



Published in final edited form as:

*J Pediatr.* 2010 October ; 157(4): 594–597. doi:10.1016/j.jpeds.2010.04.040.

## CREATION AND INITIAL EVALUATION OF A STOOL FORM SCALE FOR CHILDREN

Bruno P. Chumpitazi, MD, MPH<sup>1,3</sup>, Mariella M. Lane, PhD<sup>1,2,3</sup>, Danita I. Czyzewski, PhD<sup>1,2,3</sup>, Erica M. Weidler, BA<sup>1,3,5</sup>, Paul R. Swank, PhD<sup>4</sup>, and Robert J. Shulman, MD<sup>1,3,5</sup>

<sup>1</sup>Department of Pediatrics, Baylor College of Medicine

<sup>2</sup>Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine

<sup>3</sup>Texas Children's Hospital, University of Texas Health Science Center- Houston

<sup>4</sup>Children's Learning Institute, University of Texas Health Science Center- Houston

<sup>5</sup>Children's Nutrition Research Center, Houston, Texas

### Abstract

**Objective**—To develop a pediatric stool form rating scale and determine its inter-rater reliability, intra-rater reliability, and agreement amongst pediatric gastroenterologists.

**Study design**—An ordinal stool scale with five categorical stool form types was created based on the Bristol Stool Form Scale (BSFS), and 32 color two-dimensional stool photographs were shown to 14 pediatric gastroenterologists. Each gastroenterologist rated the stool form depicted in each photograph using the modified stool scale. Ten gastroenterologists agreed to re-rate the stool form depicted in each photograph a minimum of six months after the first rating.

**Results**—448 ratings were completed; 430 (94%) of all ratings were within at least one category type of the most common (modal) rating for each photograph. Eight (25%) stool photographs had complete agreement amongst all raters. Inter-rater and intra-rater reliability was high with a single measure intraclass correlation of 0.85 (95% CI: 0.78–0.91;  $P < 0.001$ ) and 0.87 (95% CI: 0.81–0.92;  $P < 0.001$ ) respectively.

**Conclusion**—A modified pediatric BSFS provided a high degree of inter-rater reliability, intra-rater reliability, and agreement amongst pediatric gastroenterologists.

Alterations in stool form and frequency are associated with numerous gastrointestinal disorders ranging from inflammatory disorders (e.g. ulcerative colitis) to functional gastrointestinal disorders (e.g. irritable bowel syndrome). Diagnostic criteria for gastrointestinal disorders may depend, in part, on whether changes in stool form are associated with symptoms<sup>1</sup>. Stool form may also guide assessment of treatment efficacy or determine the clinical status of a disorder. Therefore, assessment of stool form may aide clinicians in diagnosis and management of underlying gastrointestinal disorders, and may serve as a measurable clinical outcome employed in the clinical research setting<sup>2,3</sup>.

© 2010 Mosby, Inc. All rights reserved.

Correspondence/Reprint Request Contact: Bruno Chumpitazi, M.D., M.P.H., Texas Children's Hospital, 6701 Fannin Street, CCC 1010.00, Houston, TX 77030, bpchumpi@texaschildrens.org.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

An often used measure of stool form is the Bristol Stool Form Scale (BSFS)<sup>4</sup>. This scale allows one to classify stool form into seven types ranging from “separate hard lumps like nuts” (type 1) to “watery, no solid pieces, entirely liquid” (type 7)<sup>5</sup>. However, the BSFS was validated in adults as a measure of stool transit rather than as a means of identifying stool form<sup>4</sup>. Despite this fact, it has been used to evaluate stool form in a variety of clinical studies including a population of healthy adults<sup>5</sup>, adults with HIV-related diarrhea<sup>3</sup>, and adults with functional bowel disorders<sup>2</sup>. It was recently adapted for the Spanish language<sup>6</sup>. Despite its lack of validation, the Rome foundation has recommended the BSFS be used for assessing stool form in adults with functional gastrointestinal disorders<sup>1</sup>.

