Iranian J Publ Health, Vol. 37, No.2, 2008, pp.64-74

Original Article

Reliability, Validity, and Feasibility of the Mayo Gastro-Esophageal Reflux Questionnaire (GERQ) in a Persian-Speaking Population"

**S Nasseri-Moghaddam[1], H Razjouyan[1], R Habibi[1], K Rafaat-Zand[1], B Ahrari[1], M Nouraie[1], R Majdzadeh[2], H Vahedi[1], R Malekzadeh[1]*

[1]*Digestive Disease Research Center, Shariati Hospital, Medical Sciences/University of Tehran, Iran*
[2]*Tehran University of Medical Sciences, School of Public Health and Institute of Public Health Research, Iran*

(Received 4 Aug 2007; accepted 16 Feb 2008)

Abstract

Background: Gastroesophageal reflux disease (GERD) is prevalent in the West. Mayo-GERQ is one of the most widely used questionnaires for screening GERD. We validated GERQ in an Iranian population.

Methods: The Mayo-GERQ was translated into Persian (P-GERQ) and reviewed and commented by two gastroenterologists. Eleven lay-people filled it in and commented on it. Reliability was assessed by test-retest within 2-6 wks in 53 hospital staff. Concurrent-validity was checked in another 53, comparing the results of the self-administered questionnaire with the questionnaires filled in by a gastroenterologist interviewing them. Weighted-kappa (κw) statistics was used. Time needed to complete the questionnaire, practicability of the directions and linguistic eloquence were checked (feasibility indices). Results were used to modify P-GERQ. The modified P-GERQ was tested in another 99 hospital employees in the same manner.

Results: Phase-one; One-hundred seventeen subjects were enrolled (46 men). Mean time for completion was 23.7 minutes. Mean *kw* for reliability was 0.47 and that for validity 0.26. Sources of poor performance were sought, P-GERQ was revised and underwent validation again (2nd phase). Phase-two: Ninety-nine individuals were enrolled (37 men). The modified P-GERQ took 15.8+/-11.9 min to complete. κ-values for concurrent-validity of major symptoms (acid-regurgitation, heart-burn) were 0.70 and 0.67 respectively. Corresponding κ-values for reliability were 0.57 and 0.80.

Conclusions: P-GERQ was not valid initially. After making appropriate technical and linguistic changes, it achieved acceptable validity, reliability and feasibility. In addition to making available a useful tool for population-based studies, our results underscore the importance of validation before adopting a translated questionnaire.

Keywords: *Gastro-Esophageal Reflux Disease (GERD), Gastro-Esophageal Reflux Questionnaire (GERQ), Feasibility, Reliability, Validity, Weighted Kappa Statistic (κw).*

Introduction

Gastroesophageal reflux disease (GERD) is a prevalent, chronic condition, with 10-30% of the Western population and 2-20% of the people in the east being affected by weekly symptoms (1-6). Recent studies have shown that GERD and its complications are increasing worldwide (6-10). GERD has already placed a significant burden on healthcare systems, as evidenced by a 3-fold increase in ambulatory care visits to primary care physicians and an almost 5-fold increase in visits to specialists for GERD between 1990 and 2001 (11). Therefore, handling GERD and its complications needs careful planning by the health-care community. Accurate epidemiological studies will be the cornerstone of this planning. Screening tools are necessary for this type of studies. There are several methods to make a diagnosis of GERD [e.g. pH metry, multi-channel intraluminal impedance measurement (MII), and upper GI endoscopy (UGIE)], but none of them is considered gold standard and most not a feasible screening tool. Currently questionnaires, especially self-administered ones, are the only feasible and reliable method of assessing GERD prevalence. They are relatively easy to administer, inexpensive, noninvasive, and do not need sophisticated tools and personal training procedures. Self-administered questionnaires have been shown to be cost effective methods in screening patients with

64 ***Corresponding author:** Tel: +98 21 88220022, Fax: +98 21 882415400, E-mail:sianm@ams.ac.ir

gastrointestinal (GI) complaints (12-18). Symptom assessment, management, and resolution remain the primary goals of medical interventions for both patients and physicians. Therefore, several questionnaires have been developed for this purpose (19), the Mayo gastro-esophageal reflux questionnaire (GERQ) being one of the popular ones (20). The GERQ is a structured questionnaire addressing major and minor GERD symptoms as well as demographic characteristics, habits, medications and past-medical and family history of major upper GI disorders. It has been shown to be valid for this purpose in other communities (21, 22).

