Enterics for Global Health (EFGH)
Statistical Analysis Plan
Version 1.0

Table of Contents

SECTION 1. ADMINISTRATIVE INFORMATION

Title: Statistical Analysis Plan for the Enterics for Global Health (EFGH) study

SAP Version: 1.0 (31 October 2023)

Protocol Version: 5.0 (09 June 2023)

SAP Revision History:

SAP version (date)	Justification for Revision	Timing of SAP in relation to interim analysis

SAP Working Group

Name	EFGH Role
Faisal Ahmmed	Co-Investigator, icddr,b
Hannah Atlas	Co-Investigator, UW
John Clemens	Co-Principal Investigator-Bangladesh, IVI
Erika Feutz	Data Manager, UW
Sean Galagan	Senior Data Manager, UW
Jahangir Hossain	Co-Principal Investigator-the Gambia, MRCG
Muhammed Jawwad	Co-investigator, AKU
Karen Kotloff	Co-Principal Investigator-Mali, Kenya, the Gambia, UMB
Donnie Mategula	Co-Investigator, MLW
Christine McGrath	Co-investigator, UW
Billy Ogwel	Co-investigator, KEMRI
Dr. Patricia Pavlinac	Co-Principal Investigator-Coordination, UW
James Platts-Mills	Co-Investigator, UVA
Elizabeth Rogawski-McQuade	Co-Investigator, UVA
Francesca Schiaffina	Co-Investigator, UVA
Mili Tapia	Co-Investigator, UMB
Tahir Yousufzai	Co-investigator, AKU

Abbreviations

AKU	The Aga Khan University
CI	95% confidence interval
CLSI	Clinical and Laboratory Standards Institute
EFGH	Enterics for Global Health
GEMS	Global Enteric Multicenter Study
HAZ	Height for age Z-score
HR	Hazard ratio
HUS	Healthcare Utilization Survey
icddr,b	International Centre for Diarrhoeal Disease Research, Bangladesh
IR	Incidence rate
IRB	Institutional Review Board
IQR	Interquartile range
KEMRI	Kenya Medical Research Institute
LAZ	Length for age Z-score
LRTI	Lower respiratory tract infection
LSD	Less-severe-diarrhea
LTFU	Lost to follow-up
MAD	Medically-attended diarrhea
MAL-ED	Malnutrition and Enteric Disease Study
mBGS	Modified buffered glycerol saline
MDR	Multidrug-resistant
MLW	Malawi-Liverpool-Wellcome Trust
MRCG	The Medical Research Council Unit - The Gambia
MSD	Moderate-to-severe diarrhea
MUAC	Mid-upper arm circumference
MVS	Modified Vesikari score
NPV	Negative predictive value
PPV	Positive predictive value
ROC	Receiver operating characteristics
RR	Risk ratio
UMB	University of Maryland, Baltimore
UVA	The University of Virginia
UW	University of Washington
WAZ	Weight for age Z-score
WHO	World Health Organization
WLZ	Weight for length Z-score
XDR	Extensively drug-resistant

SECTION 2. INTRODUCTION

Background and Rationale

In low- and middle-income countries, nearly one third of children experience at least one episode of *Shigella*-attributable diarrhea during their first two years of life.[1] In addition to being a leading cause of diarrhea, this enteric bacterium is also associated with linear growth faltering, a precursor to stunting.[2,3] Stunting is a marker of vulnerability to childhood infection, decreased vaccine efficacy and lifelong morbidity. Currently, several promising *Shigella* vaccines are in development. Eventual Phase 2b/3 *Shigella* vaccine trials will require a consortium of potential vaccine trial sites in settings with a high incidence of *Shigella*-attributed medically-attended diarrhea (MAD), high participant retention, and the laboratory capacity to confirm *Shigella* infection. The Enterics for Global Health (EFGH) *Shigella* surveillance study will employ cross-sectional and longitudinal study designs to establish updated incidence rates and document consequences of *Shigella* diarrhea within seven country sites in Africa, Asia, and Latin America. Specifically, we aim to:

Aims

1. **Primary Aim**: Determine the incidence of *Shigella*-attributed MAD in children 6 to 35 months of age in each of the EFGH country sites.
2. **Secondary Aim 1:** Determine the incidence of *Shigella* MAD by serotype, severity definition, laboratory method (culture vs. qPCR), age, and by season.
3. **Secondary Aim 2:** Describe the prevalence of resistance to commonly used antibiotics in *Shigella* isolates in each EFGH country site.
4. **Secondary Aim 3:** Determine the risk of death, hospitalization, persistent diarrhea, diarrhea recurrence, and linear growth faltering in the three months following an episode of *Shigella* MAD.
5. **Secondary Aim 4:** Compare various severity definitions in their ability to distinguish *Shigella* from non-*Shigella* attributable diarrhea and ability to predict risk of death or hospitalization in the subsequent 3 months.
6. **Secondary Aim 5:** Quantify the cost incurred by families and health care systems due to *Shigella* morbidity and mortality.
7. **Secondary Aim 6:** Identify optimal laboratory methods for *Shigella* culture by:
    a. Comparing the isolation rate of *Shigella* between two transport media for rectal swabs (Cary-Blair and modified Buffered Glycerol Saline [mBGS]).
    b. Comparing the isolation rate of *Shigella* between two fecal sample types (rectal swabs and whole stool) among the subset of children who produced whole stool in The Gambia and Bangladesh country sites.

SECTION 3. STUDY METHODS

Study Design

The EFGH study will employ cross-sectional and longitudinal study designs to establish incidence and consequences of *Shigella* diarrhea. Children aged 6-35 months of age presenting with diarrhea at selected study health facilities will be recruited over a 24-month period and followed for three months. Randomly selected households will be visited within each EFGH site catchment area over the 24-month recruitment period to estimate the population of children aged 6-35 months in the catchment area. A healthcare utilization survey (HUS) will be conducted among households in the catchment area with one or more children in the 6-35-month age range to determine the proportion of diarrhea cases that present to EFGH study health facilities. The number of cases divided by the estimated number of children 6-35 months of age residing in the catchment area and time of surveillance will estimate the crude incidence rate of *Shigella* diarrhea in 6-35 month of age. This incidence rate will be adjusted by the proportion of eligible children enrolled and the proportion of eligible children with diarrhea who sought care at an EFGH recruiting facility.

Sample Size

The minimum number of confirmed *Shigella* cases (numerator) and minimum size of the source population within health facility catchment area (denominator) required to estimate *Shigella* incidences and 95% confidence intervals with specified precision (half-widths of 0.25 cases per 100 child-years) within children aged 6-35 months were estimated. To ensure appropriate precision in estimates of serotype and antibiotic susceptibility of *Shigella* isolates, we aim recruit at least 65 children with culture-confirmed *Shigella* diarrhea at each EFGH site. We assumed that 4.8% of diarrhea cases will be *Shigella* culture-positive in the EFGH study, requiring that we enroll 1400 children presenting to health facilities with diarrhea over the 24-month period (~58 per month). The size of the catchment area population (>97,000) was determined assuming an unadjusted *Shigella* incidence rate of 0.63 per 100 child-years (unadjusted for health care seeking behavior) and 95% confidence interval precision of 0.25 (half-width).

Framework

Not applicable.

Interim Statistical Analysis and Stopping Guidance

Interim "Data Readouts" presenting preliminary study results will be generated on a bi-yearly basis and shared with study investigators and the BMGF.

As this study is not a randomized trial and there is no intervention, there is no stopping guidance.

Timing of Final Analysis

The primary study publication will be prepared for the primary aim when every enrolled child has completed their three-month follow-up visit or is deemed lost to follow-up and all data for the primary aim has been cleaned (anticipated publication submission in May of 2025).

Figure A. Timeline of study activities

	2022				2023				2024				2025	
Key Activity	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**
Diarrhea case surveillance (24 months recruitment + 3 months final follow-up)														
Population enumeration activities														
Healthcare utilization survey														
Data cleaning														
Data analysis														
Publication preparations														
Dissemination of results														

Longitudinal follow-up

The schedule of study procedures is outlined in Figure A. Regularly scheduled clinical visits include those at enrollment and at follow-up (week four and month three) and are defined as a 14-day period in which visits are considered completed (24-37 days post-enrollment for the week 4 visit and 84-97 days post-enrollment for the month three visit). Additionally, visits are allowed for up to 30 days past the visit window (38-67 days post-enrollment for the week 4 visit and 98-127 days post-enrollment for the month 3 visit) though may be excluded from analyses (**Table B**).

Table B. Allowable windows for longitudinal follow-up

Visit	**Visit Window**	**Allowable visits**
Week 4 follow-up visit	24-37 days post-enrollment.	38-67 days post-enrollment.
Month 3 follow-up visit	84-97 days post-enrollment.	98-127 days post-enrollment.

SECTION 4. STATISTICAL PRINCIPLES

Confidence Intervals and P-values

Level of Statistical Significance
All statistical tests will be two-sided using a 5% significance level (alpha of 0.05).

Type I Errors
We will not adjust the alpha for multiple testing in the primary or secondary aims. For secondary aims that involve multiple hypothesis tests, we will frame the results as exploratory and acknowledge type 1 error in the limitations sections.

Confidence Intervals to be Reported
Two-sided 95% confidence intervals will be used. The specific calculation of confidence intervals are described under each specific aim.

Protocol Deviations

Definition of protocol deviations
The following are pre-defined major protocol deviations with a direct bearing on the primary outcome:

Errors in applying inclusion/exclusion criteria that are discovered after enrollment, including lack of informed consent.

The following are pre-defined minor protocol deviations:
Missed sample collection (stool/rectal swab, blood spot) due to participant refusal or other barrier to sample collection (such as visit occurring over phone).
Missed anthropometry assessment due to follow-up visit occurring over the phone

Description of which protocol deviations will be summarized
Protocol deviations will be classified as major and minor. The number (and percentage) of participants with major and minor protocol deviations will be summarized by study site in relevant analyses with details of the deviation provided. No statistical tests will be performed.

