Vaccine 30S (2012) A159–A166

Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Comparison of two clinical severity scoring systems in two multi-center, developing country rotavirus vaccine trials in Africa and Asia☆,☆☆

Kristen D.C. Lewis[a,*], Michael J. Dallas[b], John C. Victor[a], Max Ciarlet[b,1], T. Christopher Mast[b], Ming Ji[c], George Armah[d], K. Zaman[e], Aimee Ferraro[f], Kathleen M. Neuzil[a]

[a] PATH, Seattle, WA, USA
[b] Merck Research Laboratories, North Wales, PA, USA
[c] San Diego State University, San Diego, CA, USA
[d] Noguchi Memorial Institute for Medical Research, University of Ghana, Legon, Ghana
[e] International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh
[f] Walden University, Minneapolis, MN, USA

ARTICLE INFO

Article history:
Received 12 May 2011
Received in revised form 22 July 2011
Accepted 27 July 2011

Keywords:
Rotavirus
Gastroenteritis
Severity score
Vesikari scoring system
Clark scoring system
Vaccine efficacy

ABSTRACT

Background: Clinical severity scoring systems are used in rotavirus vaccine efficacy and effectiveness studies to define the primary endpoint, severe rotavirus gastroenteritis (RVGE). Understanding how scoring systems perform in diverse settings is critical for proper design and interpretation. This investigation aims to understand how the Vesikari scoring system (VSS) and Clark scoring system (CSS) categorize severe disease among children under 2 years of age using data from two Phase III efficacy trials conducted in five developing countries in Africa and Asia.
Methods: Signs and symptoms were collected on trial participants who presented to a medical facility with study-defined gastroenteritis. Severity scores were calculated using pre-established VSS and CSS criteria and compared to identify differences in the proportions of severe RVGE within regions and sites, and by gender and age.
Results: In Africa and Asia, 40.6% and 56.0% of rotavirus-positive episodes were severe according to the VSS, while 9.5% and 6.3% of episodes were severe according to the CSS (Fisher's Exact, $p \leq 0.001$). Using the mean scores in these trials (VSS: ≥10 Africa, ≥11 Asia; CSS: Africa and Asia ≥10) as the severity thresholds, agreement between scoring system severity classifications improved substantially within each region (Africa: kappa = 0.67; Asia: kappa = 0.78) as compared to the original severity classification (Africa: kappa = 0.27; Asia: kappa = 0.10). Using the mean score, 17.1% and 9.5% of severe VSS cases in Africa and Asia, respectively, were classified as not severe according to the CSS and 14.7% and 9.5% of severe CSS cases in Africa and Asia were classified as not severe according to the VSS.
Conclusion: The two scoring systems performed differently among developing country populations in Africa and Asia, with the VSS classifying more cases as severe in both regions. One accurate and reliable scoring system should be developed and implemented for all trials so that results may be more comparable.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The introduction of rotavirus vaccines, provided in infancy, should have a major impact on rotavirus gastroenteritis (RVGE) among developing country populations in Africa and Asia [1]; 5% of all-cause under-5 child mortality and up to 36% of under-5 gastroenteritis hospitalizations across the globe could be prevented by using rotavirus vaccines [2–5]. Recently published results from three developing country trials testing the efficacy of the two World Health Organization (WHO) pre-qualified rotavirus vaccines (human rotavirus vaccine [HRV] and pentavalent rotavirus vaccine [PRV]) demonstrated that, although efficacy estimates were lower than for developed countries, the absolute reduction in RVGE incidence due to these vaccines in these populations was

☆ *Disclaimer*: The views expressed in this publication are those of the authors alone and do not necessarily represent the decisions, policy, or views of Merck Research Laboratories, San Diego State University, Noguchi Memorial Institute for Medical Research, International Centre for Diarrhoeal Disease Research, Bangladesh, Walden University, Novartis Vaccines & Diagnostics, or PATH, Seattle.

☆☆ The Program for Appropriate Technology in Health (PATH) – GAVI Alliance grant and Merck were the study sponsors of the original trial generating the dataset that this secondary analysis is based on. The funding source had no involvement in the research, writing, or the decision to submit the paper for publication.

* Corresponding author at: PATH, P.O. Box 900922, Seattle, WA 98109, USA. Tel.: +1 206 285 3500; fax: +1 206 285 6619.
E-mail address: klewis@path.org (K.D.C. Lewis).
[1] Present address: Novartis Vaccines & Diagnostics, Cambridge, MA, USA.

0264-410X/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.vaccine.2011.07.126

Table 1
Vesikari and Clark clinical severity scoring systems.

Symptoms	Vesikari et al. points				Clark et al. points		
	0	1	2	3	1	2	3
Diarrhea							
No. stools/day	N/A	1–3	4–5	≥6	2–4	5–7	>7
Duration (days)	N/A	1–4	5	≥6	1–4	5–7	>7
Vomiting							
No. emesis/day	0	1	2–4	≥5	1–3	4–6	>6
Duration (days)	N/A	1	2	≥3	2	3–5	>5
Temperature							
Rectal temperature (°C)	<37	37.1–38.4	38.5–38.9	≥39	38.1–38.2	38.3–38.7	≥38.8
Temperature duration (days)	N/A	N/A	N/A	N/A	1–2	3–4	≥5
Behavior							
Dehydration	None	N/A	1–5%	≥6%	N/A	N/A	N/A
Behavioral symptoms	N/A	N/A	N/A	N/A	Irritable/ less playful	Lethargy/ listless	Seizures
Duration (days)	N/A	N/A	N/A	N/A	1–2	3–4	≥5
Treatment	None	Rehydration	Hospitalization	N/A	N/A	N/A	N/A
Severity rating scales	Mild <7	Moderate 7–10	Severe ≥11	Max score 20	Mild 0–8	Moderate 9–16	Severe ≥17	Max score 24

Note. Table adapted from rotavirus clinical trials utilizing the Clark and Vesikari clinical scoring systems [17,18,20].

substantial [6–8]. While not designed for such an analysis, the results of these clinical trials suggested a trend towards increasing efficacy with increasing episode severity [6–9]. The results of these studies informed the WHO recommendation to include rotavirus vaccines in the national immunization programs of all countries [10].

