Published in final edited form as:

Neurogastroenterol Motil. 2016 March ; 28(3): 443–448. doi:10.1111/nmo.12738.

Bristol Stool Form Scale Reliability and Agreement Decreases When Determining Rome III Stool Form Designations

Bruno P. Chumpitazi^{1,3}, Mariella M. Self^{1,2}, Danita I. Czyzewski^{1,2,3}, Sydney Cejka¹, Paul R. Swank⁴, and Robert J. Shulman^{1,3,5}

¹Department of Pediatrics, Baylor College of Medicine, Houston, TX

²Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, TX

³Section of Gastroenterology, Hepatology, and Nutrition, Texas Children's Hospital, Houston, TX

⁴University of Texas School of Public Health, Houston, TX

⁵Children's Nutrition Research Center, Houston, TX

Abstract

Background—Rater reproducibility of the Bristol Stool Form Scale (BSFS), which categorizes stools into one of seven types, is unknown. We sought to determine reliability and agreement by individual stool type and when responses are categorized by Rome III clinical designation as normal or abnormal (constipation or diarrhea).

Methods—Thirty-four gastroenterology providers from three institutions rated 35 stool photographs using the BSFS. Twenty rerated the photographs.

Key Results—1190 individual stool type ratings were completed. Though only 4 photographs had absolute agreement (all Type 1 or Type 7), general agreement was high with 1132 (95.1%) of ratings being within one category type of the modal rating. Inter-rater and intra-rater reliability of the BSFS by individual stool type was excellent with intraclass correlations of 0.88 (95% CI: 0.86–0.90), $P < 0.001$) and 0.89 (95% CI: 0.86–0.91, $P < 0.001$) respectively. However, agreement decreased when using Rome III designations, with 13 (37%) photographs having significantly diverging classifications (semi-interquartile range=0.5). These 13 photographs were rated by the majority of raters as either type 2 vs. type 3 or type 5 vs. type 6 stools, representing the boundaries of normal vs. abnormal stools. Inter-rater and intra-rater reliability of the BSFS by Rome III clinical categorization decreased with intraclass correlations of 0.75 (95% CI: 0.69 – 0.81, $P < 0.001$) and 0.65 (95% CI: 0.49 – 0.81, $P < 0.001$) respectively.

Corresponding Author: Bruno P. Chumpitazi, M.D., M.P.H., Texas Children's Hospital, 6701 Fannin Street, CCC 1010.03, Houston, TX 77030, chumpita@bcm.edu; Telephone: 832-822-4703; Fax: 832-825-3633.

CONFLICTS OF INTEREST

The authors do not have competing interests to declare.

AUTHOR CONTRIBUTION

BPC, DIC, MMS, and RJS designed the project; BPC, SC conducted the research; BPC, PRS analyzed the data; BPC, DIC, MMS, SC, PRS, and RJS wrote the paper; BPC had primary responsibility of final content. All authors read and approved the final manuscript.

Conclusions and Inferences—BSFS has excellent reliability and agreement when used to rate individual stool type by raters. However, BSFS reliability and agreement decreases when determining Rome III stool form categories.

Keywords

stool form; IBS; irritable bowel syndrome; constipation; diarrhea

INTRODUCTION

The Bristol Stool Form Scale (BSFS), a frequently used measure in gastroenterology practice and research, categorizes stools into one of seven stool types ranging from type 1 (hard lumps) to type 7 (watery diarrhea). It was validated as a measure of intestinal transit but not yet as a reproducible measure of stool form amongst raters.(1) Despite this, it has been used to assess stool form in several gastrointestinal disorders including irritable bowel syndrome (IBS) and Crohn's disease.(2–8)

Scale reproducibility, which includes both reliability and agreement, is an important measure of any scale, particularly when measuring diagnostic criteria and to determine eligibility for therapeutic trials. The BSFS was adopted by the Rome III committee for use in categorizing IBS into one of four subtypes based on the propensity toward constipation, diarrhea or presence of both.(9) Subsequently, children with IBS have been categorized using the same criteria.(10, 11), and clinical trials use the BSFS to determine eligibility and evaluate outcomes. (12, 13)

Given the wide usage of the BSFS but lack of assessment of its reproducibility, we initiated evaluation of the reliability and agreement of the BSFS when used to determine stool form by experts. We determined inter- and intra-rater reliability and agreement of the BSFS by individual stool type and when classifying stools by Rome III clinical delineation (i.e. constipation, normal, or diarrhea).