Patient-Reported Outcomes (PRO) instruments must meet scientific standards (13) and satisfy regulatory criteria, particularly from the perspective of claims for labeling and promotion (23). The regulatory criteria are linguistic, cross-cultural adaptation and psychometric documentation. Persian translation, cross-cultural adaptation and psychometric validation of the Mayo GERQ were the goals of our study. This is done as the pre-requisite of population-based studies in Iran.

Materials and Methods

The Mayo GERQ The gastroesophageal reflux questionnaire, a self-administered instrument with 80 questions that measures symptoms during the prior year and collects a medical history, has been shown to be reliable (median kappa: 0.70) on test retesting and valid (median kappa, 0.62) in comparison with a physician interview (20). The final structure of the questionnaire is as follows:

1) Four parts assessing major GERD symptoms [namely heartburn (HB), acid regurgitation (AR), chest pain (CP), and dysphagia] during the past 12 months. Each part begins with a screening entry question. Following a positive answer to the screening question, three questions assess duration, frequency, and severity of the symptom and one question seeks the most noxious symptom of the four. Whenever the answer is "No" to the entry question, the respondent is directed to ignore the rest of the questions of that section and go to the next part.

2) Another part assesses the effect of HB and AR on daily life and medical-care seeking behavior.

3) Nine questions assess various upper GI symptoms including dyspepsia, regurgitation, globus, eructation, nausea, hematemesis, hiccups and early satiety. Five other questions assess respiratory problems, cough, bronchitis, asthma, and hoarseness. These are considered as minor GERD symptoms.

4) The rest of the questions evaluate physician/ hospital referrals during the past 12 months, use of antacids, anti-secretory agents (H-2 blockers and proton pump inhibitors), aspirin, and NSAIDs, as well as pregnancy, presence of hiatal hernias, esophageal dilatation or surgery, history of any esophageal, gastric, cardiac or pulmonary diseases. Coffee and alcohol intake, smoking, general health, and history of esophageal or gastric problems in spouse.

The study was performed in two phases:

Phase-1

Translation, cultural adaptation and primary validity and reliability testing

Preparation & Sampling The permission to use GERQ was obtained from Mayo Clinic authorities, Rochester, Minnesota, USA. It was translated to Persian by a gastroenterologist with special interest in GERD. Since alcohol consumption is legally and religiously forbidden in Iran, the question addressing alcohol consumption was omitted in order to prevent loss of cooperation and compliance. This version was presented to, evaluated and approved by two other gastroenterologists (face validity).

Due to lack of an organized national data bank, picking a sample to represent the general population was impractical. The staff of Shariati Hospital, a metropolitan hospital in Tehran (the capital city of Iran), were chosen as the sampling frame (741 subjects). The advantages included better accessibility and follow-up, and presence of a detailed data bank that made randomization and matching possible.

A pilot study was conducted on 11 randomly chosen hospital staff. Subjects were asked to fill in

the questionnaire while being timed and write down any ambiguity or difficulties faced. Later, they were individually interviewed. All comments were recorded and used to revise the Persian GERQ. These modifications are detailed in the "results" section.

From the remaining 730 employees, 106 [14.5%, 41 (38.7%) male] were randomly chosen and stratified for gender and educational level. The population's educational profile was as follows: eleven (10.4%) 5th graders, nine (8.5%) with middle-school degree, 22(20.8%) with high school diploma, 12(11.3%) technicians, 37(34.9%) Bachelor of Science, and 15(14.2%) MDs and higher.

Test: Introduced as part of the hospital general health survey, the revised questionnaire was given to all 106 subjects to be completed (self-administered). A separate section was provided for any comments or suggestions. The subjects were intentionally kept unaware of the later re-test or interview step in order to avoid biased patterns of answering and mimic the actual conditions where the questionnaire is to be used. The sample was then divided into two matched groups, 53 subjects each, one for retest and the other for gastroenterologist's interview. The time between the test and the retest/interview step was 2-6 wk.

Retest: At this step, which assessed the reliability, the questionnaires were filled in our presence since we intended to time the subjects. As before, the latter was not revealed to the subjects to avoid interference. Upon completion, each subject's questionnaire was compared with the initial one in his/her presence to identify the cause(s) of any disparity between the answers.

