Critical evaluation of the Hirschsprung-associated enterocolitis (HAEC) score: A multicenter study of 116 children with Hirschsprung disease

Philip K. Frykman[a,*], **Sungjin Kim**[b], **Tomas Wester**[c,d], **Agneta Nordenskjöld**[c,d], **Akemi Kawaguchi**[e], **Thomas T. Hui**[f], **Daniel H. Teitelbaum**[g], **Anna L. Granström**[c,d], **Andre Rogatko**[b], and **for the HAEC Collaborative Research Group (HCRG)**

[a]Division of Pediatric Surgery and Departments of Surgery and Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, California, USA

[b]Biostatistics and Bioinformatics Research Center at the Cedars-Sinai Samuel Oschin Comprehensive Cancer Institute, Los Angeles, California, USA

[c]Department of Pediatric Surgery, Astrid Lindgren's Children's Hospital, Karolinska University Hospital, Stockholm, Sweden

[d]Department of Women's and Children's Health and Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden

[e]Department of Pediatric Surgery, University of Texas Health Science Center Houston, Houston, TX, USA

[f]Division of Pediatric Surgery, University of California San Francisco Benioff Children's Hospital Oakland, Oakland, California, USA

[g]Division of Pediatric Surgery, C.S. Mott Children's Hospital, University of Michigan Medical School, Ann Arbor, MI, USA

Abstract

Objective: To identify the optimal clinical criteria to diagnose Hirschsprung-associated enterocolitis (HAEC) in children with Hirschsprung disease (HSCR).

Background: HAEC is the most common life-threatening complication in HSCR patients, yet the diagnostic criteria for HAEC remain unclear. The consensus-based HAEC scoring system was not validated using patient data, thereby making its diagnostic accuracy uncertain.

Methods: From 2009 to 2015, consecutive children with HSCR underwent retrospective evaluation of their medical records, and questionnaire-directed parent interviews to identify treatment of suspected HAEC episodes and the 16 clinical criteria in the HAEC score. Logistic regression modeling was employed to identify criteria predicting suspected HAEC episodes.

*Corresponding author at: Division of Pediatric Surgery, Cedars-Sinai Medical Center, 8635 W. Third St., Suite 675 W, Los Angeles, California 90048. Tel.: +1 310 423 6235; fax: +1 310 423 5454. Philip.Frykman@cshs.org (P.K. Frykman).

Results: One hundred sixteen HSCR patients met inclusion criteria, 43 patients (37.1%) were treated for at least one suspected HAEC episode. An HAEC score of 4 maximized the sum of sensitivity (83.7%) and specificity(98.6%) while the previously established cut-off score of 10 showed lower sensitivity (41.9%) with perfect spec-ificity. Multivariable analysis identified four criteria utilized to create a new HAEC Risk score with performance characteristics similar to the HAEC score cutoff of 4.

Conclusion: When using the HAEC score, a cutoff of 4 should be used rather than 10, which under-diagnosed patients with HAEC. Alternatively, the new HAEC Risk score could be employed.

Level of Evidence: Diagnostic Study, Level 3.

Keywords

Hirschsprung-associated enterocolitis (HAEC) is the most frequent potentially life-threatening complication of Hirschsprung disease (HSCR) patients [1].[1] Reported incidence varies widely, ranging from 17% to 50% with most contemporary series reporting approximately 30% incidence in HSCR patients [2,3]. One of the major factors explaining this wide variation is lack of an agreed upon definition of what constitutes HAEC, given there is significant overlap of symptoms with other conditions. One attempt to standardize the diagnostic criteria for HAEC, Pastor et al. [4] developed a scoring system using a Delphi analysis to gain consensus from a panel of experts comprising pediatric surgeons and pediatric gastroenterologists. While a significant step forward in better defining the most important criteria, preliminary validation of the HAEC score was limited to case scenario applications, not patient data. Further, the authors stated, the HAEC scoring system was intended to be used as a standardized and reproducible outcome measure to compare studies and not primarily for clinical diagnosis. To date, only two published studies (by our group) have applied the HAEC scoring system [5,6] to 20 HSCR patients (who are included in this cohort). Through these prior studies we began to recognize limitations in the HAEC scoring system to accurately diagnose patients with HAEC.

In this study, we sought to define the optimal set of clinical criteria to improve accuracy of HAEC diagnosis through studying 116 HSCR patients from five centers, all of whom had completed definitive surgery to treat HSCR. We applied the 16 HAEC score criteria to our cohort of HSCR patients and then performed robust statistical analysis.

1. Materials and methods

1.1. Patients and study design

Inclusion criteria for enrollment were children less than 18 years of age with histopathological diagnosis of Hirschsprung disease who had completed definitive pull-through surgery. Exclusion criteria included colonic pseudoobstruction and intestinal neuronal dysplasia. This research was approved by the Cedars-Sinai Medical Center IRB

(Protocol# 00020809) as a multi-center study, and has been conducted according to the principles expressed in the Declaration of Helsinki. Written informed consent was obtained from a parent by the attending surgeons, research fellows, or research nurses at each site. From 2009 to 2015, children were enrolled by five member institutions of the HAEC Collaborative Research Group (HCRG): Cedars-Sinai Medical Center, Los Angeles, California; Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm, Sweden; Children's Hospital Los Angeles, Los Angeles, California; Children's Hospital of Oakland, Oakland, California; C.S. Mott Children's Hospital, University of Michigan, Ann Arbor, Michigan. We enrolled 116 children with HSCR. We performed a two-step process to collect clinical data for this study. In the first, we performed a retrospective review of all available medical records using standardized questionnaires that included demographic, medical history, surgical history, radiographic findings, histopathology, complications and information regarding HAEC signs, symptoms (including the 16 items in the HAEC score by Pastor et al. [4]) and treatments of suspected HAEC episodes. To augment the medical record review, our second step was to perform patient and/or parent/guardian interviews to clarify any discrepancies or ambiguities in the medical records and focused on past HAEC episodes, the 16 items in the HAEC score (in each episode) and the treatment of suspected HAEC episodes.