METHODS

Study Design

Thirty-five color, two-dimensional stool photographs ranging from liquid to hard pellets were obtained from publicly accessible areas of the internet. Only photographs that were focused, close-ups of entire bowel movements with white backgrounds were chosen. Photographs were randomized into one of 6 sequences to minimize the potential effect of one photograph biasing the rating of a subsequent photograph.

Thirty-four health care providers providing gastroenterology clinical care from three institutions were asked to rate the photographs using the BSFS (Figure). Each page of the survey contained an individual stool photograph with the BSFS graphic scale beneath. The individual stool type ratings were then classified based on the Rome III criteria, with stool types 1 and 2 designated as constipation, types 3, 4, and 5 as normal stool form, and types 6 and 7 as diarrhea.(9) Twenty raters agreed to repeat the survey for intra-rater analyses a

minimum of 2 weeks after the initial ratings. The study was approved by the Baylor College of Medicine Institutional Review Board.

Statistical Analysis

Parameters of agreement and reliability were both determined to assess reproducibility.⁽¹⁴⁾ Statistical measures included determination of inter- and intra-rater reliability (i.e., reproducibility of ratings) using intraclass correlation coefficient (ICC) two-way random effects model with absolute agreement of single measures, and determination of sources of variance.⁽¹⁵⁾ ICC interpretation parameters were: <0.40=Poor; 0.40–0.59=Fair; 0.60–0.74=Good; and 0.75=Excellent.^(16, 17)

Agreement (i.e., similarity of ratings for each photograph) was assessed using percent exact agreement and percent within one rating type of the most common (modal) rating chosen for each photograph. Semi-interquartile ranges were calculated to assess variability and agreement of the Rome III clinically delineated categories (constipation, normal, or diarrhea) with scores greater than zero representing notable disagreement (more than half responses in different categories).

SAS/STAT (version 9.3., SAS Institute, Cary, NC) was used for statistical analyses.

RESULTS

Characteristics of Raters

Raters were 8 adult gastroenterologists, 17 pediatric gastroenterologists, 8 pediatric gastroenterology fellows, and one pediatric nurse practitioner. The practice settings varied from community centers to tertiary care centers. Twenty (59%) were male. All initial raters were asked to re-rate the stools; twenty raters (59%) agreed. No significant differences in this subgroup were found from the original 34 with respect to experience (trainee vs. faculty), gender, or institution (data not shown).

Stool Photograph Ratings (Table)

Rating 35 photographs, the 34 raters made a total of 1190 BSFS ratings. The Table presents the distribution of stool types selected for each photograph.

Etiology of Inter-rater Variance in Ratings

The variance in inter-rater ratings due to the raters themselves (e.g., rater tendency to rate all photographs more toward one end of the scale or the other) was very low at 0.03. Inter-rater variance due to the interaction of raters by photographs (e.g., inconsistent choices of a rater by photograph) was also low at 0.47. The variance in ratings due to the photographs themselves was much higher at 3.71. Therefore, the primary driver of the ratings was the photographs themselves with an overall ratio of 7.35:1.

Inter- and Intra-rater Reliability and Agreement

We found 852 (71.6%) ratings were in agreement with the most commonly chosen (modal) photograph rating. Four (11.4%) stool photographs had complete agreement amongst all the

raters; two type 1 and two type 7 photographs. Most (95.1%) of the ratings on the BSFS were within at least one category type of the modal rating for each photograph.

Inter- and intra-rater reliability of the BSFS was excellent with a single measure ICC of 0.88 (95% CI: 0.86–0.90, $P<0.001$) and 0.89 (95% CI: 0.86–0.91, $P<0.001$), respectively.

Inter- and Intra-rater Reliability and Agreement by Rome III Clinical Delineation (Constipation, Normal, and Diarrhea)

When evaluating ratings by clinical delineation 12 (34.3%) photographs had complete agreement, 10 (28.6%) had the majority of ratings in agreement (semi-interquartile range=0), and 13 (37.1%) had more than half the responses in different categories (semi-interquartile range=0.5). All 13 photographs with decreased agreement were rated by the majority of raters as either type 2 vs. type 3 ($n=5$) or type 5 vs. type 6 ($n=8$) stools.