Studies characterizing stool form in children have been conducted, but the pediatric scales used in these studies were never validated<sup>7, 8</sup>. Given the lack of a scale validated to assess stool form for use in adults or children, we sought to develop such a scale, and assess its inter-rater reliability, intra-rater reliability, and agreement when used by expert raters.

## METHODS

This study was approved by the Baylor College of Medicine Institutional Review Board.

The original BSFS was reviewed by a group consisting of two pediatric gastroenterologists, two pediatric psychologists, and three research assistants experienced with clinical trials in which children had reported stool form<sup>9, 10</sup>. Since the ultimate goal was to develop a scale that could be used by children to rate their own stools, consideration of childhood cognition deemed that the seven categories of the BSFS be reduced. Thus, Type 3 “like a sausage or snake but with cracks on its surface” and Type 5 “soft blobs with clear cut edges” were eliminated as options (Figure). The final scale consisted of line drawings of the five stool forms accompanied by short descriptors.

Thirty-two color, two-dimensional photographs of stool forms ranging from liquid to formed to hard pellets were obtained from the public domain (publicly accessible areas of the internet) to be used as stimuli for the ratings. Stools were in various real-world settings (e.g. within a toilet or within a diaper). Only photographs that were focused, close-ups of entire bowel movements with white backgrounds were chosen.

Fourteen pediatric gastroenterologists from Baylor College of Medicine/Texas Children’s Hospital were asked to rate the 32 photographs for stool form based on the modified stool scale. Each page of the survey contained an individual stool photograph, with the modified stool scale beneath. All physicians who were asked to participate completed the initial survey. No special instruction in the use of the modified stool scale was given to any of the physicians. However, the gastroenterologists were instructed not to assist one another or discuss the ratings among themselves. The initial surveys were performed between August 2008 and August 2009. All previous participants were asked to again complete the survey at least 6 months after the initial assessment to measure intra-observer reliability. Ten of the original 14 physicians agreed to complete the repeat evaluation.

### Statistical Analysis

The stool rating for each stool photograph was not pre-determined and the most commonly chosen (modal) rating by the physicians for each photograph was noted. Further statistical measures included determination of inter-observer and intra-observer reliability using intraclass correlation coefficient (two-way random effects model with absolute agreement) of single measures, percent exact agreement, variance in the ratings due to the raters, variance in the ratings due to the photographs themselves, variance in the rating due to the interactions of rater per photograph, and percent within one rating type of the most common (modal) rating

chosen by the physicians for each photograph. We chose intraclass correlation coefficients to measure reliability as we felt this method would take into account the fact that measurements were exchangeable (i.e., future stool photographs could be treated the same as past stool photographs), and this type of evaluation more appropriately accounts for the ordinal nature (e.g. stools are looser as one increases in stool category type) of the modified stool scale as adjacent categories are more similar than non-adjacent ones 11. Parameters of agreement and reliability were both determined as each measures a related but different aspect of reproducibility 12. A Pearson correlation was calculated to evaluate the relation between clinical experience and number of ratings deviating from the most commonly chosen (modal) rating for each stool photograph.

If not otherwise specified, data are presented as mean  $\pm$  standard deviation. Statistical Analysis Software (SAS) was used for statistical calculations.

## RESULTS

The 14 pediatric gastroenterologists who completed the ratings had been trained at six different pediatric gastroenterology fellowship programs. The gastroenterologists had a median of 17.5 years of clinical experience after completing fellowship (range 1 – 38 years) and a mean age of  $50.4 \pm 11.5$  years (range 33 – 71). Nine of the 14 were male.

The raters made a total of 448 stool type selections using the modified stool scale (Figure); the Table presents the distribution of stool types selected for each photograph. Three photographs most commonly received a rating of type 1, seven photographs as type 2, nine as type 3, ten as type 4, and three as type 5. Years of clinical experience was not significantly related to the number of ratings deviating from the most commonly chosen (modal) rating for each stool photograph ( $r= 0.16$ ,  $P=0.58$ .)