Interview: To test the concurrent validity, the other 53 subjects were interviewed within 2-6 wk of the initial test by the same gastroenterologist who translated GERQ and who was unaware of the subjects' previous responses or medical history. After a structured interview, based on the data obtained, a questionnaire was filled on behalf of the subject by the interviewer.

Linguistics: The linguist team was first asked to analyze the translated GERQ per se. After retest they were provided with all the comments, suggestions and ambiguities faced. The linguists' report was then generated based upon both the independent review and the information obtained.

Statistical Analysis Kappa statistics was used to assess reliability and validity of each question and the questionnaire as a whole. Kappa (κ) statistics and weighted kappa (κw) are used to compare the inter- and intra-rater variability of the reports. Conceptually, κ removes the agreement by chance and informs the clinician of the extent of the possible agreement over and above chance. If the raters agree on every judgment, the total possible agreement is always 100%. Therefore, κ values may vary from -1 (complete disagreement) to +1 (complete agreement), with a zero value pointing to agreement by chance only. Kappa values over 0.8 are considered excellent agreement and those less than 0.2 are considered as very poor. Values of 0.6-0.79 point to good, 0.4-0.59 to moderate, and 0.2-0.39 to weak agreement (24). Kappa statistics fails to differentiate smaller degrees of disagreement from the larger ones. This was compensated for by calculating the weighted kappa statistic (κ_w) where responses could be assigned an ordinal scale.

Data were analyzed using the SPSS program and descriptive statistics (means, standard deviation, and percentage) were calculated. Validity and reliability of the questionnaire was assessed by STATA version 7.0 using kappa statistics. Questions with more than two options were analyzed using weighted kappa statitics. STAT transfer 7.0 software was used to transfer SPSS data to STATA.

Agreement between answers in the test and retest represented the reliability of each question. The same stood for concurrent validity, which was estimated by comparing the results of the test against the interview.

Results of this phase showed that the Persian translation was neither valid nor reliable in this population (details explained in the "results" section). The sources of incongruity were sought by analyzing the available data and input by the linguist and the questionnaire was reworded and revised. This revised questionnaire was used in the second phase for validation.

Phase-2

The Final Questionnaire

Considering the input from phase-1, the following modifications were made and the validation process was repeated. Changes included question appearance, using color papers for each of the "entry questions" (pink, blue, green, and yellow for pages 3, 5, 7, and 9, respectively), changes of some options of some questions (Q2, 9, 10, 17, 21, 24, 30, 32, 43, 44, 57, 58, 59, 72, 73, 77, 78), deletion of some questions (Q15, 55, 56, 68, 79) and page 18, change in some statements to make them more clear, addition of a question to assess OCP usage in women and two more checking questions to assess AR and HB frequency during the last three months. The final questionnaire was assessed by two gastroenterologists and their comments were considered.

The final questionnaire, the "Modified Persian Mayo GERQ" contained 79 questions and asks the subject to rate how often 16 common symptoms occur and how bothersome they are. Symptom frequency was measured on a scale of 1 to 6 in the following categories: none in the past year, one to ten times in the past year, about once a month, about once a week, several times a week, or daily. Severity of symptoms was classified as mild, moderate, severe, and very severe. We asked about the number of aspirin as well as nonsteroidal anti-inflammatory drug (NSAIDs) tablets taken on average each week in the past year.

Cigarette use was ascertained as a history of cigarette smoking (yes vs. no) and, for current smokers, the number of packs smoked per day. Current coffee or tea (the commonly used drink in Iran) use (yes or no) and the number of cups per day were measured. Questions assessed whether any of the subject's immediate family members (mother, father, brothers, sisters, and children) or spouse had significant heartburn or disease of the esophagus or stomach. Self-reported current weight (in kilograms) and height (in centimeters) were also asked.

Study population Two sets of 50 stratified random subjects from Shariati hospital employees were recruited for the validation of the revised "Persian GERQ". The same process of checking reliability and concurrent validity described in the first phase were repeated.

Data analysis The same procedure as fore phase-1 was repeated.

Results

Phase-1

To express the results, we have divided the questionnaire into two functionally distinct sections: Questions 1-38 and questions 39 to 76.

The first 38 questions, in 4 clusters, investigate the main symptoms of GERD (heartburn, acid regurgitation, chest pain, and dysphagia) and contain 4 entry (lead) questions. Each lead question is preceded by a definition of the symptom, and followed by a "go to" direction. Next, there are individual questions that assess GERD's minor & extra-esophageal symptoms, as well as past medical and health status history of the subject.