Given the retrospective design of the study, we created multiple levels of oversight and review to optimize consistency, accuracy and completeness of data collection to achieve the best quality possible. First, detailed instructions (encompassing 19 pages) regarding which data to collect and how it should be recorded were given to all study staff, along with a live tutorial over Skype with the lead site (CSMC). Following this training, attending pediatric surgeons (individual site PI's), pediatric surgery research fellows and experienced research nurses with at least 2 years of experience with Hirschsprung disease research performed the first step retrospective medical record review. At each site, pediatric surgery research fellows and experienced research nurses reviewed each subject's medical record with the site PI (an attending pediatric surgeon) to maintain consistency of interpretation, accuracy and completeness of data. Only site PI's and pediatric surgery research fellows conducted the parent/guardian/patient interviews. Ultimately, the assignment of values of each variable was made by the site PI's who had access to the overall study PI (PKF) at CSMC to provide guidance on an ad hoc basis. The final layer of oversight was the overall study PI and study coordinator at CSMC, who reviewed all entries into the database during the study period to maintain consistency with interpretation and ensure data completeness for each subject. When questions arose regarding entered data or interpretation of data, there was a discussion between the overall study PI and site PI's to resolve the question. In addition, regular HCRG meetings via Skype were held every 2 months over the study period to discuss study progress and any issues regarding data acquisition and interpretation. When information was not obtainable from the medical record it was noted as "missing". The HCRG data was stored in a secure SQL relational database at the data-coordinating center at CSMC where central review of cases was performed; however, central review did not include radiographs.

1.2. Primary outcome

The primary outcome was identification of clinical predictors for the presence or absence of a suspected episode of HAEC based on history of documented prior treatment (inpatient or outpatient) for suspected HAEC. Predictors included the following 16 clinical criteria composing HAEC score described in Pastor et al. [4] during a suspected episode: distended abdomen, diarrhea with explosive stool, diarrhea with foul smelling stool, lethargy, explosive discharge of gas and stool on rectal exam, fever, leukocytosis, decreased peripheral perfusion, previous history of suspected enterocolitis, "left shift" on complete blood count, diarrhea with bloody stool, dilated loops of bowel, multiple air-fluid levels, "cutoff sign" in rectosigmoid region, "sawtooth" appearance with irregular mucosal lining, and pneumatosis intestinalis on abdominal radiograph. A secondary objective was determination of similarity between the 16 criteria.

1.3. Statistical analysis

Data were presented as frequency (percentage, %) for categorical variables and median (interquartile range, IQR) for continuous variables. Cohen's kappa coefficient (κ) [7] was used to measure the similarity between all possible pairs of the 16 criteria. Hierarchical clustering analysis was conducted for grouping similar criteria according to a distance defined as $1 - \kappa$ between a pair of criteria using mean linkage clustering, which finds all possible pairwise distances for criteria belonging to two different clusters, and then averages them [8]. The number of clusters was chosen with the largest average silhouette width criterion [9]. A logistic regression model was employed in univariate and multivariable analyses to identify criteria that predict the outcome. Firth's penalized maximum likelihood estimation along with a profile-likelihood confidence interval from the penalized likelihood ratio test was employed in cases of separation in logistic regression [10,11]. Variable selection was carried out as outlined by Collett [12] and the possibility of multicollinearity was assessed by tolerance and the variance inflation factor (VIF). To create a risk score, points associated with variables in the constructed multivariable model were calculated by dividing parameter estimates (estimates of the regression coefficients) for the variables in the model by the natural logarithm of 2 so that each one-point increase in the risk score corresponds to a 2-fold increase in the risk of HAEC episode. The points were rounded to the nearest integer to simplify the risk score. The final risk score for each patient was calculated by summing the points for each variable in the model. To determine the optimal threshold values to discriminate patients with and without HAEC episode for both the new risk and the established HAEC scores [4], we examined their sensitivity and specificity. Note that the sum of sensitivity and specificity is comparable with Youden Index[13], a well-known measure for classification performance. Receiver operating characteristic (ROC) curves along with the areas under the ROC curve (AUCs) were generated for the risk score and the established HAEC score to further assess discrimination [14]. Calibration of the model prediction was graphically assessed with predicted versus observed probability of HAEC episode based on the loess algorithm [15]. Internal validation of the model was performed by estimating and correcting possible overfitting and optimism in the model performance estimates using the bootstrap method with 1000 replicates [15–17]. All analyses were done using SAS 9.3 (SAS Institute, Inc., Cary, North Carolina) and R package

version 3.2.2 (cluster, pROC, rms, and Hmisc library; The R Foundation for Statistical Computing) with two-sided tests and a significance level of 0.05.

2. Results

2.1 . Patient characteristics

The cohort of 116 HSCR patients studied median age of 6 (IQR 3–8), 99 (85.34%) male and 7 (6%) had trisomy 21 (Table 1). Eighty-five patients (85%) had aganglionosis restricted to the rectosigmoid colon, with 15 (15%) aganglionosis extending proximal to the sigmoid colon at the time of definitive pull-through operation, and 16 had missing data. We found 43 (37.1%) of 116 had at least one suspected episode of HAEC; 38 (32.8%) had 1–4 episodes, 3 (2.6%) had 5–9 episodes and 2 (1.7%) had 10 episodes or more. In 9.8%, the first HAEC episode occurred pre-diversion or pre-pullthrough procedure, while the remainder occurred post-pullthrough procedure.

2.2. Clinical criteria composing the HAEC score

The HAEC score for each subject used in this analysis was a compilation of the 16 criteria identified in each episode of suspected HAEC for each subject. Twenty-three of 43 patients had a single suspected HAEC episode documented. In the remaining 20 patients who experienced greater than one episode of HAEC, 16 (80%) of 20 had a consistent HAEC score between episodes. Hence compilation of criteria from multiple episodes into a single score per patient accurately represents most individual's presentation of HAEC.

When we evaluated the cohort of HSCR patients for the 16 clinical criteria composing the HAEC score, we found that 31% had a distended abdomen, 24% had diarrhea with explosive stool, 23% had diarrhea with foul smelling stool, 19.8% had lethargy and 19% had explosive discharge of gas and stool on rectal exam and fever, respectively (Table 2). Additionally, 15.5% had dilated loops of bowel, 12.9% had leukocytosis, 11.2% had decreased peripheral perfusion, 8.6% had multiple air-fluid levels, 7.8% had a previous history of suspected enterocolitis or "left shift" on complete blood count. Less prevalent findings were diarrhea with bloody stool(2.6%), "cutoff sign" in rectosigmoid region (1.7%), and "sawtooth" appearance with irregular mucosal lining or pneumatosis intestinalis each (0.9%). Not surprisingly, the median HAEC score was 0, with an interquartile range of 0–6.5. However, when the HAEC score cutoff value of 10 was applied as proposed in Pastor et al. only 18 (15.5%) of 116 patients had a score of 10 or greater. With less than half of patients in the cohort with suspected HAEC meeting the score of 10, it raised the possibility that the cutoff may be overly restrictive, thereby missing some patients who have HAEC.

