Chinese Version of the Gastrointestinal Symptom Rating Scale: A Comparative Study of Psychometric Properties Across Different Patient Populations - Postprint

Authors: Qin Yuning, Zhao Tianyi, Liu Fengbin, Wang Xin, Cao Xue, Sun Minglin, Lai Keyun, Di Luyao, Gě Zhìshǎn, Song Liu, Xing Ying, Yang Lei, Yue Lihong, Zou Meimei, Hongjiao Li, He Liyun, Hongjiao Li, He Liyun

Date: 2022-12-01T00:00:00+00:00

Abstract

Background: Gastrointestinal symptoms, as common clinical evaluation indicators, require universal assessment tools. The Gastrointestinal Symptom Rating Scale (GSRS) has been widely utilized in domestic research; however, the versions employed are not standardized Chinese translations, and their psychometric properties have not been reported.

Objective: To investigate the psychometric properties of the Chinese version of the GSRS, thereby providing objective evidence for its expanded application.

Methods: From November 2021 to March 2022, patients visiting the gastroenterology or spleen-stomach departments of 45 hospitals nationwide were recruited as study subjects. Gastrointestinal symptoms were evaluated using the Chinese version of the GSRS. Reliability and validity analyses were performed. The paired Wilcoxon signed-rank test was applied to compare score changes before and after treatment, while effect size (ES), standardized response mean (SMR), and change rate (CR) were employed to assess the scale's responsiveness.

Results: The Chinese version of the GSRS exhibited a Cronbach' s α coefficient of 0.896, a Guttman split-half coefficient of 0.920, a Spearman-Brown coefficient of 0.926, an intraclass correlation coefficient (ICC) of 0.589 for test-retest reliability, and a Spearman correlation coefficient of 0.662. The item-level content validity index ranged from 0.78 to 1.00, the scale-level universal agreement content validity index was 1, and the average content validity index was 0.96. Exploratory factor analysis extracted three common factors with eigenvalues greater than 1, accounting for a cumulative variance contribution of 60.721%. Confirmatory factor analysis indicated poor fit between the data sample and the

initial model M0; however, after modification based on modification indices, all fit indices of model M1 fell within acceptable ranges. Comparison of total GSRS scores before and after treatment revealed a statistically significant difference (P<0.001), with ES=1.03, SMR=1.01, and CR=74.32%.

Conclusion: The Chinese version of the GSRS demonstrates high measurement performance with satisfactory reliability, validity, and responsiveness, making it suitable for assessing gastrointestinal symptoms in the general population and evaluating treatment outcomes.

Full Text

Preamble

Comparison of Measurement Properties of the Chinese Version of the Gastrointestinal Symptom Rating Scale in Patients with Different Diseases

QIN Yuning[1], ZHAO Tianyi[1], LIU Fengbin[2], WANG Xin[1], CAO Xue[1], SUN Minglin[1], LAI Keyun[3], DI Luyao[3], GE Zhishan[4], LIU Song[3], XING Ying[5], YANG Lei[3], YUE Lihong[4], ZOU Meimei[3], LI Hongjiao[1], *HE Liyun*[1]

[1]Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China
[2]Department of Spleen and Stomach Disease, The First Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou 510405, China
[3]The First Clinical Medical College of Shaanxi University of Chinese Medicine, Xianyang 712000, China
[4]Acupuncture Research Center, Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China
[5]Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China

Corresponding authors: HE Liyun, Professor, Doctoral supervisor; E-mail: hely3699@163.com
LI Hongjiao, Associate Professor; E-mail: lhjiao2013@163.com

