Vaccine 39 (2021) 4391–4398

Contents lists available at ScienceDirect

Vaccine

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

WHO Report

Global diarrhoea-associated mortality estimates and models in children: Recommendations for dataset and study selection

Egle Butkeviciute [a], Holly J. Prudden [b], Mark Jit [c], Peter G. Smith [c], Gagandeep Kang [d], Mark S. Riddle [e], Benjamin A. Lopman [f], Virginia E. Pitzer [g], Claudio F. Lanata [h], James A. Platts-Mills [i], Robert F. Breiman [j], Birgitte K. Giersing [b], Mateusz Hasso-Agopsowicz [b,*]

[a] Department of Infection Biology, London School of Hygiene & Tropical Medicine, London, United Kingdom
[b] Immunization, Vaccines and Biologicals, World Health Organization, Geneva, Switzerland
[c] Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, United Kingdom
[d] The Wellcome Trust Research Laboratory Division of Gastrointestinal Sciences, Christian Medical College, Vellore, India
[e] University of Nevada Reno School of Medicine, Reno, Nevada, United States
[f] Department of Epidemiology, Emory University, Atlanta, United States
[g] Department of Epidemiology of Microbial Diseases, Yale School of Public Health, Yale University, New Haven, United States
[h] Instituto de Investigacion Nutricional, Lima, Peru
[i] Division of Infectious Diseases & International Health, University of Virginia, Charlottesville, United States
[j] Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, United States

ARTICLE INFO

Article history:
Received 20 May 2021
Accepted 25 May 2021
Available online 13 June 2021

Keywords:
Enteric pathogens
Burden
Vaccine value
Mortality estimates

ABSTRACT

Background: Multiple factors contribute to variation in disease burden, including the type and quality of data, and inherent properties of the models used. Understanding how these factors affect mortality estimates is crucial, especially in the context of public health decision making. We examine how the quality of the studies selected to provide mortality data, influence estimates of burden and provide recommendations about the inclusion of studies and datasets to calculate mortality estimates.
Methods: To determine how mortality estimates are affected by the data used to generate model outputs, we compared the studies used by The Institute of Health Metrics and Evaluation (IHME) and Maternal and Child Epidemiology Estimation (MCEE) modelling groups to generate enterotoxigenic *Escherichia coli* (ETEC) and *Shigella*-associated mortality estimates for 2016. Guided by an expert WHO Working Group, we applied a modified Newcastle-Ottawa Scale (NOS) to evaluate the quality of studies used by both modelling groups.
Results: IHME and MCEE used different sets of ETEC and *Shigella* studies in their models and the majority of studies were high quality. The distribution of the NOS scores was similar between the two modelling groups. We observed an overrepresentation of studies from some countries in SEAR, AFR and WPR compared to other WHO regions.
Conclusion: We identified key differences in study inclusion and exclusion criteria used by IHME and MCEE and discuss their impact on datasets used to generate diarrhoea-associated mortality estimates. Based on these observations, we provide a set of recommendations for future estimates of mortality associated with enteric diseases.

1. Introduction

Estimates of the global burden of morbidity and mortality are important to help inform decisions on vaccine prioritisation, development, introduction and use. Two modelling groups (The Institute of Health Metrics and Evaluation (IHME) and the Maternal and Child Epidemiology Estimation group (MCEE)) have historically reported mortality estimates for enterotoxigenic *Escherichia coli* (ETEC) and *Shigella,* with modelling results for the same year tending to show a degree of overlap within those estimates [1–3]. However, there is some variation in the estimates, both between the models and iterations within the models [2]. For 2013, MCEE estimated that 42,000 (20,000–76,000) under five year old (U5) ETEC deaths and 28,000 (12,000–53,000) U5 *Shigella*

* Corresponding author.
E-mail address: hassoagopsowiczm@who.int (M. Hasso-Agopsowicz).

https://doi.org/10.1016/j.vaccine.2021.05.086
0264-410X/

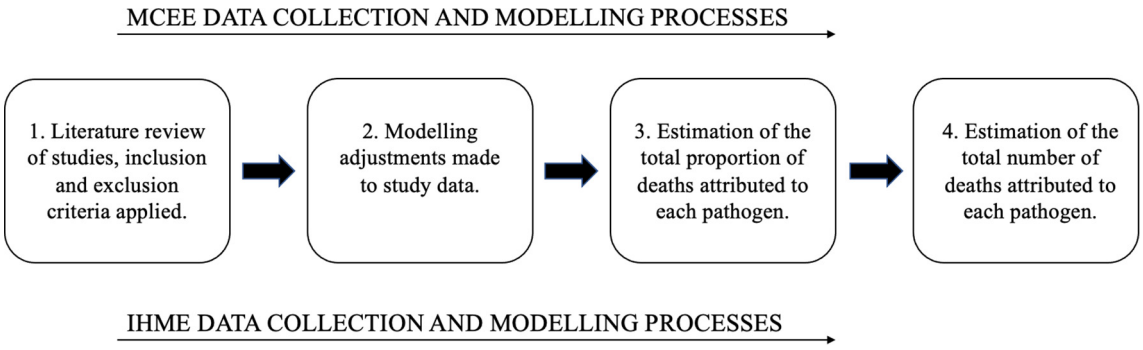

Fig. 1. Steps in IHME and MCEE data collection and processing to generate the estimates for U5 diarrhoea-associated mortality.

deaths occurred [4], with similar estimates calculated again for the year 2016 (data are currently unpublished). In 2013, IHME was more consistent with MCEE, estimating that there were 23,100 (17,000–30,400) ETEC- and 33,400 (24,900–43,500) *Shigella*-associated U5 deaths [3]. However, in 2016, modelling estimates for these pathogens by IHME began to show a divergence from those of MCEE, with 18,669 (9800–30,659) U5 ETEC deaths and 63,713 (41,191–93,611) U5 Shigella deaths [3,5]. Such variations in pathogen mortality estimates can have implications for future decision making on vaccine development funding, policy and uptake and therefore, must be more thoroughly investigated and understood.