Phase II and Phase III trials are currently underway or being planned to evaluate new rotavirus vaccine candidates [11]. Moreover, following vaccine introduction into countries, post marketing surveillance studies can help monitor the effectiveness of routine vaccine use [5,11–13]. For developing countries, the main outcome of public health interest will be severe RVGE, in addition to safety [6–8,14,15]. Thus, for optimal study design and interpretation, as well as in potential future studies examining the benefits of therapeutic interventions like probiotics [16], it is important to improve understanding of how rotavirus clinical severity scoring systems used for measuring RVGE severity compare and perform in diverse settings.

Two main RVGE clinical severity scoring systems, the Clark scoring system (CSS) and the Vesikari scoring system (VSS), first developed for use in developed country settings [17–20], have been used in developing country rotavirus vaccine efficacy and effectiveness studies [6–8,14,15]. The scoring systems have different point spreads and contain different categorical and continuous signs and symptoms (i.e. items) for measuring severe RVGE (Table 1). The CSS includes a maximum of 24 points, with scores between 17 and 24 classified as severe (33.3% of the point spread). In contrast, the VSS includes a maximum of 20 points, with scores between 11 and 20 classified as severe (50.0% of the point spread).

Both scoring systems assess the magnitude and duration of vomiting and diarrhea and the maximum temperature. The CSS also assesses the magnitude and duration of behavioral symptoms and the duration of a temperature greater than 38.0 °C, while the VSS also assesses dehydration by measuring acute weight loss, although it is now common for studies to assess dehydration using WHO Integrated Management of Childhood Illness (IMCI) dehydration criteria [21], and treatment (i.e. rehydration or hospitalization). The categorical items in both scoring systems are assigned point scores ranging from 1 to 3. Similarly, the continuous variables are classified into categories that are also assigned point scores, with an increasing point score indicating increasing severity of that item (Table 1). With the exception of dehydration and treatment in the VSS, which have two possible scores, all scoring system items have three possible scores (i.e. 1, 2, or 3). The use of different point assignments and thresholds for assigning categorical scores (e.g. VSS temperature ≥37.1 °C, CSS ≥38.1 °C), as well as overall scales (i.e. 20-point VSS, 24-point CSS scales) indicate that the two scoring systems do not generate identical individual scores [17–20,22]. Additional information regarding the development and use of these scoring systems is provided by Ruuska and Vesikari [20] and Clark et al. [17].

Recently, Givon-Lavi et al. [23] highlighted the differences between the CSS and the VSS when used in an observational prospective hospital-based surveillance study among children less than 5 years of age in southern Israel, concluding that the two scoring systems were not comparable in that population, and that efficacies against severe RVGE cannot be directly be compared between trials using different scoring systems, especially with dissimilar study designs and locations. However, a comparison using clinical trial data has not been previously described. The severity of RVGE was measured using both the modified VSS and CSS using data collected in the recent large Phase III clinical efficacy trials of PRV among developing country populations less than 2 years of age in Africa and Asia [7,8]. In order to determine how the two scoring systems performed in these trials, we compared the VSS and the CSS post-hoc as used in these two trials.

2. Materials and methods

Two multi-center, double-blind, placebo-controlled, randomized trials were conducted in five developing country sites using a common protocol and data collection forms from March 29, 2007, to March 31, 2009, to assess efficacy of three doses of PRV (RotaTeq™: Merck, Whitehouse Station, NJ, USA) against severe RVGE in infants less than 2 years of age [7,8]. Participating sites were located in rural Kassena-Nankana district, Ghana; rural Karemo division, Siaya district, Nyanza province, Western Kenya; urban Bamako, Mali; rural Matlab, Bangladesh; and urban and periurban Nha Trang, Vietnam. The design and efficacy results of these trials have been previously reported [7,8].

In summary, participants were randomly assigned to receive three doses of PRV or placebo in a 1:1 ratio at approximately 6, 10 and 14 weeks of age. Following the first dose of study vaccine, participants were visited at home at least monthly by field workers through up to 24 months of age to remind parents to present to a study medical facility if their child experienced an

episode of acute gastroenteritis (AGE; defined as 3 or more looser-than-normal stools and/or forceful vomiting within a 24-h period). A common study protocol, symptom collection standard operating procedure (SOP), and data collection forms were used across all study sites. At the medical facility, signs and symptoms (i.e. those items contained within the VSS and CSS) from the start of the episode through discharge were collected by a trained study clinical staff (Table 1). Because the scoring systems require capture of signs and symptoms since the beginning of an episode, the information collected by study clinical staff was based on a combination of parental recall of symptoms before presentation and clinical staff examination and parental recall while at the medical facility. In previous trials [6,24], diary cards were provided to parents at enrollment so that they could record AGE symptoms of enrolled children if an episode occurred after vaccination. However, in these trials, parental diary cards were not utilized due concerns that limited literacy in certain trial sites would prevent accurate data collection.

In these trials, the VSS was modified in three ways. First, the score for "treatment" was modified from responses of "Hospitalization (score = 2)" and "Rehydration (score = 1)" in the original VSS to the revised "hospitalized or received IV rehydration (score = 2)" and "received oral rehydration medication (score = 1)", respectively. Secondly, dehydration was measured using the WHO IMCI dehydration criteria, rather than based on measuring acute weight loss. The guidelines include clinical signs that are used to evaluate the level of dehydration in children: appearance, sunken eyes, thirst, skin pinch and respiration. Although guidelines no longer advocate use of respiration, this parameter was included in this study since it was of historical importance in previously reported WHO assessments of dehydration. Finally, an axillary temperature was measured and this was converted to rectal during analysis.

