Renzi score for obstructed defecation syndrome – validation of the Portuguese version according to the COSMIN checklist

Ana Celia **CAETANO**[1,2], Sara **DIAS**[2], André **SANTA-CRUZ**[2,3] and Carla **ROLANDA**[1,2]

ABSTRACT – Background – Recently, the Obstructed Defecation Syndrome score (ODS score) was developed and validated by Renzi to assess clinical staging and to allow evaluation and comparison of the efficacy of treatment of this disorder. **Objective** – Our goal is to validate the Portuguese version of Renzi ODS score, according to the Consensus based Standards for the selection of the Health Measurement Instruments (COSMIN) checklist. **Methods** – Following guidelines for cross-cultural validity, Renzi ODS score was translated into the Portuguese language. Then, a group of patients and healthy controls were invited to fill in the Renzi ODS score at baseline, after 2 weeks and 3 months, respectively. We assessed internal consistency, reliability and measurement error, content and construct validity, responsiveness and interpretability. **Results** – A total of 113 individuals (77 patients; 36 healthy controls) completed the questionnaire. Seventy and 30 patients repeated the Renzi ODS score after 2 weeks and 3 months respectively. Factor analysis confirmed the unidimensionality of the scale. Cronbach's α coefficient of 0.77 supported item's homogeneity. Weighted quadratic kappa of 0.89 established test-retest reliability. The smallest detectable change at the individual level was 2.66 and at the group level was 0.30. Renzi ODS score and the total (-0.32) and physical (-0.43) SF-36 scores correlated negatively. Patient and control's groups significantly differed (11 points). The change score of Renzi ODS score between baseline and 3 months correlated negatively with the clinical evolution (-0.86). ROC analysis showed minimal important change of 2.00 with AUC 0.97. Neither floor nor ceiling effects were observed. **Conclusion** – This work validated the Portuguese version of Renzi ODS score. We can now use this reliable, responsive, and interpretable (at the group level) tool to evaluate Portuguese ODS patients.

HEADINGS – Constipation, classification. Defecation. Psychometrics. Surveys and questionnaires.

INTRODUCTION

Chronic constipation is one of the five most common gastro-intestinal disorders. It consumes substantial health care resources due to the high prevalence and specificity of the diagnostic tests and treatments involved. It also has a considerable impact on patients' quality of life and mental state[1].

Patients with functional constipation (FC) according to the Rome IV criteria, fall into one of three categories: normal transit constipation, slow transit constipation and obstructed defecation syndrome (ODS)[2,3]. The ODS is the most frequent form of FC and it can be purely functional (anismus), mechanical (retocele, rectal prolapse) or both[4].

To diagnose ODS, clinical criteria must be complemented with at least two alterations related to impaired evacuation in anorectal tests – balloon expulsion test (BET), anorectal manometry (ARM), imaging or electromyography (EMG)[5-7]. Therefore ODS diagnosis involves great complexity and important direct and indirect costs. A clinical score is a useful tool to categorize the severity of symptoms and evaluate post-treatment results[8,9]. Besides, it can be studied as first line evaluation instrument before step up to a more complex diagnostic approach. In 1996, Agachan et al. published the first constipation scoring system useful for the evaluation and management of constipated patients[10]. Altomare score and Renzi score are the only two published and validated tools to evaluate ODS patients.

Altomare score includes "time spent at the toilet" and "stool consistency", items with a strong cultural influence and diet effect in our opinion. On the other hand, Renzi score is a simple questionnaire (FIGURE 1) that discriminates between healthy controls and ODS patients with sensitivity of 92% and specificity of 96%[9].