The variance in initial inter-rater ratings due to the raters themselves (e.g. assessment of one rater rating all photographs more toward one end of the scale or another) was very low at 0.009. The inter-rater variance in ratings due to the interaction of raters by photographs (e.g. assessment of inconsistent choices of a rater by photograph) was also low at 0.196. The variance in ratings due to the photographs themselves was much higher at 1.194. As such, the signal to noise ratio (variance in ratings due to photographs themselves versus other causes of variance in ratings) was high at 6.1.

### Inter-rater Reliability and Agreement

Of the 448 ratings that were made, 373 (83.3%) were in agreement with the most commonly chosen (modal) rating. Of the 448 stool type selections made, 430 (96%) were within at least one form type of the most commonly chosen (modal) rating for each stool photograph. Eight of thirty two (25%) stool photographs received unanimous stool type assignment. The overall inter-item correlation for all physicians across photographs was 0.842 (range 0.634 – 0.977). The single measures intra-class correlation for inter-rater reliability was 0.85 (95% CI: 0.77 – 0.91),  $P<0.001$ .)

### Intra-rater Reliability

Ten of the 14 initial raters agreed to retake the survey a minimum of 6 months after the initial survey was conducted. No significant differences between the original group of raters and the subgroup that retook the survey occurred with respect to mean initial age ( $50.4 \pm 11.5$  versus  $48.7 \pm 11.9$  years), initial mean clinical experience ( $17.2 \pm 12$  versus  $15.2 \pm 11.6$  years), and mean number of selections per physician that differed from the most commonly selected for each stool photograph ( $5.4 \pm 2.6$  versus  $5.1 \pm 2.6$ ) respectively. No significant difference in

sex composition between the original group of raters and the subgroup of raters taking the survey twice (35.7% versus 30% women) was present respectively.

The single measures intra-class correlation comparing the initial survey results with those of the repeat survey was 0.87 (95% CI: 0.81–0.92),  $P < 0.001$ .)

## DISCUSSION

When used by experts to evaluate photographs of a wide variety of stool forms, we have demonstrated that a modified pediatric BSFS has a high degree of overall inter-rater reliability, intra-rater reliability, and agreement. Clinical experience did not influence rating selections and the vast majority of the variance in ratings was found to be due to the photographs themselves.

Surprisingly, despite the utility of knowing a patient's stool form for both clinical and research purposes, few have attempted to validate such scales. Despite its widespread use to describe stool character, the BSFS was never validated against actual stools (pictures or otherwise)<sup>4</sup>. Rather, the diagrams on the BSFS were shown to correspond to stool transit time in adults<sup>4</sup>. Bekkali et al used color photographs of infant stools in diapers to validate a new infant stool form scale<sup>13</sup>. We also chose stool photographs as we felt this would allow for a uniform assessment, would be more feasible, and in comparison with diagrams would more closely resemble the actual experience of evaluating stool form. Future validation of stool form scales using actual stool or three-dimensional images may be helpful but much more onerous.

In addition to using stool photographs as the stimuli used by the raters, our approach to the development of the modified pediatric BSFS has the strength of using a large number of expert raters in the initial evaluation. This is in comparison with the two raters (medical student and a senior attending physician) used in creating the Bekkali et al infant stool scale<sup>13</sup>. Demonstrating high reproducibility with physicians of various training and clinical backgrounds suggests the modified pediatric BSFS is a strong measure and may be useful in various clinical practice and research settings.

The adaptation of the BSFS was made based on previous experience<sup>9, 10</sup> and desire ultimately, to maximize the feasibility and appropriate use of the scale by children. As such, pictorial representations were retained as graphic scales are thought to offer children more information on how to grade answers on self-report measures<sup>14</sup>. Pictorial representations have also been shown to reduce child demands on memory, to maintain attention, and to avoid reliance on verbal or reading skills that have not fully developed<sup>15</sup>. Decreasing categories is supported by research taking into account the cognitive developmental abilities of children, particularly young children, when creating scales and inventories<sup>16</sup>. Further, the five stool types chosen likely encompass the clinically relevant differentiations.

