seen to have a clear and consistently negative impact on the daily lives of patients, and this impact increased with increasing symptom severity (Table 4). There was a significant and marked reduction in GERD symptoms in response to the 4-week PPI therapy. Improvements in FD symptoms and daily living status were also significant, though to a lesser extent than the amelioration of GERD symptoms. Thus, the responsiveness of the GERD-TEST to these improvements was gratifying. A comparison between responders and non-responders according to three definitions of responders (a residual symptom rate ≤ 50%, a patient's impression that was "improved" or better, and an NRS score ≤ 5) revealed significant and substantial differences in GERD symptoms between these two groups, thereby indicating that the GERD-TEST has a satisfactory concurrent validity.

The GERD-TEST enabled a multifaceted evaluation not only of the severity of symptoms, but also of the impact of the symptoms on daily life, the therapeutic response as assessed by the patient. The GERD-TEST is expected to be a useful diagnostic/treatment tool for both clinical research and in daily clinical practice settings, since it consists of relatively few items and subscales that are readily understandable and enable the detection of concurrent FD symptoms.

Symptoms of FD are often seen in patients with GERD [7-11]. The present study results showed that concurrent FD symptoms were noted in as many as 76% of the patients with GERD who met the Montreal definitions, and this finding is consistent with previous reports [7-11]. Symptoms of GERD are generally known to affect various aspects of daily living [5, 6], and symptoms of FD have similarly been reported to interfere with the daily living status of patients [12-14], resulting in a reduction in QOL. In the present study, the results of a correlation analysis revealed that both GERD and FD symptoms impair the daily life of patients, affecting eating, sleeping, daily activity and mood (Table 4); these results support those reported by others[5, 6, 12-14]. Even if a patient presents with a chief complaint of GERD symptoms at the time of their first visit, the possibility that the patient's QOL might be lowered because of concurrent FD and GERD symptoms still exists. Therefore, cases should be carefully selected by observing both FD symptoms and GERD symptoms, and appropriate treatment aimed at treating the former condition should also be administered simultaneously.

~~The 4-week PPI therapy enabled a significant and marked improvement of GERD symptoms. There also was a significant improvement in FD symptoms, concurrent with the GERD symptoms, in response to the PPI therapy; however, the effect size in terms of Cohen's *d* was somewhat smaller than that for the GERD symptoms and was limited, particularly with respect to early satiation (Table 5). These findings indicate a need to strengthen treatment with additional prokinetic regimens in GERD patients presenting~~

~~with concurrent severe early satiation.~~

~~Regarding DS, a significant improvement was observed following the 4-week PPI therapy for all items of DS as well as for the daily life subscale parameters. A comparison between responders and non-responders according to the three responder definitions disclosed a significantly greater change in the scores for the DS-SS and practically all items in the responder group, suggesting that the improvement in GERD symptoms enabled an improvement in the daily living status.~~

Inasmuch as it is often difficult to identify concurrent FD symptoms in patients with GERD, the use of an appropriate PRO might enable such symptoms to not be overlooked, allowing appropriate treatment to proceed. Based on the assumption that GERD and FD are diseases with a spectrum of overlapping symptoms [7-11], the GERD-TEST may allow clinicians to use only one PRO instrument to measure health-related QOL outcomes in patients with GERD, FD, or overlapping symptoms of both conditions.

The use of an appropriate PRO tool for which both reliability and validity have been verified is recommended to ensure evidence-based evaluations of the usefulness of a treatment for disorders such as GERD and FD, where the treatment is primarily aimed at symptomatic improvement [15]. Many PRO tools have been developed and applied in various clinical trials as well as in daily clinical practice settings for the diagnosis of GERD and for evaluating therapeutic responses [17, 19]. ~~In daily clinical practice settings as well, it is important to select a treatment program suited to each individual patient after appropriately diagnosing GERD, evaluating the degree of disease severity and its impact on the patient's daily life, assessing the efficacy of the treatment from the patient's viewpoint, and reflecting on the desired outcome of the treatment (i.e., making decisions regarding treatment fortification, modification, continuation, stepping down, or completion according to patient-based evaluations). Thus, a PRO can be a clinically relevant outcome measure of disease impact and treatment response in both clinical trials and primary care.~~ The practical use and dissemination of PRO as a diagnostic and evaluation tool is anticipated; however, most PROs are lengthy and complicated, and a simple and effective PRO was previously unavailable. The GERD-TEST was developed for this reason.

The goal of treatment for NERD, FD and irritable bowel syndrome (IBS) lies in improving symptoms and signs characteristic of each of these disorders and thereby lessening a patient's sense of burden and impairment of daily living activities. A variety of sets of criteria have been used to evaluate responses to pharmacotherapies for those disorders. Global binary endpoints (a method in which an alternative response to each question is provided, i.e., whether an adequate or satisfactory relief of symptoms has or has not been obtained) and a "residual symptom rate ≤ 50%" have both exhibited an intense convergent validity and are capable of detecting clinically significant but minimal changes [26]; therefore, these variables are recommended [18, 27-29].