Abstract
Background: The Gastrointestinal Symptom Rating Scale (GSRS) is widely used in domestic research; however, the Mandarin translation of the current Chinese version is not standardized, and its measurement performance has not been reported in any literature. **Objective:** To evaluate the reliability, validity, and responsiveness of the Chinese version of the GSRS and to provide an objective basis for its clinical and scientific application. **Methods:** Patients who visited gastrointestinal departments in 45 hospitals across China from November 2021 to March 2022 were selected using convenience sampling. Their gastrointestinal symptoms were evaluated with the Chinese version of GSRS. Reliability and validity analyses were conducted, and the paired Wilcoxon rank sum test was

applied to compare changes in patients' scale scores before and after treatment. Effect size (ES), standardized response mean (SRM), and score change rate (CR) were used to assess responsiveness. **Results:** For reliability analysis, the Cronbach' s alpha, Guttman split-half coefficient, and Spearman-Brown coefficient of the Chinese version of GSRS were 0.896, 0.920, and 0.926, respectively. The intraclass correlation coefficient (ICC) for the two retest results was 0.589, and the Spearman correlation coefficient was 0.662. The item-level content validity index (I-CVI) ranged from 0.78 to 1.00, the scale-level universal agreement S-CVI (S-CVI/UA) was 1, and the average S-CVI (S-CVI/Ave) was 0.96. Exploratory factor analysis extracted three common factors with eigenvalues greater than 1, with a cumulative variance contribution rate of 60.721%. Confirmatory factor analysis showed that the data samples did not fit the initial model M0 desirably; however, after correction using modification indices, all relevant indicators for the new model M1 were within acceptable ranges. There was a significant difference in GSRS scores before and two weeks after therapy ($P < 0.001$). The ES and SRM of the GSRS were 1.03 and 1.01, respectively, and the CR was 74.32%. **Conclusion:** The Chinese version of the GSRS is well suited for measuring the general population presenting with gastrointestinal symptoms and evaluating treatment efficacy due to its high measurement properties, including good reliability, validity, and responsiveness.

[Key words] Gastrointestinal Symptom Rating Scale; Reliability; Validity; Responsiveness; Scale

---

Gastrointestinal symptoms are among the most common clinical manifestations and are closely related to public health. In addition to being primary symptoms of gastrointestinal diseases, typical gastrointestinal symptoms may also appear in psychiatric conditions such as insomnia, depression, and anxiety, as well as in cardiovascular and cerebrovascular diseases and kidney disease. Consequently, gastrointestinal symptoms often serve as evaluation indicators in clinical research on these conditions. However, current clinical studies typically employ disease-specific scales, quality-of-life measures such as SF-36, or psychological assessment tools to evaluate gastrointestinal symptoms. Some scholars have even developed self-designed gastrointestinal symptom questionnaires based on clinical experience. In traditional Chinese medicine (TCM) research, investigators tend to select TCM syndrome assessment scales. Overall, there is currently a lack of a universal gastrointestinal symptom assessment tool.

The Gastrointestinal Symptom Rating Scale (GSRS), developed by Jan Svedlund et al. in 1988, is an integrated scale designed to assess gastrointestinal symptoms. It has been widely used internationally, translated into more than 10 languages, and has demonstrated satisfactory performance. In domestic research, many investigators have also used it as an evaluation index for interventions targeting gastrointestinal symptoms, applying it extensively in clinical studies. However, most studies have exhibited citation inconsistencies, with varying versions and linguistic expressions of the scale, and no research team

has thoroughly examined its reliability, validity, and responsiveness.

This study applied the Chinese version of GSRS to assess populations presenting with gastrointestinal symptoms and validated the scale's reliability, validity, and responsiveness to provide further objective evidence for its clinical and scientific application in China.

1.1 Clinical Data

The study population comprised patients who visited the spleen-stomach or gastroenterology departments of 45 hospitals (including community health service centers) such as Jilin Provincial People' s Hospital and Beijing Changping District Urban Community Health Service Center between October 2021 and March 2022. After screening by attending physicians according to inclusion and exclusion criteria, eligible patients voluntarily participated in the study.