Unlike in the previously reported efficacy analyses [7,8], this post-hoc investigation to compare the VSS and CSS utilized all rotavirus AGE episodes (mostly wild-type) with a subsequent rotavirus-positive stool sample occurring at anytime following study vaccine Dose 1, regardless of whether the episode occurred 14 days following study vaccine Dose 3 or the participant received all study vaccinations, and included all (multiple) episodes for each study participant. Scores for the VSS and CSS were calculated by applying a uniform computer program code across all episodes in the dataset.

Because the trials were originally planned and conducted as two regional trials in Africa and Asia, this analysis focused on each region separately, with sub-analyses conducted by site. Within each region the two clinical scoring systems were compared similar to what was done by Givon-Lavi et al. [23] and Ruuska and Vesikari [20]. Demographic and clinical information such as site (i.e. country), gender, hospitalization status (i.e. hospitalization or receipt of IV therapy), and age was compared between each scoring system for rotavirus and non-rotavirus gastroenteritis cases. Mean scores and proportions of participants meeting severe criteria according to each scoring system were calculated. To demonstrate the differences between each item score for the two scoring systems, the item scoring distributions for each sign/symptom commonly included in the clinical scoring systems were compared and the VSS to CSS ratio of the numbers of participant episodes with each item point score calculated. Chi-Square or, when appropriate, Fisher's Exact tests, Student's *t*-tests, or ANOVAs were used to test for statistical significance of contingency tables and continuous variables, respectively.

The scoring system severity classifications were compared between the VSS and the CSS based on the "original" and two "modified" severity classifications. The original classification is based on the mild, moderate, and severe cut points historically used for defining severity; VSS: <7 mild, 7–10 moderate, and ≥11 severe, CSS: <9 mild, 9–16 moderate, and ≥17 severe. The original classification is based on consistency with the original severity classification method used by Ruuska and Vesikari [20], where the threshold was selected as the mean score (i.e. severe ≥11), also corresponding to the median score in the scoring distribution for this study. Modified classifications were also used in this study. One modified classification used the mean VSS severity score observed among rotavirus-positive participants in these trials in Africa (≥10) and Asia (≥11) as the severity threshold and compared these to a CSS severity threshold based on the mean in each region (Africa and Asia: ≥10). A second modified classification comparison set the severity threshold at the median of the scoring distribution (VSS: ≥11/20 points; CSS ≥13/24 points). The degree of severity classification agreement between the two scoring systems for the original and modified classifications was determined using Cohen's kappa for non-unique raters with a kappa greater than 0.60 identified as at least good agreement [25]. All analyses were completed using Intercooled Stata 11.1 for Windows (Version 11.1 College Station, TX; StataCorp LP; 2011).

3. Results

In Africa and Asia, of 3814 and 906 participants, respectively, with stool specimen results and clinical data, approximately 14.7% (559/3814) and 22.8% (207/906) of AGE episodes, respectively, were rotavirus-positive; 16.3% (139/854) in Ghana, 11.6% (50/430) in Kenya, 14.6% (370/2530) in Mali, 22.0% (166/753) in Bangladesh, and 26.8% (41/153) in Vietnam. In Africa, approximately 66% (370/559) of the rotavirus-positive cases were from Mali, 25% (139/559) from Ghana, and 9% (50/559) from Kenya. In Asia, 80% (166/207) of rotavirus-positive cases were from Bangladesh and 20% (41/207) from Vietnam. Less than 5% of participants experienced more than one rotavirus-positive episode (i.e. two or three episodes).

Overall, VSS and CSS mean scores within each region and each scoring system were significantly higher for RVGE cases as compared to non-rotavirus GE cases (Africa: VSS, 10.1 vs. 7.5; CSS, 9.9 vs. 7.2; Asia: VSS, 10.9 vs. 7.8; CSS, 10.3 vs. 7.1; p-value ≤ 0.001). Proportionally more rotavirus-positive episodes were captured in Africa as compared to Asia, but, based on similar distributions between regions, participant episodes were just as likely to receive a severe score in Asia as they were in Africa for the CSS, but not the VSS (Fig. 1, Table 2). When compared within gender and age, the mean VSS and CSS for rotavirus-positive episodes did not differ statistically, while within hospitalized cases and site there was a significant difference (Table 2). The Mali site had a lower mean score for both the VSS and the CSS than the other sites. The mean score for hospitalized cases was lower for both the VSS and CSS in Asia as compared to Africa.

Among the five common items contained within both scoring systems, the VSS provided proportionally higher scores for each item in Africa and Asia as compared to the CSS, with the exception of temperature (Table 3). The VSS to CSS ratio of the number of gastroenteritis episodes with an item score of 3 was greater than 1.0 for every scoring system item, except maximum temperature, indicating that it was easier to gain a higher item score for these symptoms using the VSS. This is consistent with how the scoring system would have been expected to perform given that, in the VSS, a value of 3 is reached with a lower frequency of episodes or number of days of duration (Table 1).

The CSS and VSS did not result in uniform categorization of severe gastroenteritis among rotavirus-positive gastroenteritis episodes in either trial. Using the traditional definitions for severity, within Africa and Asia, respectively, 40.6% (227/559) and 56.0% (116/207) of rotavirus-positive episodes, respectively, were