Pediatric gastroenterologists were chosen as the initial group for evaluation of this modified stool scale because they frequently diagnose and manage children with disorders involving abnormalities in stool form. Therefore this group is more likely to have clinical expertise in stool form assessment. As such, if pediatric gastroenterologists were unable to successfully use the modified stool scale to reach agreement on stool form, we felt it would be unlikely that other groups (e.g. children themselves) would be able to use the scale reliably.

One limitation for this study is a lack of a multi-center evaluation with physicians at various institutions. In theory, the fact that all physician raters in the study practiced in one institution increases the likelihood that the physicians would acquire common practices including evaluation of stools. We feel this was potentially ameliorated by including as many raters of

various backgrounds as possible. Nonetheless, future validation of this and other stool scales would benefit from a multi-institutional evaluation.

Another limitation is that not all of the original raters agreed to participate in the second survey. However no significant differences between the original group of raters and the subgroup that retook the survey occurred with respect to sex composition, age, years of clinical experience, and number of ratings that differed from that most commonly selected for each stool photograph. As such, we feel that the subgroup captures the full range of variability from the original group, and hence was able to measure intra-observer reliability well.

It is our hope that this modified pediatric BSFS will be adopted and used in both clinical and research settings as an objective measure to record stool form relatively quickly by those caring for children. Future studies should include evaluation of the psychometric properties of this scale when used by children of various age groups. Evaluation of the reliability of parents using the modified pediatric BSFS should be conducted while comparing the results with those obtained from the children themselves. Future studies using the modified pediatric BSFS may include capturing defecation stool form patterns in healthy children of various ages and in those with disorders related to changes in stool form (e.g. irritable bowel syndrome) to better understand normal patterns and/or aide in diagnosis of pathologic states.

## Acknowledgments

We thank the physicians that agreed to participate in this study.