Inter-rater reliability of defining a stool as reflecting constipation, normal form, or diarrhea based on intraclass correlation decreased but remained excellent with ICC: 0.75 (95% CI: 0.69 – 0.81, $P<0.001$). However, intra-rater reliability by clinical delineation decreased to good with ICC: 0.65 (95% CI: 0.54 – 0.77, $P<0.001$).

DISCUSSION

Despite its widespread use, to our knowledge, our study is the first to evaluate the BSFS for rater reproducibility in assessing stool form. We found that the BSFS has very high reliability and agreement when used by gastroenterology providers to assess individual stool types. However, when categorizing the stool types into clinically meaningful categories of constipation, normal, or diarrhea per the Rome III standard, reliability and agreement decreased even for these expert ratings. These decrements appeared to occur particularly when raters attempted to differentiate type 2 from type 3 or type 5 from type 6 stools, which represent the boundaries of normal versus constipation or diarrhea, respectively.

Our findings suggest that the reproducibility of the BSFS depends on its application. The BSFS has previously been validated as a measure of stool transit.(1, 18) When assessing stool form, the BSFS appears to be useful to determine a specific type (1–7) of stool a subject is passing, and given the high level of agreement within one specific individual stool type, in determining differences of two or more category types in stool form. However, because of the difficulty in differentiating between type 2 vs. type 3 or type 5 vs. type 6 stools, it is more problematic to designate a stool as representing constipation vs. normal or diarrhea vs. normal, respectively, as defined by the Rome III standard.(9)

A strength of our study approach is the large number of expert raters from several institutions that were included. Raters practiced in several different clinical settings and had expertise in diagnosing and managing patients with disorders affecting stool form. We feel this increases the applicability and generalizability of the findings. Experts were chosen as the initial study population, similar to our previous effort using a pediatric stool scale(19). Future studies assessing reproducibility of the BSFS in different patient populations (e.g. children, adults and the elderly) and clinical entities of interest (e.g. IBS) are a needed next

step. These results can be compared to previously developed stool scales to determine the most reproducible and helpful measure.(19–22). An additional strength was the use of photographs of stool rather than drawings. This more closely resembles the actual experience of evaluating stool form while allowing for uniform assessment. Future evaluation of the BSFS using actual stool or three-dimensional representations, though more onerous, may be helpful in the future.

Limitations to the study included having only a subpopulation of the initial raters agree to re-take the survey for intra-rater analysis. However as noted, we did not identify significant differences between the original sample and those who took the survey again.

In summary, our findings indicate that the use of the BSFS to differentiate normal from abnormal (i.e., constipation, diarrhea) stool form as defined by the Rome III criteria is compromised compared to its ability to be used to differentiate one stool type from another. These findings may alter the manner in which the BSFS is used to determine clinical status, eligibility and outcomes in clinical practice and clinical trials. Future studies investigating whether differentiation between normal from abnormal stool forms can be improved by education of the rater and/or revision of the scale itself are needed.

Acknowledgments

We thank the gastroenterology providers who participated.

FUNDING

Financial and/or intellectual support during the conduct of the study was provided by NIH K23 DK101688 (BPC) and NIH R01 NR05337 and the Daffy's Foundation (RJS), the USDA/ARS under Cooperative Agreement No. 6250-51000-043 (RJS), and P30 DK56338 which funds the Texas Medical Center Digestive Disease Center (BPC, RJS).