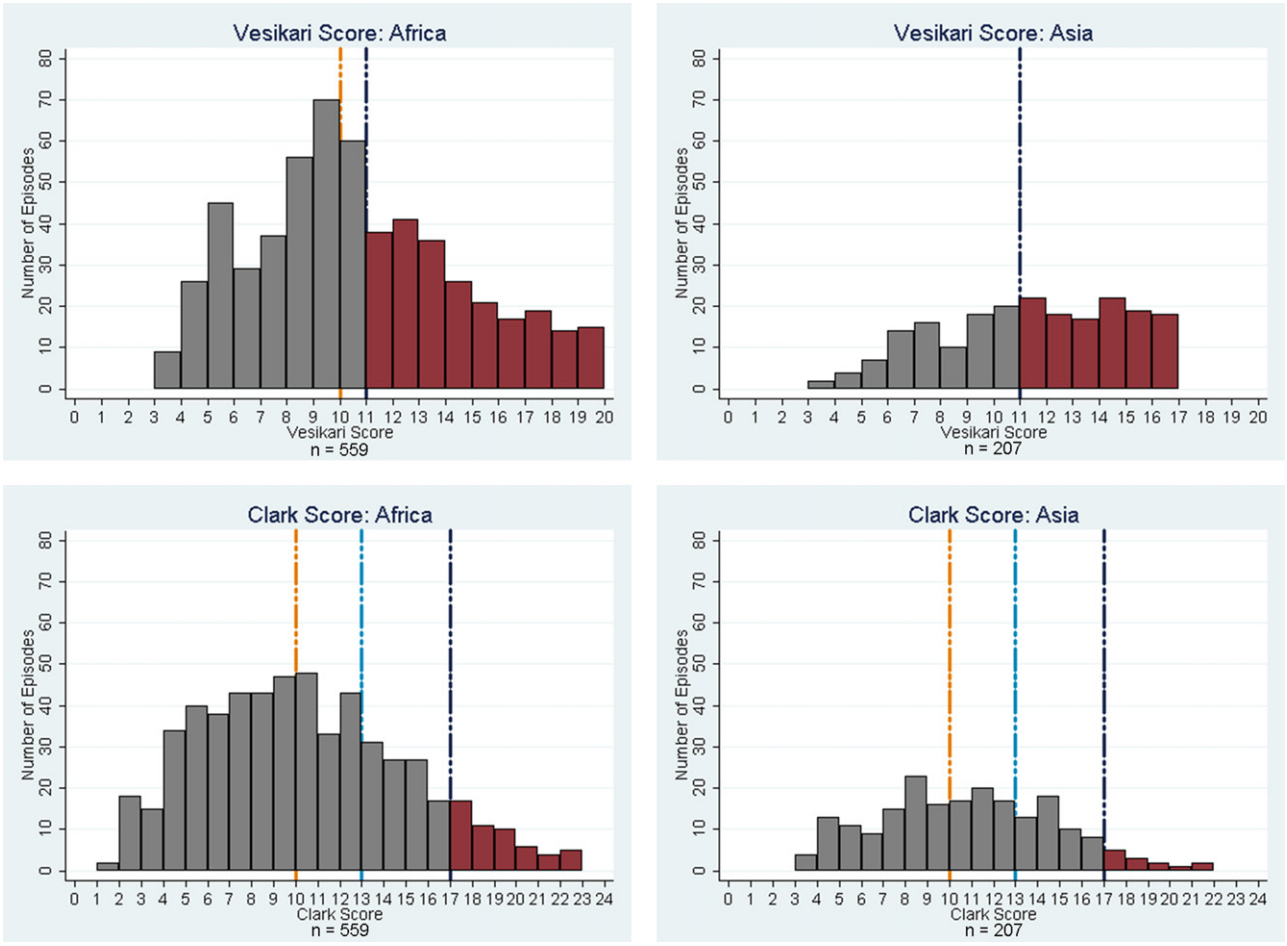

Fig. 1. Vesikari and Clark scoring system distributions among rotavirus-positive episodes by trial. Orange line = mean score in this trial (≥10 CSS and VSS Africa); light blue line = median scoring distribution for CSS (≥13 Clark); maroon bars/dark blue line = original severity scoring classification (≥11 Vesikari, ≥17 Clark). Where the scoring distribution did not change from the original classification (i.e. Vesikari Asia all cutoffs were based on a score ≥11), only a dark blue line is shown.

identified as severe according to the VSS (≥11), while only 9.5% (53/559) and 6.3% (13/207) of episodes were identified as severe by the CSS (≥17) (Fisher's Exact, $p \leq 0.001$) (Table 4). This pattern remained across sites, gender and age group.

The results in Table 5 demonstrate poor agreement in categorizing severe gastroenteritis between the two scoring systems when using the original severity classifications, but that agreement improves substantially when using modified severity classifications. When using the original scoring classification, every episode categorized as severe according to the CSS was also classified as severe according to the VSS; 76.7% (174/227) and 88.8% (103/227) of severe VSS in Africa and Asia, respectively, were identified as not severe according to the CSS. When a modified scoring classification based on the mean scores (VSS: ≥10 Africa, ≥11 Asia;

Table 2
Mean (±SD) scores among rotavirus-positive gastroenteritis episodes by region, site, gender, age, and hospitalized cases according to scoring system.

Scoring system	Mean score (±SD)				
Region	*Africa* (*n* = 559)	*Asia* (*n* = 207)			
Vesikari*	10.1 (4.0)	10.9 (3.5)			
Clark	9.9 (4.6)	10.3 (4.0)			
Site	*Bangladesh* (*n* = 166)	*Vietnam* (*n* = 41)	*Ghana* (*n* = 139)	*Kenya* (*n* = 50)	*Mali* (*n* = 370)
Vesikari*	10.8 (3.2)	11.3 (4.4)	11.7 (4.0)	10.9 (5.0)	9.4 (3.6)
Clark*	10.4 (4.1)	10.0 (3.6)	12.2 (4.1)	10.4 (6.1)	8.9 (4.2)
Gender	*Male* (*n* = 413)	*Female* (*n* = 353)			
Vesikari	10.5 (3.9)	10.2 (3.8)			
Clark	10.2 (4.5)	9.8 (4.4)			
Age (*months*)	*<6* (*n* = 76)	*6–11* (*n* = 225)	*12–17* (*n* = 343)	*18–23* (*n* = 119)	*≥24* (*n* = 3)
Vesikari	9.8 (4.5)	10.5 (3.9)	10.5 (3.8)	9.9 (3.6)	10.3 (2.9)
Clark	9.0 (5.2)	10.3 (4.4)	10.3 (4.4)	9.3 (4.2)	9.3 (2.5)
Hospitalized cases, by region	*Africa* (*n* = 151)	*Asia* (*n* = 136)			
Vesikari*	14.8 (2.8)	12.5 (2.9)			
Clark*	14.7 (3.9)	11.9 (3.7)			

* Comparison between means: $p \leq 0.01$ between the VSS or CSS between regions, sites, or hospitalized cases.