Symptoms / variables	0	1	2	3	4
Excessive straining	Never	Rarely	Sometimes	Usually	Always
Incomplete rectal evacuation	Never	Rarely	Sometimes	Usually	Always
Use of enemas / laxative	Never	Rarely	Sometimes	Usually	Always
Vaginal/perineal digital pressure	Never	Rarely	Sometimes	Usually	Always
Abdominal discomfort / pain	Never	Rarely	Sometimes	Usually	Always

Sintomas / variáveis	0	1	2	3	4
Esforço excessivo	Nunca	Raramente	Às vezes	Frequentemente	Sempre
Esvaziamento rectal incompleto	Nunca	Raramente	Às vezes	Frequentemente	Sempre
Uso de clisters / laxantes	Nunca	Raramente	Às vezes	Frequentemente	Sempre
Pressão digital vaginal / perineal	Nunca	Raramente	Às vezes	Frequentemente	Sempre
Desconforto / dor abdominal	Nunca	Raramente	Às vezes	Frequentemente	Sempre

FIGURE 1. Renzi ODS score (English and Portuguese version).

Recently, a Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist was developed in an international Delphi study to evaluate the methodological quality of studies on measurement properties of health-related patient reported outcomes[11].

To be applied in Portuguese-speaking countries, there must be a validated Portuguese version of the Renzi ODS score, what did not exist to date. Our goal is to validate the Renzi ODS score, in the Portuguese language according to the COSMIN checklist.

METHODS

Translations and cross cultural validity

After receiving authorization from the developers of the original ODS score, Renzi ODS score was translated forward (into Portuguese) by two bilingual translators. Then, two Gastroenterology's experts synthetized the results of the translation into a unique score that suffered a backward translation (into English) by two bilingual translators[12]. All of them worked independently from each other to allow detection of errors, divergent interpretation or ambiguous items[11]. A panel of two gastroenterologists, with command of English, evaluated the obtained Portuguese version to achieve a consolidated version without discrepancies. The protocol of research had a positive report from the Ethics Committee for Health of the Hospital of Braga and the Ethics Subcommittee for Life and Health Sciences (SECVS). All data were collected anonymously.

Patients' selection

From July to October 2016, outpatients over 18 years old attending Gastroenterology Department of *Braga's* hospital, with ODS diagnosis were invited to participate in the study. Healthy individuals, mostly patients accompanying persons, were randomly selected as control group. Exclusion criteria were previous colorectal surgery or cancer, Hirschsprung's disease and/or inability to answer the questionnaire. A written informed consent was obtained from those who agreed to participate.

Study design and data colection

It was a prospective, observational, and longitudinal study carried out at the Gastroenterology Department of *Braga* Hospital.

At baseline, there was a face-to-face interview. Socio-demographic and clinical data were collected[13]. The Renzi ODS score and the validated Portuguese version 2 of the Sort Form – 36 items (SF-36) were applied. SF-36 is a generic measure of health that is clear and easy to fill, with good psychometric properties[14-16]. It comprises 36 items considering two main components (physical and mental). Furthermore, it is not specific for any age or disease. Higher SF-36 scores define more favorable health status. After 2 or more weeks, patients were contacted by telephone for a second Renzi ODS score application. Three months later, patients were contacted by telephone for a third interview to apply the Renzi ODS score again and to assess clinical evolution. Clinical evolution was rated in a Global Perceiving Effect (GPE) score of seven categories: "Much better", "Better", "Somewhat better", "Unchanged", "Somewhat worse", "Worse" and "Much worse", in a scale ranging from 3 to -3, respectively.

Statistical analysis

There is no consensus on the number of patients required to run a factor analysis, with rules of thumb varying from 4 to 10 subjects per variable, with a minimum of 100 subjects[17]. At least 50 patients are needed to assess reliability[11]. Collected data were gathered in Statistical Package for Social Sciences (SPSS)® - version 23.0 and all statistical analysis was performed using this program. Normality of the variables was evaluated throw Skewness and Kurtosis values (normality considered for values between -1 and 1), Komolgorov-Smirnov test and Shapiro-Wilk test[18]. If not specified, statistical significance was accepted for $P<0.05$. Acceptability depends on the proportion of Renzi-ODS score and SF-36 questionnaires with incomplete filling and missing items.