Diagnostic criteria were as follows: Irritable bowel syndrome (IBS) was diagnosed using the C1 criteria from the Rome III diagnostic criteria for functional gastrointestinal disorders. Chronic enteritis was defined based on clinical practice as: (1) presence of varying degrees of abdominal pain, diarrhea, bloating, and increased stool frequency for more than 6 weeks, and (2) blood routine examination indicating an inflammatory response. Chronic gastritis was diagnosed according to the 2011 edition of the "Guidelines for the Diagnosis and Treatment of Chronic Gastritis" recommended by the China Association of Chinese Medicine.

Inclusion criteria: (1) age 18–80 years; (2) seeking medical care for gastrointestinal discomfort and diagnosed with one of the three common gastrointestinal diseases mentioned above; (3) able to cooperate with scale assessment.

Exclusion criteria: (1) concurrent diagnosis of two or more gastrointestinal diseases; (2) communication or cognitive difficulties preventing completion of the scale.

The Chinese version of GSRS contains 15 items. According to general principles of multivariate analysis, the minimum sample size should be 10 times the number of variables (items). Assuming a 20% attrition rate based on clinical and follow-up experience, the sample size was calculated as N = 180. Considering that structural validity testing requires dividing the sample into two datasets for different factor analyses, with an absolute sample size greater than 200 cases, the minimum sample size was determined to be 400 cases. To facilitate factor analysis, as many samples as possible were included beyond this minimum.

This study was approved by the Ethics Committee of the Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences (Approval No.: P21009/PJ09).

1.2 Research Methods

Attending physicians screened eligible patients according to the inclusion and exclusion criteria, explained the purpose and content of the GSRS assessment, collected general clinical information, and administered the scale. The Chinese version of GSRS used in this study was previously translated and culturally adapted by our team, with authorization obtained from the original author, Prof. Jan Svedlund. The translated scale contains 15 items, all scored using a 4-point Likert scale with forward scoring; higher scores indicate greater symptom severity, while lower scores indicate better health status.

Patients received oral Xingpi Capsules (Guizhou Jianxing Pharmaceutical Co., Ltd., National Drug Approval No.: Z20028021) at a dosage of 5 capsules daily for a 14-day treatment course; concurrent use of other medications or non-pharmacological treatments was permitted. The Chinese version of GSRS was administered on days 0, 3, 7, and 14 of treatment. An independent follow-up team conducted telephone follow-ups with patients. Baseline survey data were used to evaluate the reliability and validity of the Chinese version of GSRS, while follow-up data were used to assess responsiveness.

1.3.1 Reliability Testing

Scale reliability is typically measured using multiple indicators including internal consistency, split-half reliability, and test-retest reliability. This study used Cronbach' s alpha coefficient to assess the internal consistency of GSRS, with Guttman split-half coefficient and Spearman-Brown coefficient representing split-half reliability. A reliability coefficient above 0.7 is considered acceptable. The recommended time interval for test-retest reliability calculation is 2–14 days; therefore, the intraclass correlation coefficient (ICC) and Spearman correlation coefficient (r) between the first and second test scores were calculated. Correlation coefficients were interpreted as follows: $r < 0.400$ indicates low correlation, $0.400 \leq r \leq 0.600$ indicates moderate correlation, and $r > 0.600$ indicates high correlation.

1.3.2 Validity Testing

Content validity was assessed using the expert judgment method. A panel of nine experts (four from Beijing, three from Tianjin, one from Hubei, and one from Jilin; four with senior professional titles, four with associate senior titles, and one with intermediate title; two methodology experts, five gastroenterology experts, and two TCM internal medicine experts) was convened to evaluate the content validity of the Chinese version of GSRS through an online questionnaire. All 15 items of the scale were rated using a 4-point Likert scale ranging from 1 (“not relevant”) to 4 (“very relevant”). The item-level content validity index (I-CVI) and scale-level content validity index (S-CVI), including universal agreement S-CVI (S-CVI/UA) and average S-CVI (S-CVI/Ave), were calculated based on expert ratings. Evaluation criteria for content validity indices were:

I-CVI $\geq$ 0.780, S-CVI/UA $\geq$ 0.800, and S-CVI/Ave $\geq$ 0.900. The K* evaluation criteria were: 0.40–0.59 as fair, 0.60–0.74 as good, and >0.74 as excellent.