Table 3
Comparison of each item score for the Clark and Vesikari scoring systems among rotavirus-positive gastroenteritis episodes.

	Item score (*n* (%))					
	Africa (*n* = 559)			Asia (*n* = 207)		
	VSS	CSS	Ratio (VSS/CSS)	VSS	CSS	Ratio (VSS/CSS)
Looser than normal stools duration*						
0	42 (7.5)**	47 (8.4)	0.89	1 (0.5)	2 (1.0)	0.50
1	416 (74.4)	424 (75.9)	0.98	166 (80.2)	165 (79.7)	1.00
2	34 (6.1)	69 (12.3)	0.49	14 (6.8)	34 (16.4)	0.41
3	67 (12.0)	19 (3.4)	3.53	26 (12.5)	6 (2.9)	4.33
Maximum number of looser than normal stools*						
0	42 (7.5)	47 (8.4)	0.89	1 (0.5)	2 (1.0)	0.50
1	144 (25.7)	265 (47.4)	0.54	14 (6.8)	30 (14.5)	0.47
2	234 (41.9)	197 (35.3)	1.19	40 (19.3)	64 (30.9)	0.63
3	139 (24.9)	50 (8.9)	2.78	152 (73.4)	111 (53.6)	1.37
Vomiting duration*						
0	98 (17.5)	200 (35.8)	0.49	47 (22.7)	108 (52.2)	0.44
1	102 (18.3)	149 (26.6)	0.68	61 (29.5)	49 (23.7)	1.24
2	149 (26.6)	185 (33.1)	0.81	49 (23.7)	41 (19.8)	1.19
3	210 (37.6)	25 (4.5)	8.40	50 (24.1)	9 (4.3)	5.56
Maximum number of vomiting episodes*						
0	98 (17.5)	98 (17.5)	1.00	47 (22.7)	47 (22.7)	1.00
1	41 (7.3)	200 (35.8)	0.21	17 (8.2)	60 (28.9)	0.28
2	242 (43.3)	201 (36.0)	1.20	68 (32.9)	50 (24.2)	1.36
3	178 (31.9)	60 (10.7)	2.96	75 (36.2)	50 (24.2)	1.50
Maximum temperature (°C)*						
0	12 (2.2)	222 (39.7)	0.05	2 (1.0)	22 (10.7)	0.09
1	327 (58.5)	58 (10.4)	5.64	93 (44.9)	40 (19.3)	2.32
2	104 (18.6)	128 (22.9)	0.81	52 (25.1)	70 (33.8)	0.74
3	116 (20.7)	151 (27.0)	0.77	60 (29.0)	75 (36.2)	0.80

* Fisher's Exact significant at ≤0.001 within each region.
** 40/42 cases with a VSS score of 0 for the number of looser-than-normal stools were from Mali, see explanation in Section 4.

CSS: Africa and Asia ≥10) is used, the proportion of severe VSS cases classified as not severe by the CSS was reduced to 17.1% (49/287) in Africa and to 9.5% (11/116) in Asia, with 14.7% and 9.5% of CSS severe cases in Africa and Asia, respectively, classified as not severe according to the VSS. As compared to the original classification, when the modified scoring classification based on a threshold set at the median of the scoring distribution (VSS: ≥11; CSS ≥13) was used, the proportion of severe VSS cases classified as not severe by the CSS was reduced to 35.7% (81/227) in Africa and 48.3% (56/116) in Asia, with 5.8% (9/155) and 3.2% (2/62) of CSS severe cases in Africa and Asia, respectively, classified as not severe according to the VSS. Notably, while there were still differences in severe gastroenteritis categories when using either of the modified classifications, the agreement between the two scoring systems improves substantially as compared to the original severity classification; from kappa = 0.27 and kappa = 0.10 in Africa and Asia using the original severity classifications to kappa = 0.68 and kappa = 0.78 using the mean score modified classification and kappa = 0.65 and kappa = 0.47 using the median of the scoring distribution modified classification.

4. Discussion

In these randomized, controlled efficacy trials of PRV in low-resource settings in Africa and Asia, the VSS and CSS performed differently, with the VSS classifying more cases as *severe* in both regions. Using the VSS as compared to the CSS resulted in approximately four and nine times the number of severe cases in Africa and Asia, respectively (Table 4). These results are consistent with those identified by Givon-Lavi et al. [23] in a study conducted using a different design – a prospective hospital-based observational study – and among a different population – children less than 5 years of age in Israel.

The developing country trials of PRV, analyzed here, relied on active collection of symptom data from the beginning of an episode using parental recall at the time of presentation to a clinic and a combination of clinical staff examination and parental recall to collect symptoms from clinic presentation through discharge. Thus, the general similarities in findings to the Givon-Lavi et al. are particularly interesting, given that their study collected severity score information based on a reporting system in which completion of symptom collection occurred 8 days following the initial assessment based on parental recall and review of the medical chart. However, the relative proportions of severe cases captured using the CSS as compared to the VSS in the Givon-Lavi et al. study were somewhat lower than in this Africa study. This may be due to the fact that the CSS relies more on symptom duration for scoring than the VSS, and the full duration of symptoms may have been more difficult to capture using the reporting system in the Givon-Lavi et al. study.