Methodological testing according COSMIN checklist
• Internal consistency

This measurement property evaluates the extent to which different items in a questionnaire are correlated, thus measuring the same concept[17]. A principal component analysis (PCA) was performed to assess the unidimensionality of the scale. The scree plot of eigenvalues was used to determine the number of components and items with factor loading <0.5 were considered for deletion[19]. The homogeneity of the items is determined by Cronbach's α, item–total, and item–item correlations. Cronbach's α was calculated using a one-at-a-time deletion procedure to assess the impact of each item on internal consistency separately. Cronbach's α between 0.70 and 0.95 was considered an indicator of adequate homogeneity of the items[7]. Inter-item correlations should be under 0.7 and corrected item–total correlations should exceed 0.4[17,20].

• Reliability

To evaluate reliability, a second interview with application of Renzi ODs score was conducted, after 2 weeks. It was independent from the first, as neither the patients nor the investigator were aware of the scores achieved in the initial evaluation. This time interval was considered appropriate since it is long enough to prevent recall bias and short enough to ensure that there were no changes in the construct to be measured[11]. Since it is an ordinal score, simple kappa coefficient (κ) for each item and weighted quadratic kappa for the total score were assessed to evaluate test-retest reliability[21]. Kappa is "acceptable" when it is >0.40, "fair to good" if it is between 0.40 and 0.75 or "excellent" if it is >0.75[22].

• Measurement error

Measurement error includes the systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured[11]. The Standard Error of Measurement (SEM) was calculated using the formula $SEM = SD \sqrt{1-r}$ (SD=standard deviation; r=reliability). The Smallest Detectable Change (SDC) reflects the smallest individual change in score that can be interpreted as a real change in score, above measurement error, in one individual ($SDC_{ind}= SEM \times 1.96 \times \sqrt{2}$) and in a group of people ($SDC_{group} = SDC_{ind}/\sqrt{n}$)[17].

• Content validity

Content validity was defined as the degree to which the content of a questionnaire is an adequate reflection of the construct to be measured and it should be assessed by judging the relevance and comprehensiveness of the items[23]. An indication that the comprehensiveness of the items was assessed could be the indication of missed items: large floor and ceiling effects can indicate a scale that is not comprehensive[11].

• Validity (construct validity and structural validity)

This measurement property refers to the extent to which scores, in a particular instrument, relate to other measures and its consistency with theoretically derived hypotheses concerning the concepts that are being measured[18]. It includes structural validity and construct validity. Structural validity is evaluated by performing exploratory factor analysis, more specifically PCA (see Internal Consistency above). Construct validity was assessed using the 'hypotheses testing' method that tests predefined hypothesis, for example, expected correlations between measures or expected differences in scores between groups[17]. A higher percentage of correct predictions indicates stronger support for construct validity. Convergent evidence was assessed by calculating Spearman rank correlation coefficient between the scores of the Renzi ODS score and the SF-36, obtained at the first interview, based on the assumption that patients with more severe ODS would have lower quality of life; therefore it is expected a negative correlation, at least moderate, between the total Renzi ODS score and the total, physical and mental scores of SF-36[9]. A Spearman's correlation <0.30 is classified as "weak", between 0.30 and 0.50 as "moderate" and >0.50 as "strong"[24]. As Renzi score showed significant differences between the median score for patients (median=4.5) and controls (median=12.5)[9], it is expected a difference of, at least, 6 points between the Renzi ODS scores of the two groups.

• Responsiveness

Responsiveness is the ability of a questionnaire to detect clinically important changes over time in the construct to be measured, even if these changes are small, considering it a measure of longitudinal validity[11,17]. To evaluate responsiveness, a third interview was conducted by telephone, 3 months after the first one, to apply Renzi ODS score, under the same instructions and by the same initial interviewer, and to apply a GPE score to assess patient's perceived change about the construct that are being measured. This time interval was considered appropriate since it is long enough to expect clinical change in patients with ODS, but still short enough to assume that patients would be able to recall whether any changes in their condition occurred during this period[25]. It is expected that changes in Renzi ODS score between the first and the third interview negatively correlates with the GPE score with a correlation coefficient <-0.5[17].