2.3. Univariate analysis of the presence or absence of suspected HAEC episode

Univariate analyses evaluating the 16 clinical criteria with presence or absence of suspected HAEC found that lethargy (OR: 168.5; 95% CI:21.71–21,732.18; p < .001), distended abdomen (OR: 134.07; 95% CI:27.46–654.53; p < .001), diarrhea with explosive stool (OR: 121.50; 95% CI: 15.36–961.06; p < .001), diarrhea with foul smelling stool (OR: 110.12; 95% CI: 13.95–869.21; p < .001), leukocytosis (OR: 79.95; 95% CI: 10.16–10,330.11; p < .001), fever or explosive discharge of gas on rectal exam (OR: 68.73; 95% CI: 8.74–540.38;

p < .001), decreased peripheral perfusion (OR: 65.07; 95% CI: 8.18–8421.86; p < .001), "left shift" or previous history of enterocolitis (OR: 40.48; 95% CI:4.89–5273.84; p < .001), dilated loops of bowel on abdominal film (OR: 21.04; 95% CI: 4.53–97.67; p < .001), multiple air-fluid levels on abdominal film (OR: 19.06; 95% CI: 2.32–156.55; p = .006), and diarrhea with bloody stool (OR: 12.70; 95% CI: 1.19–1724.37; p = .034) were associated with an increased Incidence of suspected HAEC while presence of radiographic findings of "cutoff sign", "sawtooth" appearance with irregular mucosal lining, and pneumatosis intestinalis were not associated with the presence of suspected HAEC (Table 2).

Threshold values for HAEC scores to discriminate patients with and without suspected HAEC episode were determined (Table 3). We found that a threshold value of HAEC score of 4 maximized the sum of sensitivity 83.72% (95% CI: 69.30–93.19) and specificity 98.63% (95% CI: 92.60–99.97) with AUC of 0.91 (95% CI: 0.85–0.97) (Table 3 and Fig. 1), while a cut-off value of HAEC score of 10, which is the pre-established threshold value described in Pastor et al. [4], showed a relatively lower sensitivity of 41.86% (95% CI: 27.01–57.87) with 100% spec-ificity (95% CI: 95.07–100) and AUC of 0.71 (95% CI: 0.63–0.78). Additionally, we found that an HAEC score of 2 maximized sensitivity of 86.05% (95% CI: 72.07–94.70) while still maximizing specificity of 95.89% (95% CI: 88.46–99.14) with AUC of 0.91 (95% CI: 0.85–0.97) and that an HAEC score of 9 maximized specificity of 100% (95% CI:95.07–100) while still maximizing sensitivity of 55.81% (95% CI:39.88–70.92) with AUC of 0.78 (95% CI: 0.70–0.85).

2.4. Associations between 16 criteria and hierarchical clustering analysis

Cohen's *kappa coefficients* [7] *as similarity measures between 16 clinical criteria showed that* "distended abdomen" was associated with "Diarrhea with Explosive Stool" (κ = .743), "Diarrhea with Foul Smelling Stool" (κ = .719), "Explosive Discharge of Gas and Stool on Rectal Exam" (κ = .684) and lethargy" (κ = .665) (Table A.1.). Not surprisingly, "Diarrhea with explosive stool" was highly associated with "diarrhea with foul smelling stool" (κ = .786) and "explosive discharge of gas and stool on rectal exam" (κ = .797). Additionally, the hierarchical relationship among the 16 criteria where seven clusters were determined using 1-κ as a distance and the average linkage method is presented (Fig. A.1). The following 6 criteria: "decreased peripheral per-fusion", "lethargy", "distended abdomen", "diarrhea with foul smelling stool", "diarrhea with explosive stool", and "explosive discharge of gas and stool on rectal exam" were close to each other. "History of suspected enterocolitis", "multiple air-fluid levels", and "dilated loops of bowel" formed another cluster. Similarly, "Fever", "leukocytosis", and "shift to the left" were classified as close items, and the remaining 4 criteria were placed into 4 single clusters.

2.5. Multivariable analysis of the presence or absence of suspected HAEC episode and development of HAEC risk score

On multivariable analysis "diarrhea with explosive stool" (OR:31.09; 95% CI: 5.11–346.27; p < .001), "decrease peripheral perfusion" (OR: 27.83; 95% CI: 1.13–4393.16; p = .042), "lethargy" (OR: 27.60; 95% CI: 1.71–4281.63; p = .016), and "dilated loops of bowel" (OR: 14.59; 95% CI: 2.55–102.61; p = .003) were significant independent predictors associated with suspected HAEC episodes (Table 4). These four variables were then used to create a

new HAEC risk score. The parameter estimates presented in Table 4 were used to assign points for each level of the variables. As a result, the presence of "diarrhea with explosive stool", "decrease peripheral perfusion", and "lethargy" each carried a point of 5 while the presence of "dilated loops of bowel" carried a point of 4 (Table 4). Table 5 also showed the distribution of patients across the possible risk scores ranged from 0 to 19, and the crude HAEC episode incidence rates for each score. Of the 116 patients,65.5% had score of 0, the remaining 34.5% had scores in the range of 4 to 15, and none had a score of 19. Crude HAEC episode incidence rates were 7.89% for patients with the score of 0, 60% for those with the score of 4, 85.71% for those with score of 5, and 100% for those with scores of 9 or greater. On univariate analysis, higher level of the new HAEC risk score derived from the four criteria was associated with an increased risk of suspected HAEC episode with odds ratio of 2.26 (95% CI:1.63–3.14; p < .001; Table 6), which indicated roughly a doubling of risk of suspected HAEC episode on average as every 1-point increase in the new risk score.