Abbreviations

BSFS	Bristol Stool Form Scale
IBS	Irritable Bowel Syndrome

References

1. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. *Scand J Gastroenterol.* 1997; 32:920–924. [PubMed: 9299672]
2. Garsed K, Chernova J, Hastings M, et al. A randomised trial of ondansetron for the treatment of irritable bowel syndrome with diarrhoea. *Gut.* 2014; 63:1617–1625. [PubMed: 24334242]
3. Shin A, Acosta A, Camilleri M, et al. A Randomized Trial of 5-Hydroxytryptamine-Receptor Agonist, YKP10811, on Colonic Transit and Bowel Function in Functional Constipation. *Clin Gastroenterol Hepatol.* 2015; 13:701–708. [PubMed: 25148765]
4. Ballard A, Parker-Autry C, Lin CP, Markland AD, Ellington DR, Richter HE. Postoperative bowel function, symptoms, and habits in women after vaginal reconstructive surgery. *Int Urogynecol J.* 2015; 26:817–821. [PubMed: 25672646]
5. Caroff DA, Edelstein PH, Hamilton K, Pegues DA. Program CDCPE. The Bristol stool scale and its relationship to *Clostridium difficile* infection. *J Clin Microbiol.* 2014; 52:3437–3439. [PubMed: 25031446]

6. Nolan JD, Johnston IM, Pattni SS, Dew T, Orchard T, Walters JR. Diarrhea in Crohn's Disease: investigating the role of the ileal hormone Fibroblast Growth Factor 19. *J Crohn's Colitis*. 2015; 9:125–131. [PubMed: 25518063]
7. Bharucha AE, Low P, Camilleri M, et al. A randomised controlled study of the effect of cholinesterase inhibition on colon function in patients with diabetes mellitus and constipation. *Gut*. 2013; 62:708–715. [PubMed: 22677718]
8. Paramor KA, Ibrahim QI, Sadowski DC. Clinical parameters and symptom severity in males with fecal leakage and incontinence. *Neurogastroenterol Motil*. 2014; 26:361–367. [PubMed: 24329987]
9. Longstreth GF, Thompson WG, Chey WD, Houghton LA, Mearin F, Spiller RC. Functional bowel disorders. *Gastroenterology*. 2006; 130:1480–1491. [PubMed: 16678561]
10. Giannetti E, de'Angelis G, Turco R, et al. Subtypes of irritable bowel syndrome in children: prevalence at diagnosis and at follow-up. *J Pediatr*. 2014; 164:1099–1103. [PubMed: 24485818]
11. Self MM, Czyzewski DI, Chumpitazi BP, Weidler EM, Shulman RJ. Subtypes of irritable bowel syndrome in children and adolescents. *Clin Gastroenterol Hepatol*. 2014; 12:1468–1473. [PubMed: 24486406]
12. Rao S, Lembo AJ, Shiff SJ, et al. A 12-week, randomized, controlled trial with a 4-week randomized withdrawal period to evaluate the efficacy and safety of linaclotide in irritable bowel syndrome with constipation. *Am J Gastroenterol*. 2012; 107:1714–1724. [PubMed: 22986440]
13. Videlock EJ, Cheng V, Cremonini F. Effects of linaclotide in patients with irritable bowel syndrome with constipation or chronic constipation: a meta-analysis. *Clin Gastroenterol Hepatol*. 2013; 11:1084–1092. [PubMed: 23644388]
14. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006; 59:1033–1039. [PubMed: 16980142]
15. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psych Bulletin*. 1979; 86:420–428.
16. Cicchetti D, Bronen R, Spencer S, et al. Rating scales, scales of measurement, issues of reliability: resolving some critical issues for clinicians and researchers. *J Nerv Ment Dis*. 2006; 194:557–564. [PubMed: 16909062]
17. Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Mental Defic*. 1981; 86:127–137.
18. Degen LP, Phillips SF. How well does stool form reflect colonic transit? *Gut*. 1996; 39:109–113. [PubMed: 8881820]
19. Chumpitazi BP, Lane MM, Czyzewski DI, Weidler EM, Swank PR, Shulman RJ. Creation and initial evaluation of a Stool Form Scale for children. *J Pediatr*. 2010; 157:594–597. [PubMed: 20826285]
20. Bekkali N, Hamers SL, Reitsma JB, Van Toledo L, Benninga MA. Infant stool form scale: development and results. *J Pediatr*. 2009; 154:521–526 e521. [PubMed: 19054528]
21. Saps M, Nichols-Vinueza D, Dhroove G, Adams P, Chogle A. Assessment of commonly used pediatric stool scales: a pilot study. *Rev Gastroenterol Mexico*. 2013; 78:151–158.
22. Lane MM, Czyzewski DI, Chumpitazi BP, Shulman RJ. Reliability and validity of a modified Bristol Stool Form Scale for children. *J Pediatr*. 2011; 159:437–441 e431. [PubMed: 21489557]