• Interpretability

Interpretability is defined as the degree to which one can assign qualitative meaning to quantitative scores[11]. This can be assessed through distribution of the total scores in the study population and relevant subgroups, floor and ceiling effects and Minimal Important Change (MIC). Floor and ceiling effects are considered to be present if more than 15% of respondents achieved the lowest or highest possible score, respectively[17]. Study sample was dichotomized, according to GPE, in groups of "importantly improved" (includes "much better" and "better" options) versus "stable" (includes "somewhat better", "unchanged" and "somewhat worse" options). Subsequently, receiver operating characteristic (ROC) analyses were performed. The area under the curve (AUC), a measure of responsiveness, was considered as the probability of correctly classifying the "importantly improved" from "stable" patients, and a good discrimination is considered if AUC >0.70[17,26]. For determining MIC, optimum ROC cut-off point of the change score was obtained by weighting sensitivity and specificity equally. MIC for improvement was calculated instead of MIC for deterioration because one of the main goals of Renzi ODS score is to assess change after treatment interventions. If the measurement error (SDC) is smaller than MIC, it is possible to distinguish clinically important change from measurement error with much confidence[17,27].

RESULTS

The gastroenterologists compared the two translated versions and the original version and no changes were applied to the final translation. A total of 113 individuals signed the consent and participated in the study with no missing items in the questionnaire. From the total of 113 individuals interviewed at baseline, 77 (68.14%) enrolled the patient's group. The patients' group had more women ($P<0.05$). No other significant differences were found between patients and controls.

Internal consistency

After performing PCA, one leading eigenvalue emerged in the scree plot of eigenvalues (FIGURE 2) that explained 53.37% of total variance, conferring unidimensionality to the scale. This means that all items are associated to a single construct. All items analyzed had factor loadings >0.5, which means high correlation with the principal component, so all items were kept. This was also taken into account when assessing structural validity.
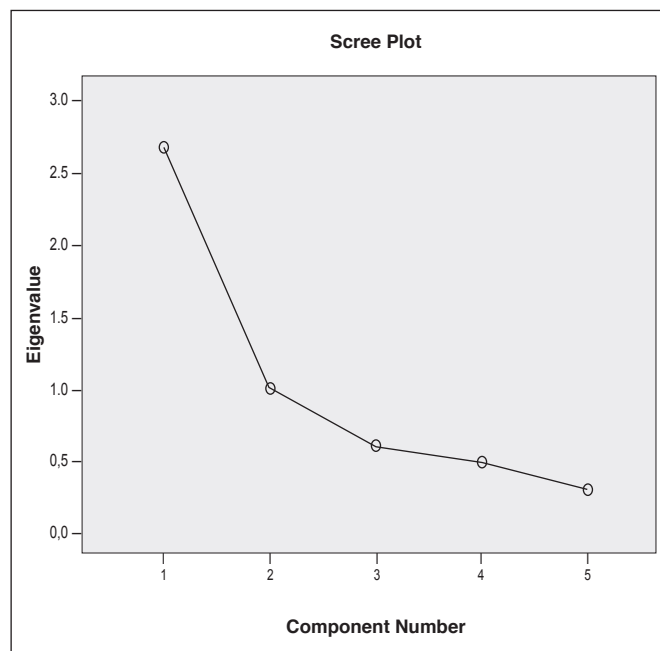

FIGURE 2. Scree plot of eigenvalues from PCA (n=113).

Cronbach's α coefficient calculated for the 5 items was 0.77, thus all items have enough homogeneity (TABLE 1). Cronbach's α was also calculated considering deleting each item separately, which resulted in a range of 0.68 to 0.78, therefore the exclusion of any item would not substantially affect reliability. There is no redundancy between the items since inter-item correlations were <0.7. Corrected item total correlation lowest value was 0.39, which it's not considerably different from 0.40.