To explore potential diagnostic threshold values for the new HAEC risk score, possible cut-off values for the risk score were examined along with sensitivity and specificity on predicting presence or absence of suspected HAEC episodes. Table 5 presented the estimates of sensitivity and specificity at different threshold values for the risk score. HAEC risk score cut-off value of 4 maximized the sum of sensitivity of86.05% (95% CI: 72.07–94.70) and specificity of 95.89% (95% CI:88.46–99.14) with AUC of 0.910 (95% CI: 0.853–0.967) (Table 5 and Fig. 1), while an HAEC risk score of 9 maximized specificity of 100% (95% CI: 95.07–100) while still maximizing sensitivity of 65.12% (95% CI: 49.97–78.99) with AUC of 0.826 (95% CI: 0.754–0.898). On univariate analysis with the risk score cut by 4 (or <), patients with any of the four criteria were more likely to have a suspected HAEC episode compared to those who did not experience any of the criteria (OR: 143.88; 95% CI: 34.02–608.49; p < .001) (Table 6).

2.6. Performance and internal validation of the HAEC risk score

The models with the new HAEC risk score as a continuous measure (AUC: 0.925; 95% CI: 0.865–0.974) and with a cut-off value of 4 maximizing the sum of sensitivity and specificity (AUC: 0.910; 95% CI:0.853–0.967) (Fig. 1 and Table 5) performed well in predicting the presence or absence of suspected HAEC episodes as the model with the established HAEC score as a continuous measure (AUC: 0.922; 95% CI:0.862–0.970) and with a cut-off value of 2 (AUC: 0.910; 95% CI:0.852–0.967) and 4 (AUC: 0.912; 95% CI: 0.854–0.969) (Fig. 1 and Table 3). However, when the established HAEC score was dichotomized by a cut-off value of 10 as proposed in Pastor et al. [4], AUC decreased to0.709 (95% CI: 0.635–0.784). Internal validation by bootstrapping method showed that after correcting a possible optimism, the new HAEC risk score as a continuous measure (optimism-corrected AUC: 0.925; 95% CI:0.871–0.979) and at a cut-off value of 4 (optimism-corrected AUC:0.910; 95% CI: 0.854–0.967) remained predicting the outcome as well as the established HAEC score on a continuous scale (optimism-corrected AUC: 0.923; 95% CI: 0.868–0.978) with cut-off values of 2 (optimism-corrected AUC: 0.910; 95% CI: 0.854–0.967) or 4 (optimism-corrected AUC: 0.913; 95% CI: 0.856–0.970). Further, the HAEC risk score was significantly better at predicting the outcome compared to the established HAEC score at the cut-off value of 10 (optimism-corrected AUC: 0.708; 95% CI: 0.634–0.781).

Calibration plots for investigated models with smooth functions of observed incidences of suspected HAEC episode versus predicted probabilities of suspected HAEC episodes were created using a loess method and we found very good calibration for the investigated models predicting suspected HAEC episode (Fig. A.2).

3. Discussion

This is the first study to evaluate the HAEC score criteria in a large multicenter cohort of HSCR patients. We found that the HAEC score derived from 16 clinical criteria with a cutoff of 4 maximized both sensitivity and specificity to detect HAEC episodes, while the HAEC score of 10 proposed by Pastor et al. [4], significantly reduced sensitivity to detect HAEC episodes, while maximizing specificity. In other words, a cutoff score of 10 appears to be too restrictive, and would exclude more than half of patients with suspected HAEC in our cohort. This finding is not surprising given that the preliminary validation of the HAEC score was limited to 10 clinical case scenarios given to the panel of 27 experts, three of whom (P.F., T.W., and D.T.) are authors on the current study [4]. One potential explanation may be that the case scenarios included a more robust set of items within the 16 score criteria than was typically found in our patient cohort.

Further, our study identified high levels of similarity and clustering of criteria, most notably in "decreased peripheral perfusion", "lethargy", "distended abdomen", "diarrhea with foul smelling stool", "diarrhea with explosive stool", and "explosive discharge of gas and stool on rectal exam" (Table A.1 and Fig. A.1). When each element of a group of criteria is highly associated, only one of them can be used for predicting HAEC without loss of prediction power. After all, a desirable characteristic of a set of predictors is to be pairwise independent.

Our multivariable analyses identified that the four criteria: "diarrhea with explosive stool", "decreased peripheral perfusion", "lethargy", and "dilated loops of bowel" were most closely associated with HAEC episodes (Table 4). Perhaps not surprisingly, the identi-fied criteria overlap with the 5 most frequent presenting symptoms of HAEC reported in the seminal paper by Elhalaby et al.: "abdominal distension", "explosive diarrhea", "vomiting", "fever" and "lethargy"[18] further supporting the importance of this subset of criteria in making the diagnosis of HAEC. Conversely, the lack of association of the radiologic findings of "sawtooth" appearance of mucosa and pneumatosis intestinalis is likely the result of there being a low incidence in this cohort. We found this somewhat surprising, given that when pneumatosis intestinalis or "sawtooth" appearance is present on imaging in a HSCR patient, this is typically severe HAEC. One possible explanation for these findings may be that our cohort had fewer severe cases than other cohorts [18].

Our findings demonstrated that reduction of the HAEC score cut off from 10 to 4, required fewer clinical criteria that lead to a doubling in HAEC diagnoses, increasing sensitivity to 83.7% from 41.9%. This is especially important given that this significant increase in HAEC diagnosis rate would have a major impact on how HAEC is reported in future studies. One could argue that lowering the cut-off score to 2, to maximize the sensitivity (86.1% with a

negligible decrease in specificity from 98.6% to 95.9%) would be an even more conservative approach in that it would capture patients with mild or subtle signs and symptoms of HAEC.

Limitations of the study include the retrospective design with reliance on medical record; while mitigated by use of parental interviews regarding HAEC symptoms, treatment and score criteria, these too are subject to recall bias of the interviewees. Another potential limitation is variation in radiologic interpretation and management of HAEC between participating centers. Although state-of-the-art statistical methods were used to construct and compare predictive scores, these methods were applied on a collection of data where each patient has been included by nonuniform selection process. Further, although bootstrapping is recommended for internal validation because it gives reasonably valid estimates of the expected optimism in predictive performance provided that any selection of predictors is taken into account [19], it still relies only in the present study sample. External validation with a prospective cohort with well-defined inclusion/exclusion criteria would be the next step to ascertain the clinical usefulness of such a predictive score.