Our findings suggest that the differences in severity score classification are at least partially due to the severity threshold chosen. To be categorized as severe using the CSS, one needed a value in the upper-third of all possible total values (17 points or higher out of a possible 24), while in the VSS on needed a value in upper half of all possible values (11 points or higher out of a possible 20). For this reason, the VSS more frequently scores gastroenteritis episodes as severe as compared to the CSS. By setting the severity thresholds at different points along the two scales in this investigation, the degree of inconsistency in severity classifications was reduced. As presented, when the severity threshold for the CSS and VSS was set equivalent to the mean score observed in these trials, similar to the threshold used in the development of the VSS [20], fewer cases identified as severe according to the VSS were identified as not severe according to the CSS in Africa and Asia. When the severity threshold for both scoring systems was set at the median of the distribution, the number of severe VSS cases classified as not severe by CSS increased as compared to the mean severity threshold, although was reduced as compared to the original severity classifications. This increase in severity classification agreement

Table 4
Severity classification by scoring system among rotavirus-positive gastroenteritis episodes.

Vesikari scoring system (≤10 not severe, ≥11 severe)	Clark scoring system, *n* (%)	
	Not severe (≤16)	Severe (≥17)
Region		
Africa*		
Not severe	332 (100.0)	0 (0.0)
Severe	174 (76.7)	53 (23.4)
Asia*		
Not severe	91 (100.0)	0 (0.0)
Severe	103 (88.8)	13 (11.2)
Site		
Ghana*		
Not severe	53 (100.0)	0 (0.0)
Severe	63 (73.3)	23 (26.7)
Bangladesh*		
Not severe	75 (100.0)	0 (0.0)
Severe	80 (87.9)	11 (12.1)
Kenya*		
Not severe	26 (100.0)	0 (0.0)
Severe	15 (62.5)	9 (37.5)
Mali*		
Not severe	253 (100.0)	0 (0.0)
Severe	96 (82.1)	21 (17.9)
Vietnam		
Not severe	16 (100.0)	0 (0.0)
Severe	23 (92.0)	2 (8.0)
Gender		
Male*		
Not severe	218 (100.0)	0 (0.0)
Severe	154 (79.0)	41 (21.0)
Female*		
Not severe	205 (100.0)	0 (0.0)
Severe	123 (83.1)	25 (16.9)
Age		
<6 months*		
Not severe	48 (100.0)	0 (0.0)
Severe	19 (67.9)	9 (32.1)
6–11 months*		
Not severe	112 (100.0)	0 (0.0)
Severe	93 (82.3)	20 (17.7)
12–17 months*		
Not severe	186 (100.0)	0 (0.0)
Severe	126 (80.3)	31 (19.8)
18–23 months*		
Not severe	76 (100.0)	0 (0.0)
Severe	37 (86.1)	6 (13.1)
≥24 months		
Not severe	1 (100.0)	0 (0.0)
Severe	2 (100.0)	0 (0.0)

* Fisher's Exact significant at ≤0.002.

Table 5
Agreement between scoring system severity classifications among rotavirus-positive gastroenteritis episodes (mild, moderate, severe).

Clark Severity Scale	Vesikari Severity Scale		
	<11 (not severe)	≥11 (severe)	Total
Original classification[a]			
Africa			
<17 (not severe)	332	174	506
≥17 (severe)	0	53	53
Total	332	227	559
kappa = 0.27; $p \leq 0.001$			
Asia			
<17 (not severe)	91	103	194
≥17 (severe)	0	13	13
Total	91	116	207
kappa = 0.10; $p \leq 0.001$			
	<10 (not severe)	≥10 (severe)	Total
Modified classification: mean score observed in trials by region			
Africa			
<10 (not severe)	231	49	280
≥10 (severe)	41	238	279
Total	272	287	559
kappa = 0.68; $p \leq 0.001$			
Asia	<11 (not severe)	≥11 (severe)	Total
<10 (not severe)	80	11	91
≥10 (severe)	11	105	116
Total	91	116	207
kappa = 0.78; $p \leq 0.001$			
	<11 (not severe)	≥11 (severe)	Total
Modified classification: median scoring distribution			
Africa			
<13 (not severe)	323	81	404
≥13 (severe)	9	146	155
Total	332	227	559
kappa = 0.65; $p \leq 0.001$			
Asia			
<13 (not severe)	89	56	145
≥13 (severe)	2	60	62
Total	91	116	207
kappa = 0.47; $p \leq 0.001$			

[a] Table 4 presents the original classification according to severe and not severe categories.

between the two scoring systems using modified severity cutoffs is not unexpected; assuming that each scoring system is classifying severity relatively accurately, the modified cut offs standardized the two distributions relative to each other for the purposes of severity classification.

In this investigation, we lowered the CSS severity threshold based on utilizing mean scores for rotavirus-positive episodes observed in these trials and the median of the scoring distribution to make it more similar to the VSS. In contrast, the Givon-Lavi et al. study utilized different modified scoring categories; in that study, when the severity cutoff for the VSS was modified, a higher severity cutoff was used to make it more similar to the CSS. The differences in severity threshold classifications resulted in more similarity (i.e. higher agreement) between the scoring systems in this investigation when the severity thresholds were changed to match the mean score in these trials as compared to the median scoring distribution or the original severity thresholds. However, this greater agreement may not be generalizable. It is based on mean scores internal to these clinical trials which may not translate into the same level of agreement between scoring systems in other studies using different methods for symptom collection, such as more frequent home visits by field workers or diary cards for real-time parental collection of symptoms.

The CSS identified 9.5% and 6.3% of cases as severe in Africa and Asia, respectively. This is much lower than one-third of scores classified as severe according to the severity scoring distribution, while the VSS captured about 40.6% and 56.0% of cases as severe in Africa and Asia, respectively, similar to the one-half of cases captured as severe by Ruuska and Vesikari [20] in the case population in which it was originally designed. This reduction in identification of severe cases relative to the proportion of the scoring distribution classified as severe when using the CSS raises the question as to whether it was operating in these trial populations as it was originally intended and how this may relate to measurement of vaccine efficacy. Due to a lack of published information on CSS development, it is difficult to know for certain what percentage of participants were expected to be captured as severe.