The version of GSRS used in this study was not dimensionally structured in its original publication. In a subsequent 1993 study, the author divided the 15 items into three dimensions (reflux syndrome, dyspepsia syndrome, and intestinal dysfunction syndrome) based on clinical experience. Considering that modifications to the language version may affect the target population and cultural context, the research team, after expert discussion, randomly divided all patients into two samples (Sample 1 and Sample 2, each with 277 cases) for exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) in structural validity testing.

The KMO and Bartlett' s sphericity tests were used to determine whether Sample 1 data were suitable for EFA. Principal component analysis with varimax orthogonal rotation was employed to extract common factors with eigenvalues > 1, requiring a cumulative variance contribution rate above 50%. Amos 23.0 software was used for confirmatory factor analysis to evaluate the rationality of the factor structure model derived from EFA. When the observed data in Sample 2 did not fit the theoretical model well, appropriate modifications were made to the model.

1.3.3 Responsiveness Testing

Responsiveness is an indicator that reflects a scale' s ability to detect minimal changes over time and is one of the most important metrics in scale application research, directly influencing treatment evaluation and selection. This study assessed the responsiveness of the scale by comparing patients' GSRS scores before and after treatment using three responsiveness coefficients: effect size (ES), standardized response mean (SRM), and change ratio (CR).

1.4 Statistical Analysis

SPSS 22.0 software was used for statistical processing and analysis. Normality and homogeneity of variance tests were conducted for measurement data. Consistency tests, Spearman correlation coefficients, and intraclass correlation coefficients (ICC) were used in reliability testing. The Wilcoxon signed-rank test was used to compare differences in total GSRS scores before and after treatment, with a two-sided test at $\alpha = 0.05$. ES, SRM, and CR values were calculated to evaluate scale responsiveness.

2.1 Demographic Characteristics

A total of 554 patients presenting with gastrointestinal symptoms and diagnosed with different gastrointestinal diseases were enrolled and followed up in this study (Table 1). The cohort included 206 males (37.18%) and 348 females (62.82%), with ages ranging from 18 to 80 years and a mean age of 34.35 $\pm$

12.42 years. The majority (475 cases, 85.74%) had education levels above junior high school. Disease diagnoses included IBS in 127 cases (22.92%), chronic enteritis in 244 cases (44.04%), and chronic gastritis in 183 cases (33.03%). The research team strictly implemented telephone follow-up during the study, ensuring research quality; consequently, there were no missing data.

2.2 Scale Reliability

The reliability analysis results for the Chinese version of GSRS are presented in Table 2 .

2.2.1 Internal Consistency

Cronbach' s alpha coefficient is the most widely used method for internal consistency testing. The Cronbach' s alpha coefficient for all 554 samples was 0.896, and all three disease subgroups showed $\alpha > 0.8$.

2.2.2 Split-Half Reliability

Split-half reliability measures scale reliability by dividing the scale into odd and even items and calculating their respective reliabilities and correlation. Since GSRS contains 15 items, Guttman split-half coefficient and Spearman-Brown coefficient for unequal lengths were used. The Guttman split-half reliability coefficient was 0.883 and the Spearman-Brown coefficient was 0.886 in the IBS subgroup, while both coefficients were above 0.9 in the total patient group and other subgroups.

2.2.3 Test-Retest Reliability

Test-retest reliability results showed that Spearman correlation coefficients for all groups were greater than 0.6 and reached statistical significance. The intra-class correlation coefficient (ICC) indicated fair but acceptable test-retest reliability, suggesting adequate stability of the scale measurement. The relatively low ICC values may be related to substantial symptom changes in patients during the interval between the two tests due to pharmacological treatment.