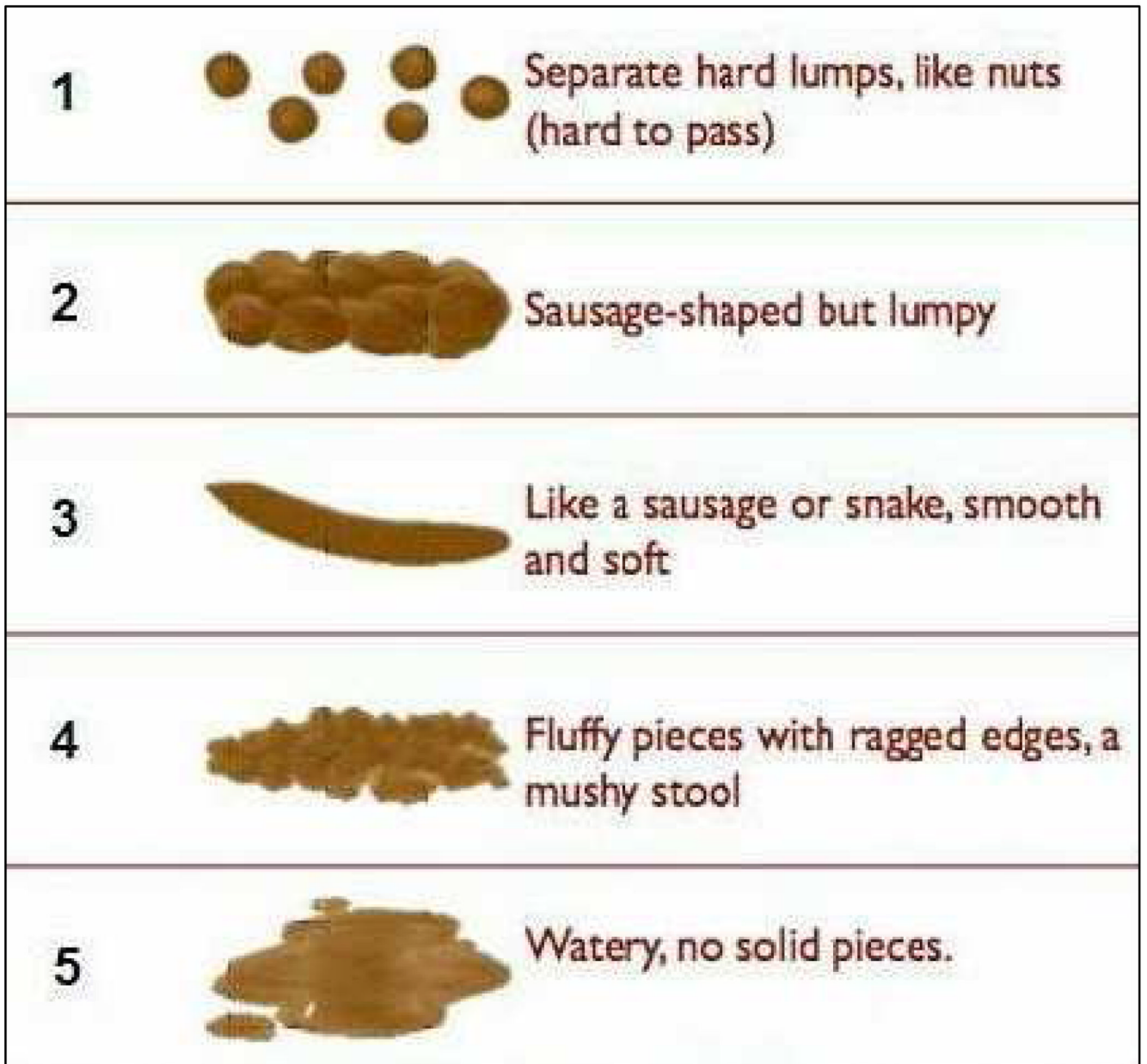
Supported primarily by an investigator-initiated grant from Takeda Pharmaceuticals. Salary support to one or more of the authors during the conduct of this study has been provided by R01 NR05337, UH2 DK083990, and RC2 NR011959 from the National Institutes of Health, the Daffy's Foundation, the USDA/ARS under Cooperative Agreement No. 6250-51000-043, and P30 DK56338, which funds the Texas Medical Center Digestive Disease Center. Sponsors were not involved with: determining the study design; the collection, analysis, and interpretation of data; the writing of the report; or the decision to submit the paper for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The contents of this publication do not necessarily reflect the views or policies of the USDA, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

## REFERENCES

1. Longstreth GF, Thompson WG, Chey WD, Houghton LA, Mearin F, Spiller RC. Functional bowel disorders. *Gastroenterology* 2006;130:1480–1491. [PubMed: 16678561]
2. Austin GL, Dalton CB, Hu Y, Morris CB, Hankins J, Weinland SR, et al. A very low-carbohydrate diet improves symptoms and quality of life in diarrhea-predominant irritable bowel syndrome. *Clin Gastroenterol Hepatol* 2009;7:706–708. e1. [PubMed: 19281859]
3. Tinmouth J, Tomlinson G, Kandel G, Walmsley S, Steinhart HA, Glazier R. Evaluation of Stool frequency and stool form as measures of HIV-related diarrhea. *HIV Clin Trials* 2007;8:421–428. [PubMed: 18042507]
4. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. *Scand J Gastroenterol* 1997;32:920–924. [PubMed: 9299672]
5. Heaton KW, Radvan J, Cripps H, Mountford RA, Braddon FE, Hughes AO. Defecation frequency and timing, and stool form in the general population: a prospective study. *Gut* 1992;33:818–824. [PubMed: 1624166]
6. Pares D, Comas M, Dorcaratto D, Araujo MI, Vial M, Bohle B, et al. Adaptation and validation of the Bristol scale stool form translated into the Spanish language among health professionals and patients. *Rev Esp Enferm Dig* 2009;101:312–316. [PubMed: 19527076]
7. Sandhu B, Steer C, Golding J, Emond A. The early stool patterns of young children with autistic spectrum disorder. *Arch Dis Child* 2009;94:497–500. [PubMed: 19329445]
8. Steer CD, Emond AM, Golding J, Sandhu B. The variation in stool patterns from 1 to 42 months: a population-based observational study. *Arch Dis Child* 2009;94:231–233. [PubMed: 18676434]