Analysis Populations

All children with non-missing outcome data (for each relevant aim) will be included in primary and secondary analyses.

Missing data for symptom duration variables will be assumed as absence of a given symptom in primary analyses and in secondary analyses assumed to be presence of the given symptom.

SECTION 5. STUDY POPULATIONS

Diarrhea Case Surveillance Screening

The total number pre-screened, screened, and enrolled in diarrhea case surveillance will be reported along with summary of reasons for exclusion into the study.

Diarrhea Case Surveillance Eligibility

Children aged 6 to 35 months old presenting at an EFGH facility with diarrhea (three or more abnormally loose or watery stools per 24 hours) (Table C).

Table C. Description of study population and criteria for the inclusion and exclusion of Diarrhea Case Surveillance enrollment.

Study Population	Children aged 6 to 35 months old presenting at an EFGH facility with diarrhea (3 or more abnormally loose or watery stools in the previous 24 hours).

Inclusion Criteria	• Primary caregiver and child plan to remain at their current residence for at least the next 4 months
	• Primary caregiver is able to provide informed consent (legal age or emancipated minor) and provides consent within a common language for which translations are available
	• Child resides within the pre-defined study area
	• Fewer than four hours have passed since the child presented to a health facility
	• Diarrhea episode is:
	  o Acute (onset within seven days of study enrollment) and
	  o Represents a new episode (onset after at least two diarrhea-free days)
	• Caregiver is willing to have child participate in follow-up visits at week four and month three
	• Willingness to have samples collected from the child (rectal swabs at enrollment)
	• Site enrollment cap has not been met
	• Child is not being referred to a non-EFGH facility at the time of screening

Exclusion Criteria	• Children younger than 6 months or 36 months or older • Diarrhea does not meet the study definition (three or more abnormally loose or watery stools in the previous 24 hours) • Primary caregiver unable to provide informed consent or refuses to provide consent • Primary caregiver refuses verbal consent to screening procedures • Child does not reside in the study catchment area • four or more hours have passed since child presented to the study facility • Diarrhea episode is not acute (>7 days) or does not represent a new episode (<2 diarrhea free days) • Caregiver unwilling to have child participate in follow-up at four weeks and three months • Caregiver unwilling to have samples collected at enrolment (rectal swabs) • Site enrolment cap has been met • Child is being referred to a non-study facility

Population Enumeration and HUS Eligibility

All households residing in the study catchment area and where an adult household member provides verbal consent will be eligible to participate in population enumeration activities. Children who are (1) 6-35 months of age, (2) had diarrhea (three or more unusually loose or watery stools during a 24-hour period) in the past 14 days, and (3) whose primary caretakers provides written or verbal informed consent (per site procedures and IRB requirements) will complete the HUS.

Recruitment

Children will be passively recruited from study outpatient facilities for Diarrhea Case Surveillance enrollment. Prior to the screening process, potential participants will be pre-screened for eligibility by the study staff.

Per CONSORT guidelines, we will report the number of individuals who (**Figure 1**):

1. Underwent pre-screening
2. Underwent screening
3. Did not undergo screening (and reasons)
4. Met inclusion criteria
5. Did not meet inclusion criteria (and reasons)
6. Enrolled in the study
7. Not enrolled in study (and reasons)
8. Completed week four and month three follow-up visits

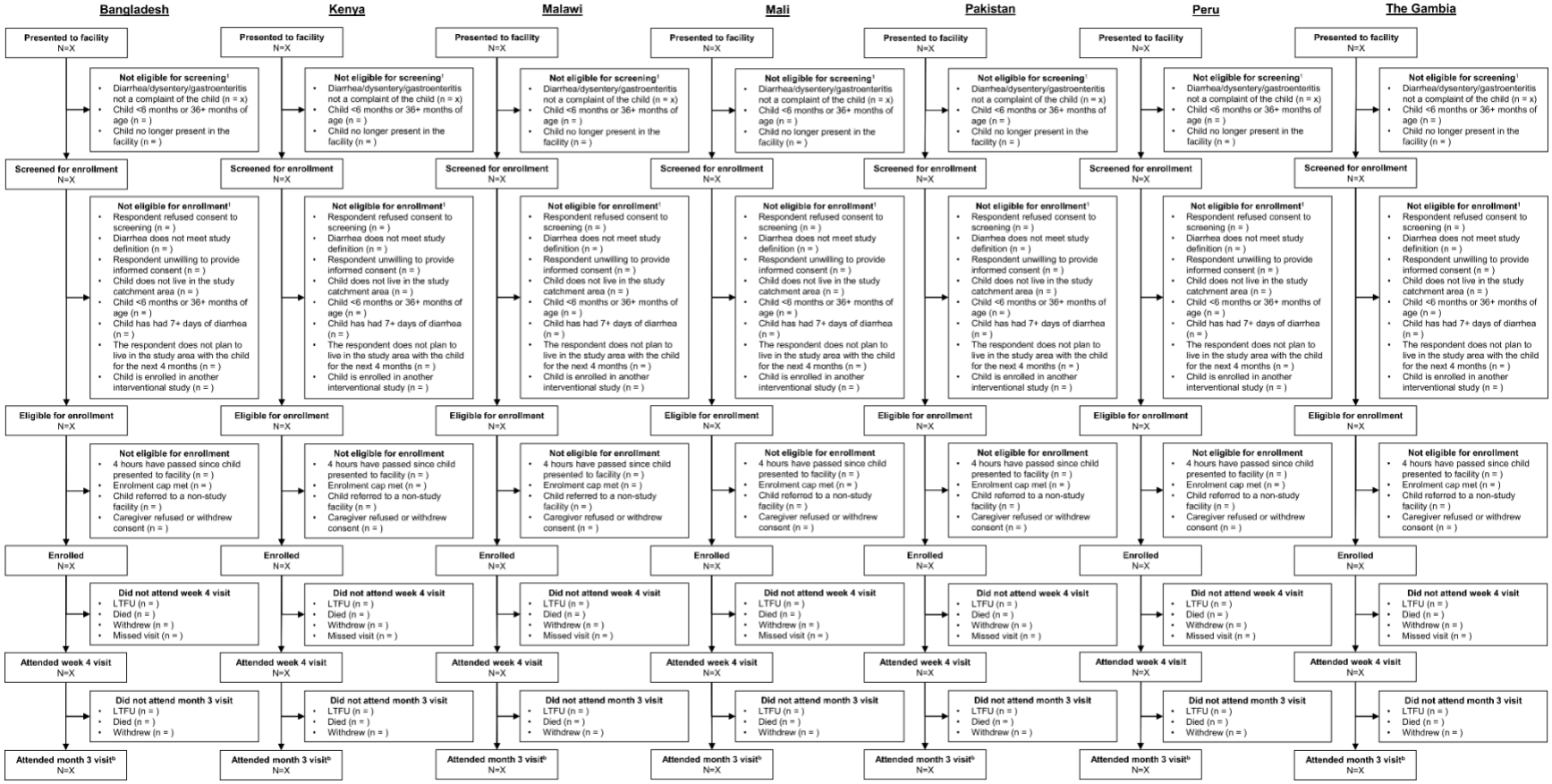

Figure 1. EFGH CONSORT diagram. [a] Reasons participants were not screened or did not meet eligibility do not sum to total as participants could have screened out for multiple reasons. [b] Among enrolled participants who attended the Week Four visit or missed the Week Four visit but not due to death, withdrawal or lost to follow-up (LTFU).

Withdrawal/Follow-Up

Withdrawal

Withdrawal of consent will be tabulated by country site using the following categories: withdrawal but allow prior collected data/ samples to be used and withdrawal and disallow already collected data/samples to be utilize.

Missed Visits/Loss to follow-up

Week four visits (defined as study visits 24 to 67 days after enrollment) and month three visits (defined as study visits 84 to 127 days after enrollment) that were missed will be tabulated in categories of completely missed, visit occurred but outside of window, visit occurred over phone by country site. Lost to follow-up (LTFU) will be defined as missing both the week four and month three visit.

Baseline Participant Characteristics

Table 1. Demographic and clinical characteristics of participants enrolled to Diarrheal Case Surveillance, by study site.

	Bangladesh	Kenya	Malawi	Mali	Pakistan	Peru	The Gambia	Total
Demographic indicators								
Total enrolled: N								
Female sex: n (%)								
Age (months): n (%)								
6 – 11								
12 – 23								
24 – 35								
Age in months: median (IQR)								
Highest maternal education achieved: n (%)								
Less than primary school								
Koranic school only								
Primary school or greater								
Unknown or declined								
Accompanying caregiver employment: n (%)								
Not employed								
Employed								
Unknown								
Children <5 years in household: median (IQR)								
Wealth index: median (IQR)								
Clinical characteristics								
Symptoms at enrollment: n (%)								
Dysentery[a]								
Oedema								
Signs of LRTI[b]								
Stiff neck								
Generalized rash								
Convulsions								
Lethargy or unconscious								
Palmar pallor								
Dehydration[c]: n (%)								
None								
Some								
Severe								
Diarrhea severity classifications								
Modified Vesikari score (MVS)[d]: median (IQR), n (%)								
Mild illness (0-8 points)								
Moderate illness (9-10 points)								
Severe illness (11+ points)								
Unknown								
Moderate or severe diarrhea by MVS or dysentery[e]: n (%)								
GEMS[f]: median (IQR), n (%)								
MSD								
LSD								
GEMS-*Shigella* score[g]: median (IQR), n (%)								
Mild (<6 points)								
Moderate (6-8 points)								
Severe (9+ points)								
Unknown								
Clark score[h]: median (IQR), n (%)								
Mild (2-8 points)								
Moderate to severe (9+ points)								
Unknown								
MAL-ED score[i]: median (IQR), n (%)								
Non-severe (<6 points)								
Severe (6+ points)								
Unknown								
CODA score[j]: median (IQR), n (%)								
Mild (0 points)								
Moderate to severe (1+ points)								
Anthropometry								
Stunted[k]: n (%)								

Table 1. Demographic and clinical characteristics of participants enrolled to Diarrheal Case Surveillance, by study site.								
	Bangladesh	Kenya	Malawi	Mali	Pakistan	Peru	The Gambia	Total
Severe								
Moderate								
None								
Wasted[l]: n (%)								
Severe								
Moderate								
None								
Underweight[m]: n (%)								
Severe								
Moderate								
None								

IQR: interquartile range, LAZ: length for age Z-score, LRTI; lower respiratory tract infection, LSD: less-sever-diarrhea, MSD: moderate-to-severe diarrhea, MUAC: mid-upper arm circumference, MVS: modified Vesikari score, WAZ: weight for age Z-score, WHO: world health organization, WLZ: weight for length Z-score.