The results presented in this article were generated to evaluate differences in the modelling outputs of IHME and MCEE groups for ETEC and *Shigella,* as recommended by the World Health Organization (WHO) Product Development for Vaccines Advisory Committee (PDVAC) in 2018 [7]. In this article, we look at the studies that are used to generate the estimates for the two pathogens by comparing inclusion criteria applied by both groups (Fig. 1, *Step 1*), investigating the type and quality of studies. We conclude our findings with a discussion on how studies should be selected for inclusion in modelling the burden of enteric diseases for future iterations of mortality estimates. We also present recommendations for investigators and modellers to further improve study design and outcome measures.

2. Methods

2.1. Comparison of studies used by IHME and MCEE to calculate U5 mortality estimates

We obtained the lists of studies used by MCEE and IHME to generate ETEC and *Shigella* U5 mortality estimates for 2016. We validated the included studies by re-applying the selection criteria used by the modelling groups. We cross-checked the included studies between both groups and noted studies that overlapped.

2.2. The assessment of the quality of studies

Following consultation by a WHO-appointed expert working group, we generated a grading template by using a modified version of the Newcastle-Ottawa Scale (NOS) to assess the quality of ETEC and Shigella studies included by IHME and MCEE for estimates of etiology-specific U5 mortality. The NOS was designed to evaluate the quality of non-randomised cohort or case-control studies based on three main criteria: (1) selection of the study population, (2) comparability of the study population and (3) methods to ascertain exposure in cohort studies or study outcome in case-control studies [8]. We modified the NOS to: (1) allow for the assessment of cross-sectional studies, (2) to account for the fact that few studies addressed the exposure to factors associated with risk for diarrhoeal disease, and (3) to account for the fact that most of the studies considered the presence of diarrhoeal pathogens as one of the study outcomes rather than exposure, hence predetermining diarrhoeal disease (Table S1).

The study quality scores were normalised and presented as a percentage of the maximum possible score. We excluded publications or datasets that could not be obtained at a level of data stratification required for the analysis, were unpublished at the time of conducting the analysis, or were published in non-English languages. The study design of those datasets generated by The Etiology, Risk Factors and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health and Development cohort study (MAL-ED) and the Global Enteric Multicenter Study (GEMS) that were unpublished at the time the ETEC and *Shigella* burden of disease estimates for 2016 were made, was graded using the following publications: *Platts-Mills et al., 2015* [9] and *Kotloff et al., 2019* [10]. We visualised the findings using GraphPad Prism 8 and ArcMap 10.5.

3. Results

3.1. Comparison of studies used by IHME and MCEE to calculate etiology-specific U5 mortality estimates

Comparison of the study selection criteria used by IHME and MCEE to generate model inputs for 2016 U5 mortality estimates found that seven out of ten study selection criteria were unique to a single group (Table 1). In addition, the groups analysed different sets of diarrhoea-associated pathogens. Both groups included selected pathogens that were associated with diarrhoea in GEMS qPCR re-analysis [12], however IHME also included *Clostridium difficile* (Table 1).

We found that IHME included 44, and MCEE included 7 studies to generate the ETEC mortality estimates for 2016, and that there was no overlap in the selection of ETEC studies included by both groups (Table 2). For *Shigella*, IHME included 61 studies, while MCEE selected 31 studies, with only one study used by both groups being in common (Table 2). This may be attributed to the periods from which the studies were selected: IHME used studies conducted in 2011–2016 and MCEE included studies published between 1990 and 2014, the time windows only overlapping for 2011–2014.

3.2. The assessment of the quality of studies

We used the modified NOS to examine the quality of studies used by both groups. We were unable to assess the quality of 14

Table 1
Overview of criteria used by MCEE and IHME for study selection for 2016 burden of enteric disease estimates.