The main reason for this separation is that we believed that the answering patterns in the first section were additionally influenced by the provided definitions and directions.

The Pilot study: The "Go to" directions were modified, complemented by more explanations to render them more user-friendly. A graphic presentation of chest wall, sternum and the area where heartburn is felt was added to the first question to make the heartburn definition more comprehensible. Three out of eleven subjects did not know the organ referred to by the word "esophagus", so the proper definition was added prior to the related questions. Sore throat was confused with dysphagia in two cases, and an explanation was offered to avoid confusion. The symptom checklist was omitted from the questionnaire, because the directions were intriguing and difficult to follow. The mean and median for completing the questionnaires were 25.7 and 23.0 minutes respectively (Range: 13.2-43.3 min).

Feasibility: The mean and median for completing the questionnaires were 23.7 min and 20.0 min respectively (range: 10.5-66.3 min). Table

1 details this variable for different educational groups. The translated directions were not understood in the same way by different respondents, therefore aberrant patterns of answering were observed. The linguists assessed the questionnaire as inappropriate for an educational level below "middle- school" (8 yr of academic education).

Table 1: Mean time for completing the questionnaire in different educational groups

	5th graders	middle-school	High school Diploma	Technician	B.Sc.	M.D. & higher
Mean time (min.)	39	26	24	20	20	16

Reliability: Table 2 compares some statistics between the original and translated questionnaires while a similar comparison between the two sections of GERQ are presented in table 3. Table 4 illustrates the separate calculations for each major symptom's question cluster.

Table 2: Reliability and validity statistics of the original versus translated GERQ

	$\kappa_{(w)}$ for reliability		$\kappa_{(w)}$ for validity	
	Translated GERQ	Original GERQ	Translated GERQ	Original GERQ
Median	0.48	0.70	0.29	0.62
25th percentile	0.36	0.60	0.08	0.49
75th percentile	0.63	0.81	0.41	0.74

Table 3: Reliability and validity statistics for the two sections of GERQ

	$\kappa_{(w)}$ for reliability		$\kappa_{(w)}$ for validity	
	1st section (qq. 1-38)	2nd section (qq. 39-76)	1st section (qq. 1-38)	2nd section (qq. 39-76)
Mean (SD)	0.36 (0.22)	0.57 (0.25)	0.25 (0.16)	0.26 (0.23)
Median	0.42	0.54	0.30	0.26
Range	-0.05-0.66	-0.09-1.00	-0.06-0.55	-0.09-0.72
25th percentile	0.28	0.39	0.10	0.06
75th percentile	0.51	0.69	0.36	0.48
90th percentile	0.63	1.00	0.47	0.55

Table 4: Reliability in four question clusters of GERQ

	$\kappa_{(w)}$ for reliability				$\kappa_{(w)}$ for validity			
	Heartburn	Acid regurgitation	Chest pain	Dysphagia	Heartburn	Acid regurgitation	Chest pain	Dysphagia
Median	0.37	0.38	0.53	-0.03	0.40	0.46	0.08	0.26
Range	0.24-0.47	0.33-0.49	0.45-0.66	-0.05--0.01	0.32-0.55	0.29-0.49	-0.06-0.18	0.07-0.34

For this primary study, the median and mean Kw for reliability was 0.48 and 0.47, respectively and the median and mean Kw for validity was 0.29 and 0.26. According to the results of this study, the questionnaire had moderate reliability and poor validity therefore it was not suitable for epidemiologic studies. These findings led to introducing modifications into the questionnaire as described above.

Phase-2

On average, the Modified GERQ took 15.8 min (SD: 11.9, range: 5-60 min) to complete. Ninety nine patients participated in the validation process (mean age: 35.8 +/- 8.1 yr, range: 18-57 yr) with a female: male ratio of 1.67: 1. More than 72% of participants were high school graduates or had a university degree. Concurrent validity and reliability testing results are presented in Table 5.

After considering the first 20 filled questionnaires, we found that the questions do not need further changes to make suitable for performing study. Two cases did not fill out the questionnaire properly, but changing in questionnaire did not seem to avoid the problem.

Feasibility:

Most interviewees did not have any problem with understanding the questions and responding to them as directed. Mean time for filling the questionnaire was 15.8 (SD: 11.9) min.