2.3.1 Content Validity

As shown in Table 3 , all 15 items of the Chinese version of GSRS met the criterion of I-CVI ≥ 0.78, with K* > 0.74 after adjusting for chance agreement. All 15 items received expert ratings of no less than 3 points. The S-CVI/UA was calculated as 1, exceeding 0.8, and the S-CVI/Ave was 0.96, exceeding 0.9.

2.3.2 Structural Validity

Exploratory factor analysis was conducted using Sample 1 (n = 277). The KMO and Bartlett' s sphericity test results showed a KMO value of 0.911 and

a Bartlett' s sphericity test value of 1869.145 ($P < 0.001$), indicating that the dataset was suitable for factor analysis. Factor analysis extracted three common factors with eigenvalues greater than 1.000, with a cumulative variance contribution rate of 60.721%. Factor 1 included 7 items (abdominal pain, borborygmus, increased flatulence, increased stool frequency, loose stools, urgency, incomplete evacuation). Factor 2 included 6 items (heartburn, acid regurgitation, abdominal hunger pain, nausea and vomiting, bloating, belching). Factor 3 included 2 items (decreased stool frequency, hard stools). See Figure 1 [Figure 1: see original paper] and Table 4 .

Confirmatory factor analysis was performed using Sample 2 ($n = 277$). The initial model M0 was established based on the factor structure explored in Sample 1. The fit between Sample 2 data and M0 was modest; therefore, modifications were made according to modification indices by adding correlations between error variables, resulting in the modified standardized single-factor structural equation model M1 (Figure 2 [Figure 2: see original paper], Table 5). In model M1, $^2/df < 3.000$, RMSEA < 0.800, and all fit indices were > 0.900. The standardized regression coefficients for each item ranged from 0.327 to 0.758.

2.4 Comparison of GSRS Scores Before and After Treatment

The GSRS scores of all 554 patients before and after intervention followed a normal distribution according to Kolmogorov-Smirnov test but did not meet homogeneity of variance according to Levene' s test; therefore, the Wilcoxon signed-rank test was performed, with the same statistical approach applied to the three disease subgroups. The results showed that GSRS scores decreased significantly in all groups after 14 days of treatment intervention, with statistically significant differences compared to pre-intervention scores ($p < 0.001$). See Table 6 .

2.5 Responsiveness of GSRS in Assessing Treatment Effects

Responsiveness coefficients were calculated using the difference in GSRS scores between baseline and day 14: CR = difference / pre-treatment score; ES = mean of differences / pre-treatment standard deviation; SRM = mean of differences / standard deviation of differences. The responsiveness coefficients CR, ES, and SRM values for each patient group are presented in Table 7 . The results showed that the Chinese version of GSRS scores decreased by an average of 74.32% after treatment, indicating that the scale can sensitively detect changes in patients' post-treatment status. IBS patients showed the highest change rate, with the highest effect size and standardized response mean among the three subgroups, suggesting that the scale is more responsive in IBS patients and indirectly indicating that the intervention used in this study may have better efficacy in the IBS population.

3 Discussion

The diagnosis of gastrointestinal diseases is primarily based on clinical symptoms, endoscopy, and biopsy histopathology, with the latter two being the main criteria for diagnosis and differential diagnosis. Additionally, functional gastrointestinal diseases such as IBS exhibit functional changes without histopathological abnormalities, and when gastrointestinal symptoms appear as secondary manifestations of other diseases, additional examinations to confirm specific gastrointestinal disease diagnoses are rarely performed. Therefore, using diagnostic tools alone as efficacy indicators is insufficient. In actual clinical practice, patients often refuse relevant examinations when symptoms are mild or when work schedules are inconvenient, and changes in endoscopic or pathological examination results generally require a long time to reflect treatments received by patients. Consequently, more flexible and easily implementable criteria and symptom-oriented tools are needed for efficacy evaluation. Since most gastrointestinal diseases are diagnosed primarily based on patient-reported symptoms and treatment aims at symptom relief rather than clinical cure, using scales to evaluate disease severity and degree of relief is both reasonable and feasible.