[a] Blood in the stool as reported by caregiver during the diarrheal episode or by clinical diagnosis

[b] One or more of the following signs indicative of lower respiratory infection: cough, difficulty breathing, chest in-drawing, chest auscultation, central cyanosis, oxygen saturation <90%, and severe respiratory distress.

[c] Based on WHO criteria. Severe dehydration = At least two of the following signs: lethargy, abnormally sunken eyes, drinks poorly, skin pinch >2 seconds. Some dehydration = At least two of the following signs: restless/irritable, abnormally sunken eyes, drinks eagerly, skin pinch 1-2 seconds. [4]

[d] Defined as in PATH Vesikari Clinical Severity Scoring System Manual: Duration of diarrhea: 1-4 days (1 point), 5 days (2 points) ≥6 days (3 points); Max # of stool in 24 hour period: 1-3 (1 point), 4-5 (2 points), ≥6 (3 points); Duration of vomiting: 1 day (1 point), 2 days (2 points), ≥3 days (3 points); max # of vomiting episodes in 24 hour period: 1 (1 point), 2-4 (2 points), ≥5 (3 points); Axillary temperature 36.6-37.9°C (1 point), 38.0-38.4°C (2 points), ≥38.5°C (3 points); dehydration 1-5% (2 points), ≥6% (3 points); treatment: rehydration (1 point), hospitalization (2 points).

[e] Defined as in Pavlinac, Vaccines, 2022 as a MVS of 9+ or presence of visible blood in stool. N=x could not be assessed due to missing data for diarrhea duration. [5]

[f] Defined as in Kotloff, Lancet GH, 2019. Moderate-to-severe diarrhea (MSD) defined as presenting to a health facility with diarrhea and severe or some dehydration (by WHO criteria), visible blood in stool, or inpatient admission. Less-severe-diarrhea (LSD) defined as presenting to a health facility without MSD.[6]

[g] Defined as in Pavlinac, CID, 2021. Duration of diarrhea through day of presentation: 1-3 days (0 points), 4-5 days (2 points), ≥6 days (3 points); WHO-defined dehydration categories: severe (8 points), some (4 points), none (0 points); inpatient admission (5 points).[7]

[h] Defined as in Clark, JID, 1988. Duration of diarrhea: 1-4 days (1 point), 5-7 days (2 points) >7 days 3 points); Max # of stool in 24 hour period: 2-4 (1 point), 5-7 (2 points), >7 (3 points); Duration of vomiting: 2 days (1 point), 3-5 days (2 points), >5 days (3 points); max # of vomiting episodes in 24 hour period: 1-3 (1 point), 4-6 (2 points), >6 (3 points); Duration of reported fever: 1-2 days (1 point), 3-4 days (2 points), ≥5 days (3 points); Rectal temperature 38-38.2°C (1 point), 38.3-38.7°C (2 points), ≥38.8°C (3 points); behavioral signs: Irritable/less playful (1 point), Lethargic/listless (2 points), Seizures (3 points). N=x could not be assessed due to missing data for diarrhea duration.

[i] Defined as in Lee, BMJ Open, 2014: Duration of diarrhea: 2-4 days (1 point), 5-7 days (2 points), ≥8 days (3 points); Max # of stool in 24 hour period: <5 loose stools/24 hours (1 point), 5-7 loose stools/24 hours (2 points), ≥8 stools/24 hours (3 points); Duration of vomiting: 1 day (1 point), 2 days (2 points), ≥3 days (3 points); duration of reported fever: 1+ days (1 point); Confirmed temperature: ≥37.5°C (confirmed by field worker) (2 points); Dehydration: Some (2 points), severe (3 points). N=x could not be assessed due to missing data for diarrhea duration.

[j] Defined as in Lee, BMJ Open, 2014: Max # of stool in 24 hour period: 4-5 stools (1 point), 6-7 stools (2 points); Duration of vomiting: 1-2 days (1 points), 3-4 days (2 points), ≥5 days (3 points); Duration of reported fever: 1-2 days (1 point), 3-4 days (2 points), ≥5 days (3 points); Dehydration duration: 1-2 days (1 point), 3-4 days (2 points), ≥5 days (3 points); Anorexia: 1-2 days (1 point), 3-4 days (2 points), ≥5 days (3 points) ≥8 stools (3 points).

[k] Severe stunting: LAZ < -3, Moderate stunting: -3 ≤ LAZ < -2.

[l] Severe wasting: WLZ < -3 or MUAC < 11.5, Moderate wasting: -3 ≤ WLZ < -2 or 11.5 ≤ MUAC < 12.5.

[m] Severely underweight: WAZ < -3, Moderately underweight: -3 ≤ WAZ < -2.

Table 2. Demographic and Clinical Characteristics of the Population Enumeration and Healthcare Utilization Survey, by study site.								
Indicator	**Bangladesh**	**Kenya**	**Malawi**	**Mali**	**Pakistan**	**Peru**	**The Gambia**	**Overall**
Population Enumeration								
Total clusters demarcated: n								
Study area: m^2								
Clusters enumerated: n (%)								
Cluster enumerated								
No households present in the cluster								
Could not enumerated cluster due to safety or other reasons								
Not enumerated								
Total individuals enumerated: n								
Children 6-35 months of age enumerated: n								
Caregiver reported children 6-35 months of age had diarrhea in the past two weeks: n (%)								
Yes								
No								
Don't know								
Caregiver of child 6-35 months of age with diarrhea consented to HUS: n (%)								
Yes, consented								
No, refused								
No, caregiver not available for consent no successful revisit								
Care-seeking for diarrhea: n (%)[a]								
Did not seek care								
Sought care at an EFGH facility								
Sought care at a different outpatient hospital or health center								
Sought care at an inpatient hospital or health center								
Sought care at a health outpost								
Sought care at a drug seller or pharmacist								
Sought care at a traditional or religious healer								
Other								
Demographic indicators of children in the HUS								
Female sex: n (%)								
Age (months): n (%)								
6 - 11								
12 - 17								
18 - 23								

24 - 35	
Age in months: median (IQR)	
Clinical characteristics	
Symptoms during diarrheal illness: n (%)[a]	
Blood in stool	
Irritable	
Very thirsty	
Sunken eyes	
Wrinkled skin	
Drinks eagerly, thirsty	
Unable to drink or drank poorly	
Lethargic, unconscious, or hard to stay awake	

HUS: healthcare utilization survey, IQR: interquartile range.
[a] Column does not sum to total as surveyed participants may have sought care from multiple sources or reported multiple clinical symptoms.

SECTION 6. ANALYSIS

Definitions

1. **Diarrhea** defined as three or more abnormally loose or watery stools in a 24-hour period**.**
2. **Diarrhea episode** will be defined as the period of days in which the above definition is met followed by two diarrhea-free days.
3. **Culture-confirmed *Shigella* diarrhea case** will be defined as a case of acute diarrhea presenting to an EFGH facility in which *Shigella* was isolated from either of the cultured fecal sample collected by rectal swab and transported in mBGS or Cary-Blair media.
4. **qPCR-confirmed *Shigella* diarrhea case** will be defined as a case of acute diarrhea presenting to an EFGH facility in which *Shigella* DNA was identified in the fecal sample by qPCR (cycle threshold [CT] <31.1 for rectal swab or CT<29.8 for whole stool) by the TaqMan Array Card (TAC) assay.
5. ***Shigella* – attributed diarrhea case** will be defined as a case of acute diarrhea presenting to an EFGH facility in which *Shigella* was confirmed by either culture or qPCR (as outlined in #1 and #2 above).
6. **Dysentery** will be defined for diarrhea case surveillance as: a caregiver report of blood in the stool during the index diarrhea episode or a clinician diagnosis during the enrollment procedures and for the HUS as: caregiver report of blood in the stool.
7. **Watery diarrhea** will be defined for diarrhea case surveillance as: the lack of caregiver report of blood in the stool during screening, enrollment and during the diarrhea episode, and no dysentery diagnosis by the clinician among enrolled participants in the diarrhea case surveillance, and for the HUS as: the lack of caregiver report of blood in the stool for children whose caregiver reported diarrhea in the previous 14 days.
8. **Severity of diarrhea** will be defined according to multiple definitions (dysentery/watery diarrhea, hospitalized diarrhea, modified Vesikari score [MVS], moderate or severe diarrhea by MVS or dysentery, GEMS moderate-to-severe diarrhea [MSD] or less severe diarrhea [LSD], GEMS-*Shigella,* Clark, CODA or MAL-ED) for incidence rate and consequence comparisons.
9. **Deaths** occurring within the three-month follow-up period will be assessed by caregiver report at each scheduled visit or during upcoming visit phone reminders. Date and cause of death will be obtained from caregiver history, hospital records or death certificate, when available. The death certificate will be considered the gold standard for date of death. If a child died in the three-month period, but this information was not known until up to five months (two months beyond scheduled three-month visit), this will be included as death. Deaths occurring outside of the three-month window will not be included.
10. **Hospitalization** will be assessed at four-week and three-month visits by caregiver recall and hospital records (gold standard) when available. Date and time of admission, length of hospital stay, presenting signs/symptoms, and treatment received will be obtained. Hospitalizations that are a continuation of

management from a previous hospitalization (such as referrals) will be excluded from the analysis. For the purpose of standardization across sites, hospitalization will be defined as an overnight stay (child was on the ward from at least 12am to 6am).