TABLE 1. Cronbach's α coefficient for total Renzi ODS score, item-total correlation and Cronbach's α with item deleted for each item

Cronbach's α coefficient (n=113)	0.77	
	Item-total correlation	Cronbach's alpha if item deleted
1. Excessive straining	0.69	0.68
2. Incomplete rectal evacuation	0.68	0.68
3. Use of enemas and laxatives	0.41	0.78
4. Vaginal-anal-perineal maneuvers to attempt defecation	0.39	0.78
5. Abdominal discomfort/pain	0.57	0.72

Reliability

TABLE 2 shows the results of test-retest reliability, evaluated in 77 patients. As simple values for each item ranges from 0.49 to 0.81 and weighted quadratic kappa for the total score is 0.89, test-retest reliability is considered good to excellent.

TABLE 2. Descriptive statistics, test-retest reliability and measurement error of Renzi ODS score

Test Mean (SD) (n=77)	12.0 (±2.9)
Re-test Mean (SD) (n=77)	12.4 (±3.3)
Difference test-retest mean (SD) (n=77)	-0.35 (±1.4)
Simple kappa coefficient (n=77)	
1. Excessive straining	0.75
2. Incomplete rectal evacuation	0.49
3. Use of enemas and laxatives	0.73
4. Vaginal-anal-perineal maneuvers to defecate	0.81
5. Abdominal discomfort/pain	0.71
Weighted quadratic kappa (n=77)	
Total score (95 CI)	0.89 (0.85 - 0.94)
SEM (95% CI)	0.96 (0.71-1.12)
SDCind (95% CI)	2.66 (1.97-3.10)
SDCgroup (95% CI)	0.30 (0.22-0.35)

CI: confidence interval; SD: standard deviation; SDCind: Smallest Detectable Change at the individual level; SDCgroup: smallest detectable change at group level; SEM: standard error of measurement.

Measurement error

SDC at the individual level was 2.66 and at the group level was 0.30 (TABLE 2).

Validity

As previously hypothesized, the total score of Renzi ODS score negatively correlated with the total (rs=-0.32; P<0.01) and physical scores (rs=-0.43; P<0.01) and these are considered moderate correlations. There was also a negative correlation with mental scores of SF-36 (rs=-0.14) but it did not reach statistical significance (TABLE 3). Significant differences were obtained (Mann-Whitney U test=0; P<0.01) comparing patients (median=12) and controls (median=1). Therefore, there is a difference of 11 points in the medians of both groups, greater than the 6 point previously hypothesized.

TABLE 3. Spearman's correlation coefficients determined when comparing Renzi ODS score and SF-36's scores

	SF-36 total score	SF-36 physical score	SF-36 mental score
Renzi ODS score (n=77)	-0.32*	-0.43*	-0.14

*Correlation statistically significant (P<0.01).

Responsiveness

There was a negative correlation between change score in Renzi ODS score and the GPE score with a correlation coefficient <-0.5 (rs=-0.86; P<0.01). Hypothesis regarding ROC analysis was confirmed, since AUC was 0.97 (95% CI: 0.91-1.00).

Interpretability

Neither floor nor ceiling effects were observed for any of the scores (TABLE 4). Using ROC analysis for discriminating improved versus stable patients (as assessed by the GPE score) through changes in the Renzi ODS score and by weighting sensitivity and specificity equally, the best cut-off was 2.00 (sensitivity-100%; specificity-92%). SDCgroup is smaller than MIC (0.30<2.00) and SDCind is greater than MIC (2.66>2.00).

DISCUSSION

Nowadays, ODS is still a common but complex syndrome. A clinical severity index, as proposed by Renzi et al. could be a useful tool in the diagnostic algorithm and the evaluation of meaningful changes after treatment[9].

Since Renzi ODS score was successfully validated in the English language, showing good discriminatory power to distinguish between patients and controls and changes in patients over time, we intend to validate a Portuguese version of Renzi ODS score following COSMIN checklist guidelines[9,11].

According to what we stated, we could validate the domains of reliability, validity, responsiveness and interpretability (at group level) of Renzi ODS score, following strict methodology based in the COSMIN checklist[11].

Regarding cross cultural validity, guidelines for a correct translation and cultural adaptation were fulfilled without considerable constraints, since it is a pretty straightforward questionnaire[12]. Good acceptability and absence of missing items reinforced this.