Inclusion Criteria*	MCEE	IHME	Approach
Sample size	Must include at least 100 samples.		Same
Introduction of rotavirus vaccines	Includes studies conducted before the introduction of rotavirus vaccines.	Includes studies conducted before and after the introduction of rotavirus vaccines.	Different
Inclusion of pathogens	Includes selected pathogens that were associated with diarrhoea in GEMS qPCR re-analysis: Adenovirus, *Aeromonas*, Astrovirus, *Campylobacter*, *Cryptosporidium*, *Entamoeba*, EPEC, ST-ETEC, Norovirus, Rotavirus *Salmonella*, *Shigella*, *Sapovirus* *Vibrio cholerae*.	Includes selected pathogens that were associated with diarrhoea in the GEMS qPCR re-analysis: Adenovirus, *Aeromonas*, *Campylobacter*, *Cryptosporidium*, *Entamoeba*, EPEC, ETEC, Norovirus, Rotavirus, *Salmonella*, *Shigella* *Vibrio cholerae* Also included *Clostridium difficile*.	Different
Study dates	Includes studies published between 1990 and 2014.	Includes studies with recruitment periods falling within five years before or after the year for which mortality or morbidity estimates were calculated.	Different
Exclusion Criteria			
Possibility to stratify participants by study setting or hospitalisation status	Excludes studies which do not stratify data by inpatient, outpatient or community settings.	Includes studies which do not stratify data by inpatient, outpatient or community settings. Uses non-stratified data and adjusts for sample population in the model.	Different
Participant representativeness of the general population	Excludes studies that report nosocomial, chronic, antibiotic induced or outbreak diarrhoea.	Excludes studies that report nosocomial, chronic, antibiotic induced or outbreak diarrhoea. Also excludes studies where population are non-representative. Except for *Clostridium difficile*, where nosocomial infections are captured but modelling strategy is different in this instance.	Same except for *C. difficile*
Type of laboratory methods used to diagnose the enteric pathogens	Excludes studies which used non-standard laboratory methods.	Includes studies which used all laboratory methods.	Different
Studies exploring the presence of single pathogen only	With the exception of rotavirus, excludes studies that test for a single pathogen only.	Includes all studies and applies an adjustment factor if the data are from studies that test for a single pathogen only.	Different
Differentiation of pathogen strains	Excludes studies that fail to differentiate norovirus GI/II, atypical and typical EPEC, LT/ST-ETEC.	Includes all studies regardless of whether they differentiate strains or serotypes.	Different
Study duration	Excludes studies that last less than 12 months		Same

studies (we couldn't translate ten studies and couldn't obtain four studies). Out of the assessed studies, 91% of all ETEC and 87% of all *Shigella* studies received quality scores above 70% of a maximum obtainable score and were considered high quality studies (Fig. 2A, B). Of these studies, 51.1% of all ETEC studies and 38.8% of all *Shigella* studies received the highest quality score (Fig. 2A, B). We found that the distribution of all ETEC and Shigella quality scores between IHME and MCEE was similar (two-tailed p-value = 0.25 for ETEC studies and p = 0.2813 for *Shigella*; Wilcoxon matched-pairs signed rank test). Both ETEC and *Shigella* studies were distributed across all WHO regions; however, this distribution was heterogeneous within regions, resulting in no studies selected from multiple countries. The low-quality studies (less than 70% of top obtainable score) were heterogeneously distributed across the globe.

Table 2
Comparison of studies used by IHMEE and MCEE to generate 2016 burden of enteric disease estimates.

Pathogen	Total number of studies	Number of studies used by MCEE[1]	Number of studies used by IHME[2]	Number of studies used by IHME and MCEE
ETEC	51	7	44	0
Shigella	91	31	61	1

[1] Studies published in 1990–2014 were used to generate MCEE estimates for 2016.
[2] Studies with recruitment periods overlapping with 2011–2016 were used to generate IHME estimates for 2016.

We also examined the most frequent reasons for studies losing quality scores. Among ETEC and *Shigella* studies used by IHME or MCEE to generate mortality estimates for 2016, the most frequent reason for losing a quality score was a lack of case definition and

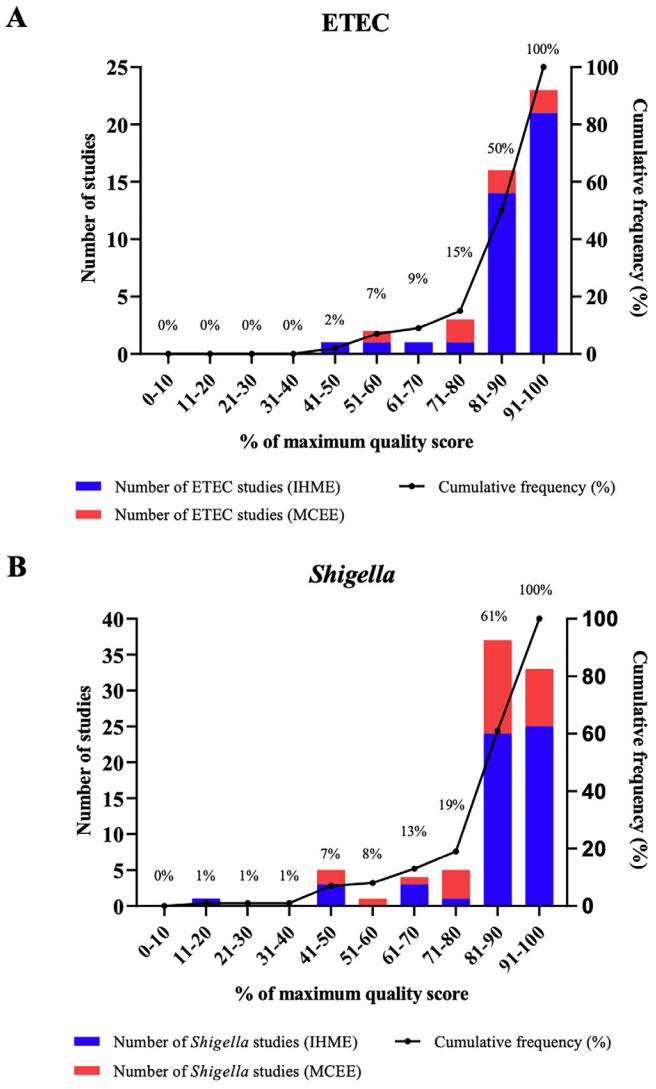

Fig. 2. Distribution of (A) ETEC and (B) Shigella study quality scores for studies used to generate 2016 burden of enteric disease estimates.

reliance on patient self-reporting of diarrhoea-associated symptoms, and inconsistent use of laboratory methods to determine the aetiology of infection (Table 3).