11. **Loss to follow-up (LTFU)** will be defined as not having attended both the four-week and three-month follow-up visits after two months of actively tracing the child.
12. **Linear growth change** will be defined as the change in mean length/height-for-age z-score (ΔLAZ/ΔHAZ) from enrollment to three months. The 2006 World Health Organization (WHO) reference population will be used to calculate HAZ from the average of two repeated length/height (cm) measures per child per time point.
13. **Duration of diarrhea/dysentery/vomiting/fever** will be determined as the number of days a child has the symptoms within an episode of diarrhea. Because an episode is defined by two diarrhea-free days, the duration of each symptom may differ from the duration of the diarrhea episode (for example if there is one diarrhea free day within an episode). Notably this is distinct from duration of the episode.
14. **Duration of diarrhea episode** will be defined as the number of days between when the caregiver reported the child's diarrhea starting and the last day of diarrhea prior to the two consecutive diarrhea-free days concluding the episode.
15. **Prolonged diarrhea** will be defined as 7 or more days of diarrhea within an episode (starting from the date at which the diarrhea first started (as opposed to date at presentation).
16. **Persistent diarrhea** will be defined as 14 or more days of diarrhea within an episode (starting from the date at which the diarrhea first started (as opposed to date at presentation).
17. **Chronic diarrhea** will be defined as 30 or more days of diarrhea within an episode (starting from the date at which the diarrhea first started (as opposed to date at presentation).
18. **Diarrhea/dysentery recurrence** will be defined as new diarrhea/dysentery episodes (>48 hours after a diarrhea-free period).
19. **Cost per episode treated** will be calculated using the direct and indirect costs of *Shigella*-associated MAD per outpatient and inpatient episode, from the household, health system, and societal perspectives.
20. **Antibiotic resistance** will be based on zone size (from disc diffusion) values for each tested antibiotic and categorized as susceptible, intermediate, or resistant according to the most recent Clinical and Laboratory Standards Institute (CLSI) interpretive standards at the time of data analysis. Resistant and intermediate categories will be collapsed into a non-susceptible category to create a dichotomous variable. Multidrug-resistant (MDR) will be defined as resistance to at least three of the following medications: Ampicillin, Azithromycin, Ceftriaxone, Ciprofloxacin, Nalidixic Acid, Pivemicellinam, and Trimethroprim-Sulfamethoxazole. Extensively drug-resistant (XDR) will be defined as resistance to Ampicillin, Azithromycin, Ciprofloxacin, Ceftriaxone and Trimethoprim-Sulfamethoxazole. Resistance to all World Health Organization (WHO) recommended treatments will be defined as resistance to Azithromycin, Ciprofloxacin and Ceftriaxone.

Analysis Methods

Statistical Analysis

1. **Aim 1.** *To determine the incidence of* Shigella*-attributed diarrhea in children 6 to 35 months of age in each of the EFGH country sites.*
    a. *Statistical Analysis:* The adjusted incidence of *Shigella* will be calculated as the sum of total confirmed *Shigella* diarrhea cases divided by the child-years at risk of children 6-35

months of age in the defined catchment area, adjusting for healthcare seeking behavior (or the A-factor) and children who were eligible but not enrolled (the B factor). Crude incidence rates as well as incidence rates adjusted for enrollment but not care-seeking will also be presented. Analyses will be conducted separately using both culture-confirmed *Shigella* and *Shigella* confirmed by molecular methods (TAC). All incidence rates will first be calculated using the number of *Shigella* cases and defined population size at the facility level, adjusted for the facility-specific enrollment adjustment, summed to determine the total enrollment-adjusted incidence rate, and then adjust for care-seeking to determine the final adjusted incidence (**Table 3**).

$$\textbf{Crude Incidence} = \frac{\#\ \text{confirmed}\ Shigella\ \text{cases enrolled}}{\text{Defined population size}\ \times\ \text{Time}}$$

$$\textbf{Adjusted Incidence} = \frac{1}{\text{Defined population size}} \times \sum \frac{I_{ij}}{A_i \times B_{ij} \times Time_j}$$

Where I_{ij} is an indicator for confirmed *Shigella* case *i* in enrollment facility *j*

A_i=Episode-specific proportion who seek care at EFGH facility (e.g., among those with comparable age and SES and diarrhea of comparable severity)

B_{ij}=Episode-specific proportion enrolled (among those who sought care at the EFGH facility with diarrhea)

$Time_j$=Facility-specific period of time recruiting takes place (i.e. 24 months)

Defined population size=Estimated # of children aged 6-35 months living in the pre-defined catchment area

Child-years at risk to *Shigella* is determined by the defined population size and the facility-specific period of follow-up ($Time_j$; approximately two years). The first is the estimated number of children 6-35 months of age in the catchment area which will be determined by totaling the number of children in this age range enumerated during population enumeration activities and extrapolating this estimate to study clusters not visited and households that were not reached.

Among children 6-35 months of age who had diarrhea in the past 14 days and whose caregiver consented the HUS, the healthcare seeking adjustment, A_i is defined as the proportion of children with a similar syndrome who sought care for diarrhea at an EFGH-facility. The healthcare seeking adjustment (or A_i) will be estimated for each enrolled case using the estimated coefficients from a 'propensity to seek care' weighting model fit to data from children reporting diarrhea in the previous two weeks in the HUS. At a minimum, A_i will be computed separately for dysentery and watery diarrhea (i.e., the weighting model will include dysentery as a covariate as ascertained by the caregiver answering the HUS). The weighting model may also include other characteristics of diarrhea severity, socioeconomic status (SES), age and other variables of importance depending on the numbers of comparable cases observed in the HUS. The propensity to seek care model in the HUS will use any care-seeking at a primary healthcare facility and will then be adjusted for the site-specific difference in overall care-seeking and care-seeking to EFGH facilities.

In sensitivity analyses, we will restrict the weighting model to cases of diarrhea from the HUS within the past seven days instead of 14 days as seven-day recall is likely less biased than 14-day recall (although will inevitably be less precise because of fewer episodes included). Because children who have more severe diarrhea are more likely to seek care,

we expect to upweight diarrhea cases enrolled with less severe diarrhea more heavily than diarrhea cases enrolled with more severe diarrhea.

The enrollment adjustment (or B_{ij}) will be episode and facility-specific based on Diarrhea Case Surveillance enrolment procedures. Specifically, it will be computed separately for dysentery and watery diarrhea and for each facility. The enrollment adjustment is defined as the percentage of children who were enrolled among the estimated number of children six to 35 months of age living in the catchment area who met the study definition of diarrhea and therefore would have been eligible for study participation but screened out due to lack of consent, no longer being at the facility, or other administrative reasons (i.e. presented overnight or during a weekend when EFGH screening did not occur). This adjustment is split into two parts: *prescreening* and *screening*. The *screening* adjustment determines the percentage enrolled of those who were theoretically eligible (6-35 months, living in catchment area, and met diarrhea definition) to account for children who were not enrolled due to caregiver refusal or withdrawal of consent, limited staff of clinic capacity, the respondent not planning to remain in the study area four months, the child being enrolled in another interventional study, more than four hours having passed since the child began screening procedures, the child being referred to a non-EFGH facility or the child not being enrolled due to the enrollment cap. The *prescreening* adjustment determines the percentage of children 6-35 months of age whose caregiver or clinical records indicated they presented with diarrhea, dysentery or gastroenteritis and who were not already enrolled in EFGH who were screened to account for children who did not get screened due to verbal refusal by the caregiver or the child not being present in the facility. Because it is unknowable whether or not these children would have been eligible had they undergone full screening, we will apply the observed eligibility proportion from screening to those pre-screened. The final enrollment adjustment is taken by multiplying the screening and prescreening adjustments.

Confidence intervals around the primary aim of adjusted incidence of *Shigella* (as well as secondary aims presenting stratified *Shigella* incidence by age, serogroup, serotype, disease severity and seasonality) will be generated using M-estimation. Specifically, estimating equations will be constructed for each component of the incidence estimate (the number of confirmed *Shigella* cases, A_i, B_{ij}, defined population size, and incidence) and solved in a single step using the *geex* package in R. The sandwich variance estimator will be used to calculate standard error and generate Wald-based confidence intervals (estimate +/- 1.96*standard error).

b. Sensitivity analyses:
   i. We will restrict the care seeking adjustment to cases of diarrhea within the past seven days instead of 14 days as seven-day recall is likely less biased than 14-day recall (although will inevitably be less precise because of fewer episodes included).
   ii. As the enrollment adjustment does not consider, as potentially eligible, children who are currently enrolled in EFGH and in the three-month follow-up period, we will estimate the potential impact of this exclusion on the diarrhea incidence by estimating, through follow-up visit and unscheduled visit questionnaires, the proportion of enrolled cases who presented to EFGH facilities with diarrhea who would otherwise meet the EFGH eligibility criteria. We will estimate incidence rates that include these diarrhea recurrences to estimate the upper bound of incidence rates (such as might be observed in a prospective cohort study).

iii. Attributable *Shigella* by TAC will be defined as a rectal swab ipaH CT<31.1, however a sensitivity analysis using a cutoff of CT<30.0 will be conducted.