4. Conclusions

This study provides patient-based validation of the HAEC scoring system, which revealed a markedly different cutoff score from the original study using case scenarios. Moving forward, our findings recommend that when the HAEC scoring system is employed, a cutoff score of 4 should be used to maximize sensitivity and specificity instead of a cutoff score of 10, further optimizing the clinical criteria to diagnose HAEC. Alternatively, our novel HAEC Risk Score employing 4 of the 16 criteria could be applied, with the added benefit of requiring less clinical data with similar performance characteristics to the HAEC score with a cutoff of 4.

Acknowledgements

Appendix A

Table A.1

kappa Coefficients [7] between 16 criteria as similarity measures.

	Distended abdomen	Diarrhea with explosive stool	Diarrhea with foul smelling stool	Lethargy	Explosive discharge of gas and stool on rectal exam	Fever	Dilated loops of bowel on AXR	Leukocytosis	Decreased peripheral perfusion	Multiple air fluid levels	Previous history of suspected enterocolitis	Left shift on Complete blood count	Diarrhea With bloody stool	Cutoff sign In rectosigmoid region	Sawtooth appearance with irregular mucosal lining
Diarrhea with explosive stool	*0.743*														
Diarrhea with foul smelling stool	*0.719*	*0.786*													
Lethargy	*0.665*	*0.574*	*0.542*												
Explosive discharge of gas and stool on rectal exam	*0.684*	*0.797*	*0.665*	*0.476*											
Fever	*0.594*	*0.441*	*0.458*	*0.531*	*0.383*										
Dilated loops of bowel on AXR	*0.44*	*0.303*	*0.372*	*0.321*	*0.397*	*0.337*									
Leukocytosis	*0.448*	*0.469*	*0.372*	*0.438*	*0.393*	*0.521*	*0.33*								
Decreased peripheral perfusion	*0.438*	*0.395*	*0.411*	*0.481*	*0.435*	*0.368*	–0.15	0.107							
Multiple air fluid levels	*0.296*	*0.216*	*0.227*	*0.208*	*0.291*	0.149	*0.598*	*0.331*	–0.012						
Previous history of suspected enterocolitis	*0.264*	*0.296*	*0.308*	*0.297*	*0.239*	*0.311*	*0.463*	*0.354*	–0.101	*0.484*					
Right shift on complete blood count	*0.315*	*0.234*	*0.308*	*0.297*	*0.239*	*0.529*	*0.298*	*0.539*	–0.001	0.026	*0.398*				
Diarrhea with bloody stool	*0.111*	0.086	0.091	*0.113*	*0.12*	*0.204*	*0.153*	*0.187*	0.087	0.119	0.133	0.133			
Cutoff sign in rectosigmoid region	*0.075*	0.036	0.038	*0.132*	0.053	*0.139*	*0.174*	–0.031	–0.031	–0.03	–0.029	*0.158*	–0.021		
Sawtooth appearance with irregular mucosal lining	0.038	0.053	–0.017	–0.017	*0.072*	–0.017	*0.09*	*0.111*	–0.016	*0.169*	–0.016	–0.016	–0.013	–0.012	
Pneumatosis	0.038	0.053	0.056	*0.068*	*0.072*	*0.072*	*0.09*	–0.016	–0.016	*0.169*	*0.187*	–0.016	–0.013	–0.012	–0.009

Note: Bold-italic values were significant at 0.05 significance level. Bold and underlined values denote moderate to strong associations.
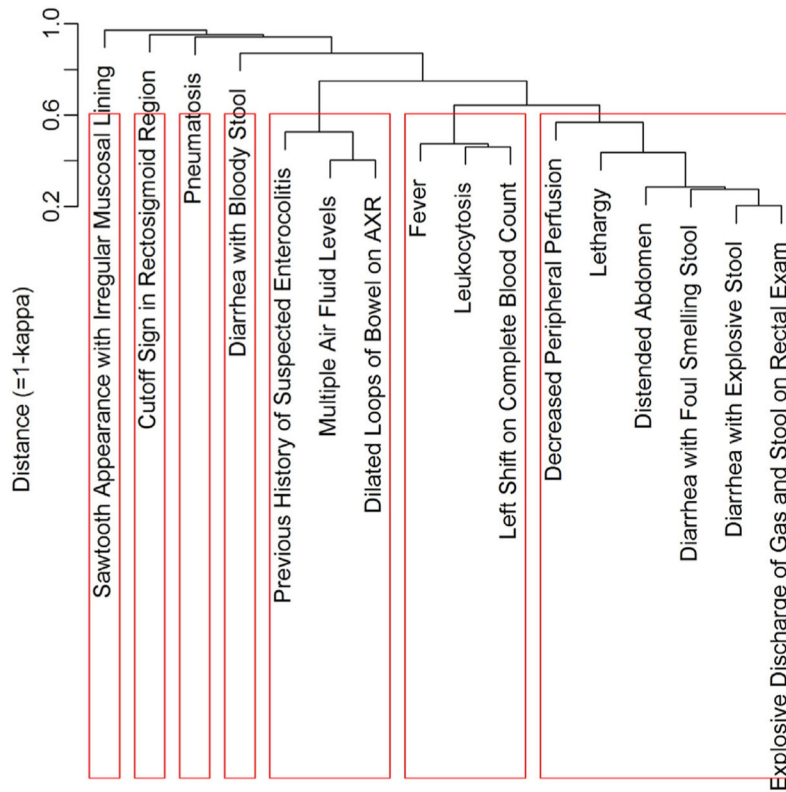
Fig. A.1.