The results showed good internal consistency assessed by PCA procedure, confirming that this is a unidimensional scale. The reliability of Renzi ODS score was confirmed by high internal consistency (Cronbachs's α=0.77 supporting homogeneity of all items in the scale), test–retest (weighted quadratic kappa ᵡ=0.89) and inter-rater (0.96) reliabilities.

Validity was assessed through structural validity and hypothesis testing. Structural validity refers to PCA, already explained above. Of the following a-priori-formulated hypotheses in correlation analysis: the Renzi ODS score is negatively and moderately correlated with the total and physical scores of SF-36 (rs=-0.32 and -0.43, respectively). Although also negatively correlated, the same assumption failed on the statistical association of Renzi ODS score with mental score of SF-36. Probably, a greater study sample, as it was used in the original validation study, would lead to a more significant correlation values[9]. There was a significant

TABLE 4. Distribution of total scores, floor and ceiling effects

	Mean	SD	Median	Range	Floor effects	Ceiling effects
Renzi ODS score at first interview (n=113)	8.71	5.51	10.00	0.00-19.00	9 (7.96%)	0
Female (n=78)	10.63	4.44	12.00	0.00-18.00	-	-
Male (n=35)	4.43	5.29	2.00	0.00-19.00	-	-
Renzi ODS score at second interview (n=77)	12.39	3.27	13.00	5.00-19.00	0	0
Female (n=66)	12.55	3.18	13.00	5.00-19.00	-	-
Male (n=11)	11.46	3.83	11.00	6.00-19.00	-	-
Renzi ODS score at third interview (n=30)	12.10	4.11	12.50	4.00-20.00	0	1 (3.33%)
Female (n=27)	12.56	4.02	13.00	4.00-20.00	-	-
Male (n=3)	8.00	2.65	7.00	6.00-11.00	-	-
SF-36 at first interview (n=77)	48.42	16.71	43.61	19.72-87.17	0	0
Female (n=66)	47.51	16.61	43.06	19.72-85.50	-	-
Male (n=11)	53.85	17.06	52.72	34.44-87.17		

SD: standard deviation.

difference between the median of Renzi ODS score considering patients and controls, being this difference greater than the 6 points previously hypothesized.

Responsiveness was assessed by a similar fashion as construct validity, but in this case, hypothesis about correlation of Renzi ODS score changes with GPE score were formulated, considering an "anchor based" method. A correlation coefficient <-0.5 was obtained, as expected. AUC was 0.97, supporting a good responsiveness of Renzi ODS score. SDCind >2.66 can be considered real changes at the individual level and large SDCind values in the current study are somewhat expected, suggesting questionnaire's inability to detect minimal but still clinically important changes. Changes in SDCgroup >0.30 can be detected with 95% confidence.

Interpretability considered assessment of floor and ceiling effects, and none of them were observed in any of the administrations of Renzi ODS score. A MIC of 2,00 was defined for discriminating "importantly improved" patients. Comparing measurement error with MIC, MIC is higher than SDCgroup and smaller than SDCind, implying that Renzi ODS score is ineffective at detecting minimal, but still clinically important changes at individual level[27]. Thus, in future studies, the sample size should be adjusted accordingly to minimize measurement error.

Although the main goal of this work was achieved, there are some limitations. In the original validation of ODS score, Renzi et al. used 100 patients and 100 controls, which is considerably greater than the sample size used by us. A greater sample size, would strengthen statistical analysis in the validation of the score and would probably minimize measurement error, in order to achieve interpretability at the individual level[27]. An even better option would be a multicentric study in order to validate a score culturally well adapted to the Portuguese population. Questionnaires applied at baseline were done by face-to-face interview and score application at 2 weeks and 3 months occurred by telephone. Although applied by the same interviewer under the same instructions, this might lead to bias on the obtained scores[11]. Anchor based methods, resorting a GPE scale, used to analyze responsiveness validity and to achieve MIC value, was reported as a reliable and valid measure of health change perceived by patients and considered the best measure from individual perspective. However, a retrospective rating of change obtained over an extended period of time are susceptible to recall bias[28,29]. Furthermore, the 3 months' follow-up to evaluate clinical perceived change by patients is probably not enough, since this is a chronic disability which may not change much in such a short period of time.