Table 3. Adjusted incidence of *Shigella* by EFGH country and method of ascertainment.

	Confirmed Shigella cases: n			*Shigella* MAD incidence		
EFGH Country	*Watery diarrhea*	*Dysentery*	*Total*	*Crude IR (CI)*	*Enrollment-adjusted IR[a] (CI)*	*Adjusted IR[b] (CI)*
Culture-confirmed *Shigella*[c]						
Bangladesh						
Kenya						
Malawi						
Mali						
Pakistan						
Peru						
The Gambia						
Total[d]						
***Shigella* attributable by molecular diagnostics[e]**						
Bangladesh						
Kenya						
Malawi						
Mali						
Pakistan						
Peru						
The Gambia						
Total[d]						

CI: 95% confidence interval, CT: cycle threshold, MAD: medically-attended diarrhea, IR: incidence rate.
[a] The number of confirmed *Shigella* watery diarrhea cases per 100 child-years adjusted for enrollment plus the number of confirmed *Shigella* dysentery cases per 100 child-years at risk adjusted for enrollment (# confirmed *Shigella* watery diarrhea/[child-years*B_1] + # confirmed *Shigella* dysentery/[child-years*B_2]).
[b] The number of confirmed *Shigella* watery diarrhea cases per 100 child-years, adjusted for care-seeking and enrollment plus the number of confirmed *Shigella* dysentery cases per 100 child-years at risk, adjusted for care-seeking and enrollment (# confirmed *Shigella* watery diarrhea/[child-years*A_1*B_1] + # confirmed *Shigella* dysentery/[child-years*A_2*B_2]). As the healthcare seeking adjustment is further stratified beyond dysentery/watery diarrhea, the healthcare seeking adjustment is not shown in this table and is shown in supplementary table 1.
[c] Includes isolates from rectal swabs transported in mBGS or Cary-Blair media.
[d] Totals are the sum of country-level estimates weighted to the proportion of child-years contributing to the combined denominator.
[e] Defined as an ipaH cycle threshold (CT) < 31.1.

Statistical Analysis of Secondary Aims

1. ***Secondary Aim #1:*** Determine the incidence of *Shigella* by serotype, severity definition, laboratory method (culture vs. qPCR), age, and by season.
    a. *Statistical Analysis:* Enrollment- and healthcare seeking-adjusted *Shigella* incidence following the protocol incidence calculation as measured by culture and qPCR will be stratified by the following factors (**Tables 4-5, supplementary table 1**):
        i. **Serogroup and Serotype**:
            1. *Shigella* species (*S. boydii*, S. *dysentariae, S. flexneri, S. sonnei* or undetermined), *S. flexneri* serotype (1a, 1b, 1d, 2a, 2b, 3a, 3b, 4a, 4b, 5a, 5b, 6, 7a, X, Y, non-typable and other) and combined bivalent and quadrivalent vaccine target indicators (defined as *Shigella* positive for *S. flexneri* 2a or *S. sonnei, and S. flexneri* 2a, 3a or 6 or *S. sonnei*, respectively).
            2. The denominator for this sub-analysis will be the overall denominator for the primary endpoint.
        ii. **Age (Table 5)**:

1. Stratified by age at enrollment visit (6-11 months, 12-17 months, 18-23 months or 24-35 months).
2. The denominator for this sub-analysis will be the estimated child-years at risk in the catchment area for each age stratification.

iii. **Diarrhea severity definition:** *Shigella* incidence will be stratified by the following diarrhea severity definitions (**Table 5**):

1. **Dysentery vs. watery diarrhea** will be defined according to whether or not dysentery was present during the diarrheal episode
2. **Hospitalized diarrhea** defined as the child being admitted to hospital (overnight stay) during the diarrheal episode
3. **Modified Vesikari** Score (MVS) defined as in PATH Vesikari Clinical Severity Scoring System Manual[8]: Duration of diarrhea: 1-4 days (1 point), 5 days (2 points) ≥6 days (3 points); Max # of stool in 24 hour period: 1-3 (1 point), 4-5 (2 points), ≥6 (3 points); Duration of vomiting: 1 day (1 point), 2 days (2 points), ≥3 days (3 points); max # of vomiting episodes in 24 hour period: 1 (1 point), 2-4 (2 points), ≥5 (3 points); Axillary temperature 36.6-37.9°C (1 point), 38.0-38.4°C (2 points), ≥38.5°C (3 points); dehydration 1-5% (2 points), ≥6% (3 points); treatment: rehydration (1 point), hospitalization (2 points). The scores will be summed and categorized as severe illness (11+ points), moderate illness (7-10 points), mild illness (0-6 points). If pre-post rehydration weights are not known to calculate % dehydration, WHO dehydration categories of some and severe as per IMCI guidelines will be used.
4. **Moderate or severe diarrhea by MVS or dysentery** will be defined as a MVS of 9+ points or presence of visible blood in stool (Pavlinac, Vaccines, 2022)[9]. Secondary analyses will use a more stringent cut-off of 11 points and a less stringent cut-off of 7 points.
5. **GEMS MSD and LSD** will be defined as in Kotloff, Lancet GH, 2019.[6] Moderate-to-severe diarrhea (MSD) defined as presenting to a health facility with diarrhea and severe or some dehydration (by WHO criteria), visible blood in stool, or inpatient admission. Less-severe-diarrhea (LSD) defined as presenting to a health facility without MSD.
6. **GEMS-*Shigella*** will be defined as in Pavlinac, CID, 2021.[7] Duration of diarrhea through day of presentation: 1-3 days (0 points), 4-5 days (2 points), ≥6 days (3 points); WHO-defined dehydration categories: severe (8 points), some (4 points), none (0 points); inpatient admission (5 points). The scores will be summed and categorized as mild (<6 points), moderate (6-8 points), and severe (9+ points).
7. **Clark score** will be defined as in Clark, JID, 1988[10]. Duration of diarrhea: 1-4 days (1 point), 5-7 days (2 points), >7 days (3 points); Max # of stool in 24 hour period: 2-4 (1 point), 5-7 (2 points), >7 (3 points); Duration of vomiting: 2 days (1 point), 3-5 days (2 points), >5 days (3 points); max # of vomiting episodes in 24 hour period: 1-3 (1 point), 4-6 (2 points), >6 (3 points); Duration of reported fever: 1-2 days (1 point), 3-4 days (2 points), ≥5 days (3 points); Rectal temperature 38-38.2°C (1 point), 38.3-38.7°C (2 points), ≥38.8°C (3 points); behavioral signs: Irritable/less playful (1 point), Lethargic/listless (2 points), Seizures (3 points). The scores will be summed and categorized as mild (2-8 points), moderate to severe (9+ points).
8. **MAL-ED score** will be defined as in Lee, J Pediatr Gastroenterol Nutr., 2016[11]: Duration of diarrhea: 2-4 days (1 point), 5-7 days (2 points), ≥8

days (3 points); Max # of stool in 24 hour period: <5 loose stools/24 hours (1 point), 5-7 loose stools/24 hours (2 points), ≥8 stools/24 hours (3 points); Duration of vomiting: 1 day (1 point), 2 days (2 points), ≥3 days (3 points); duration of reported fever: 1+ days (1 point); Confirmed temperature: ≥37.5°C (confirmed by field worker) (2 points); Dehydration: Some (2 points), severe (3 points). The scores will be summed and categorized as non-severe (<6 points) or severe (6+ points)

9. **Modified CODA** will be defined as in Lee, BMJ Open, 2014.[12] Max # of stool in 24 hour period: 4-5 stools (1 point), 6-7 stools (2 points); Duration of vomiting: 1-2 days (1 points), 3-4 days (2 points), ≥5 days (3 points); Duration of reported fever: 1-2 days (1 point), 3-4 days (2 points), ≥5 days (3 points); Dehydration duration: 1-2 days (1 point), 3-4 days (2 points), ≥5 days (3 points); Anorexia: 1-2 days (1 point), 3-4 days (2 points), ≥5 days (3 points) ≥8 stools (3 points). The scores will be summed and categorized as mild (0 points), moderate (1-6 points), severe (+7 points).
10. The denominator for each subgroup of this sub-analysis will be the overall child-years at risk of *Shigella* in the catchment area (same as the primary analysis). Participants who are missing one or more component of scores will be presented as "Unknown".

iv. **Diarrhea season**:

1. Seasonality will be presented graphically (**Figure 1**, not shown) as monthly or quarterly incidence.
2. The denominator for this sub-analysis will be the estimated child-years at risk in the catchment area for each study month or quarter.

The adjusted incidence rate and 95% confidence interval will be reported for each stratification both overall and by country.

Table 4. **Adjusted *Shigella* incidence stratified by *Shigella* species and serotype.**

		Enrollment- and Healthcare Seeking-Adjusted Shigella *MAD Incidence Rate (CI)*															
		Bangladesh		*Kenya*		*Malawi*		*Mali*		*Pakistan*		*Peru*		*The Gambia*		*Total*	
	Serotype	*Culture*[a]	*TAC*[b]	*Culture*[a]	*TAC*[b]	*Culture*[a]	*TAC*[b]	*Culture*[a]	*TAC*[b]	*Culture*[a]	*TAC*[b]	*Culture*[a]	*TAC*[b]	*Culture*[a]	*TAC*[b]	*Culture*[a]	*TAC*[b]
S. flexneri and *S. sonnei*	*S. flexneri*																
	1a																
	1b																
	1d																
	2a																
	2b																
	3a																
	3b																
	4a																
	4b																
	5a																
	5b																
	6																
	7a	-		-		-		-		-		-		-		-	
	X																
	Y																
	Non-typable																
	S. sonnei																
	Bivalent vaccine targets[c]																
	Quadrivalent vaccine targets[d]																
Other	*S. boydii*	-		-		-		-		-		-		-		-	
	S. dysentariae	-		-		-		-		-		-		-		-	
	Undetermined[e]																

CI: 95% confidence interval, CT: cycle threshold, TAC: TaqMan array card.