4. Discussion

4.1. Inclusion and exclusion criteria for modelling ETEC and Shigella mortality estimates

To address the factors that determine the variation in etiology-specific U5 mortality estimates generated by BoED models using ETEC and *Shigella* cases as proxies for diarrhoeal deaths, we examined the studies used by IHME and MCEE modelling groups and conducted quality grading analysis.

We found that both modelling groups used different ETEC and *Shigella* studies to inform the BoED models, with no commonality in studies used to generate estimates for ETEC and only one study used by both groups to model *Shigella*-attributed U5 mortality in 2016 (Table 2). We also observed that MCEE used fewer ETEC and *Shigella* studies compared to IHME. These differences were likely to reflect the individual strategies on data selection and processing used by each group. While both modelling groups included studies that examined at least 100 samples with a timeframe of 12 months or more (to reduce the chances of pathogen seasonal variation), other study selection or exclusion criteria differed between the two groups (Table 1).

The groups applied different approaches towards timeframes from which the studies were selected. While MCEE included studies published in 1990–2014 to generate the estimates for 2016, IHME used studies that had the recruitment period fully or partially overlapping the 5 years immediately before and after the year for which mortality was estimated (i.e. for 2016 estimates, studies with a recruitment period falling between 2011 and 2016, until the year the estimates were being made were considered for inclusion; Tables 1–2). This divergence in timeframes from which the studies were selected was possibly a critical factor that contributed to the use of different datasets by both groups (study periods only overlapped for 2011–2014). Furthermore, while MCEE used only peer-reviewed published studies, IHME also included some datasets that were unpublished or published in languages other than English (which we did not include in the analysis). We acknowledge that such datasets can contain valuable information and suggest identifying unpublished datasets or ongoing studies where possible to reduce the risk of publication bias, as per current Cochrane recommendations (Table 4) [13].

Both groups applied different approaches towards data stratification. MCEE only selected publications that report inpatient data to extrapolate information on diarrhoea-associated mortality, whereas IHME used both inpatient and outpatient data and adjusted the outpatient data to be representative of the inpatient

Table 3
Reasons for studies losing a quality score. Numbers in the table indicate number of studies used to generate 2016 burden of enteric disease estimates that lost scores for the indicated criteria.

Criterion	IHME		MCEE	
	ETEC	*Shigella*	ETEC	*Shigella*
Case definition not provided, based on record linkage or patient self-reporting	10	23	4	15
Exposed group is not representative of the general population (e.g. includes patients with chronic diseases, malignancies)*	1	1	–	–
Cases are not representative of the general population (e.g. include patients with chronic diseases, malignancies)	3	4	–	–
Study does not describe how controls were selected (e.g. community, hospital or mixed)♣	–	1	1	1
Sample size <100	–	–	–	–
Information on missing data or controls not described	2	3	1	1
Laboratory diagnostic methods are not described or microbiological cause of diarrhoea is determined by symptomatic reporting, or patient self-reporting	1	3	–	1
Laboratory methods are not consistent	7	10	3	13
Length of follow-up less than 12 months*	–	–	–	1
Cohort loss to follow-up more than 20% or not reported*	–	–	–	1

* Cohort studies only.
♣ Case-control studies only.

Table 4
Recommendations for inclusion of studies and datasets when conducting mortality estimates for enteric infections.

Purpose of recommendation	Audience	Recommendation	Justification
Reduce the risk of publication bias	Modellers	Identify unpublished datasets or ongoing studies if possible using systematic methods such as searching pre-print servers or funders' databases. Consider risks associated with use of unpublished data (e.g. datasets that have not been peer-reviewed). Consider using studies published in local languages, especially if they pertain to regions with high burden of enteric diseases or areas where data is limited.	4.1, 4.3
Estimate the contribution of the inherent properties of BoED models to estimate variation	Modellers	Conduct sensitivity analyses.	4.1
Reduce the risk of bias associated with selection of cases or overrepresentation of aetiologies	Investigators	Use broad case definitions in multiple-pathogen studies. Record information on aetiology-associated symptoms or complications as outcomes of investigation in multiple-pathogen studies to allow stratification of data*.	4.2
	Modellers	Where possible, minimise the need to extrapolate data between populations. If needed, the extrapolation should be well documented, transparent and tested using various scenarios.	4.1
Reduce the risk of bias due to the inclusion of participants that are considered non-representative	Investigators	If study includes participants that are not considered representative of the general population (e.g. patients with chronic conditions), the data should be presented in a way that would allow its stratification by factors that affect representativeness.	4.2, 4.3
	Modellers	Consider subregional or subnational heterogeneity when extrapolating data for areas where transmission of enteric pathogens may occur in low SES settings localised to particular areas.	
Reduce the risk of bias or variation due to use of different diagnostic methods	Investigators	Use consistent laboratory diagnostic techniques (e.g. use standardised protocols across multiple study sites; tests compatible across strains or serotypes). Provide full information on which diagnostic techniques were used to extract the data. Consider use of additional validation methods to differentiate pathogen species sharing virulence factors (e.g. use of pathogen-selective growth media or serotyping to differentiate EIEC/*Shigella* co-detected by PCR).	4.2
	Modellers	Consider stratifying enteric pathogens by serotypes, strains or sub-strains. Consider strategies used to compile data obtained through conventional and molecular diagnostic techniques (e.g. model adjustments or weighting of data acquired through molecular testing). Consider the effect of enteric pathogen serotypes, strains or sub-strains to estimates on burden associated with pathogen of interest.	