Currently, researchers use various assessment tools including disease-specific scales, quality-of-life scales, psychological measurement scales, self-designed questionnaires, and TCM syndrome scales, all of which have limitations. Although disease-specific scales have high sensitivity, they are not applicable in scenarios without corresponding disease diagnoses. Quality-of-life scales assess overall patient conditions and are unsuitable for designs requiring only gastrointestinal symptom evaluation. Psychological measurement scales cannot directly reflect symptom changes. Self-designed scales often lack rigorous reliability and validity testing. TCM-specific scales are limited by "syndrome" diagnosis and cannot be universally applied. The spleen-stomach disease PRO scale developed by Liu Fengbin' s team can be applied to patients with multiple diseases, but its development was also based on TCM spleen-stomach theory, and as a PRO scale, its comprehensive content makes it lengthy with limited application scenarios. Therefore, we considered formally introducing GSRS into China based on actual conditions. GSRS has already been widely used in domestic clinical research. As a symptom measurement scale for patients' subjective reporting of disease severity and feelings, and originally constructed with reference to the Comprehensive Psychopathological Rating Scale (CPRS), it can also reflect the impact of symptoms on daily work and life to some extent, making it highly suitable for current clinical research needs for gastrointestinal symptom evaluation tools—that is, a universal scale not limited by disease diagnosis. Although the original GSRS target population was patients with peptic ulcer disease and IBS, practical application has shown that it can also be used to evaluate patients with gastrointestinal symptoms but without the above diseases, such as those with dyspepsia or patients with other diseases exhibiting secondary gastrointestinal symptoms. Based on this, and considering that the scale was constructed with reference to the universal

symptom scale CPRS from functional psychiatric disease research, we believe GSRS can similarly become such a universal evaluation tool. Previous Chinese versions have been used extensively but with numerous translation versions; however, no team has examined their measurement properties, and issues with unclear target populations have limited GSRS application in clinical research. Therefore, our team previously introduced and conducted translation research on GSRS to standardize its application.

The U.S. Food and Drug Administration (FDA) guidelines on PRO scales state that measurement tools should possess reliability, validity, and the ability to detect change. This study recruited patients seeking care for gastrointestinal discomfort from 45 centers nationwide and comprehensively evaluated the measurement properties of the Chinese version of GSRS from three aspects—reliability, validity, and responsiveness—to clarify its application scope and expand its target population, thereby creating conditions for large-scale application of the gastrointestinal symptom rating scale in domestic clinical and scientific research.

3.1 Reliability

Reliability represents the consistency and stability of scale measurement results. This study evaluated reliability using multiple indicators including Cronbach' s alpha, split-half reliability, and test-retest reliability. The Chinese version of GSRS showed a Cronbach' s alpha coefficient of 0.896, Guttman split-half coefficient of 0.920, and Spearman-Brown coefficient of 0.926—all above 0.7, indicating high internal consistency in the general population presenting with gastrointestinal symptoms. The ICC for two scale administrations was 0.589, and the Spearman correlation coefficient was 0.662. Statistically, correlation coefficients of 0.4–0.6 are considered moderate, and 0.6–0.8 are considered high. The test-retest reliability was within acceptable range, indicating good stability of the scale.

3.2 Validity

Validity reflects the degree to which scale results correspond to the content being measured—that is, accuracy and effectiveness. This study used the expert judgment method to examine content validity and combined exploratory and confirmatory factor analysis for structural validity. Results showed that all 15 items had I-CVI > 0.78 and K* > 0.74, indicating excellent content validity. Item 14 received "weakly relevant" ratings from 2 of 9 experts, the lowest among all items, suggesting that this item may be appropriately adjusted based on practical application if necessary. The overall S-CVI/UA exceeded 0.8, and the S-CVI/Ave exceeded 0.9, leading to the conclusion that the Chinese version of GSRS has excellent content validity.

EFA identified three common factors with eigenvalues greater than 1, with a cumulative variance contribution rate of 60.721% (exceeding 40%). Factor 1 contained 7 items, Factor 2 contained 6 items, and Factor 3 contained 2 items.