[a] *S. flexneri* 7a is not assessed by culture.

[b] Defined as an ipaH cycle threshold (CT) < 31.1. Data for Malawi is not complete and will be shown in future reports. *S. boydii* and *S. dysentariae* are not assessed by TAC.

[c] *S. flexneri* 2a or *S. sonnei.*

[d] *S. flexneri* 2a, 3a, or 6 or *S. sonnei*.

[e] Undetermined by culture means any serotypes/subserotypes not listed in the table. Undetermined by TAC means *Shigella* detected by PCR but molecular criteria to assign species/serotype to *S. sonnei* or *S. flexneri* was not met.

Table 4. Adjusted *Shigella* incidence stratified by participant age and diarrhea severity definition.

	Enrollment- and Healthcare Seeking-Adjusted Shigella *MAD Incidence Rate (CI)*															
	Bangladesh		*Kenya*		*Malawi*		*Mali*		*Pakistan*		*Peru*		*The Gambia*		*Total*	
Strata	*Culture*	*TAC*[a]	*Culture*	*TAC*[a]	*Culture*	*TAC*[a]	*Culture*	*TAC*[a]	*Culture*	*TAC*[a]	*Culture*	*TAC*[a]	*Culture*	*TAC*[a]	*Culture*	*TAC*[a]
Age at enrollment (months)																
6 - 11																
12 - 17																
18 - 23																
24 – 35																
Diarrhea severity classifications																
Dysentery																
Watery diarrhea																
Hospitalized																
Modified Vesikari score (MVS)[b]																
Mild illness (0-8 points)																
Moderate illness (9-10 points)																
Severe illness (11+ points)																
Moderate or severe diarrhea by MVS or dysentery[c]																
GEMS[d]																
MSD																
LSD																
GEMS-*Shigella* score[e]																
Mild (<6 points)																
Moderate (6-8 points)																
Severe (9+ points)																
Clark score[f]																
Mild (2-8 points)																
Moderate to severe (9+ points)																
MAL-ED score[g]																
Non-severe (<6 points)																
Severe (6+ points)																
CODA score[h]																
Mild (0 points)																
Moderate to severe (1+ points)																

CI: 95% confidence interval, CT: cycle threshold, LSD: less-severe-diarrhea, MAD: medically-attended diarrhea, MSD: moderate-to-severe diarrhea, MVS: modified Vesikari score, TAC: TaqMan array card, WHO: World Health Organization.

[a] Defined as an ipaH cycle threshold (CT) < 31.1. *S. boydii* and *S. dysentariae* are not assessed by TAC.

[b] Defined as in PATH Vesikari Clinical Severity Scoring System Manual: Duration of diarrhea: 1-4 days (1 point), 5 days (2 points) ≥6 days (3 points); Max # of stool in 24 hour period: 1-3 (1 point), 4-5 (2 points), ≥6 (3 points); Duration of vomiting: 1 day (1 point), 2 days (2 points), ≥3 days (3 points); max # of vomiting episodes in 24 hour period: 1 (1 point), 2-4 (2 points), ≥5 (3 points); Axillary temperature 36.6-37.9°C (1 point), 38.0-38.4°C (2 points), ≥38.5°C (3 points); dehydration 1-5% (2 points), ≥6% (3 points); treatment: rehydration (1 point), hospitalization (2 points).

[c] Defined as in Pavlinac, Vaccines, 2022 as a MVS of 9+ or presence of visible blood in stool.

[d] Defined as in Kotloff, Lancet GH, 2019. Moderate-to-severe diarrhea (MSD) defined as presenting to a health facility with diarrhea and severe or some dehydration (by WHO criteria), visible blood in stool, or inpatient admission. Less-severe-diarrhea (LSD) defined as presenting to a health facility without MSD.

[e] Defined as in Pavlinac, CID, 2021. Duration of diarrhea through day of presentation: 1-3 days (0 points), 4-5 days (2 points), ≥6 days (3 points);WHO-defined dehydration categories: severe (8 points), some (4 points), none (0 points); inpatient admission (5 points).

[f] Defined as in Clark, JID, 1988. Duration of diarrhea: 1-4 days (1 point), 5-7 days (2 points) >7 days (3 points); Max # of stool in 24 hour period: 2-4 (1 point), 5-7 (2 points), >7 (3 points); Duration of vomiting: 2 days (1 point), 3-5 days (2 points), >5 days (3 points); max # of vomiting episodes in 24 hour period: 1-3 (1 point), 4-6 (2 points), >6 (3 points); Duration of reported fever: 1-2 days (1 point), 3-4 days (2 points), ≥5 days (3 points); Rectal temperature 38-38.2°C (1 point), 38.3-38.7°C (2 points), ≥38.8°C (3 points); behavioral signs: Irritable/less playful (1 point), Lethargic/listless (2 points), Seizures (3 points).

[g] Defined as in Lee, J Pediatr Gastroenterol Nutr., 2016: Duration of diarrhea: 2-4 days (1 point), 5-7 days (2 points), ≥8 days (3 points); Max # of stool in 24 hour period: <5 loose stools/24 hours (1 point), 5-7 loose stools/24 hours (2 points), ≥8 stools/24 hours (3 points); Duration of vomiting: 1 day (1 point), 2 days (2 points), ≥3 days (3 points); duration of reported fever: 1+ days (1 point); Confirmed temperature: ≥37.5°C (confirmed by field worker) (2 points); Dehydration: Some (2 points), severe (3 points).

[h] Defined as in Lee, BMJ Open, 2014: Max # of stool in 24 hour period: 4-5 stools (1 point), 6-7 stools (2 points); Duration of vomiting: 1-2 days (1 points), 3-4 days (2 points), ≥5 days (3 points); Duration of reported fever: 1-2 days (1 point), 3-4 days (2 points), ≥5 days (3 points); Dehydration duration: 1-2 days (1 point), 3-4 days (2 points), ≥5 days (3 points); Anorexia: 1-2 days (1 point), 3-4 days (2 points), ≥5 days (3 points) ≥8 stools (3 points).

Supplementary Table 1. Adjusted incidence of relevant vaccine *Shigella* serotypes and species by diarrhea severity.

	Enrollment- and Healthcare Seeking-Adjusted Shigella *MAD Incidence Rate (CI)*															
	Bangladesh		*Kenya*		*Malawi*		*Mali*		*Pakistan*		*Peru*		*The Gambia*		*Total*	
Strata	*Culture*	*TAC*[a]	*Culture*	*TAC*[a]	*Culture*	*TAC*[a]	*Culture*	*TAC*[a]	*Culture*	*TAC*[a]	*Culture*	*TAC*[a]	*Culture*	*TAC*[a]	*Culture*	*TAC*[a]
Bivalent vaccine targets (*S. flexneri* 2a & *S. sonnei*)																
GEMS[b]																
LSD																
MSD																
MVS moderate or severe diarrhea or dysentery[c]																
Quadrivalent vaccine targets (*S. flexneri* 2a, 3a & 6 & *S. sonnei*)																
GEMS[b]																
LSD																
MSD																
MVS moderate or severe diarrhea or dysentery[c]																

CI: 95% confidence interval, CT: cycle threshold, GEMS: Global Enteric Multicenter Study, LSD: less-severe-diarrhea, MSD: moderate-to-severe-diarrhea, MVS: modified Vesikari score, TAC: TaqMan array card, WHO: World Health Organization.

[a] Defined as an ipaH cycle threshold (CT) < 31.1.

[b] Defined as in Kotloff, Lancet GH, 2019. Moderate-to-severe diarrhea (MSD) defined as presenting to a health facility with diarrhea and severe or some dehydration (by WHO criteria), visible blood in stool, and/or inpatient admission. Less-severe-diarrhea (LSD) defined as presenting to a health facility without MSD.

[c] Defined as in Pavlinac, Vaccines, 2022 as a MVS of 9+ or presence of visible blood in stool.

2. ***Secondary Aim #2:*** *Describe the prevalence of resistance to commonly used antibiotics in* Shigella *isolates in each EFGH country site.*
    a. *Statistical Analysis:* The proportion of *Shigella* culture positive stool samples that are resistant to the medications listed below will be calculated by EFGH region and EFGH study site (**Table 6**):
        - Ampicillin
        - Azithromycin
        - Ceftriaxone
        - Ciprofloxacin
        - Nalidixic acid
        - Pivemicellinam
        - Trimethoprim-sulfamethoxazole

        Resistance will be defined as "non-susceptibility", or resistant or intermediate zone size classifications according to the most recent Clinical and Laboratory Standards Institute (CLSI) guidelines at the time of analysis. Resistance will be computed for all of the above antibiotics for all sites, by region (South America [Peru], West Africa [The Gambia and Mali], East Africa [Kenya and Malawi], and Asia [Pakistan and Bangladesh]), and among *S. sonnei* positive samples and *S. flexneri* positive samples. Additionally, we will present the percentage of multi-drug resistant *Shigella* (MDR, resistant to three or more antibiotics), extensively drug resistant *Shigella* (XDR, resistant to all of the following: Azithromycin, Ciprofloxacin, Ceftriaxone, Trimethoprim-sulfamethoxazole, and Ampicillin), and resistance to all WHO recommended treatments (resistant to all of the following: Azithromycin, Ciprofloxacin, and Ceftriaxone).

    b. Confidence intervals around non-susceptibility proportions will be determined assuming a binomial distribution.