* Excluding single-pathogen studies, investigations on enteric pathogen strains or serotypes, studies on pathology specific to a pathogen or pathogen groups of interest, settings of interest or types of diarrhoea.

data. Both groups assumed that patients who are admitted to hospitals have more severe symptoms than non-hospitalized patients, and their distribution of pathogen aetiology reflects the distribution of aetiology in patients who died. While the extrapolation of data that represent outpatient populations to inpatient populations increases data availability, it may also lead to overrepresentation of pathogens that do not normally cause severe disease or death. Such extrapolations should be minimised where possible, or when conducted, well documented, transparent and tested using various scenarios. IHME and MCEE also use different approaches selecting datasets from studies testing for single or multiple diarrhoeal pathogens. Single pathogen studies are often conducted at sites with high pathogen prevalence and overrepresent the true pathogen aetiology [4]. As a result, MCEE excluded studies where a single pathogen (ETEC or *Shigella* only) was examined or different strains of ETEC were not differentiated. IHME included such studies, adjusting single pathogen studies for overrepresentation of aetiology (Table 1).

Differences in data obtained through application of diverse diagnostic techniques can contribute to variation in diarrhoea-associated mortality estimates [2,12,14]. While standard laboratory methods, such as bacterial culture, microscopy or antigen immunoassays may lack sensitivity compared to molecular, PCR-based diagnostics, the latter, despite their sensitivity, may not distinguish between viable and dead microorganisms [15]. IHME and MCEE both take different approaches towards use of data obtained using different laboratory techniques. While MCEE excludes studies that use non-standard detection methods (accepting both traditional microbiologic and newer molecular methods), IHME includes studies regardless of the detection methods, adjusting data relative to qPCR findings from GEMS and MAL-ED (Table 1) [1]. The adjustment for the variation in the use of detection methods is not conducted by MCEE.

The overall pathogen-specific diarrhoea-associated mortality can also be affected by pathogen strains or serotypes. This may be particularly important when generating estimates for ETEC-associated mortality, as ST-ETEC but not LT-ETEC was found to be associated with moderate-to-severe diarrhoea [11]. In addition, inclusion of strains producing different ST toxins was shown to modify the estimates on proportion of cases attributable to ST-ETEC [12]. To address the possible issue of estimate variation due to differential contribution of diarrhoeal pathogen strains, the modelling groups applied different approaches. While MCEE excludes studies that do not differentiate between atypical and typical enteropathogenic *E. coli* or LT- or ST-ETEC, IHME includes studies irrespective of whether they differentiate serotypes or strains (Table 1), adjusting for variation [1].

Considering these different approaches – use of only those epidemiological datasets that require as little processing as possible (generally the MCEE approach) or more diverse data that would need normalisation through adjustments before using it to generate the estimates on diarrhoea associated mortality (generally the IHME approach), it is not surprising that the modelling groups used different datasets to generate ETEC- and *Shigella*-associated mortality for 2016. It is possible that the use of such divergent datasets may contribute to differences in BoED estimates, although other factors, such as different modelling strategies, could add to estimate variation as well. Sensitivity analyses, for instance, testing