Factors 1–3 were named diarrhea symptom cluster, reflux symptom cluster, and constipation symptom cluster, respectively. Since the original scale developer did not propose a dimensional structure in the 1988 researcher-rated version, our results differ from previous foreign language versions, possibly due to differences in race, dietary habits, and disease characteristics among included populations. Further in-depth research on scale dimensionality can be conducted subsequently. For CFA, ^{2}/df should generally be as small as possible (typically ≤ 3.000), RMSEA ≤ 0.08, and other fit coefficients above 0.9. The modified M1 model in this study met all relevant criteria, with acceptable fit. Therefore, we conclude that the Chinese version of GSRS has good structural validity.

3.3 Responsiveness

In clinical research, especially interventional studies, scales used as efficacy evaluation indicators should possess not only reliability and validity but also the ability to detect clinically meaningful changes over time—that is, responsiveness. Current considerations of scale responsiveness primarily involve: (1) whether it can detect changes in quality of life over time in the same population, and (2) whether it can differentiate quality of life between different populations. For temporal considerations, CR, ES, and SRM are typically used. CR reflects the degree of score increase or decrease after treatment; ES and SRM represent the magnitude of responsiveness effect, with values of 0.8 or above indicating high responsiveness. There is no consensus in the academic community regarding which better reflects responsiveness. Differentiation considerations include comparisons between patients and healthy individuals, between patients with different disease diagnoses, and between different domains within the scale.

In this study, ES and SRM values for the Chinese version of GSRS were above 0.9 in both the total patient group and the three disease subgroups, with CR values exceeding 70%. As a symptom integration measurement tool, the Chinese version of GSRS demonstrated excellent responsiveness in both the general population presenting with gastrointestinal symptoms and populations diagnosed with specific gastrointestinal diseases. It can sensitively reflect symptom changes over time and with interventions, and can also assess current symptom severity through scale scores, making it suitable as an outcome measure in efficacy studies. However, this also indicates that scale scores can only distinguish between patients and healthy individuals and measure symptom severity, but cannot differentiate between different gastrointestinal diseases or serve as diagnostic criteria.

3.4 Limitations and Future Directions

This study has certain limitations: due to the lack of an available "gold standard," criterion validity analysis was not performed; and only the three most common gastrointestinal disease diagnoses were included in population selection, without exploration in a broader population. In future efficacy evaluation research, the Chinese version of GSRS can be used for symptom evaluation

regardless of whether a gastrointestinal disease diagnosis is confirmed, particularly suitable for use by general practitioners and medical institutions in clinical research, with expected better application prospects in nursing, rehabilitation, and TCM fields, thereby further verifying the scale' s scientific validity and practical utility in larger populations.

This study demonstrates that the Chinese version of the Gastrointestinal Symptom Rating Scale has high measurement performance in the general population presenting with gastrointestinal symptoms, with good reliability and validity and excellent responsiveness, making it suitable for treatment efficacy evaluation and enabling large-scale application in domestic clinical research.

Author Contributions: QIN Yuning drafted the manuscript, collected and organized data, performed statistical analysis, and interpreted the results. ZHAO Tianyi conceptualized and designed the study and drafted and translated the manuscript. LIU Fengbin provided methodological guidance for scale localization and performance testing. WANG Xin managed the project, operated software, and governed data. CAO Xue, SUN Minglin, LAI Keyun, DI Luyao, GE Zhishan, LIU Song, XING Ying, YANG Lei, YUE Lihong, and ZOU Meimei conducted the investigation and patient follow-up. HE Liyun and LI Hongjiao proposed the research concept, analyzed feasibility, supervised implementation, and critically reviewed the overall manuscript.

Conflict of Interest Statement: The authors declare no conflicts of interest.

Acknowledgments: We thank the 45 medical institutions, clinical physicians, and all patients who cooperated with patient recruitment during the study period.

Note: Figure translations are in progress. See original paper for figures.