Table 6. Country- and region-specific antibiotic non-susceptibility to culture-confirmed *Shigella* isolates

	Antibiotic non-susceptibility: % (CI)										
Antibiotic	**Kenya**	**Malawi**	**East Africa total**	**Mali**	**The Gambia**	**West Africa total**	**Bangladesh**	**Pakistan**	**South Asia total**	**Peru**	**Total**
Ampicillin											
All *Shigella* isolates											
S. flexneri											
S. sonnei											
Azithromycin											
All *Shigella* isolates											
S. flexneri											
S. sonnei											
Ceftriaxone											
All *Shigella* isolates											
S. flexneri											
S. sonnei											
Ciprofloxacin											
All *Shigella* isolates											
S. flexneri											
S. sonnei											
Nalidixic acid											
All *Shigella* isolates											
S. flexneri											
S. sonnei											
Pivemicellinam											
All *Shigella* isolates											
S. flexneri											
S. sonnei											
Trimethoprim-sulfamethoxazole											
All *Shigella* isolates											
S. flexneri											
S. sonnei											
Resistance to all WHO recommended treatments[1]											
All *Shigella* isolates											
S. flexneri											
S. sonnei											
Multidrug resistance (MDR)[2]											
All *Shigella* isolates											
S. flexneri											
S. sonnei											
Extensive drug resistance (XDR)[3]											
All *Shigella* isolates											
S. flexneri											
S. sonnei											

CI: 95% confidence interval, MDR: multidrug resistance, XDR: extensive drug resistance, WHO: World Health Organization.
[1] WHO recommended treatments include Azithromycin, Ciprofloxacin, and Ceftriaxone.
[2] Resistant to any three or more antibiotics.
[3] Resistant to all of the following: azithromycin, ciprofloxacin, ceftriaxone, trimethoprim-sulfamethoxazole, and ampicillin.

3. ***Secondary Aim #3:*** *Determine the risk of all-cause mortality, hospitalization, persistent diarrhea, diarrhea recurrence, and change in linear growth in the three months following an episode of MAD.*
    a. <u>*Statistical Analysis:*</u>

    The cumulative incidence of all-cause mortality will be plotted using the Kaplan-Meier survival function. Participants will be censored at date of death, last known alive date, or latest follow-up date. Cox proportional hazards regression models adjusted for site and age will be used to estimate hazard ratios (HR) and 95% confidence intervals (CIs) for the associations between *Shigella*-attributed vs. non-*Shigella*-attributed MAD, and other sociodemographic and clinical covariates, with all-cause mortality. In a supplementary analysis, Kaplan-Meier plots and Cox proportional hazards regression will be used to model all-cause mortality by species of *Shigella*-attributed MAD. Risk ratios (RRs) will be calculated when sample sizes are small. For descriptive purposes, we will present the overall and site-specific prevalence of cause-specific deaths, including diarrheal diseases, acute respiratory infections and pneumonia, malaria, and severe malnutrition.

    Similar to the methods above, Kaplan-Meier survival analysis and Cox proportional hazards models will be used to determine the cumulative risk of first admission to hospital during the three-month follow-up period. Supplementary analysis will use an Anderson-Gil model to evaluate recurrent hospitalizations. Hospitalizations occurring as part of the initial diarrhea episode (for which the child was recruited) will be excluded from this analysis.

    Generalized linear mixed-effects regression models with random intercepts and slopes for time will be used to determine the influence of *Shigella*-associated MAD and sociodemographic and clinical characteristics on mean ΔHAZ in the three-month follow-up period among surviving children. We will compare the mean ΔHAZ in children with *Shigella* MAD to non-*Shigella* MAD. Univariate models will be adjusted for site, age, and enrollment HAZ (baseline). A supplementary table will include the mean (95% CI) HAZ at enrollment and the mean ΔHAZ from enrollment to three-months by age for each site. In the subset of children not stunted (HAZ $\geq$-2) at recruitment, we will determine the incidence of stunting among children with and without *Shigella* MAD. The cumulative incidence of stunting will be plotted using the Kaplan-Meier survival function.

    The proportion of children with diarrhea or dysentery lasting 14 or more days after the beginning of the episode will be compared between children with and without *Shigella*-attributed diarrhea using log-binomial regression. Multivariable models will include adjustment for site and age.

    The incidence of new diarrhea episodes will be defined as total number of new diarrhea episodes (>48 hours after a diarrhea-free period), divided by the child-time at risk during the three-month follow-up period. Time at risk will be censored at the date of last follow-up for children who have died or are lost to follow-up. Child-time at risk during the enrollment MAD episode, and the 48 hours after, will not be included in the denominator. The RR comparing *Shigella* attributed to non-*Shigella* attributed episodes will be determined using a Poisson model with number of episodes as the outcome and time at risk (defined above) as the model offset and Wald chi-square tests of the two-way comparison. We additionally stratify results by whether or not the diarrhea was medically-attended. Multivariable models will include adjustment for site and age.

4. ***Secondary Aim #4:*** *Compare various severity definitions in their ability to distinguish* Shigella *from non-*Shigella *attributable diarrhea and ability to predict risk of death or hospitalization in the subsequent three months.*

   Based on signs and symptoms collected from EFGH cases, we will construct diarrhea severity definitions as described above in secondary aim 1. Scores with be dichotomized into moderate or severe vs. mild diarrhea. We will calculate proportions of children with *Shigella*-attributed diarrhea by severity score categories. We will calculate the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of severe vs. non-severe diarrhea (by each definition) in identifying children with *Shigella*-attributed diarrhea. We will also calculate sensitivity, specificity, PPV, and NPV of severe vs. non-severe diarrhea (by each definition) in identifying children who die or were hospitalized during the three-month follow-up period. In secondary analyses, we will explore varying cut-offs used for score-based severity definitions by conducting Receiver Operating Characteristics (ROC) analyses to identify cut-points that maximally distinguish those children who had experienced a death or hospitalization from those that did not.

5. ***Secondary Aim #5****: Quantify the cost incurred by families and health care systems due to* Shigella *morbidity and mortality.*

For each child enrolled in the study, we will collect data on medical and non-medical resources used to treat the episode of diarrhea, using child medical records and standardized questionnaires administered to caregivers. *Direct medical costs* include medical resources used in the treatment of diarrhea. We will estimate the costs of drugs and diagnostics used in diarrhea management, by applying a standardized country-specific unit cost from National Price Lists to each treatment administered or prescribed in medical records. In situations where data is missing from National Price Lists, we will estimate unit costs using a mean value from local pharmacies and health facilities. We will also estimate service delivery costs: operational costs for medical visits, such as health workers' salaries or maintenance of facility equipment. To estimate these costs, we will apply standardized country-specific unit costs from WHO-CHOICE[13] to each visit based on the level of care (inpatient vs. outpatient) and number of days admitted (inpatient only). Standardized questionnaires include questions regarding whether medical resources were paid for by the caregiver or funded by the public health system. *Direct non-medical costs* include non-medical fees incurred by caregivers, such as costs to travel to the health facility. We will estimate non-medical fees through caregiver direct reports from standardized questionnaires. *Indirect (non-medical) costs* include caregiver time lost from work while caring for their sick child. To estimate indirect costs, we will apply a national average wage rate from the International Labor Organization (https://www.ilo.org) to caregiver-reported time lost from work from standardized questionnaires.

We will estimate the average cost per episode of *Shigella*-attributable MAD, referred to henceforth as the "cost per episode", by averaging costs across *Shigella*-positive cases, with analyses conducted separately using both culture-confirmed *Shigella* and *Shigella* confirmed by molecular methods (TAC). We will present the mean and standard deviation of costs per episode, from the household perspective (including out-of-pocket medical and non-medical fees incurred by caregivers), the health system perspective (medical resources funded by the public health system), and the societal perspective (both household and health system costs). We will stratify societal costs by visit characteristics, as described in Table 6. Costs will be presented in 2023 USD to allow for comparability across sites. We will inflate costs incurred in earlier years to 2023 values using GDP price deflators from World Bank and convert local currencies to equivalent USD values using midyear currency values.

To identify the effects of visit characteristics on costs per episode, we will use multivariate Generalized Linear Models (GLMs). GLMs require explicit specification of the distribution of the dependent variable and the link function describing how independent variables are functionally related to the dependent variable. We will use the modified Park test[14], common to health econometrics, to identify the GLM distribution and link function. Our final model will be determined from the modified Park test results but will be initially tested using a gamma distribution with log link:

$$\ln(Y_i) = \beta_0 + \beta_1 Country_i + \beta_2 Level_i + \beta_3 Inpatient_i + \beta_4 Severity_i$$

where the cost per episode (Y) for a given child, i, is a function of the country in which the child was treated (*Country)*, the level of health facility treated at (*Level*); whether the child was admitted as an inpatient or treated as an outpatient (*Inpatient),* and the severity of diarrhea *(Severity)*. Models will be run separately per costing perspective, and separately for children with *Shigella* confirmed by culture and TAC. Coefficients, confidence intervals, and p-values will be presented for all independent variables (Table 7).

We will estimate the annual economic burden of *Shigella* MAD in each EFGH country site (c). The estimated incidence of *Shigella* MAD and the estimated population sizes of children 6-35 months (converted to average person-years per site, per year) from the primary aim will be used to estimate the

number of annual *Shigella* MAD episodes per country site. We will then apply the average cost per episode of *Shigella* MAD to the number of estimated episodes per year. The annual economic burden will be calculated separately per costing perspective and by *Shigella* confirmation method (Table 8).

$$\text{Annual } \textit{Shigella} \text{ MAD costs}_c = \frac{\textit{Shigella} \text{ MAD episodes}_c}{100 \text{ person years}} \times \text{person years}_c \times \text{cost per episode }_c$$

Table 7. Cost per episode of medically attended *Shigella* diarrhea, stratified by visit characteristics.