Dendrogram showing hierarchical relationship among 16 criteria with 1-kappa coefficient as a distance and average linkage method. Red boxes represent clusters where the number of clusters was chosen with the largest average silhouette width criterion [9].
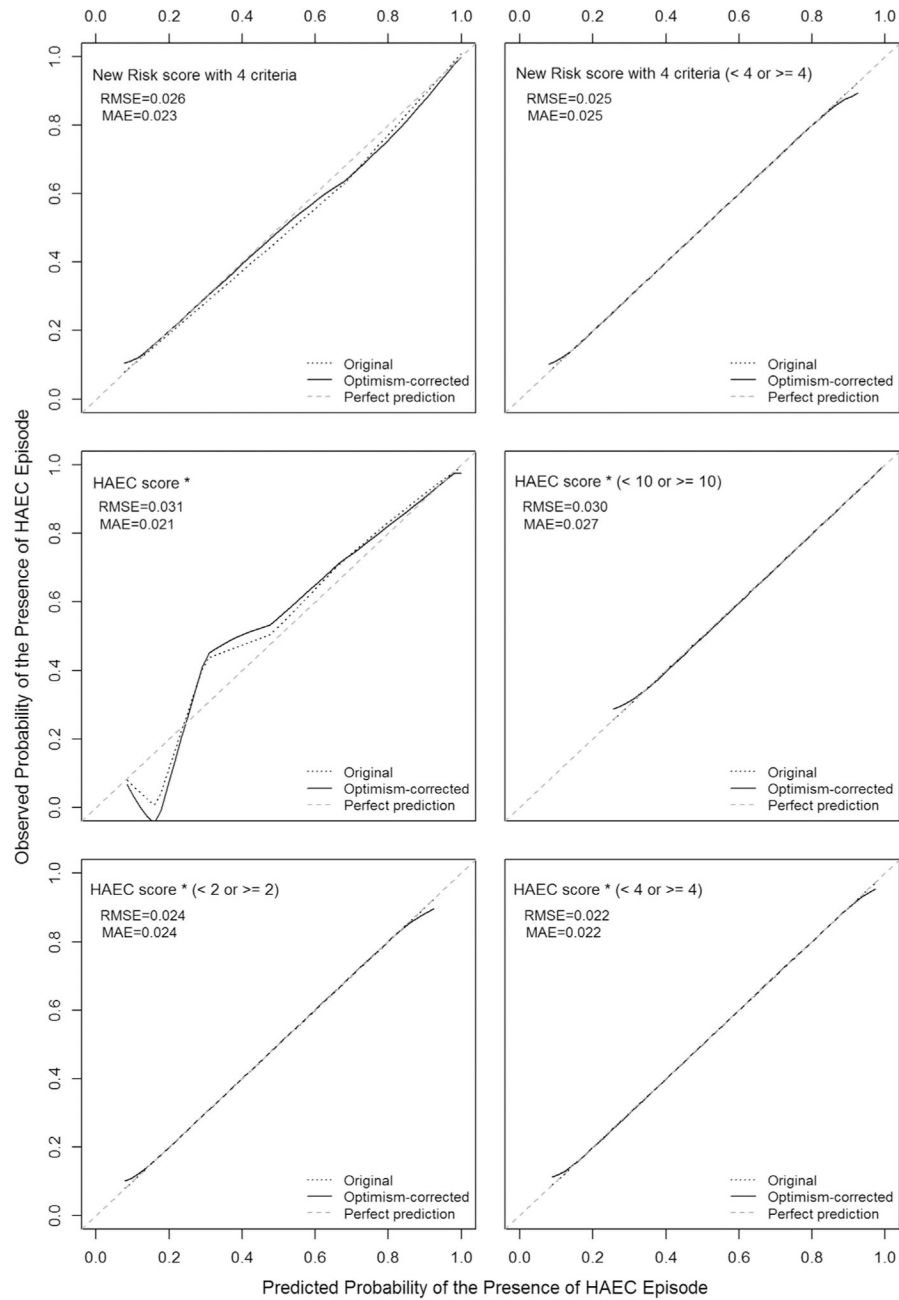

Fig. A.2.

Calibration plots with and without optimism-correction for investigated models with 45-degree line of perfect prediction. Note: *HAEC score described in Pastor et al. [4].

Abbreviations:

HSCR	Hirschsprung disease (as described in OMIM)
HAEC	Hirschsprung-associated enterocolitis
HCRG	HAEC Collaborative Research Group
IQR	inter-quartile range
VIF	variance inflation factor
ROC	receiver operating characteristic
AUC	area under the ROC curve
RMSE	root mean squared error
MAE	mean absolute error

References

[1]. Frykman PK, Short SS. Hirschsprung-associated enterocolitis: prevention and therapy. Semin Pediatr Surg 2012;21:328–35. [PubMed: 22985838]

[2]. Austin KM. The pathogenesis of Hirschsprung's disease-associated enterocolitis. Semin Pediatr Surg 2012;21:319–27. [PubMed: 22985837]

[3]. Gosain A. Established and emerging concepts in Hirschsprung's-associated enterocolitis. Pediatr Surg Int 2016;32:313–20. [PubMed: 26783087]

[4]. Pastor AC, Osman F, Teitelbaum DH, et al. Development of a standardized definition for Hirschsprung's-associated enterocolitis: a Delphi analysis. J Pediatr Surg 2009;44:251–6. [PubMed: 19159752]

[5]. Frykman PK, Nordenskjold A, Kawaguchi A, et al. Characterization of bacterial and fungal microbiome in children with hirschsprung disease with and without a history of enterocolitis: a multicenter study. PLoS One 2015;10:e0124172. [PubMed: 25909773]

[6]. Demehri FR, Frykman PK, Cheng Z, et al. Altered fecal short chain fatty acid composition in children with a history of Hirschsprung-associated enterocolitis. J Pediatr Surg 2016;51:81–6. [PubMed: 26561246]

[7]. Cohen J A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20: 37–46.

[8]. Everitt BL, Leese M, Stahl D. Cluster Analysis. John Wiley & Sons; 2011.

[9]. Rousseeuw P Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Comput Appl Math 1987;20:53–65.

[10]. Firth D Bias reduction of maximum-likelihood-estimates. Biometrika 1993;80: 27–38.

[11]. Venzon DM. SH: A method for computing profile-likelihood based confidence intervals. Appl Stat 1988;37:87–94.

[12]. Collett D Modeling survival data in medical research. London, UK: CRC; 2003.

[13]. Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32–5. [PubMed: 15405679]

[14]. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21: 128–38. [PubMed: 20010215]

[15]. Harrell FE, Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361–87. [PubMed: 8668867]

[16]. Harrell F. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis; 2001.