The efficacy of rotavirus vaccines in more developed populations has been shown to increase with increasing disease severity [26,27]. In these trials of PRV in the developing world, we would expect a higher efficacy against severe disease as measured by the CSS as compared to VSS, given that the CSS score distribution was shifted such that only the highest severity cases would have met

the CSS severity threshold. However, the point estimates of efficacy measured in these trials were in fact similar using the two scoring systems' original thresholds, indicating that the CSS may not have performed as expected in these trials or that there may not be as strong of a relationship between severity and efficacy in these settings [6–9]. In the CSS, the definitions of behavior used (i.e. irritable, lethargic, and seizure) are subjective and do not have the same meaning or may be perceived differently in developing, as compared to developed, country settings leading to a reduction in the total CSS score. Additionally, since parents were not provided with thermometers and did not commonly have thermometers available at home, the full duration of fever may not have been captured, resulting in a reduction in the total CSS score.

In the development of the original VSS, items were scored by breaking the score for each item into thirds [20]. It is not clear how mild, moderate, and severe cutoffs were created for the CSS [17,22]. Among the common items contained within the scoring systems, with the exception of temperature, the VSS had a higher proportion of episodes with item scores of 2 and/or 3 for the five items common to both the VSS and the CSS. These differences indicate that the remaining severity classification discrepancies between the VSS and the CSS may be due, not only to the severity threshold chosen, but also to the differences in individual item scoring. In order to obtain equivalent severity cutoffs between the two scoring systems, item cutoffs should be reconsidered. While better consistency between severity score cutoffs could be achieved, due to the differences in items included in each scoring system and because the CSS is affected more by missing a symptom than the VSS (i.e. CSS does not provide a point score for the number of diarrhea episodes until two episodes have occurred and for the duration of vomiting until 2 days of vomiting have passed), it is unlikely that the severity scores would ever identify the exact same proportions of severe disease in any population.

Weaknesses of this post-hoc analysis included that the trials were designed to capture moderate to severe cases and, as explained in the main efficacy manuscript for Africa [8], despite common case capture methods, success in capturing cases differed between sites and regions. The challenges in capturing and scoring cases for the Mali site are described in this supplement [28]. Despite this, scoring distributions for the VSS and the CSS appeared normal in each region. Additionally, diary cards were not used to collect symptoms at home in these trials and, depending on healthcare seeking behaviors, the average time from symptom onset to clinic assessment varied by participant and site, thus leaving some sites more dependent on parental recall than others and allowing episode severity to develop further before seeking treatment at a healthcare facility. Larger discrepancies were identified between the two scoring systems in Asia as compared to Africa; the scoring systems, originally developed for use in middle- to high-income countries, did not perform similarly across low-income regions. For the CSS, this may be due to differences between regions in interpretation and understanding of subjective items, like behavior and temperature duration. For the VSS, this may be due to differences in rehydration and hospitalization patterns between regions. It was also observed that, based on the number of participants enrolled at each site, some sites captured an increased number of cases as compared to other sites which may have been due to differences in medical facility utilization by site, indicating a challenge of running any multi-center trial and trying to ensure that case capture methods are identical, regardless of cultural differences in health care seeking behaviors.

5. Conclusion

These results add to the body of evidence that the VSS scores more RVGE episodes as severe than the CSS and that, for consistency between studies, one severity scoring system should be used. Researchers should also be cognizant that the study implementation methods (i.e. the use/non-use of a diary card, frequency of home follow-up, passive vs. active reporting, provision of thermometers), and local perceptions of symptoms will have an impact on severity scores, and potentially, vaccine efficacy estimates. In order to better understand the scoring systems and how they categorize severe disease, as well as to prepare for additional rotavirus vaccine trials, future rotavirus clinical severity scoring system research should focus on understanding the ideal mild, moderate, and severe cut points for these scoring systems, identifying the scoring system items contained within the VSS and CSS that are most indicative of severe disease, and identifying an ideal single severity scoring system for use in developing country populations less than 2 years of age in Africa and Asia.

Acknowledgements

The efficacy trials were conducted with funding from PATH's Rotavirus Vaccine Program under a grant from the a GAVI Alliance, and Merck Research Laboratories. Thank you to the study participants, mothers, and families who participated in the research that made this analysis possible and to each site Principal Investigator (Dr. D. Anh, Dr. G. Armah, Dr. R. Breiman, Dr. S. Sow, and Dr. K. Zaman) and his team for the diligence and care in implementing and collecting severity score information from the efficacy trials. Finally, thank you to Erin Kester, Joyce Erickson, and Megan Le from PATH who were instrumental in coordinating the journal supplement.

Contributors: KDCL was involved in reviewing all relevant literature, developing the study methods specific to this comparative analysis of the two scoring systems, conducting data analysis and preparing the first and subsequent drafts of the manuscript. KMN, MJ, and AF were involved in developing the study methods specific to this comparative analysis of the two scoring systems, reviewing the data, and preparing the first draft of the manuscript. MJD, JCV, MC, TCM, GA, and KZ were involved in acquisition of the data, and critical review of the data analysis and manuscript. *Conflict of interest statement:* MJD and TCM are employees of Merck Research Laboratories, which manufactures RotaTeq, and MC was an employee of Merck Research Laboratories when the clinical trial was conducted, and all own/owned equity interest in the company. *Disclosure:* All authors have approved the final article.