~~The evaluation of therapeutic responses using a Likert scale can be performed not only for individual symptoms [30-32], but also as a comprehensive evaluation procedure [27,~~

~~32-34]. One report purportedly showed that a 5-grade or 7-grade Likert scale was more sensitive than a 4-grade scale [35]. Assessments of the overall treatment effect (OTE) using a comprehensive evaluation with a Likert scale, in which the degree of symptomatic amelioration is assessed using a 5-grade or 7-grade Likert scale in comparison with baseline data, are recognized as effective evaluation procedures and have recently been used in clinical trials.~~

A NRS, which is mainly used to evaluate therapeutic responses in patients with chronic pain [36], has been proposed by the FDA as a provisional scale for evaluating abdominal pain in patients with intractable bowel syndrome [37]. An NRS has been recognized as having "higher compliance rates, better responsiveness and ease of use, and good applicability relative to a visual analogue scale."

~~A variety of sets of criteria for evaluating therapeutic responses have been proposed, but opinion as to which set is best remains divided. Since various reports on such criteria and on the evaluation of therapeutic responses have been published in recent years, the use of a set or sets of criteria recommended by these reports seems reasonable.~~

For evaluating the burden by the symptoms as well as the response to the therapy, ~~T~~the GERD-TEST can be applied using three definitions: i.e., a 7-grade Likert scale for individual symptoms, the patient's impression of the therapy (which corresponds to the OTE), and the NRS (as recommended by various reports and guidelines), and interestingly, the global assessments of the GERD symptoms using patient's impression of the therapy (Q11) and NRS (Q12) well differentiated the FD symptoms (Table 9). Therefore, evaluations of patient burden arising from various symptoms and of the comprehensive therapeutic response using this tool are thought to be appropriate. ~~An actual comparison of the degree of amelioration of GERD symptoms as assessed using the three therapeutic response evaluation definitions between responders and non-responders revealed that all three definitions resulted in significant findings with a sufficient effect size, indicating the suitability of these definitions for evaluations of therapeutic response. The respective Cohen's *d* values, i.e., the effect sizes of the 4-week PPI therapy, for the residual symptom rate, NRS, and patient's impression were as follows: 1.76, 0.90, and 0.78 for Q1 (heartburn); 1.51, 0.75, and 0.79 for Q2 (regurgitation); and 1.89, 0.91, and 0.87 for GERD-SS, respectively. Hence, the values were largest for the residual symptom rate, followed by NRS and patient's impression. Whether these three therapeutic response evaluation definitions depict the same aspect of treatment-produced symptomatic amelioration or instead reflect different aspects remains uncertain. Our previous report stated that significant, but not so strong, correlations were observed among these three therapeutic response evaluation definitions (Pearson's *r*: 0.46-0.61) [22]. Furthermore, the present data showed differences in the percentages of treatment responders among the three therapeutic response evaluation definitions (Figs. 2-4), and differences between responders and non-responders were also observed in terms of the degree of changes in the score and the effect size for individual items/subscales of the GERD-TEST (Table 6-8). From these findings, the three evaluation definitions were considered to be not entirely identical in quality to patient-reported evaluations, suggesting that patients interpret their symptoms in different ways.~~

Of the plurality of therapeutic response evaluation definitions currently available, none have been shown to be optimal for the evaluation of therapeutic responses during the management of GERD. It is thus considered preferable to report data obtained and analyzed using two or more therapeutic response evaluation definitions, rather than any single definition~~, since the available definitions are thought to depict different aspects~~.

The limitations of this study were, firstly, the clinical responses in terms of the GERD symptoms can be evaluated using three definitions in the GERD-TEST; these definitions were formulated chiefly for the diagnosis and treatment of GERD. Concurrent FD symptoms, however, can only be evaluated using a residual symptom rate. The GERD-TEST should be modified to include the patient's impression and NRS items, similar to the GERD symptom evaluations, to make this definition even more useful for the diagnosis and treatment of FD. Secondly, it is generally recognized that patients with GERD or FD present with diverse symptoms. Among patients with GERD, non-typical symptoms such as esophageal symptoms (e.g., chest pain) and extraesophageal symptoms (e.g., chronic cough, chronic laryngitis, asthma or dental erosion [1]) are often seen. Symptoms such as bloating, belching or nausea also develop among patients with FD. Clinical evaluation using the GERD-TEST is focused primarily on the cardinal symptoms of GERD and FD, and the evaluation does not cover patient burden from other symptoms or the impacts of such symptoms on daily life. Further investigation and clarification of these matters is also needed.

In conclusion, the psychometric characteristics of the GERD-TEST were excellent, demonstrating good validity and reliability. The GERD-TEST is simple~~,~~ and easy to perform and is a multifaceted PRO instrument that appears to be useful for evaluating disease-specific health-related QOL in GERD patients in both clinical trial and primary care settings.

I should like to express my appreciation to editor in chief and reviewer for suggesting how best to improve our papers.

Sincerely,

Koji Nakada, MD. PhD. Associate Professor

Invited manuscript ID (03262754)

Department of Laboratory Medicine, Daisan Hospital Jikei University School of Medicine

4-11-1, Izumihoncyo, Komae-shi, Tokyo, 201-8601 Japan

Telephone: +81-3-3480-1151 ext. 3401 FAX: +81-3-3480-5700

E-mail: nakada@jikei.ac.jp