[17]. Steyerberg EW, Harrell FE, Jr, Borsboom GJ, et al. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 2001;54:774–81. [PubMed: 11470385]

[18]. Elhalaby EA, Coran AG, Blane CE, et al. Enterocolitis associated with Hirschsprung's disease: a clinical-radiological characterization based on 168 patients. J Pediatr Surg 1995;30:76–83. [PubMed: 7722836]

[19]. Steyerberg EW, Bleeker SE, Moll HA, et al. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. J Clin Epidemiol 2003;56:441–7. [PubMed: 12812818]
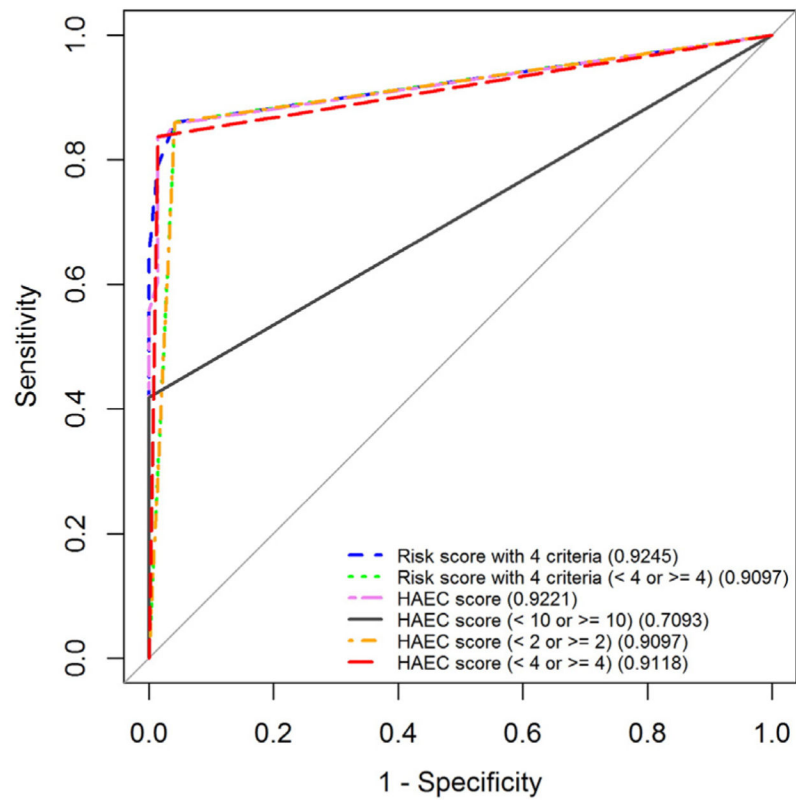
Fig. 1.
Receiver operating characteristic (ROC) curves along with the areas under the ROC curve (AUCs) for the HAEC score described by Pastor et al. [4] as a continuous measure, cut by 10 (established value), 2 (maximizing sensitivity) and 4 (maximizing sum of sensitivity and specificity) on predicting the presence versus absence of HAEC episodes; and the Risk score derived from the model with 4 criteria as a continuous measure and cut by 4. Note: The 45° line denotes a reference.

Table 1

Patient characteristics.

Variable	Total N = 116
Age at evaluation (years), median (IQR)	6 (3–8)
Gender	
Female	17 (14.66)
Male	99 (85.34)
Trisomy21	
Yes	7 (6.03)
No	109 (93.97)
Extent of aganglionosis at time of pullthrough operation	
Rectosigmoid	85 (85.0)
Descending	4 (4.0)
Transverse	4 (4.0)
Ascending	0 (0)
Ileum	7 (7.0)
Missing	16
Episodes of suspected HAEC	
Yes	43 (37.07)
No	73 (62.93)
Estimated number of episodes of suspected HAEC	
0	73 (62.93)
1–4	38 (32.76)
5–9	3 (2.59)
10	2 (1.72)
First HAEC episode	
Pre-pullthrough or diversion	10(9.8)
Post-pullthrough	92 (90.2)
Missing	14

Data are presented as number of patients (%) or median (IQR, interquartile range).

Table 2

Univariate analysis of 16 criteria associated with presence versus absence of HAEC episodes.

16 Criteria	Number of patients (%)	Odds ratio (95% CI)	*P*- value
Distended abdomen			
Yes	36 (31.03)	134.07 (27.46–654.53)	<.001
No	80 (68.97)	1 (Reference)	
Diarrhea with explosive stool			
Yes	28 (24.14)	121.50 (15.36–961.06)	<.001
No	88 (75.86)	1 (Reference)	
Diarrhea with foul smelling stool			
Yes	27 (23.28)	110.12 (13.95–869.21)	<.001
No	89 (76.72)	1 (Reference)	
Lethargy [*]			
Yes	23 (19.83)	168.51 (21.71–21,732.18)	<.001
No	93 (80.17)	1 (Reference)	
Explosive discharge of gas and stool on rectal exam			
Yes	22 (18.97)	68.73 (8.74–540.38)	<.001
No	94 (81.03)	1 (Reference)	
Fever			
Yes	22 (18.97)	68.73 (8.74–540.38)	<.001
No	94 (81.03)	1 (Reference)	
Dilated loops of bowel on AXR			
Yes	18 (15.52)	21.04 (4.53–97.67)	<.001
No	98 (84.48)	1 (Reference)	
Leukocytosis [*]			
Yes	15 (12.93)	79.95 (10.16–10,330.11)	<.001
No	101 (87.07)	1 (Reference)	
Decreased peripheral perfusion [*]			
Yes	13 (11.21)	65.07 (8.18–8421.86)	<.001
No	103 (88.79)	1 (Reference)	
Multiple air fluid levels			
Yes	10 (8.62)	19.06 (2.32–156.55)	0.006
No	106(91.38)	1 (Reference)	
Previous history of suspected enterocolitis [*]			

16 Criteria	Number of patients (%)	Odds ratio (95% CI)	P- value
Yes	9 (7.76)	40.48 (4.89–5273.84)	<.001
No	107 (92.24)	1 (Reference;	
Left shift on complete blood count *			
Yes	9 (7.76)	40.48 (4.89–5273.84)	<.001
No	107 (92.24)	1 (Reference)	
Diarrhea with bloody stool *			
Yes	3 (2.59)	12.70 (1.19–1724.37)	0.034
No	113 (97.41)	1 (Reference)	
Cutoff sign in rectosigmoid region *			
Yes	2 (1.72)	8.85 (0.70–1232.85)	0.096
No	114(98.28)	1 (Reference)	
Sawtooth appearance with irregular mucosal lining *			
Yes	1 (0.86)	5.19 (0.27–764.08)	0.274
No	115 (99.14)	1 (Reference)	
Pneumatosis intestinalis *			
Yes	1 (0.86)	5.19 (0.27–764.08)	0.274
No	115 (99.14)	1 (Reference)	

*
Firth's penalized maximum likelihood estimation along with profile-likelihood con-fidence interval and p-value from penalized likelihood ratio test [10,11] was reported to reduce bias in the parameter estimates since quasi-complete separation of data points was detected and the maximum likelihood estimate may not exist.