	Cost per episode, USD 2023: Mean (SD)															
	Bangladesh		*Kenya*		*Malawi*		*Mali*		*Pakistan*		*Peru*		*The Gambia*		*Total*	
Strata	*Culture*[1]	*TAC*	*Culture*	*TAC*	*Culture*	*TAC*	*Culture*	*TAC*	*Culture*	*TAC*	*Culture*	*TAC*	*Culture*	*TAC*	*Culture*	*TAC*
Costing perspective																
Household[2]																
Health system[3]																
Societal[4]																
Visit characteristics[5]																
Visit type																
Inpatient																
Outpatient																
Level of facility treated at																
Primary																
Secondary																
Tertiary																
Sex																
Female																
Male																
Age at enrollment (months)																
6—11																
12—17																
18—23																
24—35																
Dysentery																
Watery diarrhea																
GEMS[6]																
Less severe																
Moderate to severe																
Severity by Modified Vesikari score (MVS)[7]																
Mild (0-8 points)																
Moderate (9-10 points)																
Severe (11+ points)																
MVS moderate or severe diarrhea and/or dysentery[8]																

[1] Children may be represented in both Culture and TAC columns if they tested positive for *Shigella* with both diagnostics.
[2] Household perspective includes caregiver out-of-pocket costs and indirect costs (caregiver time lost from work).
[3] Health system perspective includes costs of medical treatment funded by the public health system.

[4] Societal costs include all costs presented in the household and health system perspectives.
[5] Costs stratified per visit characteristic category are societal costs.
[6] Defined as in Kotloff, Lancet GH, 2019. Moderate-to-severe-diarrhea (MSD) defined as presenting to a health facility with diarrhea and severe or some dehydration (by WHO criteria), visible blood in stool, or inpatient admission. Less-severe-diarrhea (LSD) defined as presenting to a health facility without MSD.
[7]Defined as in PATH Vesikari Clinical Severity Scoring System Manual: Duration of diarrhea: 1-4 days (1 point), 5 days (2 points) ≥6 days (3 points); Max # of stool in 24 hour period: 1-3 (1 point), 4-5 (2 points), ≥6 (3 points); Duration of vomiting: 1 day (1 point), 2 days (2 points), ≥3 days (3 points); max # of vomiting episodes in 24 hour period: 1 (1 point), 2-4 (2 points), ≥5 (3 points); Axillary temperature 36.6-37.9°C (1 point), 38.0-38.4°C (2 points), ≥38.5°C (3 points); dehydration 1-5% (2 points), ≥6% (3 points); treatment: rehydration (1 point), hospitalization (2 points). N=46 individuals could not be assessed due to missing diarrhea duration data.
[8]Defined as in Pavlinac, Vaccines, 2022 as a modified Vesikari score of 9+ or presence of visible blood in stool.

Table 8. Annual costs of medically-attended *Shigella* diarrhea in country sites.

	Annual costs per site, USD 2023													
	Bangladesh		*Kenya*		*Malawi*		*Mali*		*Pakistan*		*Peru*		*The Gambia*	
Strata	*Culture*[1]	*TAC*	*Culture*	*TAC*	*Culture*	*TAC*	*Culture*	*TAC*	*Culture*	*TAC*	*Culture*	*TAC*	*Culture*	*TAC*
Household costs[2]														
Health system costs[3]														
Societal costs[4]														

[1] Children may be represented in both Culture and TAC columns if they tested positive for *Shigella* with both diagnostics.
[2] Household costs includes caregiver out-of-pocket costs and indirect costs (caregiver time lost from work).
[3] Health system costs includes costs of medical treatment funded by the public health system.
[4] Societal costs include all costs presented in the household and health system perspectives.

6. ***Secondary Aim #6****: To identify optimal laboratory methods for Shigella culture by:*

    a. *comparing the isolation rate of* Shigella *between two transport media for rectal swabs (Cary-Blair and modified BGS)*

    We will calculate the proportion of rectal swab samples from Cary-Blair and mBGS media from which *Shigella* was isolated overall and per site as well as stratified by serotype (**Table 9)**. A two-sided 95% confidence intervals around each proportion will be calculated assuming a Binomial distribution. Proportions will be compared using a McNemar's test of superiority to determine if one media is superior to the other in terms of *Shigella* isolation rates.

    b. *comparing the isolation rate of* Shigella *between two fecal sample types (rectal swabs and whole stool) among the subset of children who produced whole stool in the Gambia and Bangladesh country sites.*

    We will calculate the proportion of rectal swab samples transported with Cary-Blair from which *Shigella* was isolated and the proportion of whole stool samples transported with Cary-Blair from which *Shigella* was isolated, overall and per site (**Table 10)**. A two-sided 95% confidence interval for the absolute difference in proportions will be calculated assuming a Binomial distribution. When assessing non-inferiority, the lower bound of the 95% confidence interval will be compared to the non-inferiority margin.

Table 9. Shigella culture positivity proportion by transport media type and EFGH study site.

		***Shigella* culture positivity[a]: % (CI)**		
Country	**N**	**Cary Blair**	**mBGS**	**p-value[b]**
Bangladesh				
Kenya				
Malawi				
Mali				
Pakistan				
Peru[c]				
The Gambia				
Overall				

CI: 95% confidence interval, mBGS: modified buffered glycerol solution.

[a] Culture *Shigella* culture positivity is defined as the percentage of enrolled participants with *Shigella* identified out of total participants with both rectal swabs cultured using both Cary-Blair and mBGS transport media.

[b] McNemar's test of superiority.

[c] The only participants included in this table are those enrolled after January 19, 2023 due to a laboratory error that made previous results incomparable across media types.

Table 10. *Shigella* culture-positivity from rectal swab and whole stool samples, across sites involved in the whole stool/rectal swab comparison sub-study.

		***Shigella* culture positivity[a]: % (CI)**		
Country	**N**	**Rectal swab[b]**	**Whole stool[c]**	**Difference**
Bangladesh				
The Gambia				
Overall				

CI: 95% confidence interval.

[a] *Shigella* culture positivity is defined as the percentage of participants with *Shigella* identified out of total participants with both rectal swab and whole stool collected for the *Shigella* culture comparison.

[b] Includes only *Shigella* isolates from rectal swabs in Cary-Blair media.

[c] Whole stool for culture placed in Cary-Blair media. Whole stool is collected among children who produce a sample while still at the enrollment facility.

Statistical Software
All analyses will be conducted using STATA or R and the software used reported in all analysis write-ups.

References
References to be Provided for Non-standard Statistical Methods
All methods being proposed are standard.

Data Management Plan
Procedures relating to data entry, management, and QA/QC are outlined in the EFGH Data Management Plan.

Statistical Master File
The Statistical Master File is maintained by the Data Lead

Referenced Literature

1. Rogawski McQuade ET, Shaheen F, Kabir F, et al. Epidemiology of Shigella infections and diarrhea in the first two years of life using culture-independent diagnostics in 8 low-resource settings. *PLoS Negl Trop Dis* 2020; **14**(8): e0008536.
2. Nasrin D, Blackwelder WC, Sommerfelt H, et al. Pathogens associated with linear growth faltering in children with diarrhea and impact of antibiotic treatment: The Global Enteric Multicenter Study. *J Infect Dis* 2021.
3. Anderson JDt, Bagamian KH, Muhib F, et al. Burden of enterotoxigenic Escherichia coli and shigella non-fatal diarrhoeal infections in 79 low-income and lower middle-income countries: a modelling analysis. *Lancet Glob Health* 2019; **7**(3): e321-e30.
4. Integrated management of childhood illness: conclusions. WHO Division of Child Health and Development. *Bull World Health Organ* 1997; **75 Suppl 1**: 119-28.
5. Pavlinac PB, Rogawski McQuade ET, Platts-Mills JA, et al. Pivotal Shigella Vaccine Efficacy Trials—Study Design Considerations from a Shigella Vaccine Trial Design Working Group. *Vaccines* 2022; **10**(4): 489.
6. Kotloff KL, Nasrin D, Blackwelder WC, et al. The incidence, aetiology, and adverse clinical consequences of less severe diarrhoeal episodes among infants and children residing in low-income and middle-income countries: a 12-month case-control study as a follow-on to the Global Enteric Multicenter Study (GEMS). *Lancet Glob Health* 2019; **7**(5): e568-e84.
7. Pavlinac PB, Platts-Mills JA, Tickell KD, et al. The Clinical Presentation of Culture-positive and Culture-negative, Quantitative Polymerase Chain Reaction (qPCR)-Attributable Shigellosis in the Global Enteric Multicenter Study and Derivation of a Shigella Severity Score: Implications for Pediatric Shigella Vaccine Trials. *Clin Infect Dis* 2021; **73**(3): e569-e79.
8. PATH. Vesikari Clinical Severity Scoring System Manual, 2011.
9. Pavlinac PB, Rogawski McQuade ET, Platts-Mills JA, et al. Pivotal Shigella Vaccine Efficacy Trials-Study Design Considerations from a Shigella Vaccine Trial Design Working Group. *Vaccines (Basel)* 2022; **10**(4).
10. 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**(3): 570-87.

11. Lee GO, Richard SA, Kang G, et al. A Comparison of Diarrheal Severity Scores in the MAL-ED Multisite Community-Based Cohort Study. *J Pediatr Gastroenterol Nutr* 2016; **63**(5): 466-73.
12. Lee G, Penataro Yori P, Paredes Olortegui M, et al. An instrument for the assessment of diarrhoeal severity based on a longitudinal community-based study. *BMJ Open* 2014; **4**(6): e004816.
13. Stenberg K, Lauer JA, Gkountouras G, Fitzpatrick C, Stanciole A. Econometric estimation of WHO-CHOICE country-specific costs for inpatient and outpatient health service delivery. *Cost Eff Resour Alloc* 2018; **16**: 11.
14. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? *J Health Econ* 2001; **20**(4): 461-94.