Table 5: Concurrent validity and reliability of questions

	Reliability N= 49			**Validity N= 50**		
	Overall agreement (%)	**K or Kw**	**95% CI**	**Overall agreement (%)**	**K or Kw**	**95% CI**
Heartburn, presence	95	0.80	0.54-1.00	85	0.67	0.39-0.95
Frequency (12Ms)	96	0.79	0.24-1.00	88	0.50	0.01-0.99
Frequency (3Ms)	96	0.78	0.22-1.00	89	0.56	-0.01-1.00
Severity	94	0.00*	-0.51-0.51	93	0.56	0.01-1.00
Nocturnal	85	0.57	0.20-1.00	83	0.66	0.63-0.68
Extension to neck	75	0.47	0.09-1.00	92	0.80	0.25-1.00
Response to antacid	85	0.75	0.37-1.00	73	0.48	0.10-0.78
Duration	100	1.00	0.41-1.00	96	0.77	0.19-1.00
Acid regurgitation, presence	77	0.57	0.32-0.82	85	0.70	0.13-1.00
Frequency (12Ms)	91	0.72	0.29-1.00	83	0.38	-0.05-0.81
Frequency (3Ms)	95	0.68	0.26-1.00	86	0.38	-0.03-0.79
Severity	89	0.76	0.31-1.00	88	0.23*	-.10-0.55
Nocturnal	86	-0.07*	-0.47-0.33	85	0.48	0.05-0.91
Duration	96	0.83	0.38-1.00	91	0.69	0.23-1.00
Heartburn and Acid regurgitation, either condition present	87	0.72	0.45-0.98	85	0.68	0.41-0.94
either condition interrupted daily activities,	96	0.74	0.34-1.00	89	0.42	0.09-0.76
either condition prompted visit to physician	89	0.44	0.01-0.89	86	0.67	0.24-1.00

Table 5: continued...

No. of physician visits for either condition	N.A	N.A	N.A	N.A	N.A	N.A
Reason for physician visits for either condition	N.A	N.A	N.A	40	0.00	N.A
Test done for either condition	N.A	N.A	N.A	100	1.00	0.12-1.00
Chest pain, presence	89	0.68	0.41-0.94	70	0.26	-0.02-0.54
Frequency	90	0.66	-0.05-1.00	94	0.76	0.06-1.00
Severity	82	0.34*	-0.08-0.76	85	-0.22*	-0.86-0.41
Duration of individual episodes	85	0.72	0.01-1.00	86	0.42*	-0.32-1.00
Exacerbation by cold or warm drinking	100	1.00	0.31-1.00	71	0.30*	-0.44-1.00
Concomitant dysphagia	100	1.00	0.31-1.00	80	0.55*	-0.24-1.00
Pleuritic pain	71	0.46	-0.16-1.00	33	0.00*	NA
Provoked by heavy exertion	75	0.33*	-0.36-1.00	57	0.16*	-0.55-0.87
Provoked by light exertion	75	0.47*	-0.23-1.00	57	-0.24*	-0.91-0.44
Identified by physician as cardiac	0.62	0.44*	-0.01-0.89	67	0.50	0.01-0.99
Duration	98	0.92	0.20-1.00	83	0.00*	0.00-0.00
Dysphagia, presence	96	0.86	0.60-1.00	96	0.78	0.50-1.00
Frequency	82	0.07*	-0.56-0.60	87	0.50*	-0.15-1.00
Severity	94	0.65	-0.04-1.00	75	0.50*	-0.35-1.00
Odynophagia	75	0.50	-0.09-1.00	50	0.20*	-0.39-0.79
Solids, liquids, or both	87	0.78	0.27-1.00	50	0.33*	-0.07-0.73
Progressive	100	1.00	0.31-1.00	N.A	N.A	N.A
Intermittent (y/n)	100	1.00	0.31-1.00	75	0.50*	-0.35-1.00
Duration	100	0.69	0.06-1.00	72	0.00*	-0.88-0.88
Upper gastrointestinal complain						
Abdominal pain	77	0.57	0.32-0.81	70	0.24	-0.02-0.50
Pain more than 6 times/yr	67	0.28*	-0.14-0.70	83	0.67	-0.15-1.00
Upper or lower abdomen	65	0.45	0.15-0.76	80	0.69	0.09-1.00
Severity of pain	86	0.13*	-0.32-0.58	90	0.00	NA
Food regurgitation	91	0.78	0.49-1.00	65	0.22	-0.03-0.46
Nausea	98	0.75	0.45-1.00	96	0.61	0.33-0.89
Vomiting	97	0.69	0.40-0.99	98	0.63	0.39-0.89
Hematemesis	98	0.66	0.36-0.95	N.A	N.A	N.A
Main symptom	67	0.58	0.95-0.72	52	0.38	0.24-0.52
Respiratory symptoms, chest sounded wheezy	96	0.66	0.37-0.95	84	-0.05*	-0.29-0.21