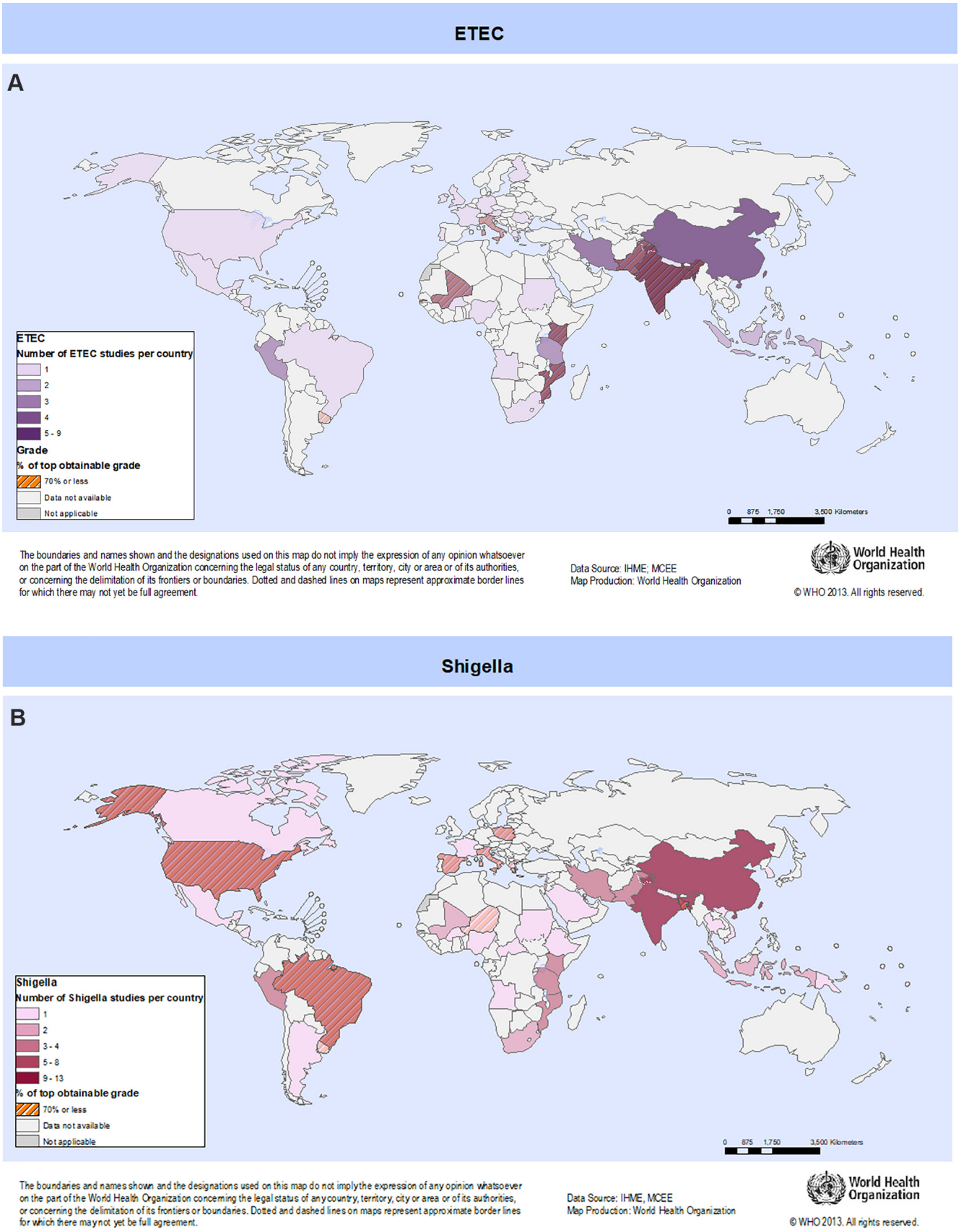

Fig. 3. Distribution of studies used in IHME and MCEE models. The map shows countries where (A) ETEC and (B) Shigella studies used by both modelling groups were conducted to generate estimates for 2016. Colour intensity indicates the number of studies conducted in that country. Striped pattern indicates countries where studies that received less than 70% of maximum quality score were conducted. Studies conducted at multiple sites (countries) were counted as individual studies per country.

how diarrhoea-associated mortality estimates are affected by use of different datasets in BoED models, would allow determining the extent to which differences in datasets or modelling strategies contribute to estimate variation.

4.2. Key insights from the grading analysis and associated recommendations

To address and compare the quality of studies used by both groups to generate ETEC and *Shigella*-attributable mortality in U5 estimates for 2016, we applied the modified NOS and conducted the grading analysis. No significant differences in the overall distribution of scores were found among ETEC and *Shigella* studies for both groups, likely due to the rigorous selection process defined by the inclusion and exclusion criteria used by each of the modelling groups. The most frequent reasons for losing score for study quality were a lack of case definition (considered important for comparability of data) or reliance on patient self-reporting (considered subjective and susceptible to recall bias); and instances where a study could not be graded for diagnostic consistency because of poor description of methods or because data was extracted from database records (Table 3). Some ETEC studies lost a point for the consistency of the laboratory methods because of substantial technological differences in the LT- and ST- toxin detection assays available at the time the studies were conducted (e.g. adrenal cell microtiter assay vs DNA hybridisation). We encourage the investigators to use standardised protocols, especially, in multi-site studies, report their laboratory methods extensively, use supplementary techniques that allow differentiation of pathogens between species and present their data in a way that would allow stratification by serotypes, strains or sub-strains (Table 4).

Although most studies used by IHME and MCEE models for ETEC or *Shigella*-associated U5 mortality based their definitions of diarrhoea on the WHO guidelines [16] and examined validated cases, additional symptoms, such as blood or mucus or consistency of diarrhoeal stools, their frequency or overall duration were also often considered when defining inclusion criteria for cases. Such variation may reflect diversity in study setting or purpose, including aetiologies or pathologies of interest [17]. For instance, in GEMS, moderate-to-severe cases of diarrhoea were recruited, and patients included those with dysentery diarrhoea [11]. While associated with certain types of enteric infections [18], these additional criteria are highly heterogeneous and may result in bias recruiting patients with selected aetiologies in multiple-pathogen studies. We also found that among studies that examined severity of ETEC or *Shigella* infections, factors considered associated with severity of diarrhoea varied, including different combinations of symptoms, such as fever, vomiting, dysentery, dehydration or requirement for oral rehydration therapy. Some studies, however, exploited tools to determine severity of diarrhoea based on clinical symptoms, such as the Vesikari scale [19]. While this tool is frequently used in identifying moderate-to-severe cases of rotavirus-associated diarrhoea, it can underestimate severity of invasive diarrhoea, e.g. *Shigella* infections as they present with different symptoms [14,20]. We encourage the investigators to use broad case definitions in multiple-pathogen studies and avoid pre-selecting cases based on symptoms or complications where it may introduce over- or underrepresentation of some aetiologies unless that is necessitated by the hypothesis. We would also encourage development of more standardised and aetiology-specific approaches to distinguish severe and milder cases of diarrhoea, in the form of guidelines or graded scales (Table 4). Such tools would allow for higher comparability of data from across different studies and could be useful to measure and compare the impact of enteric vaccines in trials [14].