CONCLUSION

According to COSMIN checklist, a validated Portuguese version of ODS score was developed, showing good internal consistency, content and construct validity, responsiveness and interpretability, at the group level. So, Renzi ODS score arises as a simple and concise instrument, designed to be a preliminary tool to approach patients with ODS, to help diagnose and grade ODD severity, and to evaluate treatment effectiveness more precisely.

Authors' contributions

Caetano AC and Rolanda C were involved in the study concept and design; Caetano AC and Dias S were involved in the acquisition of data and completed the analysis and interpretation of data; Santa-Cruz S reviewed the statistical analysis; Caetano AC and Dias S contributed to the drafting of the manuscript; Rolanda C and Santa-Cruz A made a critical revision and English editing of the manuscript.

Caetano AC, Dias S, Santa-Cruz A, Rolanda C. Score de Renzi para distúrbios evacuatórios – validação da versão portuguesa de acordo com a checklist COSMIN. Arq Gastroenterol. 2018;55(1):55-60.

RESUMO – Contexto – Recentemente, o Score de Distúrbios Evacuatórios (SDE) foi desenvolvido e validado por Renzi para avaliação e comparação da eficácia do tratamento dos doentes com esta patologia. **Objetivo** – O nosso objetivo é validar uma versão portuguesa do SDE de acordo com as orientações da checklist de COSMIN. **Métodos** – O SDE foi traduzido para o português, cumprindo as orientações para validação cultural. Indivíduos com distúrbio evacuatório e controlos saudáveis foram convidados a responder ao SDE numa fase inicial, 2 semanas e 3 meses depois, respetivamente. Foi avaliada a consistência interna, confiabilidade, erro de medição, validade de conteúdo e constructo, responsividade e interpretabilidade. **Resultados** – Foram entrevistados 113 indivíduos (77 doentes; 36 controlos saudáveis) na fase inicial. O SDE foi aplicado novamente aos 77 doentes, 2 semanas depois, e a 30 doentes, 3 meses depois. Relativamente à consistência interna, a análise fatorial confirmou a unidimensionalidade e o coeficiente α de Cronbach foi 0,77, suportando homogeneidade dos itens. O kappa quadrático ponderado de 0,89 estabeleceu a reprodutibilidade teste-reteste. Considerando o erro de medição, a mudança mínima detectável a nível individual foi 2,66 e a nível de grupo foi 0,30. A validade do constructo foi avaliada através do coeficiente de correlação de Spearman entre o SDE e o score total (-0,32) e físico (-0,43) do SF-36. Em termos de validação clínica, verificou-se uma diferença significativa de 11 pontos entre as médias dos doentes e controlos. A responsividade foi confirmada pelo coeficiente de correlação de -0,86 entre a mudança do score e a evolução clínica, avaliados após 3 meses. Através da curva ROC, a mudança mínima importante foi 2,00 e a AUC foi 0,97. Não foram observados efeito-chão efeito-tecto. **Conclusão** – Este projeto permitiu validar a versão portuguesa do SDE de Renzi. É possível agora utilizar esta ferramenta na avaliação de distúrbios evacuatórios em doentes falantes de língua portuguesa.

DESCRITORES – Constipação intestinal, classificação. Defecação. Psicometria. Inquéritos e questionários.