Table 5: continued...

wheezing leading to shortness of breath	90	0.57	0.36-0.79	77	-0.04*	-0.17-0.10
Dyspnea on exertion	73	0.35	0.06-0.64	71	0.17	0.01-0.33
Hoarseness	82	0.50	0.23-0.79	76	0.13	-0.15-0.41
Globus sensation	96	0.64	0.35-0.93	84	0.15*	-0.12-0.42
Burping	74	0.36	0.06-0.64	77	0.26	0.03-0.48
Early satiety	85	0.67	0.39-0.95	78	0.05*	-0.23-0.33
Hiccups	87	0.62	0.33-0.90	90	0.26	0.07-0.95
Cough	83	0.53	0.25-0.81	85	0.32	0.11-0.52
Cough frequency	100	1.00	0.31-1.00	100	1.00*	-0.39-1.00
Nocturnal Cough	62	0.25*	-1.42-1.00	50	0.00	0.00-0.00
Medications, use of aspirin	96	0.34	0.07-0.61	97	0.62	0.35-0.90
Nonsteroidal anti-inflammatory agent	96	0.63	0.34-0.93	88	0.40	0.11-0.68
Use of other medication	81	0.51	0.22-0.81	87	0.55	0.28-0.83
Past Dx and Rx						
Hiatal Hernia	100	1.00	0.71-1.00	N.A	N.A	N.A
Disease of the esophagus or stomach	100	1.00	0.71-1.00	91	0.55	0.28-0.83
Dilation of esophagus	100	1.00	0.70-1.00	N.A	N.A	N.A
Operation of esophagus or stomach	100	1.00	0.70-1.00	N.A	N.A	N.A
Heart disease Diagnosed	100	1.00	NA	91	0.67	0.40-0.94
Heart disease treated	98	0.90	0.61-1.00	0.94	0.69	0.42-0.97
Asthma	NA	NA	NA	93	0.00	0.18-0.79
Miscellaneous factors						
Spouse with heartburn	78	0.56	0.28-0.85	76	0.46	0.17-0.73
Family member with heartburn	84	0.68	0.36-1.00	73	0.43	0.15-0.71
In general, overall health	97	0.76	0.48-1.00	97	0.45	0.58-1.00
Elevate head of bed	100	0.96	0.67-1.00	98	0.79	0.50-1.00
Use of cigarettes	98	0.95	0.67-1.00	96	0.86	0.58-1.00
Use of tea	98	0.00	NA	93	0.00	0.00-0.00
Use of coffee	97	0.48	0.22-0.67	79	0.53	0.26-0.79
Female specific						
Pregnancy	90	0.63	0.36-0.89	100	1.00	0.64-1.00
OCP	73	0.56	0.29-0.89	100	1.00	0.64-1.00

* Not significant *P* values

Validity Fifty of 55(91%) invited participants in the validity process accepted to complete the study process. All were interviewed within 2-6 wk after the initial questionnaire was completed. The majority of validity values for the questionnaire were in the fair to excellent range. There was fair to good concordance between the physician interview and the questionnaire for the presence of Heartburn and all the other related aspects of the symptom. The kappa value for the major symptoms including AR, HB, CP, and Dysphagia were 0.70, 0.67, 0.26, and 0.78, respectively.

Reliability Forty nine patients participated in the reliability phase of this study. The mean time between completion of the first and second questionnaires was 2-6 wk. Most items had reliability values of good to excellent. The kappa values or the presence of four major symptoms i.e. AR, HB, CP, Dysphagia were 0.57, 0.80, 0.68, and 0.86, respectively.

General assessment The mean kappa values for reliability and validity of the questionnaire as a whole were 0.64 (SD: 0.26, 95% CI: 0.58-0.70) and 0.44 (SD: 0.29, 95% CI: 0.37-0.50), respectively. Median and mean of the questions related to the four major symptoms are shown in Table 6.