While both groups excluded studies reporting data from participants with nosocomial, chronic and antibiotic-induced or outbreak-associated diarrhoea (Table 1), some studies used to generate ETEC and *Shigella* estimates included patients with chronic conditions, including malignancies and gastrointestinal conditions or were conducted in settings otherwise not representative of the general population (e.g. military bases or low socio-economic status (SES) communities with lack of access to clean water). It should be noted that studies on participants with underlying health conditions might be of particular value in settings with high rates of nosocomial infections; however, if mixed participant populations are investigated (e.g. in a hospital setting), the data should be presented in a way that would allow for stratification by the participant's status. Regarding studies conducted in low SES settings, it should be noted that while such communities may be at a higher risk of enteric disease transmission and contribute to diarrhoea-associated mortality, their representativeness of region may not always be appropriate. Indeed, a recent study in 11 eastern and central African countries showed high sub-national heterogeneity in ETEC and *Shigella*-associated mortality, representing potential for diversity within SES [6].

4.3. Geographical representation on ETEC and Shigella-associated mortality

We examined the distribution of ETEC and *Shigella* studies used by IHME and MCEE models to generate the estimates for 2016 (Fig. 3A, B) and found that for both ETEC and *Shigella,* most studies were conducted in India, China and Bangladesh in South-East Asian (SEAR), African (AFR) and Western Pacific (WPR) rerions (Fig. 3, Table S2). With the exception of China, these countries and regions have been estimated to have high or moderately high diarrhoea-associated mortality and benefit from ETEC or *Shigella* vaccines [21].

Most countries, however, did not have any ETEC or *Shigella* studies included in either IHME or MCEE estimates. Both modelling groups extrapolate data for regions or countries from which data is missing: IHME uses GEMS to extrapolate odds ratios and MCEE uses global medians for pathogen specific proportions [1]. It should be noted that estimates of ETEC and *Shigella*-associated mortality vary not only across countries or regions, but at sub-national level as well [6], raising concern about a possible risk of under- or overextrapolation of regional estimates. To increase global BoED estimate coverage for 2016, both modelling groups used studies published in local languages. Where appropriate, we encourage using such studies in line with recommendations for Cochrane systematic reviews [22] (Table 4).

5. Limitations

In this descriptive study, we were unable to determine the impact of multiple factors on the final U5 etiology-specific diarrhoea mortality estimates and their variation such as: inclusion of unpublished datasets, studies published in non-English language, use of low-quality datasets, or change in mortality trends for ETEC or *Shigella*. We were also unable to determine the influence of GEMS or MAL-ED datasets in these models, especially, where regional diarrhoea-associated U5 mortality estimates were extrapolated. Similarly, both models included high numbers of studies from countries with high or moderate burden of diarrhoea, but we were not able to determine whether abundance of these studies could bias the global or regional estimates. Some limitations of this work also include the use of the modified NOS as, despite being based on a template, this tool requires the users to make subjective interpretations on adequacy of case definitions,

representativeness of controls or cases participating in the study, and has been reported to lack power in distinguishing confounding factors [8].

6. Conclusions

This work presents discussion of two different strategies for modelling U5 diarrhoea-associated mortality estimates: use of BoED studies that have been through a strict selection process (MCEE) or through a more inclusive selection of studies, adjusting for confounding (IHME). We found that both groups used different sets of studies in their models. While examining the inclusion and exclusion criteria used by the groups, we found some similarities, such as required minimal sample size and study duration, although most criteria defining data stratification, pathogens of interest or their strains, were different. The most striking difference was the timeframes from which the studies were selected, overlapping for years 2011–2014. We predict that this may have contributed to differences in IHME and MCEE ETEC- and Shigella-associated U5 mortality estimates in 2016. We also examined the quality of studies used in these models and found that it was similar. The most frequent reasons for studies losing quality scores were lack of case definitions, lack of consistency in laboratory methods or representativeness of the population or poor description of study details. We explored the geographical distribution of studies used in BoED models and found that although datasets from countries from all WHO regions were used, multiple countries were missing data. Based on these findings, we prepared a set of recommendations for the modellers, investigators and policy makers involved in future BoED estimate generation. Although we used ETEC- and *Shigella*-associated U5 mortality estimates as examples, these findings extend to modelling estimates for burden of other infectious diseases.

7. Disclaimer

The authors alone are responsible for the views expressed in this article and they do not necessarily represent the views, decisions or policies of the institutions with which they are affiliated.

CRediT authorship contribution statement

MHA was responsible for conceptualization, data curation, formal analysis, supervision, writing- original draft, review & editing. **BKG** was responsible for conceptualization, funding acquisition, supervision, writing - review & editing. **EB** was responsilbe for data curation, formal analysis, methodology, visualization, writing - original draft. All other authors were responsible for conceptualization, writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge the funding from the Bill & Melinda Gates Foundation to WHO (OPP1135836).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.vaccine.2021.05.086.