9. Shulman RJ, Eakin MN, Czyzewski DI, Jarrett M, Ou CN. Increased gastrointestinal permeability and gut inflammation in children with functional abdominal pain and irritable bowel syndrome. *J Pediatr* 2008;153:646–650. [PubMed: 18538790]
10. Shulman RJ, Eakin MN, Jarrett M, Czyzewski DI, Zeltzer LK. Characteristics of pain and stooling in children with recurrent abdominal pain. *J Pediatr Gastroenterol Nutr* 2007;44:203–208. [PubMed: 17255832]
11. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–428. [PubMed: 18839484]
12. 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]
13. Bekkali N, Hamers SL, Reitsma JB, Van Toledo L, Benninga MA. Infant stool form scale: development and results. *J Pediatr* 2009;154:521–526. e1. [PubMed: 19054528]
14. Cremeens J, Eiser C, Blades M. Characteristics of health-related self-report measures for children aged three to eight years: a review of the literature. *Qual Life Res* 2006;15:739–754. [PubMed: 16688506]
15. Salmon K, Yao J, Berntsen O, Pipe ME. Does providing props during preparation help children to remember a novel event? *J Exp Child Psychol* 2007;97:99–116. [PubMed: 17328907]
16. Varni JW, Waldron SA, Gragg RA, Rapoff MA, Bernstein BH, Lindsley CB, et al. Development of the Waldron/Varni pediatric pain coping inventory. *Pain* 1996;67:141–150. [PubMed: 8895242]





**Figure.** Modified stool form scale used that was used during the study. There are five ordinal categories.

**Table**

Distribution of Stool Form Ratings (Type 1 Through Type 5) by Percentage for Each Stool Photograph

Stool Photograph	Type 1	Type 2	Type 3	Type 4	Type 5
1	0%	0%	100%	0%	0%
2	0%	0%	0%	100%	0%
3	0%	71.4%*	21.4%	7.1%	0%
4	0%	7.1%	0%	64.3%	28.6%
5	0%	0%	0%	0%	100%
6	100%	0%	0%	0%	0%
7	0%	0%	0%	92.9%	7.1%
8	0%	7.1%	92.9%	0%	0%
9	0%	0%	0%	92.9%	7.1%
10	0%	7.1%	0%	92.9%	0%
11	100%	0%	0%	0%	0%
12	100%	0%	0%	0%	0%
13	0%	0%	100%	0%	0%
14	0%	7.1%	78.6%	14.3%	0%
15	0%	0%	100%	0%	0%
16	0%	7.1%	0%	85.7%	7.1%
17	7.1%	21.4%	28.6%	42.9%	0%
18	0%	28.6%	71.4%	0%	0%
19	0%	0%	71.4%	28.6%	0%
20	0%	0%	0%	92.9%	7.1%
21	0%	71.4%	21.4%	7.1%	0%
22	0%	0%	0%	92.9%	7.1%
23	0%	0%	57.1%	42.9%	0%
24	7.1%	71.4%	7.1%	14.3%	0%
25	0%	0%	0%	7.1%	92.9%
26	0%	42.9%	50%	7.1%	0%



Stool Photograph	Type 1	Type 2	Type 3	Type 4	Type 5
27	0%	0%	0%	7.1%	92.9%
28	7.1%	92.9%	0%	0%	0%
29	0%	0%	0%	71.4%	28.6%
30	0%	50%	28.6%	21.4%	0%
31	0%	85.7%	7.1%	7.1%	0%
32	0%	78.6%	7.1%	14.3%	0%

\* Values are rounded to the nearest decimal