Table 6: The median and mean values for reliability and validity of major symptoms

	Reliability		Validity	
	Mean	**Median**	**Mean**	**Median**
Heartburn	0.64	0.76	0.62	0.61
Acid Regurgitation	0.61	0.70	0.53	0.58
Chest pain	0.64	0.66	0.23	0.26
Dysphagia	0.70	0.74	0.40	0.50

Discussion

We have translated and tested the Mayo GERQ for use in the general Iranian population to assess GERD. The initial translation was not reliable; therefore modifications were made according to the input from the first phase of the study to make the "Persian GERQ" understandable and more feasible. The general validity and reliability of the questionnaire were fair to good. Kappa values for major GERD symptoms were acceptable, therefore it seems that the modified "Persian GERQ" in its present format is suitable for population-based studies in Iran. The kappa value for chest pain was rather poor in the initially translated questionnaire. Although it improved with modifications (from 0.07 to 0.27), but it still was not considered valid. This may be due to the low prevalence of this symptom, hence making it unfamiliar to most subjects. In addition, it may have been mistaken with heartburn i.e. heartburn overshadows the chest pain moiety.

The reproducibility of minor GERD symptoms was not so good in our final questionnaire. Again this may be due to the fact that theses symptoms are less common and more transient than the major GERD symptoms; therefore assessing their reproducibility in a relatively small sample may be problematic. This study was done in a population with low prevalence of GERD. Had the study been performed in patients referring to a GI clinic, better kappa values may have been achieved and minor symptoms assessed as well. The Modified GERQ is an instrument with fair to excellent validity and reproducibility for the features of GERD. Although a number of validated instruments have been introduced, yet a succinct instrument is lacking (19). No currently available instrument quantitates a number of key GERD elements (including when symptom was first noted, its frequency and severity). The modified GERQ takes sixteen minutes on average to be completed. This is acceptable, but if could be even shortened, it may have been more feasible. Overall, the κ-values of the modified GERQ were high and well acceptable for both reproducibility and validity. Nonetheless, the range between the highest and lowest values was broad, likely due to the sample size of the study. Given that the perception of major symptoms is a very personal experience, and given that the κ ranges of many items were similar to, if not better than, those demonstrated by other validated instruments, we anticipate the Modified GERQ will allow researchers and clinicians to create a uniform language for characterizing patients with GERD symptoms.

GERQ is considered a rather valid and useful tol for assessing GERD in population based studies (20). As demonstrated by our data and supported by others findings, simple translation and peer review of a questionnaire is not adequate to adopt a questionnaire developed for another language and culture regardless of its weight in the original language. Therefore, it is absolutely necessary to test it for appropriateness in measuring the interested items. A self-administered questionnaire should be understood and completed in a reasonable time by the target population, besides being reliable and valid.

In summary, the Persian GERQ is a comprehensive instrument with acceptable validity and reproducibility for the diagnosis and measurement of GERD related symptoms. This can be used for population based studies in Persian speaking populations in Iran. Studies to shorten the questionnaire and making it more user-friendly may be warranted.

Acknowledgments

The authors would like to express their thankfulness to Dr GR Locke from the Mayo clinic for providing the questionnaire and helping in obtaining institution permit to translate and use the GERQ. In addition we hereby express our gratitude to our linguistic team, Dr Kourosh Safavi, Dr Kambiz Mahmoodzadeh and Dr Azita Afrashti for their generous contribution to this work. In addition, we are indebted to Dr Shahin Merat, Mr Mehrdad Shahin-Mehr, and Miss Mina Minapour for their contribution in various parts of the study including the statistical analysis. The present study was supported by a grant from the Digestive Disease Research Center, Shariati Hospital, Tehran University of Medical Sciences.

The authors declare that they have no Conflict of Interests.