References

[1] Prudden HJ, Hasso-Agopsowicz M, Black RE, Troeger C, Reiner RC, Breiman RF, et al. Meeting Report: WHO Workshop on modelling global mortality and aetiology estimates of enteric pathogens in children under five. Cape Town, 28–29th November 2018. Vaccine 2020. doi: http://dx.doi.10.1016/j.vaccine.2020.01.054.
[2] Hosangadi D, Smith PG, Giersing BK. Considerations for using ETEC and Shigella disease burden estimates to guide vaccine development strategy. Vaccine 2019;37:7372–80. https://doi.org/10.1016/j.vaccine.2017.09.083.
[3] GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet (London, England) 2015;385:117–71. doi: http://dx.doi.10.1016/S0140-6736(14)61682-2.
[4] Lanata CF, Fischer-Walker CL, Olascoaga AC, Torres CX, Aryee MJ, Black RE. Global Causes of Diarrheal Disease Mortality in Children <5 Years of Age: A Systematic Review. PLoS One 2013;8. doi: http://dx.doi.10.1371/journal.pone.0072788.
[5] Khalil IA, Troeger C, Blacker BF, Rao PC, Brown A, Atherly DE, et al. Morbidity and mortality due to shigella and enterotoxigenic Escherichia coli diarrhoea: the Global Burden of Disease Study 1990–2016. Lancet Infect Dis 2018;18:1229–40. https://doi.org/10.1016/S1473-3099(18)30475-4.
[6] Bagamian KH, Anderson IV JD, Muhib F, Cumming O, Laytner LA, Wierzba TF, et al. Heterogeneity in enterotoxigenic Escherichia coli and shigella infections in children under 5 years of age from 11 African countries: a subnational approach quantifying risk, mortality, morbidity, and stunting. Lancet Glob Heal 2020;8:e101–12. https://doi.org/10.1016/S2214-109X(19)30456-5.
[7] World Health Organization. WHO's 5th Product Development for Vaccines Advisory committee (PDVAC) meeting 26–28 June 2018. Geneva: 2018.
[8] Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses n.d.
[9] Platts-Mills JA, Babji S, Bodhidatta L, Gratz J, Haque R, Havt A, et al. Pathogen-specific burdens of community diarrhoea in developing countries: a multisite birth cohort study (MAL-ED). Lancet Glob Heal 2015;3:e564–75. https://doi.org/10.1016/S2214-109X(15)00151-5.
[10] Kotloff KL, Nasrin D, Blackwelder WC, Wu Y, Farag T, Panchalingham S, 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 St. Lancet Glob Heal 2019;7:e568–84. https://doi.org/10.1016/S2214-109X(19)30076-2.
[11] Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet 2013;382:209–22. https://doi.org/10.1016/S0140-6736(13)60844-2.
[12] Liu J, Platts-Mills JA, Juma J, Kabir F, Nkeze J, Okoi C, et al. Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the GEMS case-control study. Lancet 2016;388:1291–301. https://doi.org/10.1016/S0140-6736(16)31529-X.
[13] Lefebvre C, Glanville J, Briscoe S, Littlewood A, Marshall C, Metzendorf M-I, et al. Chapter 4: Searching for and selecting studies. Cochrane Handb. Syst. Rev. Interv. version 6.1 (updated Sept. 2020)., 2020.
[14] Kotloff KL, Platts-Mills JA, Nasrin D, Roose A, Blackwelder WC, Levine MM. Global burden of diarrheal diseases among children in developing countries: Incidence, etiology, and insights from new molecular diagnostic techniques. Vaccine 2017;35:6783–9. https://doi.org/10.1016/j.vaccine.2017.07.036.
[15] Platts-Mills JA, Liu J, Houpt ER. New concepts in diagnostics for infectious diarrhea. Mucosal Immunol 2013;6:876–85. https://doi.org/10.1038/mi.2013.50.
[16] The World Health Organization. Diarrhoeal disease; 2017.
[17] Levine GA, Walson JL, Atlas HE, Lamberti LM, Pavlinac PB. Defining Pediatric Diarrhea in Low-Resource Settings. J Pediatric Infect Dis Soc 2017;6:289–93. https://doi.org/10.1093/jpids/pix024.
[18] Ahs JW, Tao W, Löfgren J, Forsberg BC. Diarrheal diseases in low- and middle-income countries: incidence, prevention and management. Open Infect Dis J 2010;4:113–24. https://doi.org/10.2174/1874279301004010113.
[19] Ruuska T, Vesikari T. Rotavirus disease in finnish children: use of numerical scores for clinical severity of diarrhoeal episodes. Scand J Infect Dis 1990;22:259–67. https://doi.org/10.3109/00365549009027046.
[20] Wierzba TF, Bourgis A. Defining cases of severe pediatric diarrhea for an efficacy trial of an enterotoxigenic Escherichia coli (ETEC) vaccine: Report on an international workshop, Washington DC, March 2016. Vaccine 2017;35:503–7. https://doi.org/10.1016/j.vaccine.2016.12.006.
[21] Anderson JD, Bagamian KH, Muhib F, Baral R, Laytner LA, Amaya M, et al. Potential impact and cost-effectiveness of future ETEC and Shigella vaccines in 79 low- and lower middle-income countries. Vaccine X 2019;2. doi: http://dx.doi.10.1016/j.jvacx.2019.100024.
[22] Boutron I, Page MJ, Higgins JPT, Altman DG, Lundh A, Hróbjartsson A. Chapter 7: Considering bias and conflicts of interest among the included studies. Cochrane Handb. Syst. Rev. Interv. version 6.1 (updated Sept. 2020), 2020.