Table 3

Estimates of sensitivity and specificity at different threshold values for HAEC score described by Pastor et al. [*] on predicting the presence versus absence of HAEC episodes.

Cut-off value (versus <)	Sensitivity (exact 95% CI) (%)	Specificity (exact 95% CI) (%)	Area under the ROC curve (AUC) (95% CI)
1	86.05 (72.07–94.70)	94.52 (86.56–98.49)	0.903 (0.844–0.962)
2	86.05 (72.07–94.70)	95.89 (88.46–99.14)	0.910 (0.852–0.967)
3	83.72 (69.30–93.19)	97.26 (90.45–99.67)	0.905 (0.846–0.964)
4	83.72 (69.30–93.19)	98.63 (92.60–99.97)	0.912 (0.854–0.969)
5	81.40 (66.60–91.60)	98.63 (92.60–99.97)	0.900 (0.840–0.961)
6	74.42 (58.83–86.48)	98.63 (92.60–99.97)	0.865 (0.798–0.933)
7	65.12 (49.07–78.99)	98.63 (92.60–99.97)	0.819 (0.745–0.892)
8	60.47 (44.41–75.02)	98.63 (92.60–99.97)	0.796 (0.720–0.871)
9	55.81 (39.88–70.92)	100.00 (95.07–100)	0.779 (0.704–0.854)
10 (established cut-off value)	41.86 (27.01–57.87)	100.00 (95.07–100)	0.709 (0.635–0.784)
11	27.91 (15.33–43.67)	100.00 (95.07–100)	0.640 (0.572–0.707)
12	18.61 (8.39–33.40)	100.00 (95.07–100)	0.593 (0.534–0.652)
13	13.95 (5.30–27.93)	100.00 (95.07–100)	0.570 (0.517–0.622)
14	6.98 (1.46–19.06)	100.00 (95.07–100)	0.535 (0.496–0.573)

[*] HAEC score described by Pastor et al. [4] = 2 × distended abdomen +2 × diarrhea with explosive stool +2 × diarrhea with foul smelling stool + lethargy +2 × explosive discharge of gas and stool on rectal exam + fever + dilated loops of bowel on AXR + leukocytosis + decreased peripheral perfusion + multiple air fluid levels + previous history of suspected enterocolitis + left shift on complete blood count + diarrhea with bloody stool + cutoff sign in rectosigmoid region + sawtooth appearance with irregular.

Mucosal Lining + Pneumatosis intestinalis.

Table 4

Multivariable analysis of presence versus absence of HAEC episodes.

Variable	Parameter estimate	Assigned point	Odds ratio (95% CI)	P-value
Diarrhea with explosive stool				
Yes	3.4367	5	31.09 (5.11–346.27)	<.001
No	Reference	0		
Decreased peripheral perfusion				
Yes	3.3261	5	27.83 (1.13–4393.16)	0.042
No	Reference	0		
Lethargy				
Yes	3.3177	5	27.60 (1.71–4281.63)	0.016
No	Reference	0		
Dilated loops of bowel on AXR				
Yes	2.6803	4	14.59 (2.55–102.61)	0.003
No	Reference	0		

One hundred sixteen observations were used in the multivariable model. Fourteen criteria not shown were dropped out of the model. Firth's penalized maximum likelihood estimation along with profile-likelihood confidence interval and p-value from penalized likelihood ratio test [10,11] was reported since quasi-complete separation of data points was detected and the maximum likelihood estimate may not exist.

Table 5

Risk score derived from the model with 4 criteria predicting the presence versus absence of HAEC episodes and estimates of sensitivity and specificity at different threshold values for the risk score.

Risk score [*]	Number of patients (%)	Crude HAEC episode incidence rate (%)
19	0	Not applicable
15	8 (6.9)	100
14	6 (5.17)	100
10	7 (6.03)	100
9	7 (6.03)	100
5	7 (6.03)	85.71
4	5 (4.31)	60
0	76 (65.52)	7.89

Cut-off value (≥ versus <)	Sensitivity (exact 95% CI) (%)	Specificity (exact 95% CI) (%)	Area under the ROC curve (AUC) (95% CI)
4	86.05 (72.07–94.70)	95.89 (88.46–99.14)	0.910 (0.853–0.967)
5	79.07 (63.96–89.96)	98.63 (92.60–99.97)	0.889 (0.826–0.952)
9	55.12 (49.97–78.99)	100.00(95.07–100)	0.826 (0.754–0.898)
10	48.84 (33.31–64.54)	100.00 (95.07–100)	0.744 (0.669–0.820)
14	32.56 (19.08–48.54)	100.00 (95.07–100)	0.663 (0.592–0.734)
15	18.61 (8.39–33.40)	100.00 (95.07–100)	0.593 (0.534–0.652)

[*] Risk score = 5 × diarrhea with explosive stool (1 if presence; 0 if absence) + 5 × -decreased peripheral perfusion (1 if presence; 0 if absence) + 5 × lethargy (1 if presence; 0 if absence) + 4 × dilated loops of bowel on AXR (1 if presence; 0 if absence).

Table 6

Effect of the risk score derived from the model with 4 criteria on predicting the presence versus absence of HAEC episodes.

Risk score*	Number of patients (%)	Odds ratio (95% CI)	P-value
Risk score	116	2.26 (1.63–3.14)	<.001
Risk score >= 4 (i.e., those who have any of the 4 criteria)	40 (34.48)	143.88 (34.02–608.49)	<.001
Risk score < 4 (i.e., those who have none of the 4 criteria)	76 (65.52)	1 (Reference)	

*
Risk score = 5 × diarrhea with explosive stool (1 if presence; 0 if absence) + 5 × decreased peripheral perfusion (1 if presence; 0 if absence) + 5 × lethargy (1 if presence; 0 if absence) + 4 × dilated loops of bowel on AXR (1 if presence; 0 if absence).