References

[1] Parashar UD, Hummelman EG, Bresee JS, Miller MA, Glass RI. Global illness and deaths caused by rotavirus disease in children. Emerg Infect Dis 2003;9(May (5)):565–72.
[2] Rotavirus surveillance—worldwide, 2009. MMWR Morb Mortal Wkly Rep 2011;60(April (16)):514–6.
[3] UNICEF, World Health Organization. Diarrhoea: why children are still dying and what can be done. Geneva: World Health Organization; 2009.
[4] Glass RI, Parashar UD, Bresee JS, Turcios R, Fischer TK, Widdowson MA, et al. Rotavirus vaccines: current prospects and future challenges. Lancet 2006;368(July (9532)):323–32.
[5] World Health Organization. Generic protocol for monitoring impact of rotavirus vaccination on gastroenteritis disease burden and viral strains. Geneva: World Health Organization; 2008 [Report No.: WHO/IVB/08.16].
[6] Madhi SA, Cunliffe NA, Steele D, Witte D, Kirsten M, Louw C, et al. Effect of human rotavirus vaccine on severe diarrhea in African infants. N Engl J Med 2010;362(January (4)):289–98.
[7] Zaman K, Anh DD, Victor JC, Shin S, Yunus M, Dallas MJ, et al. Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in Asia: a randomised, double-blind, placebo-controlled trial. Lancet 2010;376(August (9741)):615–23.
[8] Armah GE, Sow SO, Breiman RF, Dallas MJ, Tapia MD, Feikin DR, et al. Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in sub-Saharan Africa: a randomised, double-blind, placebo-controlled trial. Lancet 2010;376(August (9741)):606–14.
[9] Breiman RF, Zaman K, Armah GE, Sow SS, Dang AD, Victor JC, et al. Analyses of health outcomes from the 5 sites participating in the Africa and Asia

clinical efficacy trials of the oral pentavalent rotavirus vaccine. Vaccine 2012; 30(Suppl. 1):A24–9.
[10] Meeting of the immunization Strategic Advisory Group of Experts, April 2009—conclusions and recommendations. Wkly Epidemiol Rec 2009;84(June (23)):220–36.
[11] Steele AD, Patel M, Parashar UD, Victor JC, Aguado T, Neuzil KM. Rotavirus vaccines for infants in developing countries in Africa and Asia: considerations from a world health organization-sponsored consultation. J Infect Dis 2009;200(November (Suppl. 1)):S63–9.
[12] World Health Organization. Global framework for immunization monitoring and surveillance. Geneva: World Health Organization; 2011 [Report No.: WHO/IVB/07.06].
[13] WHO Expert Committee on Biological Standardization. WHO Expert Committee on Biological Standardization 56th Report: guidelines to assure the quality, safety and efficacy of live attenuated rotavirus vaccines (oral). Geneva: World Health Organization; 2007 [Report No.: 941].
[14] de Palma O, Cruz L, Ramos H, de Baires A, Villatoro N, Pastor D, et al. Effectiveness of rotavirus vaccination against childhood diarrhoea in El Salvador: case-control study. BMJ 2010;340:c2825.
[15] Patel M, Pedreira C, De Oliveira LH, Tate J, Orozco M, Mercado J, et al. Association between pentavalent rotavirus vaccine and severe rotavirus diarrhea among children in Nicaragua. JAMA 2009;301(June (21)):2243–51.
[16] Freedman SB, Eltorky M, Gorelick M. Evaluation of a gastroenteritis severity score for use in outpatient settings. Pediatrics 2010;125(June (6)):e1278–85.
[17] Clark HF, Borian FE, Bell LM, Modesto K, Gouvea V, Plotkin SA. Protective effect of WC3 vaccine against rotavirus diarrhea in infants during a predominantly serotype 1 rotavirus season. J Infect Dis 1988;158(September (3)):570–87.
[18] Clark HF, Bernstein DI, Dennehy PH, Offit P, Pichichero M, Treanor J, et al. Safety, efficacy, and immunogenicity of a live, quadrivalent human-bovine reassortant rotavirus vaccine in healthy infants. J Pediatr 2004;144(February (2)):184–90.
[19] Ruuska T, Vesikari T. A prospective study of acute diarrhoea in Finnish children from birth to 2 1/2 years of age. Acta Paediatr Scand 1991;80(May (5)):500–7.
[20] Ruuska T, Vesikari T. Rotavirus disease in Finnish children: use of numerical scores for clinical severity of diarrhoeal episodes. Scand J Infect Dis 1990;22(3):259–67.
[21] World Health Organization. IMCI handbook: integrated management of childhood illness. Geneva: World Health Organization; 2005 [Report No.: WHO/FCH/CAH/00.12].
[22] Georges-Courbot MC, Monges J, Beraud-Cassel AM, Gouandjika I, Georges AJ. Prospective longitudinal study of rotavirus infections in children from birth to two years of age in Central Africa. Ann Inst Pasteur Virol 1988;139(October (4)):421–8.
[23] Givon-Lavi N, Greenberg D, Dagan R. Comparison between two severity scoring scales commonly used in the evaluation of rotavirus gastroenteritis in children. Vaccine 2008;26(October (46)):5798–801.
[24] Vesikari T, Clark HF, Offit PA, Dallas MJ, DiStefano DJ, Goveia MG, et al. Effects of the potency and composition of the multivalent human-bovine (WC3) reassortant rotavirus vaccine on efficacy, safety and immunogenicity in healthy infants. Vaccine 2006;24(May (22)):4821–9.
[25] Byrt T. How good is that agreement? Epidemiology 1996;7(September (5)):561.
[26] Vesikari T, Matson DO, Dennehy P, Van Damme P, Santosham M, Rodriguez Z, et al. Safety and efficacy of a pentavalent human-bovine (WC3) reassortant rotavirus vaccine. N Engl J Med 2006;354(January (1)):23–33.
[27] Ruiz-Palacios GM, Perez-Schael I, Velazquez FR, Abate H, Breuer T, Clemens SC, et al. Safety and efficacy of an attenuated vaccine against severe rotavirus gastroenteritis. N Engl J Med 2006;354(January (1)):11–22.
[28] Tapia MD, Armah GE, Breiman RF, Dallas MJ, Lewis KDC, Sow SS, et al. Secondary efficacy endpoints of the pentavalent rotavirus vaccine against gastroenteritis in sub-Saharan Africa. Vaccine 2012;30(Suppl. 1):A79–85.