References

1. Dent J, El-Serag H, Wallander M (2005). Epidemiology of gastroesophageal reflux disease: a systematic review. *Gut*, 54:710-17.
2. Camilleri M, Dubois D, Coulie B, Jones M, Kahrilas PJ, Rentz AM, et al. (2005) Prevalence and socioeconomic impact of upper gastrointestinal disorders in the United States: results of the US Upper Gastrointestinal Study. *Clin Gastroenterol Hepatol*, 3:543-52.
3. Nouraie M, Radmard AR, Zaer-Rezaii H, Razjouyan H, Nasseri-Moghaddam S, Malekzadeh R (2007). Hygiene Could Affect GERD Prevalence Independently: A Population-Based Study in Tehran. *Am J Gastroenterol*, 102 (7):1353-60.
4. Hu WH, Wong WM, Lam CL, Lam KF, et al. (2002). Anxiety but not depression determines health care-seeking behavior in Chinese patients with dyspepsia and irritable bowel syndrome: a population-based study. *Aliment Pharmacol Ther*, 16:2081-88.
5. Wang Jin-Hai, Luo Jin-Yan, Dong Lei, Gong Jun, Tong M (2004). Epidemiology of Gastroesophageal reflux disease. A general population-based study in Xi'an of Northwest China. *World J Gastroenterol*, 10:1647-51.
6. Nouraie M, Razjouyan H, Assady M, Malekzadeh R, Nasseri-Moghaddam S (2007). Epidemiology of Gastroesophageal Reflux Symptoms in Tehran, Iran; a population-based Telephone Survey. *Arch Iranian Med*, 10(3): 289-94.
7. Holtmann G (2001). Reflux disease: the disorder of the third millennium. *Eur J Gastroenterol Hepatol*, 13(suppl 1):S5–S11.
8. Delaney B (2004). Review article: prevalence and epidemiology of gastroesophageal reflux disease. *Aliment Pharmacol Ther*, 20(suppl 8):2-4.
9. Goh K (2004). Changing epidemiology of gastroesophageal reflux disease in the Asian-Pacific region: an overview. *J Gastroenterol Hepatol*, 19(suppl 3):S22–S5.
10. Malekzadeh R, Nasseri-Moghaddam S, Sotoudeh M (2003). Gastroesophageal reflux disease: the new epidemic. *Arch Iranian Med*, 6:127-40.

11. Altman K, Stephens R, Lyttle C, Weiss KB (2005). Changing impact of gastroesophageal reflux in medical and otolaryngology practice. Laryngoscope, 115:1145–53.
12. McColl E (2004). Best practice in symptom assessment: a review. *Gut*, 53(Suppl.4):49–4.
13. Lohr K (2002). Assessing health status and quality-of-life instruments: attributes and review criteria. *Quality of Life Research*, 11:193-205.
14. Talley NJ, Zinsmeister AR, Van Dyle C, Melton LJ III (1991). Epidemiology of colonic symptoms and the irritable bowel syndrome. *Gastroenterology*, 101:927-34.
15. Talley NJ, Zinsmeister AR, Schleck CD, Melton LJ III (1992). Dyspepsia and dyspepsia subgroups: a population-based study. *Gastroenterology*, 102:1259-68.
16. Talley NJ, O'Keefe EA, Zinsmeister AR, Melton LJ III (1992). Prevalence of gastrointestinal symptoms in the elderly: a population-based study. *Gastroenterology*, 102:895-901.
17. Talley NJ, Weaver AL, Zinsmeister AR, Melton LJ III (1992). Onset and disappearance of gastrointestinal symptoms and functional gastrointestinal disorders. *Am J Epidemiol*, 136:165-77.
18. McColl E, Junghard O, Wiklund I, Revicki DA (2005). Assessing symptoms in gastroesophageal reflux disease: how well do clinicians' assessments agree with those of their patients? *Am J Gastroenterol*, 100:11-8.
19. Fass R (2007). Symptom Assessment Tools for Gastroesophageal Reflux Disease (GERD) Treatment. *J Clin Gastroenterol*, 41: 437-44.
20. Locke GR, Talley NJ, Weaver AL, Zinsmeister AR (1994). A new questionnaire for gastroesophageal reflux disease. *Mayo Clin Proc*, 69: 539-47.
21. Wong W, Lam K, Lai K, Hui WM, Hu WH, Lam CL, et al. (2003). A validated symptoms questionnaire (Chinese GERDQ) for the diagnosis of gastro-oesophageal reflux disease in the Chinese population. *Aliment Pharmacol Ther*, 17:1407-13.
22. Moreno Elola-Olaso C, Rey E, Rodríguez-Artalejo F, Locke GR, Díaz-Rubio M (2002). Adaptation and validation of a gastroesophageal reflux questionnaire for use on a Spanish population. *Rev Esp Enferm Dig*, 94(12):745-58.
23. Leidy NK, Revicki DA, Genesté B (1999). Recommendations for evaluating the validity of quality of life claims for labeling and promoting. *Value in Health*, 2:113-27.
24. Maclure M, Willett WC (1987). Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol*, 126:161-69.