REFERENCES

1. Bharucha A, Pemberton J, Locke III G. American Gastroenterological Association Technical Review on Constipation. Gastroenterology. 2013;144:218-38.
2. D'Hoore A, Penninckx F. Obstructed Defecation. Semin Colon Rectal Surg. 2003;22:45-9.
3. Mearin F, Ciriza C, Mínguez M, Rey E, Mascort JJ, Peña E, et al. Clinical Practice Guideline : Irritable bowel syndrome with constipation and functional constipation in the adult. 2016;108:332-61.
4. Rosen A. Obstructed defecation syndrome: Diagnosis and therapeutic options, with special focus on the STARR procedure. Isr Med Assoc J. 2010;12:104-6.
5. Chiarioni G, Kim SM, Vantini I, Whitehead WE. Validation of the Balloon Evacuation Test: Reproducibility and Agreement With Findings From Anorectal Manometry and Electromyography. Clin Gastroenterol Hepatol. Elsevier, Inc; 2014;12:2049-54.
6. Halverson A, Orkin B. Which physiologic tests are useful in patients with constipation? Dis Colon Rectum. 1998;41:735-9.
7. Rao SSC, Mudipalli RS, Stessman M, Zimmerman B. Investigation of the utility of colorectal function tests and Rome II criteria in dyssynergic defecation (Anismus). Neurogastroenterol Motil. 2004;16:589-96.
8. Altomare DF, Spazzafumo L, Rinaldi M, Dodi G, Ghiselli R, Piloni V. Set-up and statistical validation of a new scoring system for obstructed defaecation syndrome. Color Dis. 2008;10:84-8.
9. Renzi A, Brillantino A, Di Sarno G, D'Aniello F. Five-Item Score for Obstructed Defecation Syndrome: Study of Validation. Surg Innov. 2013;20:119-25.
10. Agachan F, Chen T, Pfeifer J, Reissman P, Wexner SD. A constipation scoring system to simplify evaluation and management of constipated patients. Dis Colon Rectum. 1996;39:681-5.
11. Mokkink L, Terwee C, Patrick D. The COSMIN checklist manual. 2009;1-8.
12. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976). 2000;25:3186–91.
13. Steele SR, Mellgren A. Constipation and obstructed defecation. Clin Colon Rectal Surg. 2007;20:110–7.
14. Ferreira PL. Criação da versão Portuguesa do MOS SF-36. Parte II - Testes de validação. Acta Med Port. 2000;13:119–27.
15. Hawthorne G, Osborne RH, Taylor A, Sansoni J. The SF36 Version 2: Critical analyses of population weights, scoring algorithms and population norms. Qual Life Res. 2007;16:661-73.
16. Ware JE SS. The MOS 36-Item Short-FormHealth Survey (SF-36): I. Conceptual framework and item selection. Med Care. 1992;30:473-81.
17. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60:34-42.
18. Chan YH. Biostatistics 101: Data presentation. Singapore Med J. 2003;44:280-5.
19. Floyd F, Widaman K. Factor analysis in the development and refinement of clinical assessment instruments. Psychol Assess. 1995;7:286-99.
20. Gliem JA, Gliem RR. Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. Midwest Res to Pract Conf Adult, Contin Community Educ. 2003:82-8.
21. Brenner H, Kliebsch U. Dependence of weighted kappa coefficients on the number of categories. Vol. 7, Epidemiology (Cambridge, Mass.). 1996. p. 199-202.
22. Hernaez R. Reliability and agreement studies: a guide for clinical investigators. Gut. 2015;64:1018-27.
23. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol. 2010;10:22.
24. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, New Jersey Lawrence Erlbaum Assoc Inc. 1988.
25. Thorborg K, Hölmich P, Christensen R, Petersen J, Roos EM. The Copenhagen Hip and Groin Outcome Score (HAGOS): development and validation according to the COSMIN checklist. Br J Sports Med. 2011;45:478-91.
26. Heidemann CH, Godballe C, Kjeldsen AD, Charlotte E, Johansen J, Faber CE, et al. The Otitis Media-6 questionnaire : psychometric properties with emphasis on factor structure and interpretability. Health Qual Life Outcomes. 2013;1-10.
27. Terwee CB, Roorda LD, Knol DL, De Boer MR, De Vet HCW. Linking measurement error to minimal important change of patient-reported outcomes. J Clin Epidemiol. Elsevier Inc. 2009;62:1062-7.
28. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. 2003;56:395-407.
29. De Boer MR, De Vet HCW, Terwee CB, Moll AC, Völker-Dieben HJM, Van Rens GHMB. Changes to the subscales of two vision-related quality of life questionnaires are proposed. J Clin Epidemiol. 2005;58:1260-8.