KEY MESSAGES

- We evaluated rater agreement and reliability of the Bristol Stool Form Scale (BSFS) by administering a stool form survey to 34 gastroenterologists from three different institutions.
- The BSFS has very high reliability and agreement when used by gastroenterology providers to assess individual stool form types.
- However, when categorizing the stool form types into clinically meaningful categories of constipation, normal, or diarrhea per the Rome III standard, reliability and agreement decreased.
- These decrements appeared to occur particularly when raters attempted to differentiate type 2 from type 3 or type 5 from type 6 stools, which represent the boundaries of normal versus constipation or diarrhea, respectively.

Type 1		Separate hard lumps, like nuts (hard to pass)
Type 2		Sausage-shaped but lumpy
Type 3		Like a sausage but with cracks on its surface
Type 4		Like a sausage or snake, smooth and soft
Type 5		Soft blobs with clear-cut edges (passed easily)
Type 6		Fluffy pieces with ragged edges, a mushy stool
Type 7		Watery, no solid pieces. Entirely Liquid

Figure.
Bristol Stool Form Scale that was used during the study. There are seven ordinal categories.

TABLE

Distribution of stool form ratings (type 1 through type 7) by percentage for each stool photograph

Stool Photograph	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7
1	0.0%	0.0%	26.5%	73.5%	0.0%	0.0%	0.0%
2	0.0%	2.9%	2.9%	5.9%	32.4%	55.9%	0.0%
3	0.0%	0.0%	0.0%	0.0%	0.0%	5.9%	94.1%
4	97.1%	0.0%	0.0%	2.9%	0.0%	0.0%	0.0%
5	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%
6	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
7	0.0%	0.0%	0.0%	0.0%	0.0%	91.2%	8.8%
8	0.0%	0.0%	20.6%	79.4%	0.0%	0.0%	0.0%
9	0.0%	0.0%	0.0%	0.0%	14.7%	79.4%	5.9%
10	0.0%	0.0%	0.0%	0.0%	61.8%	35.3%	2.9%
11	0.0%	0.0%	0.0%	0.0%	58.8%	32.4%	8.8%
12	79.4%	11.8%	0.0%	0.0%	8.8%	0.0%	0.0%
13	0.0%	2.9%	5.9%	91.2%	0.0%	0.0%	0.0%
14	0.0%	2.9%	0.0%	0.0%	0.0%	41.2%	55.9%
15	0.0%	20.6%	26.5%	50.0%	2.9%	0.0%	0.0%
16	0.0%	0.0%	0.0%	0.0%	2.9%	82.4%	14.7%
17	0.0%	0.0%	0.0%	0.0%	50.0%	38.2%	11.8%
18	97.1%	2.9%	0.0%	0.0%	0.0%	0.0%	0.0%
19	0.0%	5.9%	0.0%	0.0%	58.8%	32.4%	2.9%
20	0.0%	0.0%	0.0%	0.0%	41.2%	58.8%	0.0%
21	0.0%	2.9%	0.0%	0.0%	50.0%	44.1%	2.9%
22	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
23	0.0%	58.8%	23.5%	0.0%	5.9%	11.8%	0.0%
24	0.0%	20.6%	61.8%	14.7%	2.9%	0.0%	0.0%
25	0.0%	50.0%	26.5%	5.9%	11.8%	5.9%	0.0%
26	2.9%	26.5%	47.1%	8.8%	14.7%	0.0%	0.0%
27	0.0%	0.0%	23.5%	76.5%	0.0%	0.0%	0.0%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Stool Photograph	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7
28	2.9%	52.9%	41.2%	2.9%	0.0%	0.0%	0.0%
29	0.0%	85.3%	0.0%	0.0%	2.9%	11.8%	0.0%
30	0.0%	0.0%	0.0%	0.0%	35.3%	58.8%	5.9%
31	0.0%	52.9%	29.4%	11.8%	5.9%	0.0%	0.0%
32	0.0%	0.0%	0.0%	0.0%	0.0%	41.2%	58.8%
33	0.0%	0.0%	76.5%	20.6%	2.9%	0.0%	0.0%
34	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%
35	0.0%	0.0%	2.9%	0.0%	0.0%	35.3%	61.8%